An analysis of changing dietary trends and the implications for global health

Thai Hong Le

A Thesis Submitted in Partial Fulfilment of the Requirements of Bournemouth University for the Degree of

Doctor of Philosophy

September 2021

Department of Accounting, Finance and Economics
The Business School, Bournemouth University

COPYRIGHT STATEMENT

To

My parents and grandparents

who raised me, believed in me and supported me to pursue my dreams great and small

My late grandfather, in loving memory

who nursed me with love and affection

Author’s declaration

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the Bournemouth University, United Kingdom. I declare that this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that this thesis has not been previously or concurrently submitted, either in whole or in part, for any other qualification at Bournemouth University or other institutions.

Thai Hong Le
September 2021

Acknowledgements

First of all, I would like to express my deepest gratitude towards my supervisors Professor Tim Lloyd and Dr. Marta Disegna, who not only have inspired me with their vast knowledge but have shown me how to become a good researcher. Without their thorough guidance, I would not be able to produce such a complete thesis. I am enormously grateful to their patience, motivation, encouragement, and above all for believing in me more than myself.

Second, I would like to extend my gratitude to Bournemouth University and the Doctoral College for funding my PhD degree. I truly appreciate the support from the programme administrator as well as academic staff from Department of Accounting, Finance & Economics over the course of my study.

Even though the life of an international PhD student is usually deemed as an inherently lonely one, I have considered myself lucky to have some very good friends and colleagues who are always there for me, cherish and encourage me constantly. Having said that, many thanks go to Olivia, Sergio, Thuy, Kama, Kathy, Man, Tu Anh, Priya and Itai.

Next, I also wish to thank my family especially my parents who have wholeheartedly supported me in all of my pursuits. It was their passion for education that ignited in me a strong urge for knowledge, an ambition to pursue a degree abroad, and a strong determination to follow my dreams.

Finally, my acknowledgement would be incomplete without special thanks to Dorset & Hampshire Iyengar Yoga Association (DHIY), to my yoga teachers and friends. Their emotional support cannot be overestimated particularly during the COVID time.

Publications and Conference Presentations

Some material contained in this thesis has been used in the following publication:

- Le, T.H., Disegna, M. and Lloyd, T. (2021), National Food Consumption Patterns: Converging Trends and the Implications for Health. EuroChoices. https://doi.org/10.1111/1746-692X.12272.

Some contents of this thesis have been presented at the following conferences:

- Le, T.H., 2018. An econometric analysis of changing dietary trends and its implications for health around the world. International Food Marketing Research Symposium, Bournemouth, June 2018.
- Le, T.H., 2019. Convergence and heterogeneity in global diets. Agricultural Economics Society Conference, Warwick, April 2019.
- Le, T.H., 2019. Convergence and heterogeneity in global diets. Royal Statistical Society International Conference, Belfast, September 2019.
- Le, T.H., 2019. Convergence and heterogeneity in global diets. International Conference on Official Statistics: Emerging Trends in Statistical Methodologies and Data Dissemination, Sarajevo, April 2019.

Part of this research has been awarded the following prize:

- Best poster presentation award. 93rd Annual Conference of the Agricultural Economics Society (AES), Warwick, April 2019.

Abstract

Worldwide, obesity has reached epidemic proportions and has almost tripled between 1975 and 2016. Acknowledging that weight gain is a complex and multifactorial condition involving changes in both dietary and physical activity patterns, this research focuses on the dietary origin of obesity. Given the dearth of empirical literature on food consumption at the global level, the aim of this research is to explore dietary trends and dietary types around the world linked with the indications that they imply for obesity and global health. To this aim, annual food availability data collated by the Food and Agriculture Organisation of the United Nations (FAO) covering the period from 1961 to 2013 for 118 countries are scrutinised by econometric convergence tests, clustering techniques and spatial analysis. Results indicate that countries with lower levels of initial calories tend to exhibit higher growth rates of calorie consumption. However, this process is not homogeneous across countries. Low-income countries have converged at the fastest pace and the convergence rate reduces as income rises. In addition, the dietary convergence is conditioned by a range of structural indicators including agroecological, demographic and socio-economic variables. Evidence suggests that economic factors have become a more important determinant of the dietary convergence since the Millennium. Applying innovative fuzzy clustering algorithms which allow multiple diets to coexist within a single country, several dietary trends and dietary types are detected. While the identified clusters are all associated with relentlessly increasing calories and deteriorating dietary healthiness over the past half a century, the most calorific cluster has shown signs of stabilising calorie consumption. A notable contribution of this research is the examination of spatial patterns of global food consumption using both traditional and non-traditional measures of spatial proximity. Differing from the earlier literature emphasising the role of geographical closeness, this research utilises an economic indicator for proximity and finds that countries with similar income levels tend to have similar diets. Spatial convergence analysis reveals a convergence process that is about three times as fast as the non-spatial model; thus, ignoring the spatial relationship leads to biased results. Incorporating the spatial dimension in cluster analysis also affects the clustering results dramatically. Cluster profiling shows that only the segment of more educated and health-aware populations exhibits the behavioural changes towards better diets, hence underlining the importance of improved education and access to knowledge. The finding that dietary evolutions are 'spatially' dependent provides a basis for the development of group-specific interventions that target populations at risk of worsening diets. While these policy measures are often place-based, this research lays foundations for the implementation of coherent food policies beyond geographical boundaries. Overall, this research highlights that healthier diets are possible, but we need to act now. As we are living longer but not necessarily healthier, current attempts to improve diets are obviously inadequate and existing efforts need to be redoubled. This is an urgent message for policymakers considering the sobering fact that no country has been able to significantly reverse the rise in obesity.

Table of Contents

List of Figures

List of Tables

Acronyms and abbreviations

AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
ARMA	Auto Regressive Moving Average
ARIMA	Auto Regressive Integrated Moving Average
BIC	Bayesian Information Criterion
BMI	Body Mass Index
CHD	Chronic Disease
COFUST	Copula-based Fuzzy Space-Time Clustering Algorithm
COVID-19	Coronavirus Disease
CV	Coefficient of Variation
CVD	Cardiovascular Disease
DBSCAN	Density-based Spatial Clustering and Application with Noise
DHA	Docosahexaenoic Acid
DHM	Dietary History Method
DQI	Diet Quality Index
DTW	Dynamic Time Warping
EPA	Eicosapentaenoic Acid
EU	European Union
FAFH	Food Away From Home
FAO	The Food and Agriculture Organisation of the United Nations
FAOSTAT	FAO Statistical Databases on Food and Agriculture
FBS	Food Balance Sheet
FDI	Foreign Direct Investment
FFQ	Food Frequency Questionnaire
FS	Fuzzy Silhouette Index
GARCH	Generalised Autoregressive Conditional Heteroskedasticity
GATT	General Agreement on Tariffs and Trade
GDP	Gross Domestic Product
GHG	Greenhouse Gas
GIS	Geographic Information System
GNS	General Nesting Spatial Model
HCES	Household Consumption and Expenditure Survey
HDI	Healthy Diet Indicator
HEI	Healthy Eating Index

HES	Household Expenditure Survey
HIC	High-income Country
HIES	Household Income and Expenditure Survey
HSB	Household Budget Survey
ICN2	Second International Conference on Nutrition
LCFS	Living Costs and Food Survey
LISA	Local Indicators of Spatial Association
LM	Lagrange Multiplier Test Diagnostics for Spatial Dependence
LMIC	Low- and Middle-income Country
LSMS	Living Standards Measurement Study
MAI	Mediterranean Adequacy Index
MANOVA	Multivariate Analysis of Variance
MDS	Mediterranean Diet Score
MLD	Mean Logarithmic Deviation
NCD	Non-communicable Disease
NLS	Non-linear Least Square
NTM	Nutrition Transition Model
OECD	The Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
PAM	Partitioning Around Medoids
SAR	Spatial Autoregressive Model/Spatial Lag Model
SARAR	Spatial Autoregressive Combined Model
SDEM	Spatial Durbin Error Model
SDG	Sustainable Development Goals of the United Nation
SDM	Spatial Durbin Model
SEM	Spatial Error Model
SLX	Spatial Lag of X Model
SMART	Specific, Measurable, Achievable, Relevant, and Time-bound
SSB	Sugar-Sweetened Beverage
TFC	Transnational Food Corporation
UNICEF	The United Nations Children's Fund
UPC	Universal Product Code
VIF	Variance Inflation Factor
WHO	World Health Organisation
WTO	World Trade Organisation
XB	Xie-Beni Index

Chapter 1

Introduction

1.1 Research background

"Let food be thy medicine and medicine be thy food"

- *Hippocrates (400 BC)* -

The famous quote which is often ascribed to the Greek physician Hippocrates, the father of Western medicine, underlines the very intimate relationship between food and health (Witkamp and van Norren 2018). Spoken 2,500 years ago, these words remain relevant in the modern time: food and diet are critical to our health and wellbeing. In addition to its biological function, consuming food or eating is also a social and cultural activity through which people satisfy their aesthetic sensibilities and establish a communal identity (Cornil and Chandon 2016). Yet, "food is no longer a sole source of survival and enjoyment but has increasingly become a cause of concern due to its potential consequences for ill health" (de Ridder *et al.* 2017, p.908). The prime reason for this concern is the role of diet as the biggest determinant of chronic non-communicable diseases (NCDs) which are not only a common cause of disability but also responsible for 71% of global deaths annually (Bigna and Noubiap 2019). A key metabolic risk factor for NCDs is overweight and obesity (WHO 2018).

The World Health Organisation (WHO) recognised that obesity is a global epidemic in 1997 and launched this alert again in 2002 (WHO 2003). Since then, numerous calls to action have been proposed by national governments; nonetheless, overweight and obesity continue to be a pressing public health concern (OECD 2019b). Globally, the prevalence of obesity nearly tripled since 1975. The latest WHO estimates are that 39% of the world's adult population were considered overweight in 2016, and 13% were obese (WHO 2017a). To extrapolate the current estimates to the future, without policy interventions around one third of the projected global population will be overweight or obese by 2030 (Global Panel on Agriculture and Food Systems for Nutrition 2016). Once labelled as the "disease of affluence", obesity has gone global and is creeping up on even the poorest regions of the world, making the earlier labelling a misnomer (Reyes Matos *et al.* 2020). Western countries historically associated with the highest obesity rates continue to lead the race, though non-Western countries have experienced substantial increases in obesity prevalence over the past few decades (OECD 2019b). Statistics from the World Bank indicate that over 55% of the global rise in obesity originates from rural areas and the growth rate ranges from 80 to 90% in South East Asia, Latin America, Central Asia and North Africa (World Bank 2020a). While the speeds differ markedly by country and region, no country yet seems to have bucked the rising trend (Swinburn *et al.* 2011; Ameye and Swinnen 2019; Swinburn *et al.* 2019).

To further complicate the issue, over 70% of countries – the majority of which are low-income and middle-income countries (LMICs) – are witnessing the 'double burden of malnutrition' which is defined by the coexistence of overweight and obesity (*overnutrition*) alongside stunting and wasting (*undernutrition*) at all levels of the population (The Lancet 2019; World Bank 2020a). From this emerges a new nutrition reality: low-income countries are no longer characterised as undernourished nor are high-income countries only struggling with obesity (Popkin *et al.* 2020). From the public health perspective, such a dual burden will aggravate the long-term costs of 'globesity' owing to the lag in the effects of the current and past undernutrition reduction initiatives.

However, the health implications of overweight and obesity are alarming. Worldwide, at least 2.8 million adults die each year as the result of being obese or overweight and in fact overnutrition kills more people than undernutrition (WHO 2017). In addition to the high death toll, overweight and obesity are closely associated with various chronic diseases, including type 2 diabetes, cardiovascular diseases, respiratory diseases, musculoskeletal disorders, several types of cancer and depression (Bradshaw *et al.* 2019). Overweight and its related conditions can negatively affect the quality of life, measured by the loss of disability-free years in obese adults, and reduce the life expectancy by 3 years on average across OECD, EU28 and G20 countries (OECD 2019b).

Unsurprisingly, increasing health care costs linked to increasing obesity rates are a common trend across both developed and developing countries (World Bank 2020e). The economic cost of obesity approximates to that of smoking or armed conflict, driving between 2 and 7 percent of global healthcare spending (Dobbs *et al.* 2014) – figures that are predicted to double by 2030 (Giner and Brooks 2019). The economic burden of obesity also manifests in the labour market as having excess body fat (the common description of obesity) lowers productivity at work and increases early retirement as well as absenteeism, and according to the OECD these effects are translated into a decline of GDP by about 3.3% (OECD 2019b). Not only harming the economy, obesity can produce long-term detrimental impacts on the environment by generating 20% greater greenhouse gas (GHG) emissions compared to the normal-weight state (Magkos *et al.* 2020).

The oft-quoted fundamental cause of obesity is a mismatch between energy consumed (eating too much) and energy expended (moving too little). On the one hand, people have been engaging in dramatically reduced physical activity as manual and outdoors jobs are being replaced by seated and indoors ones in the service sector while improved transport means more time being spent in car and public transport than walking and cycling (Guthold *et al.* 2018). Such a heightened level of urban and sedentary lifestyles is proved to be the main contributor of the increasing rates of obesity in many countries including the UK over the past three decades (Griffith *et al.* 2016). On the other hand, there have been significant changes in food consumption patterns whereby populations move away from the traditional diets based on grains, roots and tubers to the 'Western' diet characterised by a higher consumption of salt, fat, sugars and animal-source foods (Popkin 2008; Kearney 2010; Popkin *et al.* 2012; Ronto *et al.* 2018; Popkin *et al.* 2020; Wells *et al.* 2020). At the same time, these shifts are accentuated by major transformations in the food system that have created the so-called *'obesogenic'* environment impacting food availability, food affordability and food acceptability worldwide (Lang 2009; Hawkes *et al.* 2017; Popkin 2017; Béné *et al.* 2019; Willett *et al.* 2019). The link between diet and obesity is obvious. Notwithstanding the interplay of multiple factors ranging from genetic predisposition to environmental influences that contribute to weight gain (Ralston *et al.* 2018), the focus of this thesis is on the identification of distinct dietary types as well as dietary trends around the world linked to indications that they imply for obesity and global health.

1.2 Research motivation

The extensive literature on food and nutrition finds its origins in the *nutrition transition model* (NTM) which describes five stages of changes that populations experience in the quantity and quality of dietary behaviours and patterns, which go hand in hand with large shifts in physical activity and causes of disease as countries become more economically developed, urbanised and globalised (Popkin 1993; Drewnowski and Popkin 1997; Popkin 1999, 2006b, 2021). Originally developed in the early 1990s, the concept of 'nutrition transition' has gained worldwide popularity over the last few decades. There is ample evidence for dietary patterns in most places that include higher intakes of nutrient-poor but calorie-dense foods often from animal sources and in processed form whilst intakes of fibres and vegetables remain inadequate (see, *for example*, Baker and Friel 2014; Zhai *et al.* 2014; Poti *et al.* 2015; Oberlander *et al.* 2017; Popkin and Reardon 2018; Rousham *et al.* 2020; Umberger *et al.* 2020). As a manifestation of the nutrition transition, modern societies seem to be converging on a Western-style diet high in saturated fats, sugar, and refined food, but low in fibre (Popkin and Gordon-Larsen 2004). Unlike the gradual transition that occurred in the USA and European countries, the dietary changes have been more rapid in many low- and middle-income countries (Popkin and Ng 2007; Anand *et al.* 2015). The drivers for the nutrition transition are manifold and often involve a wide variety of economic, social and cultural factors which are interconnected. While the impacts of global forces such as globalisation, rising income and urbanisation are well described in the earlier literature, the underlying mechanisms through which these take place remain ambiguous (d'Amour *et al.* 2020).

As the nutrition transition is coupled with a predicted shift in the disease burden towards an increased prevalence of overweight/obesity and the concomitant rise in NCDs (Popkin *et al.* 2012; Webb and Block 2012; Harris *et al.* 2019; Harris *et al.* 2020), it requires appropriate attention from public health policymakers. Surveillance of the evolution of dietary patterns is crucial for understanding the nutrition transition in different countries and for the development of timely interventions targeting at risk communities. Nonetheless, not enough is known about actual food consumption in many populations and the factors affecting it (Imamura *et al.* 2015; Menyhert 2020). Despite the critical importance of monitoring worse or better changing dietary patterns, surprisingly little is known regarding consumption trends over time (Walls *et al.* 2018; Ventura Barbosa Gonçalves *et al.* 2020; Juul *et al.* 2021).

That said, a number of previous studies attempt to address the current situation and make future projections of food availability (Baldos and Hertel 2014; Fukase and Martin 2017; Gouel and Guimbard 2019). The Food and Agricultural Organisation of the United Nations (FAO) has published a series of annual flagship reports titled "The state of food security and nutrition in the world" since 1999 (FAO 2021b), which again focus mainly on the current situation and attend to key challenges for achieving the sustainable development goals (SDGs) related to ending hunger, ensuring food security and improving nutrition. Historical trends of food availability are mostly examined in national/regional contexts and case studies (see, *inter alia*, Balanza *et al.* 2007; Chen and Marques-Vidal 2007; Regmi *et al.* 2008a; Sheehy and Sharma 2010; Sheehy and Sharma 2013; Ambagna *et al.* 2019; Sheehy *et al.* 2019). So far, only few studies provide in-depth analysis on global assessment of food and nutrition. Alexandratos and Bruinsma (2012) analyse global and regional trends in food availability and prevalence of undernourishment during 1970-2006, with a clear focus on developing countries. Also investigating past global trends in food availability, Porkka *et al.* (2013) probe into possible drivers behind these trends; yet, only the role of food trade is examined. In a unique attempt to assess diet quality and their trends worldwide, Imamura *et al.* (2015) explore global dietary consumption by country, age, gender and time, distinguishing between healthy and unhealthy food items. Using a range of data sources including nationally representative dietary surveys, local surveys, and food disappearance (or Food Balance Sheet) data, the authors find that dietary patterns have improved in many areas over the past 20 years with exceptions of China, India, and several sub-Saharan countries.

To the best of the author's knowledge, detailed studies focusing on how global food availability has developed during the past decades are limited. Thus motivated, this study aims to offer an improved understanding of the existing literature on the nutrition transition analysing dietary patterns and changes around the world.

While food consumption is often deemed as a matter of personal choice and many authors have conducted micro-level studies to explore drivers of individual consumption, in this research a macro-level study is chosen. This allows for detecting the impacts of structural factors common to individuals within a country but vary between countries (for example political stability, wars, access to education, weather or cultural conditions). These factors that affect all or nearly all individuals living in a country can be controlled for in a macro-level study but are not likely to be included in a micro-level study (Bansal and Zilberman 2020). In addition, results of micro-level studies better suit the purpose of individual prognoses, whereas results of macro-level studies can help better understandings of the impacts of public health policies that for instance target at reducing obesity. The global approach adopted in this research represents an important contribution to knowledge.

This research joins and extends insights from two related bodies of literature that have largely developed in parallel: nutrition transition and dietary convergence. Previous research has identified: (i) rising similarity in total caloric supplies and dietary structure across national borders (Blandford 1984; Balanza *et al.* 2007; Khoury *et al.* 2014; Di Lascio and Disegna 2017; Bentham *et al.* 2020), (ii) convergence in consumption of caloric intakes and certain food items (Regmi and Unnevehr 2006; Schmidhuber and Traill 2006; Erbe Healy 2014). Despite an extensive literature on dietary changes, there are some major gaps and shortcomings. Studies on dietary convergence are dated, loosely linked to the nutrition transition literature, and mainly focus on developed countries as well as the European Union. Little evidence is available on patterns of food consumption at the global level. Notable studies include those by Khoury *et al.* (2014), Azzam (2020), Bentham *et al.* (2020), and Bell *et al.* (2021). Nevertheless, Khoury *et al.* (2014) and Bentham *et al.* (2020) highlight similarity (instead of convergence) among global diets and both studies lack formal convergence analysis. Azzam (2020), in spite of conducting convergence tests, is more interested in measuring whether global diets are shifting towards a Westernised diet (or precisely a Western diet similarity index). In this research, the global dietary convergence is examined in the light of popular convergence methodologies using food availability data at the global level.

On the other hand, a major challenge of employing global data lies in the nature of data varying across space (countries) and time. Hence, the oft-adopted approach in previous research that aggregates country-level data by geographical regions or income levels is unable to capture the nuances of dietary changes. Importantly, combining very poor countries with high burdens of malnutrition with middle- and high-income countries “ignores the complexities and realities of the nutrition transition” since they are likely at different stages of the transition (Popkin 2021). In order to address this issue, this research complements the convergence analysis with a statistical technique called *cluster analysis* to summarise and describe global diets on the basis of historical trends. A great novelty of this study is the application of fuzzy clustering algorithms which permit the possibility of multiple diets coexisting within a single country.

Returning to the NTM, of particular interest to researchers is whether the shift beyond the nutrition transition (to Pattern 5), where individuals move back to a healthier diet, is anything more than a theoretical possibility. The behavioural changes for better diets could be induced either by increasing awareness of individuals or by policy interventions. Is the transition to Pattern 5 a reversal of the forces at play or does it require new factors to come into play? Results of the cluster analysis will help to detect any evidence for the existence of Pattern 5 in some countries and how the diets of these countries might look like at this stage, which is another key contribution to the existing literature.

1.3 Research questions and objectives

This research aims to summarise and analyse the evolution of global patterns of food consumption using past trends. Recognising several research gaps in the earlier literature and attempting to address them, this research intends to pursue three research questions:

- **RQ1:** How and toward which direction have diets in the world evolved?
- **RQ2:** How do structural factors influence the evolution of the world's diets?
- **RQ3:** Is it possible to revert to a healthier diet?

To achieve the aim and to answer the research questions, the following objectives are established:

1. To test for global dietary convergence, to estimate the speed of convergence and to measure how the convergence process is influenced by structural factors.
2. To identify common dietary trends and dietary types around the world and to assess how they differ in terms of the healthiness of diets and the structural characteristics.
3. To elucidate the role of the space on dietary convergence and dietary changes.

In doing so, this research employs the Food Balance Sheet (FBS) data collated by the Food and Agriculture Organisation of the United Nations (FAO) (FAO 2019b). Data are expressed in terms of calories (kcal/capita/day) and available as annual series dating back to 1961. Although the FBS remains one of the most widely used sources of dietary data (FAO 2019a), it is best described as average national diets and food available for human consumption instead of actual food consumption as consumer waste is not accounted for (FAO 2018; Vilarnau *et al.* 2019). The data will be analysed by a range of econometric methods and statistical techniques.

This study first examines the convergence in food availability using two popular convergence testing procedures: sigma and beta convergence. A great novelty is the examination of the convergence process for countries of different income levels, which sheds light on the role of economic development in dietary convergence. While convergence models of food consumption or food expenditures are not rare in the field (Blandford 1984; Herrmann and Röder 1995; Regmi *et al.* 2008b; Ogundari and Ito 2015), the approach adopted in this research differs in that various structural factors behind dietary changes are included in the convergence specification. In this respect, it allows for showing, for instance, that countries differ in these structural parameters and hence experience heterogeneous convergence paths.

Second, cluster analysis is applied to detect distinct dietary types and dietary trends around the world. This method classifies countries into different clusters (or groups) so that countries within a cluster exhibit similar patterns of food consumption to one another but dissimilar to countries in other clusters. Clustering is not an entirely new method in the food economics literature and it has been utilised to indicate which countries are similar in terms of food expenditures (Erbe Healy 2014; Staudigel and Schröck 2015) or food consumption behaviours (Gil *et al.* 1995; Balanza *et al.* 2007; Azzam 2020), and how this grouping evolves over time (Walthouwer *et al.* 2014). Dealing with dietary data that vary temporally, previous scholars either merge time series into one set of static data (Blandford 1984; Staudigel and Schröck 2015; Sadowski 2019), or apply clustering algorithms on discrete time periods comparing the results between a baseline and a follow-up period (usually the first and last year) (Gil *et al.* 1995; Di Lascio and Disegna 2017). In contrast to prior studies in this research strand, this analysis performs the clustering task on the whole set of time sequences by adopting an innovative time series clustering algorithm (Disegna *et al.* 2017). By grouping countries into distinct clusters (or groups) of homogeneous trajectories of calorie consumption, it is possible to assess which dietary characteristics most resemble each cluster, how the dietary healthiness differs between clusters, and how these changes over time. The clustering results can imply whether there is evidence for the better dietary changes.

Finally, given that much of the previous work in the food economics literature has been set in either a cross-sectional or a time series context, this analysis attempts to merge the two, bringing the spatial component into a time series framework demonstrating the value it adds to convergence and cluster analysis. Does the country's location matter in convergence modelling of the global diets? Since the convergence process takes place over time and across space, it is reasonable to account for the spatial dimension. As it has been shown in economics literature that rich regions tend to be surrounded by similarly rich regions (Annoni *et al.* 2019), perhaps countries with high-calorie diets might likely be located near countries with similarly high-calorie diets, implying similar convergence. As a matter of fact, the role of the space in dietary convergence and dietary shifts has been largely neglected, and hence this research provides a novel application. Countries are grouped into clusters that are coherent in the temporal evolution of caloric consumption whilst being spatially close. Considering the space as a contextual factor for dietary behaviour allows us to scrutinise the differing environments between clusters which otherwise would be hidden in a non-spatial model.

1.4 Structure of the thesis

This thesis starts with a description of research background and research motivation with which the objectives of the study are established. The remainder of this thesis comprises seven chapters and is structured as follows.

Chapter 2 introduces the theoretical framework underpinning this research – the nutrition transition model. Key stages of the nutrition transition along with its applications in the empirical literature are presented. As predicated by the nutrition transition, diets around the world are converging on the consumption of certain food items and therefore are becoming more similar over time. **Chapter 2** then provides a review on the development of convergence theories which originated in economic growth literature that poorer economies' per capita incomes tend to grow faster than richer economies and eventually "catch up" with them. This is followed by a discussion on how previous researchers in the food economics literature have employed this idea of convergence in exploring the dynamism in regional and international patterns of food consumption.

Chapter 3 serves as a literature review on cluster analysis, aiming to underline its suitability and novelty for this research. The chapter includes a description of the main steps in cluster analysis and discusses various methods for clustering static data as well as data varying over time and/or across the space. A later section of **Chapter 3** examines the application of cluster analysis in food economics literature with a focus on food consumption studies. From this arise several limitations of previous studies and the need for further research.

Chapter 4 reviews major confusions behind the concept of diet quality and the debate on constructing diet quality indices. Several definitions of *diet quality* and the nuances in quantifying it are introduced. The chapter continues with an overview of approaches in constructing pre-defined diet quality indices, the myriad of existing indices, different dietary data sources and the predictive ability of diet quality indices for health outcomes. **Chapter 4** concludes by highlighting motivations for adopting a diet quality index in this research and introducing the most appropriate one.

Chapter 5 gives a detailed description of the Food Balance Sheet data and offers some insights into the evolution of global patterns of food consumption over the past half a century. Evidence of the nutrition transition is identified in relation with changes in the consumption and composition of macronutrients as well as main food aggregates. It wraps up with the meaningful correlation between caloric changes and income growth.

Building on the evidence of the nutrition transition in the previous chapter, **Chapter 6** provides empirical results that seek to answer whether the dietary transition is a universal shift and whether the transitional speeds uniform across countries. To this end, two convergence methods are utilised to examine the converging caloric consumption across countries, and how the convergence process differs between countries at different levels of development. The concept of converging diets is then revisited through the lens of spatial analysis incorporating spatial relationship among countries with convergence econometric model. Spatial effects are detected by statistical techniques, the source of spatial effects is identified, and the convergence model is augmented to reflect this phenomenon.

Chapter 7 aims to investigate the major dietary trends and dietary types around the world. An innovative time series clustering algorithm is employed to ascertain the similarities among global diets. Bringing together the literature on spatial analysis and food economics, this chapter also examines the role of the space in dietary changes and their trends. The empirical analysis in **Chapter 7** continues with the application of a novel space-time clustering algorithm, which can account for the environment conditions of food consumption.

Chapter 8 concludes this thesis by offering some key conclusions, implications for policy, contributions to knowledge, limitations and directions for future research.

Chapter 2

Nutrition transition and the global dietary convergence

2.1 Chapter introduction

As countries develop and become on average wealthier, more urbanised and more open to global trades, individual's diets and lifestyles have changed (Popkin 2006a). Often, changes in dietary and activity patterns are paralleled by major demographic and socio-economic dynamics (Masters *et al.* 2016). Since the last decade of the 20th century, researchers across different disciplines (including food economics, nutrition, and health) have widely adopted the concept of 'nutrition transition' when referring to the global shift in dietary patterns towards energy-dense foods high in fat and sugars and in reduced physical activity. These changes are reflected in nutritional outcomes, such as average height and body composition (Popkin *et al.* 2012).

This chapter aims to provide a review of the literature on the nutrition transition. In doing so, the nutrition transition model (NTM) is introduced as the theoretical framework underpinning the main theme of this research. Critical developments of the nutrition transition theory along with its applications in the existing literature will be mapped out. As indicated by the NTM, the current state that most countries around the world are experiencing is characterised by the convergence towards an increased consumption of certain food items. The second major topic to be discussed in this chapter is convergence theories and how previous researchers have employed the concept of convergence in exploring the dynamism in regional as well as international food consumption patterns.

The rest of this chapter is organised as follows. Section 2.2 presents key patterns of the NTM, evidence of the nutrition transition around the world, implications for human health and the environment, and some comments about the usefulness of the model. Section 2.3 elaborates the role of globalisation and other global underlying forces. Section 2.4 discusses some scholarly debates on the issue of obesity around the world from food economics perspective. Section 2.5 introduces popular concepts of convergence and summarises how they have been applied in previous studies on food consumption. Section 2.6 wraps up with remarks confirming the rationale of this research.

2.2 The nutrition transition model

2.2.1 Five patterns of the nutrition transition model

The concept of 'nutrition transition' was originally coined by Barry Popkin in the early 1990s to describe major dietary changes that were observed worldwide in the last decades of the 20th century (Popkin 1993). Since then, a large stream in the literature has been shaped around the nutrition transition model (NTM), which identifies five temporal patterns (or stages) of dietary changes, namely Hunter-gatherers, Famine, Receding Famine, Degenerative Disease and Behavioural Change (Drewnowski and Popkin 1997). Most attention is drawn to the last three patterns, which are represented by most of the global population today (Popkin 2002a). Figure 2.1 outlines five patterns of the NTM accompanied by their most distinctive features. For the sake of convenience, these patterns can be thought of as historical developments; however, earlier patterns are not restricted to the periods in which they first arose but continue to characterise certain geographic and socio-economic subpopulations (Popkin 2006a).

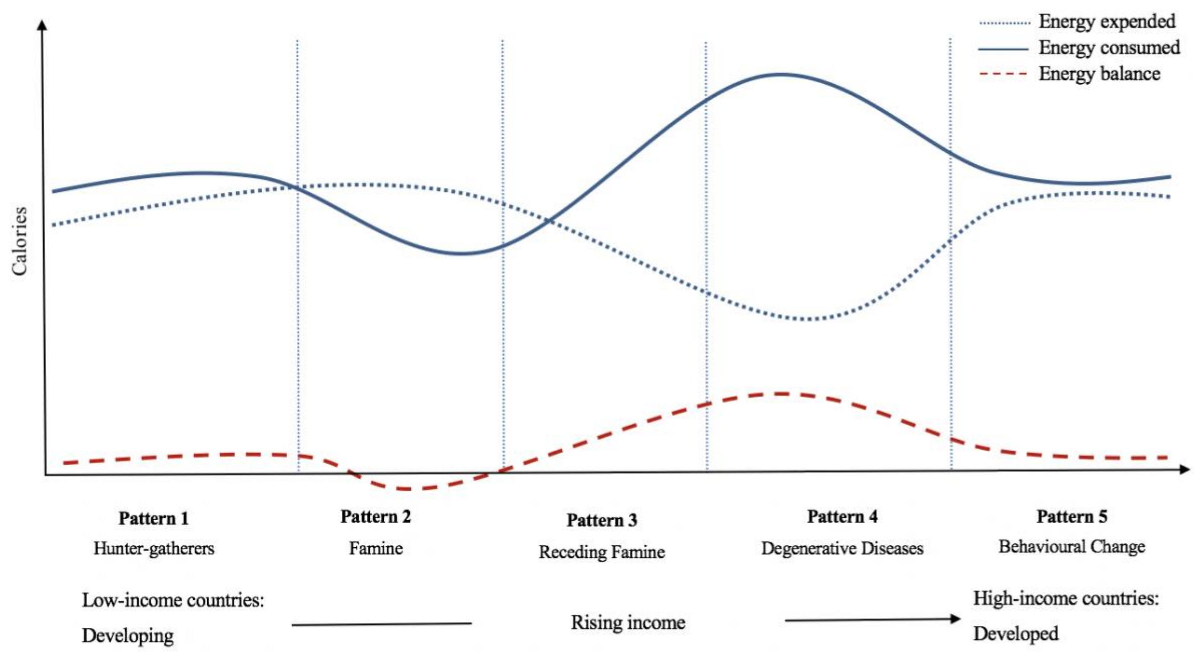


Figure 2.1 Five patterns of the nutrition transition model.

Pattern 1: Hunter-gatherers

The first stage is often termed the ‘Palaeolithic’ pattern and is closely linked with hunter-gatherer societies. Much of the research conducted for this stage is in accordance with the modern hunter-gathers due to scarce evidence for the pre-historic population. The diet characterised with this pattern is high in carbohydrates and fibre but low in fat, especially saturated fat (Truswell 1977). The percentage of polyunsaturated fat in meat from wild animals is larger than from domesticated animals (Eaton *et al.* 1988). Physical activity level is high, labour-intensive work is dominant, and obesity prevalence is low. Generally, the diet associated with this pattern is considered healthy; however, the frequent occurrence of infectious diseases and natural disasters leads to very short life expectancy (Popkin 2006a).

Pattern 2: Famine

The diet in this pattern is less varied than in the previous pattern and is impacted by periods of severe food shortages. As a consequence, chronic hunger arises. Both Eaton and Konner (1985) and Vargas (1990) attribute this dietary change with nutritional stress and a decrease in human stature (by roughly four inches). During this phase, there are larger social gaps in the society, and diet variations are subject to gender and social status (Gordon 1987). Famine fluctuates over time and space; however, some societies prove to be better able to reduce hunger, at least among upper-class citizens (Newman *et al.* 1990). This pattern of famine accompanied the development of agriculture (Popkin 2001a). Physical activity remains high and labour-intensive work is dominant, even though the types of activity vary, for example from homemaking to agricultural production and animal husbandry.

Pattern 3: Receding Famine

The diet associated with this pattern is characterised by an increase in the consumption of vegetables, fruits and animal protein. Although some earlier civilisations are partly successful in alleviating famine in the previous pattern, these changes become prevalent only in the past three centuries and are driven by rising income (Popkin 2002a, 2006a). Physical activity levels start to decline whilst inactivity gradually creeps into individual lives.

Pattern 4: Degenerative Disease

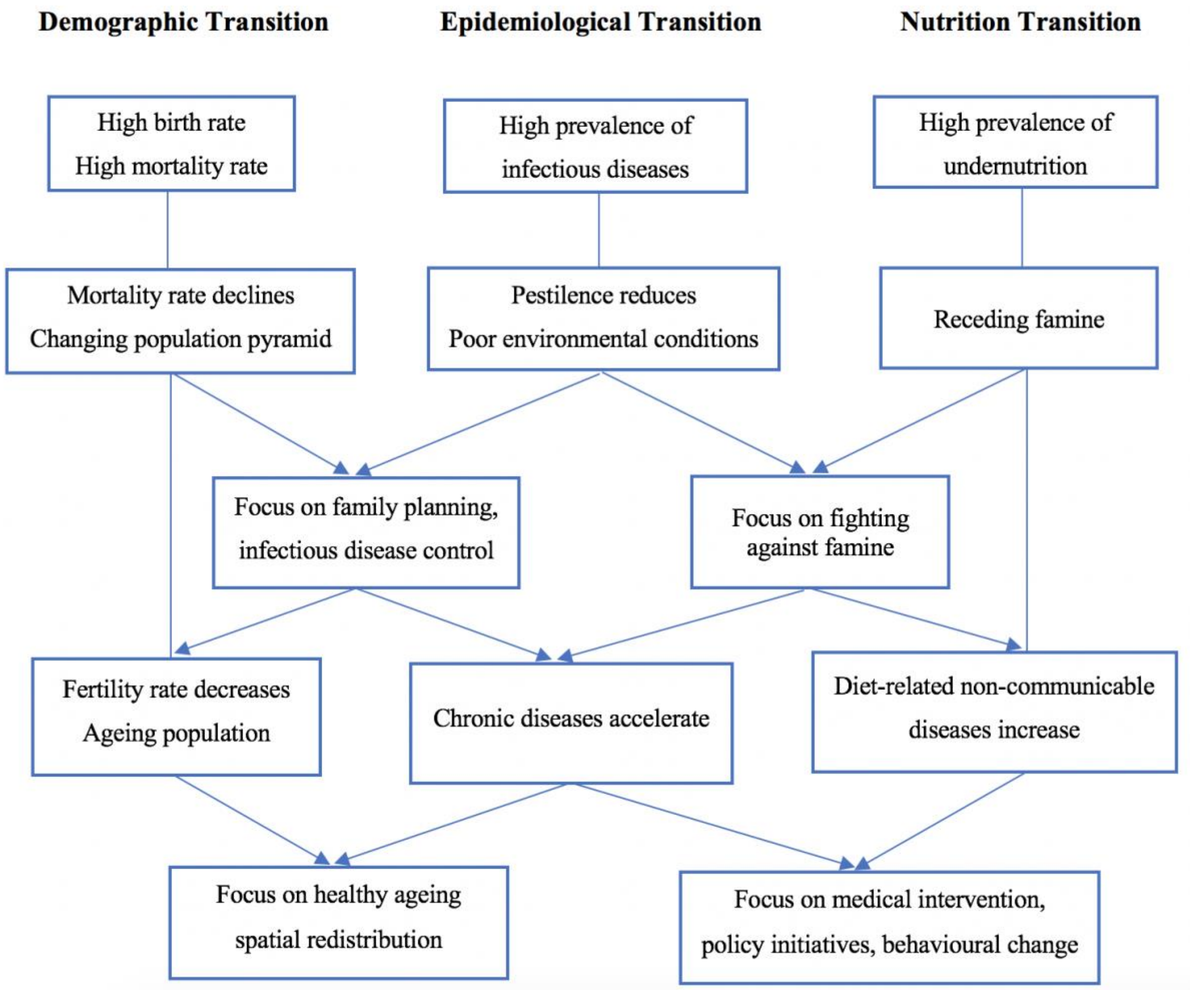
This pattern is prevailing in most developed countries and a rising portion of populations from low-income economies. The diet high in total fat, sugars, processed food but low in polyunsaturated fat and fibre leads to an accelerating rate of obesity and other degenerative diseases. Long-term negative impacts of the diet rich in fat, sugars but inadequate in fibre, have become perceptible over the past few decades (WHO 2003). Sedentary lifestyle is observed with the shift from labour-intensive work to office jobs in service sector. Most of middle- and low-income countries are experiencing a rapid move from Patter 3 to Pattern 4, and this shift is of so great concern that it is usually referred to as *nutrition transition* (Popkin 2006a).

Pattern 5: Behavioural Change

Of particular interest to researchers is the potential shift from Pattern 4 to Pattern 5, where individuals, being more concerned with their health, consciously move back to a healthier diet and increase the levels of physical activity. Is this anything more than a theoretical possibility? Swinburn *et al.* (2011) and Swinburn *et al.* (2019) claim that no country has been successful in reducing obesity and type 2 diabetes by the means of dietary change, and furthermore, over time no country has ever been able to reverse from a poor diet to a good one. Despite limited evidence, the healthier diets associated with Pattern 5 are at least plausible because of the escalating socio-economic consequences of obesity (Ezzati and Riboli 2013). Other influences (including urbanisation, economic growth, technological advance, and cultural factors) help to drive the transition (Popkin *et al.* 2012). These changes, introduced either by individuals with a high level of awareness or by governmental policies, signal a large-scale transition in dietary patterns which if it takes place will bolster a "successful ageing" process and improve healthy life expectancy (Manton and Soldo 1985; Crimmins *et al.* 1989).

Returning to the NTM, two questions need to be addressed are (i) whether the aforementioned transition applies to all countries over the world; and (ii) whether the transition speeds are uniform across countries. Perhaps, countries are all experiencing the dietary transition but are at different stages of the NTM, and as a result, global diets are changing at varying speeds. Indeed, Popkin (2002b) discovers that the nutrition transition that happened in the West for a couple of centuries took place in developing countries within just a few decades.

Many researchers subscribe to the belief that the nutrition transition is closely related to demographic and epidemiological transitions. In this regard, the latter processes occur simultaneously or precede the nutrition transition (Popkin and Gordon-Larsen 2004; McCracken and Phillips 2017). The demographic transition describes the shift from a pattern of high fertility and mortality (typical of a pre-industrial economic system) to one of low fertility and mortality (typical of modern industrialised countries) (Kirk 1996). Rural dwellers migrate to peri-urban/urban areas and the labour force is less dependent on manual jobs but demands highly skilled labour mostly in service-oriented sectors (Dyson 2011). Epidemiological transition, a phase of development ignited by medical innovations in disease mitigation, refers to the shift from a high prevalence of infectious diseases, usually associated with malnutrition, periodic famine and poor environmental sanitation, to one of a high prevalence of chronic and degenerative diseases associated with urban-industrial lifestyles (Omran 1971; Omran 2005). Figure 2.2 summarises the relationship of nutrition transition in accordance to demographic and epidemiological changes as populations move from one pattern to the next. Comprehensive details are referred to Appendix A. For example, a pattern of delayed degenerative diseases has been recognised recently along with a shift in age-specific mortality patterns and a consequent increase in life expectancy. As argued by Burggraf *et al.* (2015), epidemiological transition often differs across countries and regions since an ongoing nutrition transition is closely connected to a country's particular historic and cultural characteristics, which are clearly not transferrable to any other country.



(Source: Popkin (2002a))

Figure 2.2 Stages of health, nutritional and demographic change.

2.2.2 Evidence of the nutrition transition

While the diet consumed by populations in many countries after the World War II was often home-cooked and contained mostly basic food commodities such as vegetables and tubers, this diet has changed remarkably on the global scale over the past few decades. Blandford (1984) and Popkin (1993) are among the pioneer researchers to advocate a universal shift from diets rich in staples and vegetable-source foods to diets high in meat, saturated fat and sugars but low in fibre. Since the mid-1990s, there has been mounting evidence for such a transition in dietary composition across countries to converge on the so-called 'Western' diet characterised by high intakes of refined carbohydrates, sugars, fats, processed food and animal-source foods whilst intakes of fruits and vegetables remain inadequate (see, *among others,* Popkin 1993; Hawkes 2006; Kearney 2010; Popkin *et al.* 2012; Ronto *et al.* 2018; Popkin *et al.* 2020; Wells *et al.* 2020). It is pointed out by Khoury *et al.* (2014) that the ingredients on the plate of food around the world are becoming 36% more similar over the past 50 years.

The well-known nutrition transition portrayed theoretically by Popkin (1993) is established empirically in countries across Asia (Zhai *et al.* 2014; Harris *et al.* 2020), Africa (Bosu 2015; Harris *et al.* 2019; Rousham *et al.* 2020), and Latin America (Popkin and Reardon 2018; Tumas *et al.* 2019). Unlike the gradual transition that historically occurred in the United States and European countries, the transition speed has been more rapid in many lower-income countries (Popkin *et al.* 2012; Anand *et al.* 2015; Baker *et al.* 2020).

Coupled with the nutritional transition, there are also shifts in the way foods are sourced, prepared (cooking methods), and consumed (eating behaviours) due to changes in preference for convenience, price levels and culinary practices (Popkin 2015; Reardon *et al.* 2021). To illustrate, the traditional Chinese diet, which is considered relatively healthy with foods mainly cooked by steaming, baking or boiling, is increasingly being replaced by excessive fried foods (Wang *et al.* 2008b). Eating behaviour – a key determinant of dietary structure – has undergone a specific transition too. Snacking habit, which did not exist prior to the 1990s in China, has accelerated rapidly (Zhai *et al.* 2014). In addition, away-from-home eating has played an increasingly large role in diets of many developed as well as developing countries, even among low-income populations (Smith *et al.* 2014; Umberger *et al.* 2020).

A topical theme regarding the worldwide dietary shift is the increased consumption of processed foods, particularly in developing countries. The NOVA classification (Monteiro *et al.* 2019a) groups food products into four categories depending on the extent of processing: (i) unprocessed or minimally processed foods, (ii) processed culinary ingredients, (iii) processed foods, (iv) ultra-processed food products. The category with the highest degree of processing, ultra-processed foods, now accounts for a significant, and in some countries the largest, share of total energy intake, with estimates ranging from 30% in Mexico (Marrón-Ponce *et al.* 2018) to nearly 60% in the United States and the United Kingdom (Baraldi *et al.* 2018; Rauber *et al.* 2018). This is a great public health concern since these foods tend to be high in refined sugars, sodium, saturated fat and trans-fat (Poti *et al.* 2015), while the excessive consumption of which is closely linked to obesity and diet-related NCDs (Poti *et al.* 2017; Nardocci *et al.* 2019; Rauber *et al.* 2020). The added salt, sugars and fat, along with other sophisticated ingredients and manufacturing technologies, create 'hyper-palatable' products with higher durability (longer shelf life), better transportability and larger-scale capability for distribution (Monteiro *et al.* 2019b). These foods, propelled by mass marketing campaigns of transnational food corporations, are prepared, packaged and distributed in such a way to maximise product availability and desirability for customers across various socio-economic groups (Moss 2013; Development Initiatives 2020; OECD 2021). The expansion of processed foods has brought in several changes in terms of eating behaviours. Traditional foods are prepared and consumed at home whereas highly convenient processed foods are often designed to be prepared with minimal efforts and consumed anywhere at any time.

Despite the above universal trends across global diets, one should not lose sight of heterogeneity in dietary patterns. Firstly, both diets and dietary changes are greatly heterogeneous (Imamura *et al.* 2015) and there are many exceptions owing to the foods that drive these changes differing by region. For example, as opposed to the general pattern in the world diet, Africans are indeed consuming more starchy roots and pulses (Chauvin *et al.* 2012; Rawal and Navarro 2019) as well as fruits and vegetables (Holdsworth and Landais 2019). When it comes to animal-source foods and meat, the overall pattern for European diet has shown an inverted-U shape, and in fact, their consumption has followed a downward trend since the 1980s (Sans and Combris 2015).

Secondly, heterogeneity in dietary changes is observed across countries within a macro-continental area. Asia witnesses an increase in the consumption of animal fats and a gradual escalation of sugar and sweeteners (Kelly 2016). However, South Korea – a higher-income Asian country, has successfully retained many elements of a traditional diet (Kim *et al.* 2000). Low-fat, vegetable-centred traditional Korean dishes are still dominant, shaping a unique transition in this country. The higher intake of vegetables as compared to other countries in the region is attributed to the conservative attitude of Koreans towards food and the governmental effort in providing cooking classes as well as training housewives (Lee *et al.* 2002). Likewise evidence of some resilience of local food cultures is documented elsewhere, for instance in Colombia (Dufour *et al.* 2015), the Caribbean region (Paddock 2017), and Indonesia (Colozza and Avendano 2019). Thus, it is not without proof that dietary changes/preferences are context-specific rather than universal (Hawkes 2010).

Thirdly, heterogeneous changes can be found between urban and rural areas. The nutrition transition literature differentiates urban from rural consumers, and shows that the share of energy intake from animal-source foods, fat and sweeteners for urban residents is larger than rural peers even in the poorest areas of many low-income countries (Popkin 1999). For example, people living in urban areas of China consume more mutton and beef than in remote localities and the increase in the consumption of processed foods is more robust in urban areas (Zhou *et al.* 2014; Bai *et al.* 2020). Generally, urban dwellers consume a more diversified diet with a greater quantity of micronutrients and animal proteins than rural residents but with considerably higher intakes of refined carbohydrates, processed foods, saturated fat and lower intakes of fibre (Hawkes *et al.* 2017). Living in urban areas also raises the likelihood of consuming foods away-from-home (FAFH) (Zheng *et al.* 2019). Nonetheless, the dichotomy of urban versus rural diets does not adequately reflect the complexities of consumption, but the level of urban development also matters. d'Amour *et al.* (2020) observe some inter-urban variations: households in large metropolitan areas consume more processed foods and FAFH than households in smaller non-metropolitan urban areas.

Even though the nutrition transition signals converging trends in dietary patterns around the globe, the concept of convergence is rather nuanced. For instance, Baker and Friel (2014) find evidence for regional-level convergence in the consumption of processed foods but notable divergence among Asian countries. Obviously, the 'Westernisation' of diets is experienced differently with no single uniform transition across subpopulations. While the consumption of a small number of food commodities (often vegetable oils, meats, and processed foods) is converging worldwide, there is widening divergence in the local consumption owing to the variegated demographic, cultural, socio-economic factors as well as branding and marketing techniques at regional, national and local levels that shape dietary preferences and consumer demand (Hawkes 2006; Imamura *et al.* 2015; OECD/FAO 2020). Essentially, the assumption of identical taste across countries is over-rigid (Clements *et al.* 2006). Hence, it is too soon to jump into the conclusion of national food consumption patterns converging on one universal diet.

2.2.3 Implications of the nutrition transition

According to the NTM, the transition from Pattern 3 to Pattern 4 is generally associated with a decrease in the prevalence of diseases that are related to hunger or inaccessibility to adequate and nutritious foods. At the same time, the increasingly sedentary lifestyle together with dietary intakes of excessive and/or unhealthy calories coincides with a rise in NCDs.

Official statistics have confirmed both positive and negative health outcomes of the nutrition transition. On the one hand, food has been made more available worldwide and this has improved one pillar of food security – the production or supply of sufficient food to meet the aggregate energy requirements of the population (Kelly 2016; Gödecke *et al.* 2018). The global rate of hunger declined from 19% in 1990 to 11% in 2015 though the fall has been in reverse since then (FAO *et al.* 2020). The percentage of stunted children (under 5 years of age) dropped from 32.5% to 21.9% over the past two decades (UNICEF *et al.* 2019). In many countries, the intake of healthy food items (such as fruits, nuts, seeds and polyunsaturated fatty acid) is on the rise (Masters *et al.* 2016; Ronto *et al.* 2018). On the other hand, the increased consumption of highly calorific and energy-dense foods – the most obvious manifestation of the nutrition transition, could lead to a rise in obesity and diet-related diseases such as type 2 diabetes, coronary heart diseases and certain types of cancer (Webb and Block 2012; Harris *et al.* 2019; Harris *et al.* 2020). Poor diets are found to be associated with one in five deaths globally (Afshin *et al.* 2019).

Although the above double-edged effects of the nutrition transition are widely accepted, some researchers question the role played by the food system and argue that such impacts pertain in countries with more industrialised food system. Conversely, countries with a less modernised food system observe higher rates of stunting, underweight and micronutrient deficiencies but lower prevalence of overweight, obesity and other NCDs (IFPRI 2015).

Besides, the environmental impact of the nutrition transition has been the subject of growing empirical research, conceptual modelling, and commentaries (Aleksandrowicz *et al.* 2016; Vermeulen *et al.* 2019). Two major shifts concern the environmental sustainability include: (i) the increased consumption of animal-source products in low- and middle-income countries (LMICs) whilst the consumption level in high-income countries (HICs) has barely reduced, (ii) the trend towards higher consumption of ultra-processed foods (Popkin 2017). According to Bryngelsson *et al.* (2016), the target of limiting global warming to a two-degree increase seems impossible without further reduction in meat consumption. The environmental footprints of the trend towards higher consumption of animal-source foods are established in terms of greenhouse gas emissions, land use, water use, and biodiversity loss (see, *inter alia,* Zhai *et al.* 2014; Bodirsky *et al.* 2020; Heller *et al.* 2020). Moreover, these impacts are likely to expand as the global food demand rises rapidly (Tilman *et al.* 2011).

Due to these negative consequences on human health and the environment, the NTM predicts that demand for healthier food products will emerge in response to the needs of more affluent, health-conscious and aging consumers. Nonetheless, this is currently rare, even among HICs. Nonetheless, sales data appears to indicate such a demand in Malaysia – an upper-middle-income country with the highest rate of obesity and diabetes in Asia (Euromonitor International 2013). Centred at the core of Pattern 5 is a healthy and sustainable diet which balances the trade-off between health and sustainability. Despite some interesting proposed frameworks, researchers have not agreed on a single definition of a healthy and sustainable diet (Garnett *et al.* 2014; Bailey and Harper 2015; Steenson and Buttriss 2020). The existence of such a diet could indicate that Pattern 5 is not something fictional but at least plausible. Is the transition to Pattern 5 a reversal of the forces at play in Pattern 4 or does it require new factors to come into play? A recent LANCET report envisions that the journey to Pattern 5 is via the so-called "Great Food Transformation", which will not occur naturally but requires multi-sector and multi-level efforts to maintain healthy diets among global populations (Willett *et al.* 2019). It also forecasts that the transformation could be achieved by 2050 if actions are taken urgently without further delay.

When the gloomy picture of the nutrition transition and its undesirable consequences on human health and environment becomes clearer, actions from policymakers, food industry and societal organisations are expected. In reality, they have been slow to take up the cause (Hawkes *et al.* 2013; Swinburn *et al.* 2019). Historically, undernutrition has long been put on the discussion table and governments around the world have implemented agricultural policies to ensure that populations have access to sufficient foods. Even though greater progress is still needed in certain regions (for example in rural Africa), the focus has been gradually shifted as addressing poor dietary quality is increasingly being advocated in public health agendas (Hawkes *et al.* 2015; Hyseni *et al.* 2017; Popkin and Reardon 2018). Broadly, programmes to support healthier food choices are shaped around the "four-track" policy approach (Giner and Brooks 2019). The first track consists of demand-side interventions aiming to educate and provide information to change consumption patterns (for example via front-of-pack labelling). The second track consists of supply-side policies focusing on voluntary collaboration with the food industry. Possible initiatives include food reformulation and improvements in the food environment. The third track consists of firmer regulations when public-private incentives are misaligned (for example, rules on food advertising aimed at children). The fourth track consists of fiscal measures such as consumption taxes on less healthy products. Comprehensive reviews of government actions to improve diets can be found in studies by Gorski and Roberto (2015), Mozaffarian *et al.* (2018a), and Breda *et al.* (2020). In some way, "structural change in diets of billions of people is a primal force not easily reversed by governments" (Delgado 2003, p.3907S). In principle, choosing what to eat is a matter of personal choice and citizens experience it as a private freedom rather than a collective duty (National Food Strategy 2020). Governmental bodies are therefore disincentivised to tamper with their citizen's taste (Mason and Lang 2017).

However, it does not mean that government success is non-existent. A well-cited example in earlier studies is Korea where a combination of large-scale training of housewives in preparation of the traditional low-fat high-vegetable cuisine coupled with strong social marketing programmes has helped to retain the traditional Korean diets (Kim *et al.* 2000). Since 2002, Korean government has been promoting the increased consumption of whole grains, fruits and vegetables but the reduced consumption of fat. The so-called "Comprehensive Health Promotion Policy" includes various interventions such as the revision and dissemination of dietary guidelines, the enforcement of mandatory nutrition labelling on processed and packaged foods (Popkin and Ng 2007). In addition, initiatives to improve the dietary education of the broader public, by disseminating information on agriculture, nutrition, health, dietary habits, and on both processed and traditional foods, have been prioritised. For instance, in 2018, the country implemented programmes related to early childhood education, local foods, harvest and eco-friendly farming experience (Placzek 2021). Yet, in general there is a lack of monitoring diets in a holistic approach neutralising multifaceted dimensions related to food security, nutrition security, economic development, and the environment (Béné *et al.* 2019). At international level, the Sustainable Development Goals of the United Nations (SDGs) have not paid enough attention to nutrition and NCDs. Of relevance are only two out of 169 proposed targets of the SDGs: to reduce premature deaths from NCDs by a third, and to end malnutrition in all forms (Popkin 2017).

On the supply side, major beverage companies through affiliations such as the Healthy Weight Commitment Foundation, are avowedly changing the nutritional composition of food products via reformulations, introducing the new 'better-for-you' products and removing the less healthy ingredients (Healthy Weight Commitment Foundation 2020). It is ambiguous whether their proclaiming commitments are actually being implemented and if they are to what extent (Ng *et al.* 2014). In the meanwhile, no existing food composition database is able to keep up-to-date with the continuous reformulation of packaged foods (Ng and Dunford 2013; Traka *et al.* 2020). While multinational corporations may deliberately be engaged in such movement in some developed countries such as the US, the UK and Canada, similar efforts in developing economies are meagre (Kleiman *et al.* 2012;

Stuckler *et al.* 2012). More importantly, it remains open to question if these business endeavours will make any significant improvement on population health and for how long they will last. Nonetheless, what is plain is that multinational enterprises, in securing market share, are pandering to the lowest common denominator of public taste by highly palatable processed foods (Chandon and Wansink 2012; Moss 2013).

2.2.4 Criticism of the nutrition transition model

Despite being frequently cited in the literature, the NTM has been criticised for its oversimplification (Hawkes 2006; Lang and Rayner 2007). Pattern 1 of the NTM should be considered as an ideal vision of hunter-gatherer societies as it does not mention seasonal food shortages. In Pattern 2, less diverse diets are identified as the cause for shorter average statures; however, Mummert *et al.* (2011) propose that it might be due to infectious diseases as for a more densely packed population, diseases can hinder the body's ability to absorb nutrients from foods. Next, a key element of the nutrition transition depicted in Pattern 4 is the increased consumption of processed foods. Although a wealth of empirical evidence supports this trend (see, for example, Baker and Friel 2014; Law *et al.* 2019; Sievert *et al.* 2019), Walls *et al.* (2018) advocate that the currently used proxy for diet (which is often food availability) is not adequate to capture the nutrition transition because some highly processed foods are categorised as unprocessed foods. So far, Pattern 5 remains quite hypothetical due to the conspicuous absence of widespread behavioural or institutional changes in any country. Still, forward progress is being made. Consumption of salt, sugars, and fat from processed foods started to stagnate in few HICs (Baker and Friel 2014; Hawkes *et al.* 2017; Sievert *et al.* 2019), and taxes on sugar-sweetened beverages as well as regulations on food marketing have been introduced by some governments (Mozaffarian *et al.* 2018b; Cuadrado *et al.* 2020). Despite the reality that the target of reducing obesity is actually off course and no country has made progress in tackling obesity (Development Initiatives 2018), Popkin (2017, p.79) states that "no country has truly attempted to create programs to significantly reduce animal-food consumption".

Another criticism related to the NTM is that it assumes a uniform transition across countries from the earlier to the later stages of the model. Yet, empirical evidence reveals that LMICs are currently undergoing a different development pathway than what higher-income countries used to follow (Schmidhuber and Shetty 2005; Murray *et al.* 2015).

In addition, Lang and Rayner (2007) argue that the model places a stronger emphasis on rising income as the cause for dietary changes but underestimates the cultural context of the transition. The authors also warn that the nutrition transition should not be viewed as one transition but three overlapping transitions in diet, physical activities and culture. Of paramount importance is the conclusion drawn by Olivier *et al.* (2008) that economic integration leads to cultural divergence while social integration results in cultural convergence. The rationale is that trade in goods with a high cultural content preserves food cultures by allowing countries to produce local foods at the lowest costs. On the other hand, social globalisation, which represents social interactions between individuals of different countries through migration, tourism and communication technologies, produces shifts in food tastes. A question that should be raised is whether these aspects influence a country's level or speed in the transition process. While the contributing factors are manifold and complex, the NTM suggests that the pace of economic development is the driver for the speed of transition that countries are experiencing; yet, cultural factors determine the levels of transition.

To sum up, Section 2.2 has outlined the NTM, its main theoretical stages, and some evidence of dietary changes around the world. Due to the pressing public health issues of the nutrition-related NCDs that most developed countries and a rising proportion of developing countries are experiencing today, it is more urgent to find out solutions to help countries arrive at Pattern 5. The main challenge is that Pattern 5 so far remains rather hypothetical and no country seems to have reached that stage yet. If the underlying factors of the nutrition transition come into light, perhaps the government policy could be shaped to reverse the current trend and promote better diets (and hence better health). The next section dwells deeper into the commonly agreed drivers of the nutrition transition.

2.3 Globalisation and major underlying forces behind the nutrition transition

As the empirical evidence of the nutrition transition and its related health outcomes is piling up around the globe, researchers have expressed a growing interest in investigating the causes of the transition. Overall, dietary changes are driven by economic development as well as the availability and affordability of foods. Vegetable oils and fat became cheaper and their availability increased, leading to a higher consumption of fats among low-income countries (Drewnowski and Popkin 1997; Ng et al, 2008). In addition, government's subsidies distorted the price of grains, animal-source foods, fat, and sugars, making legumes, fruits and vegetables relatively more expensive to consume (Drewnowski and Darmon 2005; Popkin 2008; Drewnowski 2009; Lang and Heasman 2015). On the other hand, many authors connect economic globalisation, urbanisation, female labour force participation, and rising income to the modernisation of food chain (Popkin 1999; Datar *et al.* 2014; Dubois *et al.* 2014; Dave *et al.* 2016). The modern food system leads to significant structural changes in food consumption by lowering the price of processed foods as compared to traditional staples and fresh fruits and vegetables, raising the availability of less healthy foods (such as fast foods, soft drinks, and pre-cooked foods), promoting preference for the 'Western' foods through complicated marketing campaigns targeted at children, and raising individual's confidence in consuming foods from supermarkets via enhancing food safety standards (Traill 2017).

Indeed, the existing literature largely agrees on globalisation and other global forces such as rising income, urbanisation and increasing female employment as major drivers of the dietary shift.

Some detailed references include studies by Traill *et al.* (2014), Kearney (2010), and García-Dorado *et al.* (2019). Nonetheless, there is an ongoing debate on the disentangled effects of these factors as well as the transmission channel between these drivers and the nutrition-related outcomes. Some facets are relatively under-studied (for example the socio-cultural drivers of nutrition) and in general empirical results are hard to obtain due to the interconnectedness and overlaps of underlying mechanisms. The pathway illustrated in Figure 2.3 serves as the conceptual framework for the subsequent discussion.

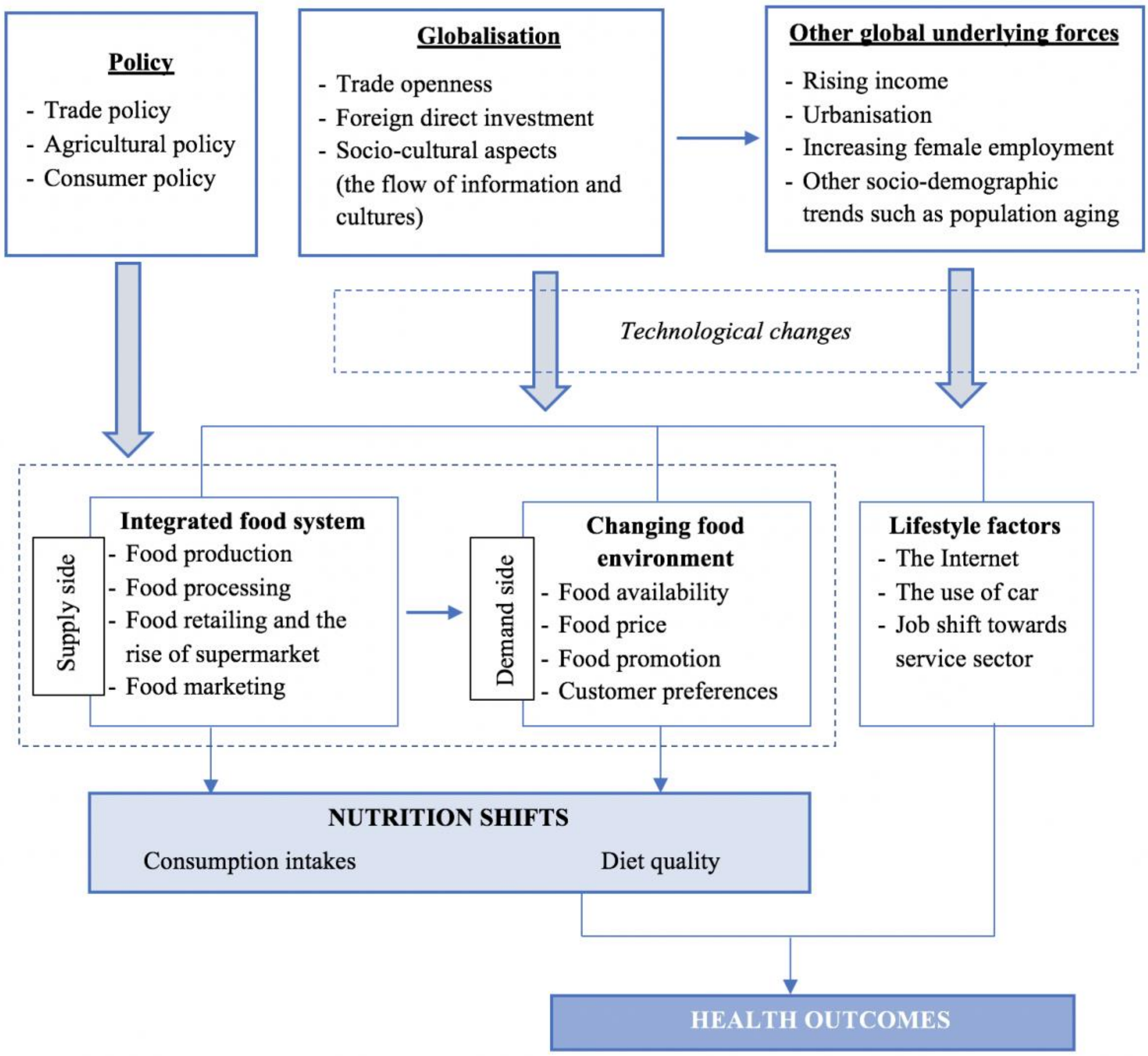


Figure 2.3 The pathway from globalisation and other underlying global forces to nutrition and health outcomes.

2.3.1 Globalisation

The proportion of international trade in global GDP has nearly doubled since 1970 and currently represents roughly 60% of the world GDP (World Bank 2020b). The international integration of markets for goods and services, with a focus on creating a freer flow of goods and capital, has transformed societies, altered economic and social relationships, and helped shape the world economy as it is today. This process, known as *globalisation*, has long been considered a key driver of the nutrition transition and its related health outcomes.

According to Jenkins (2004, p.1), globalisation is defined as "a process of greater integration within the world economy, through movements of goods and services, capital, technology and (to a lesser extent) labour, which leads increasingly to economic decisions being influenced by global conditions". This definition puts a zoom into the most easily recognised manifestations of globalisation: trade openness and economic liberalisation. Previous researchers link dietary changes with either a growing trend towards foreign direct investment – much of which went into food processing (Hawkes 2006; Gleeson and Labonté 2020; Reardon *et al.* 2021) or the expansion of free trade agreements, which enables the importation of cheaper but higher energy-dense foods from the industrialised world rather than from the local domestic production (Thow and Hawkes 2009; Thow *et al.* 2011; Porkka *et al.* 2013; Baker *et al.* 2016). This point of view emphasises how economic forces of globalisation have shaped the dietary transition; however, the globalisation of economies is not merely about integrating trade, investment and financial markets (Kónya and Ohashi 2007). Several aspects of globalisation, including both economic and non-economic dimensions, are to be explored below.

Trade openness

The second half of the 20th century witnessed a blossoming of various trade agreements and international treaties: General Agreement on Tariffs and Trade (GATT), Uruguay Round Agreement on Agriculture, World Trade Organisation (WTO) and WTO Regional Trade Agreements to name a few. These trade agreements aim to loosen government restrictions on international markets and international trades, for example, by reducing or removing tariffs on imports. Often, the lower tariffs on imports make imported produce more affordable and tend to cause countries to depend more on food imports but less on locally grown produce (Porkka *et al.* 2013; Sahal Estimé *et al.* 2014) – a phenomenon known as *dietary dependence* (Sievert *et al.* 2019). This could lead to greater availability and accessibility of imported foods that are usually less healthy (Hawkes 2006; Popkin 2006b; Baker and Friel 2014; Sievert *et al.* 2019). In Africa, for instance, trade in processed foods accounts for 30-60% of agricultural trade (Badiane *et al.* 2018). Previous studies show that having a trade agreement with the United States is associated with 63% higher soft drink consumption per capita (Stuckler *et al.* 2012) and 0.89 kilogram increase in sales of ultra-processed foods per capita per annum (Cowling *et al.* 2020). Overall, the existing literature has mostly examined individual food commodities rather than the whole diet, and less attention has been paid to the impact of trade on healthy food intake (Giles and Brennan 2014; Ravuvu *et al.* 2017).

The increased exchange of goods and services across countries is forming a global market for food products and this has impacted the availability as well as price of food commodities. From the supply side, global markets encourage specialisation in the production of export crops, which leads to greater agricultural outputs thanks to the economies of scale and eventually homogenised global food supplies are expected (Popkin 2006b; Khoury *et al.* 2014; Ogundari and Ito 2015). From the demand side, global food markets facilitate the import of goods across various ranges from healthy items such as fruits and vegetables (Huang 2004) to unhealthful processed and ultra-processed products (Stuckler *et al.* 2012; Schram *et al.* 2015). The effect of trade liberalisation on food price is more complicated. On the one hand, access to international food markets can reduce price volatility from local market's shocks; on the other hand, there is higher vulnerability to upheavals in the global demand. On average, the relative price of calorie-dense foods is lowered (Drewnowski *et al.* 2010).

Regarding health outcomes, trade openness is found to increase total calorie consumption (Zakaria 2014; Ogundari and Ito 2015), improve dietary diversity and quality (Dithmer and Abdulai 2017), reduce the odds of being underweight (Nandi *et al.* 2014), and alleviate food insecurity (Barlow *et al.* 2020). This could reflect the positive impact of trade policies aiming at improving food security and mitigating the influence of international price spikes on the domestic price of staple foods (Gilson and Fouad 2015). Trade liberalisation, by exposing consumers to less healthy foods at usually cheaper costs and reducing the government's policy space (in terms of freedom, scope and instruments) to introduce health-oriented schemes (Gleeson and Labonté 2020), is often blamed to be associated with the rising obesity prevalence and NCDs (Baker *et al.* 2014; Miljkovic *et al.* 2015; Barlow *et al.* 2017; Thow *et al.* 2017). However, no concrete conclusion has been reached so far (García-Dorado *et al.* 2019). This suggests that the availability and affordability of food products alone cannot ignite changes in lifestyles and consumption patterns associated to NCDs. Surprisingly, de Soysa and de Soysa (2017) document a negative link between trade openness and obesity rates among younger groups of the population. This is attributed to the added bonus of globalisation – higher rates of return to labour, which provide strong incentives to promote children's health, contributing to healthier diets and thus lowering the rates of obesity.

Foreign direct investment

By definition, foreign direct investment (FDI) refers to a type of cross-border investment in which an investor residing in a country buys or establishes a lasting interest in controlling over assets in another country (OECD iLibrary 2020). This distinguishes FDI with portfolio investment in which the investor does not have a direct control over the business's daily operations (Everett 2006).

It is a common belief that the agribusiness-related FDI has directed the diffusion of transnational food corporations (TFCs) into the market for processed and ultra-processed foods, particularly in lower-income countries (Blouin *et al.* 2009; Thow 2009; Popkin *et al.* 2012; Gleeson and Labonté 2020; Reardon *et al.* 2021). In turn, the food processing industry attracts an increased investment in food marketing and food advertising (Blouin *et al.* 2009; OECD 2021). Trade liberalisation with increased exports of domestic goods and imports of foreign products and the opening of national markets to foreign investment, supported by WTO, promotes the growth and power of transnational food companies (Hawkes 2005; García-Dorado *et al.* 2019). These large corporations have the ability to lower the price of fast foods or soft drinks to make them more affordable and reach a larger segment of the population. A study by Schram *et al.* (2015) uncovers a surge in the sales of sugar and sweetened beverages after restrictions to FDI were removed in Vietnam. A similar research design is applied in the context of Peru by Baker *et al.* (2016) and the authors document a diversification in the sales of soft drinks. The sales of bottled water, sports and energy drinks increase, but those of carbonated drinks stagnate. These inconsistent results could imply the role of branding and marketing in misleading customers' demand for sports drinks which are high in sugars but are often marketed as healthy products (Alsunni 2015).

Additionally, the expansion of the market for processed foods has fuelled the rise of supermarkets in LMICs. According to Reardon *et al.* (2012), the history of *'supermarket revolution'* spans over three waves. The first wave happened in Latin America, Central Europe and South Africa in early 1990s; the second wave reached Southeast Asia, Central America, and Mexico in the mid- to late 1990s; and the third wave occurred in China, Vietnam, India and Russia in the 2000s. The potential fourth wave of the diffusion of modern food retails just kicks off in many African countries (except South Africa). Conventionally considered as a shopping destination for the rich, supermarkets are now penetrating into many of the poorer areas on the planet (Reardon *et al.* 2003; Béné *et al.* 2019). It is shown that the development of supermarkets in Latin America in a single decade equates to that in the US in five decades (Reardon and Berdegué 2002). Grocery sales from supermarkets have climbed up worldwide, accounting for over half of total food sales from modern grocery retailers in 2019 (Euromonitor International 2020). The spread of supermarkets has made a more diverse diet available to a larger proportion of the population. Indeed, supermarkets offer a much wider variety of processed foods at lower costs thanks to the economies of scale in procurement (Hawkes 2008; Baker *et al.* 2020). From the stance of large food retailers, processed foods with a long shelf life are preferred to fresh seasonal food produce (Thow 2009). As such, the penetration of supermarkets in LMICs is oft-stated to create more accessibility to snacks and sugar-sweetened beverages (SSBs) compared with high-quality fruits and vegetables (Odunitan-Wayas *et al.* 2020), thereby contributing to the nutrition transition (Bailey and Harper 2015). Nevertheless, there is no conclusive evidence of a significant and direct relationship between modern market food expenditure shares and household dietary diversity as well as dietary quality (Rupa *et al.* 2019). Similarly, the link between supermarket food purchases and the rising obesity rates is not empirically established (Ford and Dzewaltowski 2008; Debela *et al.* 2020).

Overall, FDI has been found to be associated with the increased prevalence of obesity and overweight in LMICs (Nandi *et al.* 2014; Miljkovic *et al.* 2015; Schram *et al.* 2015; Baker *et al.* 2016). This suggests that FDI is the main vehicle for food system integration by allowing the greater market penetration of TFCs through both vertical and horizontal integration, transformation of distribution and retail segments, advertisement and adaptation to consumer tastes (Herforth *et al.* 2019). On the other hand, Sudharsanan *et al.* (2015) discover insignificant influence of FDI on the prevalence of diabetes after controlling for population ageing. Both Neuman *et al.* (2014) and de Soysa and de Soysa (2017) find no relationship between FDI and overweight as well as obesity rates. Other scholars show that the impacts of FDI on nutritional outcomes vary by sector (Mihalache-O'Keef and Li 2011; Djokoto 2012). Both studies agree that growing FDI in primary sector tends to threaten food security in LMICs because of resource exploitation, labour market effects and negative environmental and demographic externalities. The former study further scrutinises FDI in manufacturing sector and documents a link with technological and human capital spill-overs, increased wages, and improved nutritional outcomes (Mihalache-O'Keef and Li 2011).

Socio-cultural components of globalisation

Keohane and Nye (2000) conceptualise three dimensions of globalisation: (i) economic, (ii) political, and (iii) social. Economic globalisation was discussed previously. Political factors, which are related to the formation of regional trade blocks or participation in various international treaties, may act as a precursor to greater economic integration. Social globalisation, involving the cross-border movement of information and cultures, has become a trending topic in the literature.

The increased exchange of information (and people) has transformed cultural norms, social relationships and consumption patterns to a great extent. With the aid of communication technologies and infrastructure, information and ideas are being spread globally in a matter of seconds. Constantly influenced by mass media and marketing campaigns, individuals are exposed to the perception of foreign lifestyles as well as foreign diets. As a consequence, the social globalisation allows a smoother integration of TFCs, giving rise to demand for new products (McChesney and Schiller 2003). Here, the role of food marketing is crucial as it was pointed out that advertising, for instance via TV, of foods that are high in sugars, fat, and salt strongly influences purchase decisions of children and their parents (Hawkes 2007; Smith *et al.* 2019; Kontsevaya *et al.* 2020; Umberger *et al.* 2020). Globalisation is a contributing factor, but not the only one. Along with urbanisation, social globalisation is associated with more abundant supply and consumption of cheaper but higher-calorie food products (Drewnowski and Popkin 1997; Popkin and Gordon-Larsen 2004).

In an attempt to demystify the potentially important role played by social globalisation in nutrition and health outcomes, several researchers compare the effects of socio-cultural aspects of globalisation with the economic aspects (Goryakin *et al.* 2015; Miljkovic *et al.* 2015; Costa-Font and Mas 2016; de Soysa and de Soysa 2017; Oberlander *et al.* 2017). Nonetheless, findings are mixed. Both Goryakin *et al.* (2015) and Costa-Font and Mas (2016) suggest that the positive association of globalisation as a whole and obesity rates is mostly attributed to the social component of globalisation. The empirical evidence provided by Miljkovic *et al.* (2015) similarly indicates a link between social globalisation and higher obesity prevalence. To map the pathway between globalisation and the nutrition transition, Oberlander *et al.* (2017) show that while economic globalisation is related to higher BMI and prevalence of diabetes, only social globalisation is associated with the increased supplies of animal-source protein and sugars. The authors further argue that the transmission is channelled by the improved flow of information via television and the Internet. On the contrary, de Soysa and de Soysa (2017) do not document any significant relationship between social globalisation and obesity prevalence after controlling for the economic globalisation as well as country and time fixed-effects. These nuanced results emphasise the interrelation of relevant factors and the complication of the mechanism involved.

Interaction with other drivers of nutrition transition

The effect of trade liberalisation and market integration on nutrition outcomes is not merely transmitted through the food sector, but the globalisation process has profoundly transformed various societal dimensions which could indirectly impact the nutritional changes and nutrition-related health outcomes. Globalisation has been linked with income growth (Berg and Krueger 2003; Dreher 2006), which is likely to generate demand for processed foods. In other studies, globalisation is associated with the worsening conditions in labour market, and a switch from labour-intensive to sedentary and knowledge-based jobs (Huneault *et al.* 2011). In contrast, integration in global economy is found to improve the returns to labour, encouraging further investment in health and leading to healthier diets and better health outcomes (de Soysa and de Soysa 2017).

Other researchers add more nuance to this debate by citing the role of technological development. According to these authors, globalisation is both a product and a driver of technological changes (Popkin 2001b; Popkin *et al.* 2012), and both of them may collectively contribute to the increased obesity prevalence. These two factors are usually associated with urbanisation (living in cities offers a greater choice of foods at a lower price), increasing use of cars and mechanical aids (resulting in a decline in physical activity) (Hawkes 2006; Belasco 2008). In sum, individuals are exposed to a lower cost of calories consumed but a higher opportunity cost of calories expended. Despite the unagreed mechanism, the aforementioned structural changes could be correlated with other changes in lifestyle, social relationships and characteristics of labour market, which could ignite changes in dietary patterns. Thus, these interrelationships should be taken into account when assessing the link between globalisation and nutrition transition.

2.3.2 Other global underlying forces

Besides globalisation, any other factor that influences food production, food distribution, and food retailing (all of which comprise the so-called ‘food system’) could potentially be a determinant of the dietary shift. Environmental changes, for example the melting of Himalayan glaciers or the collapse of the Atlantic Meridional Overturning Circulation, could impact food production (Benton *et al.* 2017). Recent study shows that anthropogenic climate change has slowed global agricultural productivity growth by roughly 21% since 1961 (Ortiz-Bobea *et al.* 2021). However, this effect is conditioned by the degree of market integration and the dependency of local food market on agriculture. For instance, food market in rich countries is more integrated with international markets and therefore diversity in agricultural production is not as a strong driver of food supply as in poorer countries (Remans *et al.* 2014). The modernisation of food supply chain, from traditional mode (local, disconnected, labour-intensive) to modern (long in distance, highly concentrated, vertically integrated, capital-intensive) has been in sync with a rise in dietary diversity and greater consumption of processed foods (Herforth *et al.* 2019). In addition to the abovementioned factors, other often-cited drivers of the nutrition transition are discussed as follows.

Agricultural and food policies

As humans have been constantly fighting starvation and hunger, global agricultural policies have long focused on producing cheaper grains to meet the demand of a growing population (Fan 2020). This perspective resonates with the 'productionist' paradigm in the mid-to-late 20th century according to which problems can be resolved by producing more foods (mainly staple grains, oils, sugars and animal products) through more sophisticated methods (Foley *et al.* 2011; DeFries *et al.* 2015; Bahadur *et al.* 2018). As a consequence, there has been a substantial increase in the global production and consumption of certain foods (cereals, starchy root, meat, dairy products, oilseeds, and sugars) over the past half century (Khoury *et al.* 2014; Development Initiatives 2020). A diet comprised only of these foods would lead to negative health outcomes in the long term. Low consumption of fruits, vegetables, whole grain fibre, nuts and seeds but high consumption of sodium, processed meat and sugars will most likely exacerbate the rise of obesity and diet-related NCDs that are already accompanying undernutrition and micronutrient deficiencies, and will contribute to the double burden of malnutrition in the same country, same household or individual in the same life course (Popkin and Gordon-Larsen 2004; IFPRI 2016; Hawkes *et al.* 2020).

On the other hand, the shift in agricultural policies from state intervention which was dominant in the 1930s/1970s to market liberalisation and globalisation in the 1980s/2010s has altered the food environment by which customers make food choices in the form of food availability, food affordability and food acceptability. Hawkes *et al.* (2012) propose three hypotheses on the nexus between agricultural policies and consumer diets. First, the move away from state intervention to market liberalisation creates incentives for food producers to be more specialised in adopting certain crops, making certain types of foods more readily available in the food markets. Second, the changing paradigm of agricultural policies affects the farmgate prices in both directions, allowing food processing companies to substitute with lower priced ingredients and thus having implications for the nutritional quality of food products. Third, advancement in food research and innovation has resulted in a vast market characterised by differentiated food products with 'added value' which could satisfy varying individualised preferences, having the potential to influence customer's acceptability of the variety and quantify of food products.

Rising income

Theoretically, the relationship between income and diet is established by *Bennet's law*, which states that the share of calories from starchy staples declines as household income rises (Timmer *et al.* 1983). Empirically, Hertel *et al.* (1998) and Law *et al.* (2018) suggest that income is a significant determinant of the level of food consumption. However, the authors express their cautiousness towards reverse causation in the sense that households with better food intakes are likely to have higher work productivity and hence higher income earnings (commonly known as *efficiency wage hypotheses*).

Higher incomes also mean that people can afford a wider range of foods. For instance, rising income per capita in the 20th century allowed Western European countries to replace bread and potatoes in their diets with meat, dairy products, sugar, and oils (Grigg 1995, 1999), while higher economic resources in China in the mid-to-late 1990s led to a greater diversity of food categories (Delgado 2003; Garnett and Wilkes 2014). Evidence shows that countries with higher GDP per capita have higher total sugar intakes and lower-income countries are associated with poorer diets (Rippin *et al.* 2020). A study by Regmi and Meade (2013) reveals that an increasing demand for staple foods such as cereals continues up to a threshold of income after which a further rise in income leads to a fall in demand for cereals. Conversely, demand for animal-source protein continues to rise as income increases. In addition, increasing household incomes are likely to generate demand for animal-source foods (Cornelsen *et al.* 2016) and processed foods (Moodie *et al.* 2013; Law *et al.* 2019; Milford *et al.* 2019) since individuals are now prepared to spend some of the extra cash on other aspects of food including convenience, that were once considered luxuries (National Food Strategy 2020). This shift is particularly applicable for low- and middle-income households while high-income households increase their demand for luxury goods (including health) and as a result consumption of meat and fats declines. These patterns help to explain why obesity is more prevalent in the wealthier segment of the population in low- and middle-income countries but in low-income groups of rich countries (Cirera and Masset 2010).

Urbanisation

Worldwide, the percentage of urban population has increased from 34% in 1960 to 55% in 2017 (32% in low-income countries, 52% in middle-income countries and 82% in high-income countries) (World Bank 2020c). Urbanisation – the shift from a population that is dispersed across small rural areas in which agriculture is the dominant economic activity towards one where the population is concentrated in larger, dense urban settlements (National Research Council 2003) – is increasingly put forward as a crucial determinant of changing dietary patterns. Nonetheless, the empirical effects of urbanisation and the underlying mechanisms through which these take place remain ambiguous (d'Amour *et al.* 2020).

At cross-country level, urbanisation is found to be associated with decreased consumption of coarse grains but increased consumption of wheat and animal-source foods (Delgado 2003; Reardon *et al.* 2014) as well as sweeteners and fats (Drewnowski and Popkin 1997; Popkin 1999). Studies at within-country level report a link between urbanisation and changes in cereal consumption (Delgado 2003) and animal-source foods (Rae 1998). In some segments of the population, urbanisation is found to result in higher consumption of animal-source foods, fat, sugars, salt, and processed foods, engendering the increased prevalence of overweight and chronic diseases (Harris *et al.* 2019). However, evidence from rural-urban migration in Tanzania does not support the association between urban residency and increased consumption of animal-source foods (Cockx *et al.* 2019). These nuances demonstrate that dietary changes are not universal, but rather context-specific, meaning that many modifications in dietary preference may not be attributable to urban living alone. In Indonesia, Colozza and Avendano (2019) disclose some increases in acquisitions of animal-source and ready-made foods over the past two decades; yet, most changes occurred in parallel across rural and urban areas. Therefore, moving to urban areas is not necessarily associated with changes in food expenditures.

The pathways through which urbanisation affects diets are numerous and complicated. First, urban residence is characterised by different food supply environments, influencing the availability and affordability of food items. There are generally more choices regrading eating out or buying processed/pre-cooked meals thanks to the physical proximity of supermarkets, minimarkets and fast food outlets and the types of foods they offer (Hawkes 2008; Cockx *et al.* 2018). Locally grown produces are not easily acquired and more expensive if cities are located far from production sites (Smith 2013). On the other hand, rural markets are less integrated in national and international markets (Osborne 2005; Moser *et al.* 2009). In addition, many researchers subscribe to the idea that urbanisation characterises different socio-cultural food environment. In urban cities, higher rates of participation in labour markets shift dietary preferences towards convenience over quality (Pingali 2007; Reardon *et al.* 2021). Greater exposure to global eating patterns, mass media or improved access to nutrition knowledge all contribute to a distinctive set of food preferences and eating habits in urban areas (Huang and Bouis 2001; Regmi and Dyck 2001; Dapi *et al.* 2007; Kearney 2010; Bosu 2015). Besides, there is a literature strand advocating economic development rather than urbanisation as a fundamental driver for changes in dietary patterns (Regmi and Dyck 2001; Kearney 2010; Stage *et al.* 2010; Pandey *et al.* 2020). These authors argue that "the difference between urban and rural households' patterns of food consumption is not caused by urbanisation and cultural change but income differences".

Increasing female participation in labour market

Several items in the United Nations Millennium Development Goals (United Nations 2020) are centred around the intention to raise global awareness and promote women's employment as the means to improve population health and mitigate poverty. Since the United Nations Millennium Declaration was signed in September 2000, there has been a significant shift from part-time to full-time employment of women in LMICs (Lopez-Arana *et al.* 2014).

The additional economic earnings from maternal employment allow higher household food expenditures and the purchase of energy-dense foods, all of which could result in an overconsumption of energy (Oddo *et al.* 2017). Traditionally, women played a crucial role in meal preparation and regular shopping for fresh foods (Welch *et al.* 2009). Recent trends in workforce and changing family structures, including the growing economic participation of women, are likely to intensify time pressures and boost demand for convenience foods (Gehlhar and Regmi 2005; Popkin 2006b; Datar *et al.* 2014). Improved female employment raises the opportunity costs of food preparation, decreases the time women spend on preparing and cooking foods (Popkin and Reardon 2018), and generate preferences for processed foods that are 'ready-to-eat' or 'ready-to-heat' (Huang and David 1993; Huang and Bouis 2001; Bourne *et al.* 2002; Mutlu and Gracia 2006; Anand 2011; Reardon *et al.* 2021). Studies in the USA, China and elsewhere document a significant reduction in cooking time but an increase in leisure time (Popkin *et al.* 2012; Wang *et al.* 2012; Smith *et al.* 2014). Nonetheless, empirical evidence points to an insignificant association of maternal employment with children's diet (Nie and Sousa-Poza 2014) as well as their nutritional status (Eshete *et al.* 2017).

To sum up Section 2.3, globalisation plays an important role in driving changes in nutritional status of populations in countries of different development levels. The empirical literature provides a nuanced view of the impact, indicating that different processes and sub-components bring in different effects. Trade openness contributes to shifts in dietary patterns by broadening dietary diversity and increasing the availability of cheap calories and fat, and hence reduces undernutrition. However, trade openness alone is not adequate to explain the increase in obesity and overweight prevalence. There seems to be more associated with FDI and the global flow of information in LMICs, owing to food marketing and advertising. Socio-cultural aspects have a profound influence on dietary patterns especially regarding the consumption of calories and fats. In addition to globalisation, various development factors are proposed as drivers for the global dietary shift, including agricultural policies, rising income, urbanisation, and female employment. Nevertheless, the empirical evidence and the underlying mechanisms through which these take place remain controversial. The existence of impacts at varying degrees across sub-populations where the most vulnerable segments tend to be affected disproportionately highlights the need to reduce inequalities in access to food and to develop targeted policies which would protect the groups most vulnerable to adverse impacts of these global forces.

2.4 The literature on overweight and obesity

The last three decades have witnessed significant demographic, economic development, environmental and cultural changes. Obesity has risen at an alarming speed all over the world especially in developing countries where, despite some remarkable progress in improving the nutritional status, the prevalence of undernutrition rermains high and has increased recently (FAO 2021b). In the food economics literature, such a phenomenon is known as the *double burden of malnutrition* – the coexistence of undernutrition and overweight/obesity in the same countries, communities and households (WHO 2017b). Previous studies have analysed the double burden in different geographical regions (see, *for example*, Hanandita and Tampubolon 2015; Lowe *et al.* 2021; Nguyen *et al.* 2021). But is it a universal phenomenon or does it only inflict certain groups of population in both developing and developed countries? Shrimpton and Rokx (2012) argue that the double burden of malnutrition affects all countries, rich and poor, and is a particular concern in countries with high stunting rates. However, more evidence is needed to better understand the socioeconomic drivers of double burden in different settings as this would assist the development of food and nutrition policies.

Up til now the dual burden has been primarily described as an urban problem, that is associated with sedentary lifestyles, 'Westernisation' of diets, and 'obesogenic' food environments (Doak *et al.* 2005; Jehn and Brewis 2009; Popkin *et al.* 2012; Kimani-Murage *et al.* 2015). Nevertheless, an increasing wealth of evidence has revealed that the dual burden is also found in the rural settings where lifestyles and food environments are still more traditional (Fongar *et al.* 2019). The authors analyse the prevalence of double burden in rural Kenya and document an individual-level prevalence of 19%. This relatively high rate is likely to reflect the low dietary quality. It turns out that typical diets in the region are sufficient in terms of calories but insufficient in terms of micronutrients. Of diets in rural Kenya, most calories are derived from unprocessed foods (starchy staple foods) while quantities of fruits, vegetables, and animal products remain little. Importantly, better educated households are somewhat less affected by the double burden.

Digging into the cause of the double burden of undernutrition and overnutrition, it is a common belief that such a paradox is the result of a rapid nutrition transition (WHO 2017b). The progressive 'Westernisation' of eating patterns, represented by a strong increase in the consumption of fats, animal sourced products and processed foods, combined with an increasingly sedentary lifestyle is oft-quoted to encourage the increase in overweight and obesity.

Evidence from empirical studies shows that across countries the prevalence of undernutrition is strongly correlated with the nation's wealth (Swinburn *et al.* 2019; Nugent *et al.* 2020; Popkin *et al.* 2020). As a result, stunting and undernutrition tend to reduce as national income improves. Unfortunately, economic development is associated with rising prevalence of obesity and this has become a great concern to governments of emerging countries due to the burgeoning health care costs (Prentice 2018).

While there is an extensive literature characterising overweight and obesity in developed countries, the literature on developing countries is relatively limited besides studies that document the rising trend. In the previous literature, socioeconomic status (such as income levels and educational attainment) has been proved to be a strong predictor of obesity (Sobal 1991; Sundquist and Johansson 1998; Pickett *et al.* 2005; Costa-Font and Gil 2008; Drewnowski 2009; Font *et al.* 2010; Ralston *et al.* 2018; Mathieu-Bolh and Wendner 2020). On average, the likelihood to be obese is highest in the lower socioeconomic strata of society in rich countries and among higher income groups in low- and middle-income countries (Shrimpton and Rokx 2012). Higher economic development tends to shift obesity burden from the rich to the poor (Dogbe 2021). Across countries, weight gain tends to occur to same sort of individuals except highly educated and poor people.

While the weight of an individual can be simply a conceptualised balance between the number of calories consumed and the number of calories expended through physical activity, the determinants of weight gain are more complicated. Genetics, globalisation, economic development, technological advancement, increasing female labour market participation, and urbanisation have been proposed as factors contributing to rising body mass index (BMI) (Lopez 2004; Philipson and Posner 2008; Lakdawalla and Philipson 2009; Thow and Hawkes 2009; Welch *et al.* 2009; Nuttall 2015; Braha *et al.* 2017).

A growing body of literature has drawn attention to the role played by globalisation process, especially the ways that changes in global socio-cultural environments have led to an 'obesogenic' environment (Costa-Font *et al.* 2010; Ulijaszek and Schwekendiek 2013). In this regard, economic globalisation, particularly trade liberalisation, facilitates the diffusion of 'obesogenic' products in low- and middle-income countries (Drewnowski *et al.* 2010; Vogli *et al.* 2014). Cultural globalisation with the increasing exposure to ideas and images from the West may encourage the consumption of fast foods to sound more "modern" (Oberlander *et al.* 2017). This so-called *dependency/world systems theory* (Fox *et al.* 2019) places a stronger emphasis on underlying factors external to the country – for example, international trade regimes that allow the entry of transnational food corporations into emerging economies and thus help to promote the increased consumption of unhealthy foods and ideational lifestyle diffusion. Generally speaking, if this dependency theory holds, a greater integration into the global economy and Western culture should result in higher BMI.

Another explanation for the rising obesity rates is that in the process of modernisation countries are experiencing domestic nutrition transition (Hawkes 2006). Even without the increased exposure to global markets or images of the 'Western' diet, middle classes in countries may increase demand for a richer diet, processed foods and unhealthful local foods. Modernisation is closely associated with a set of development variables including technological advancement, urbanisation and women's empowerment. Development inexorably leads to health transitions, including the rise of unhealthy lifestyles when disposable income increases (Fox *et al.* 2019). In other words, the *modernisation theory* predicts a relationship between the growth of GDP per capita and BMI. Following this literature stream, there is a debate over whether the process of modernisation is linear or inverted-U shaped (Mathieu-Bolh and Wendner 2020). In the latter case, the overall burden of disease may increase before declining owing to a double-disease burden. It is possible that the relationship between economic development and weight gain is not linear and that at low levels of development BMI increases rapidly but decreases at higher levels of development.

2.5 Convergence theories

2.5.1 Beta versus sigma convergence

The term 'convergence' implies a dynamic mechanism moving from different initial levels towards some common outcome. Economists started to turn their attention to this concept after the historical phenomenon in the world economy in the late 1980s/early 1990s. For almost two centuries, a group of industrialised countries was the growth generator of the world economy. From 1990, a group of developing countries began to grow more rapidly than developed countries and the remarkable growth rate of developing countries has remained for the past two decades (Martin 2019). If this pattern continues, developing countries will 'catch up' with their developed peers in the future.

Several reasons have been proposed to explain why countries that were once behind the leading economies will grow more rapidly than the leaders. According to Abramovitz (1986), new technologies are adopted, and with the latest technologies, countries are able to move from the existing production possibility frontier to the new curve. Poorer countries can replicate the production methods, technologies, and institutions of the richer countries. Similarly, Baumol (1986) argues that technology is a public good and its diffusion leads to convergence. If this model holds, the *unconditional income convergence* would occur, and the growth rates would be highest among the lowest income countries. However, this convergence model proves to be far from reality during the first two centuries of the Industrial Revolution. While per the capita income increased in both industrialised and developing economies, the growth rate was substantially higher for the former group of countries (Martin 2019).

The concept of convergence was initially defined in development economics and has been examined extensively in regional studies in the context of income. Two most pertinent convergence theories include beta convergence and sigma convergence.

Beta convergence

Beta convergence has gained its popularity among economists since the seminal paper by Barro and Sala-i-Martin (1992). In the simplest terms, beta convergence refers to the process in which poor countries grow faster than rich countries at an earlier stage before converging to grow at similar rates in the long term. At the end of this transitional dynamism, the initially poorer economies would reach the per capita income level of richer economies, a concept usually referred to as the *catching-up process*. This concept is directly related to the neo-classical Solow's growth theory (Solow 1956), according to which the source of convergence is the assumed diminishing returns to capital. The growth process should enable economies to reach a long-run steady-state level characterised by a rate of growth which depends only on exogeneous factors (such as the rate of technological progress and labour force growth rate).

When all economies are assumed to converge towards the same steady-state level (in terms of GDP per capita or growth rate) regardless of their initial level of output per capita, beta convergence is said to be 'absolute' (Barro and Sala-i-Martin 1992). Absolute convergence relies on the critical assumption that structural parameters (such as saving rate, population growth, capital depreciation, and technology level) are homogeneous across countries. However, the steady-state level may depend on features specific to each country so that convergence still takes place but not necessarily at the same long-run levels for all economies. In this case, beta convergence is said to be 'conditional'. Countries that have similar structural conditions (for instance technologies, human capita, population growth rates, legal institution) tend to converge on their own steady state (Mankiw *et al.* 1992).

In addition to absolute and conditional convergence, the extant literature documents another convergence hypothesis, 'convergence club'. Countries sharing similar structural characteristics and initial factors (for example GDP per capita, human capita, preferences, public infrastructure) converge with one another in the long-term but need not converge on the same equilibrium path (Galor 1996). Instead, countries belonging to the same club move toward a club-specific steady-state equilibrium and there is no convergence across different sets of equilibria.

To sum up, absolute convergence yields one equilibrium for all countries, whereas countries approach their own equilibrium level in conditional convergence, and convergence club exhibits multiple equilibria.

In order to examine the existence of absolute beta convergence, Barro and Sala-i-Martin (1992) use the data on real per capita income, y_{it}, for a set of economies $i = 1, \dots, N$, and regress the average growth rate over the interval between any two points in time, t_0 and $t_0 + T$, on the initial level of income. Specifically, the following nonlinear equation is estimated using non-linear least square (NLS):

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B - \left(\frac{1-e^{-\beta T}}{T}\right)\log\left(y_{i,t_0}\right) + \varepsilon_{i,t_0,t_0+T} \quad (2.1a)$$

Or:
$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \left(\frac{e^{-\beta T}-1}{T}\right)\log\left(y_{i,t_0}\right) + \varepsilon_{i,t_0,t_0+T} \quad (2.1b)$$

where $\varepsilon_{i,t_0,t_0+T}$ represents an average of the error terms, ε_i, between times t_0 and $t_0 + T$. The constant term is defined as: $B = x + \left[\frac{1-e^{-\beta T}}{T}\right][\log(\hat{y}^*) + xt_0]$, which is independent of *i* under the assumptions that $\hat{y}_i^* = \hat{y}^*$ and $x_i = x$. In other words, the steady-state value, $\hat{y}_i^*$, and the rate of technological progress, x_i, are assumed to be the same across economies. The left-hand side in equations (2.1a) and

(2.1b) represents the average annual growth rate. If $\beta > 0$, then $e^{-\beta T} < 1$ and $e^{-\beta T} - 1 < 0$, indicating that the growth rate and natural log of the real per capita income in the beginning period are negatively correlated – that is indicative of absolute beta convergence.

As shown in the study by Sala-i-Martin (1996a), the following linear growth equation could be estimated using OLS approach:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B - (1 - b_T)\log(y_{i,t_0}) + \varepsilon_{i,t_0,t_0+T} \quad (2.2a)$$

Or: $$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + (b_T - 1)\log(y_{i,t_0}) + \varepsilon_{i,t_0,t_0+T} \quad (2.2b)$$

The reason to estimate equation (2.1) using non-linear least square (NLS) rather than the linear version in (2.2) using OLS is that the estimated speed of convergence β can be directly compared across economies in the data set with different time lengths. In addition, the parameter β in (2.1) can be interpreted as the speed of convergence of an economy approaching its steady-state level. Regarding the equation (2.2), the speed of convergence β could be computed by solving the equality $(1 - b_T) = \left(\frac{1-e^{-\beta T}}{T}\right)$. The OLS coefficient estimate $(1 - b_T)$ would be inversely related to T (the length of the period over which the growth rate is computed). The reason is that if convergence exists, the growth rate should fall over time as when the economy is wealthier, the growth rate is predicted to be smaller. When considering long periods of time, the early periods with large growth rates are combined with latter periods with small growth rates. Therefore, the growth rate predicted by the original low level of income is smaller, the longer the time period of analysis. As T goes to infinity, the term $(1 - b_T)$ approaches zero, and as T goes to zero, the term $(1 - b_T)$ approaches β.

To test the hypothesis of conditional convergence, a set of variables that proxy the steady state is incorporated into the regression model like (2.1) or (2.2). That is, one estimates the following model (Sala-i-Martin 1996b):

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B - \left(\frac{1-e^{-\beta T}}{T}\right)\log(y_{i,t_0}) + \psi X_{i,t_0} + \varepsilon_{i,t_0,t_0+T} \quad (2.3a)$$

Or: $$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \left(\frac{e^{-\beta T}-1}{T}\right)\log(y_{i,t_0}) + \psi X_{i,t_0} + \varepsilon_{i,t_0,t_0+T} \quad (2.3b)$$

where X_{i,t_0} is a vector of variables that hold constant the steady state. If the β estimate is positive, after controlling for X_{i,t_0}, the data set is said to exhibit conditional beta convergence.

Nonetheless, in several empirical studies, a simplified version of beta of the following growth equation is often utilised to measure beta convergence:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \beta\log(y_{i,t_0}) + \psi X_{i,t_0} + \varepsilon_i \quad (2.4)$$

where $\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right)$ represents economy i's growth rate of real per capita income between two times t_0 and $t_0 + T$; $\log\left(y_{i,t_0}\right)$ is the logarithm of economy i's GDP per capita at time t_0; X_{i,t_0} includes all other factors supposedly affecting the growth rate; ε_i is the error term. A significantly negative β is indicative of beta convergence. The estimated β also indicates the rate at which regions approach their steady state (the convergence speed). If ψ is restricted to zero, *absolute convergence* is assumed. Providing that ψ is freely estimated, *conditional convergence* is assumed. The simplified version shown in (2.4) depends on a critical assumption that economies in the data set all have the same time length.

With regard to the testing method of convergence club, previous studies have predominantly estimated a nonlinear time-varying factor model (Phillips and Sul 2007, 2009).

There is a large body of literature that attempts to detect and measure beta convergence in numerous contexts. Recent contributions (see, *among others*, Baumont *et al.* 2003; Le Gallo and Sandy 2006; Tselios 2009; Li *et al.* 2016) involve investigating equation (2.4) in the light of spatial analysis. There are good reasons to believe that the omission of a spatial dimension from the analysis of regional beta convergence process is likely to produce biased results (see, *for example*, Rey and Montouri 1999; Fingleton 2003; James and Campbell 2013; Viegas and Antunes 2013; James and Campbell 2014; Díaz Dapena *et al.* 2019; Pietrzykowski 2019). First, working with regional data requires addressing the specific issue of *spatial autocorrelation*, meaning that contiguous regions may influence each other's performance. As a consequence, regional economic variables are likely to be interdependent, invalidating the OLS assumptions under which equation (2.4) is estimated. Possible solutions involve including 'spatial lags' into the model (to account for the fact that the growth rate of one region also depends on either the growth rate or the level of income of surrounding regions) and estimating spatial error models (to account for the possible systematic measurement errors due to the spatial correlation of the variables not included in the model). Second, differences in the fundamentals of regional economies raises the problem of *spatial heterogeneity*. To put it differently, the economic relationship shown in equation (2.4) is not stable over the space, meaning that the true value of the estimated coefficients (B , β, and ψ) varies across regions or countries.

Sigma convergence

Unlike beta convergence which aims to detect the 'catching-up' process, *sigma convergence* refers to the reduction in disparities among regions over time. It is shown that beta convergence is a necessary but not a sufficient condition for sigma convergence (Young *et al.* 2008). Economies can converge towards one another but may be pushed apart due to random shocks. Since the detection of beta convergence relies on the estimation of an econometric model, some researchers prefer sigma convergence as it deals with a direct measurement of distribution among regions without the need to estimate a particular model.

(Income) convergence means that the dispersion of incomes in a set of economies reduces over time; therefore, (income) convergence refers to the decrease in the 'width' of income distribution density. All income inequality indicators, which are also sigma convergence measures, are simply statistics measuring the 'width' of income distribution in one way or another. Commonly used measures of sigma convergence include the coefficient of variation, the Gini coefficient, the Atkinson's index, the Theil index, and the mean logarithmic deviation. Each of these indices is introduced in the subsequent discussion along with their relative merits as well as caveats. A summary of the five indices is presented in Table 2.1. Being based on different concepts of inequality and due to different constructing formulas, these indices may not yield the same indication of changes in disparities over time. Therefore, if the primary interest lies on the evolution of disparities, it is crucial to analyse a variety of indices.

Coefficient of variation

The coefficient of variation is a normalised measure of dispersion of the probability distribution. It is calculated by dividing the standard deviation by the mean, thus indicating a high or low level of variability relative to the mean value. Any reduction in coefficient of variation indicates a decline in the dispersion among countries and signals sigma convergence (Baumol 1986). Also quantifying the variability of the distribution, the standard deviation is, however, less preferred than the coefficient of variation as the former is meaningless on its own unless it is accompanied by the mean.

The coefficient of variation (CV) measuring regional inequality takes the following form:

$$CV = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2}}{\bar{y}} \tag{2.5}$$

where y_i represents the per capita income of region i, $\bar{y}$ denotes the average of regional per capita incomes, and N is the number of regions.

Gini coefficient

The Gini coefficient (Gini 1936) is one of the most popular measure of inequality in the distribution of income or wealth. By definition, it varies between 0 and 1. A low value indicates more equal distribution and a high value indicates more unequal distribution. 0 corresponds to perfect equality while 1 corresponds to perfect inequality where income is concentrated in the hands of one person. The Gini index is the Gini coefficient expressed as a percentage. The Gini coefficient (G) is computed as:

$$G = \frac{\sum_{i=1}^{N}\sum_{k=1}^{N}|y_i-y_k|}{2N^2\bar{y}} \tag{2.6}$$

where y_i and y_k represent the per capita income of region i and region k $(i \neq k)$, $\bar{y}$ denotes the average of regional per capita incomes, and N is the number of regions.

It is shown in equation (2.6) that the Gini coefficient allows direct comparison of the income distribution of two populations regardless of their sizes. An important limitation is that it is influenced by the granularity of the measurements (Monfort 2008). To illustrate, a Gini coefficient computed on the basis of five 20% quantiles (low granularity) will most likely be lower than the one based on twenty 5% quantiles (high granularity) taken from the same distribution. In addition, very different income distributions can present the same Gini coefficient (Afonso *et al.* 2015).

Atkinson index

The Atkinson index (Atkinson 1970, 1983) is a popular welfare-based measure of inequality. It presents the percentage of total income that a given society would have to forego in order to have more equal shares of income between its citizens. Its feature is the ability to emphasise movements in particular segments of the distribution. The Atkinson index (A) can be calculated as follows:

$$A_\epsilon = 1 - \left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}} \qquad ,\epsilon \neq 1, \tag{2.7a}$$

$$A_\epsilon = 1 - \frac{1}{\bar{y}}\prod_{i=1}^{N} y_i^{\frac{1}{N}} \qquad ,\epsilon = 1, \tag{2.7b}$$

where y_i represents the per capita income of region i, $\bar{y}$ denotes the average of regional per capita incomes, N is the number of regions, and ϵ is an inequality aversion parameter.

The Atkinson index ranges between zero and one. Zero corresponds to complete equality, i.e. when all individuals have the same income, and one corresponds to complete inequality, i.e. when the wealth is concentrated in the hands of one individual and all others having no income at all.

As can be seen from equations (2.7a, b), the value of the Atkinson index depends on the degree of society's *aversion to inequality* (ϵ) which is a theoretical parameter set by the researcher. A higher value of ϵ involves greater willingness by individuals to accept smaller incomes in exchange for a more equal distribution (Bellu and Liberati 2006b). The value of the Atkinson index can be interpreted in terms of the equally distributed equivalent. An Atkinson index with the value of, say 0.6, means that if wealth was equally distributed, the same level of social welfare could be achieved with only 40% of actual total wealth.

Due to the inequality aversion parameter, the Atkinson index becomes more sensitive to changes at the lower end/left tail of the income distribution (low income) as ϵ increases and approaches 1. As the level of inequality aversion ϵ decreases and approaches 0, the Atkinson index becomes more sensitive to changes in the upper end/right tail of the income distribution (high income) (Afonso *et al.* 2015). An important feature of the Atkinson index is that it can be decomposed into within- and between-group inequality – a property known as *decomposability* (Shorrocks 1984).

Theil index and Mean logarithmic deviation

The Theil index and the mean logarithmic deviation are considered complex inequality measures, and both belong to the *entropy* class of inequality indices. In thermodynamics, entropy refers to a measure of disorder. Applied for income distribution, entropy can be interpreted as a measure of deviation from perfect equality (Bellu and Liberati 2006a). A generalised entropy inequality index is given as:

$$GE(\alpha) = \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)^{\alpha} - 1\right] \quad (2.8)$$

where y_i is the per capita income of region *i*, $\bar{y}$ is the average of regional per capita incomes, and N is the number of regions.

The formula (2.8) defines a class of indices since the index $GE(\alpha)$ can take different forms depending on the value of α. In theory, the parameter α can take any possible real value $[-\infty, \infty]$. However, in practice, α is often set to be non-negative, i.e. $\alpha \geq 0$ since for a negative value of α this class of indices is undefined if regions with zero incomes exist ($y_i = 0$). A positive α represents the weight assigned to distances between incomes at different parts of the distribution and captures the sensitivity of the GE index to a specific part of the income distribution. In particular, the GE index is more sensitive to changes in the upper tail of the distribution for a positive and large α, whereas it is more responsive to what happens at the lower tail of the income distribution for a positive and small α (Atkinson and Bourguignon 2015). The GE index can take values between 0 and ∞. Zero corresponds to perfect equality and a higher value corresponds to a higher level of inequality.

Of especial interest are the two inequality measures when $\alpha = 1$ and $\alpha = 0$. $GE(1)$is called "Theil's T" or commonly the *Theil's index* (Theil 1967) while $GE(0)$ is called "Theil's L" or the *mean logarithmic deviation*. Thus, the Theil's index and the mean logarithmic deviation are computed as follows:

$$GE(1) = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{\bar{y}}\ln\left(\frac{y_i}{\bar{y}}\right) \quad (2.9)$$

$$GE(0) = -\frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{y_i}{\bar{y}}\right) \quad (2.10)$$

where y_i is the per capita income of region *i*, $\bar{y}$ is the average of regional per capita incomes, and N is the number of regions.

Belonging to the entropy class, both indices share some key features. A notable advantage of these measures is the ability to be fully decomposable, meaning that inequality may be disentangled by population groups or income sources which can be useful to policymakers (Bellu and Liberati 2006a). A main limitation is that both indices are undefined if there are zero incomes.

Table 2.1 Main characteristics of five select sigma convergence measures.

Measure	Range	Main features
Coefficient of variation	[0, 1]	Sensitive to changes in the mean, particularly when the mean value is near zero.
Gini index	[0, 1]	Sensitive to changes in inequality around the median/mode.
Atkinson index	[0, 1]	Sensitive to changes in lower or upper tails of the distribution depending on the “aversion to inequality”.
Theil index	[0, ∞]	• Gives equal weights across the distribution. • Does not have a straightforward interpretation.
Mean logarithmic deviation	[0, ∞]	• Sensitive to changes at the lower end of the distribution. • Does not have a straightforward interpretation.

2.5.2 Applications of convergence theories in food economics

Being defined and examined extensively in income levels, the concept of convergence has been successfully applied in other fields to investigate for example crime rates (Cook and Winfield 2013), energy consumption (Mohammadi and Ram 2012, 2017), and obesity levels (Li and Wang 2016; Kasman and Kasman 2020; Bell *et al.* 2021). In food economics, the focus has been put on testing convergence in food demand. In terms of convergence methodologies, earlier studies have been conducted at both micro- and macro-level. Micro researchers (Angulo *et al.* 2001; Regmi and Unnevehr 2006; Regmi *et al.* 2008b; Nowak and Kochkova 2011; Erbe Healy 2014) mostly employ food expenditure data on food items in household food basket to compute the coefficient of variation (“sigma convergence”). Macro researchers (Herrmann and Röder 1995; Elsner and Hartmann 1998; Wan 2005; Borkowski *et al.* 2008; Ravallion 2012; Ogundari and Ito 2015) utilise the per capita food supply at aggregate level to estimate beta following beta convergence regression model.

Overall, studies on convergence in food consumption are not new but mainly focus on developed countries. Blandford (1984) examines dietary trends in OECD countries from 1960 to 1980 and finds that the caloric consumption not only appears to reach a ceiling but also becomes less responsive to changes in income. The author documents a growing trend in the proportion of energy from animal products and a tendency for dietary structure of OECD countries to become alike. In a similar attempt but utilising food expenditure data, Regmi and Unnevehr (2006) uncover convergence in total food as well as cereal and meat consumption among 18 OECD high-income countries. This is attributed to the highly standardised food delivery system at retail level and the increasingly homogeneous consumer’s tastes and preferences. Srinivasan *et al.* (2006) further explore the current food consumption patterns in OECD countries and find that the majority are incompatible with the WHO dietary recommendations, primarily because of excessive fat and sugar intakes. Nevertheless, some improvement is documented by Mazzocchi *et al.* (2008) and the authors report evidence of sigma convergence so that OECD countries are becoming increasingly similar in terms of dietary healthiness (measured as the degree of adherence to the WHO suggestions). Convergence, however, is not detected for the whole sample of 149 countries.

At the global perspective, Regmi *et al.* (2008b) show significant convergence in food consumption across 47 high- and middle-income countries during 1990-2004, and that the convergence speed has slowed down. The authors conclude that upper-middle-income countries are converging towards the same food expenditure level as OECD countries while lower-middle-income peers are approaching a lower steady-state level. Their findings also support the argument of Popkin (2006a) that middle-income countries are adopting less healthy food items (such as meat, dairy, sugars, stimulants, and soft drinks) that are prevalent in the diets of high-income counterparts. Furthermore, Frazão *et al.* (2008) provide an estimate ranging between 16 and 21 years for the half-life (in other words, the number of years required for progress halfway towards the steady-state level when convergence is assumed to have been achieved). It is worth noting that convergence in fast food sales is occurring at a much faster speed than any other type of food expenditure, with a half-life estimate of 9 years.

In contrast to the above studies standing from the demand-side perspective, Bentham *et al.* (2020) point out to partial convergence in the national supply of animal-source foods and sugars but divergence in the supplies of vegetables, seafood and oil crops. Also analysing the Food Balance Sheet data from the FAO, Khoury *et al.* (2014) suggest that dietary composition across 152 countries is becoming more similar and national food supplies have increasingly relied on a set of truly global crop commodities. However, the sole focus of Khoury *et al.* (2014) is to evaluate the similarity in global diets without testing any convergence theory or implying about converging diets. A recent attempt to quantify the (sigma) convergence in global patterns of dietary consumption is put forward by Bell *et al.* (2021). Comparing across various foods and nutrients, the authors conclude that the convergence of fats and animal-source foods has increased more rapidly than other foods and nutrients such as fruits and vegetables, zinc, iron and vitamin A.

Digging into the determinants of the increasingly similar patterns of food consumption, previous researchers largely agree on income growth and globalisation of the food retail and foodservices (Blandford 1984; Frazão *et al.* 2008; Gerbens-Leenes *et al.* 2010; Kearney 2010; Schneider *et al.* 2011; Brunelle *et al.* 2014). Economic theory suggests that consumers' demand for food changes in predictable ways as income rises. According to Engel's law, the income elasticity of demand for food is less than one and therefore, other things being equal, the share of income spent on food declines as income rises (Clements and Si 2018). Another important economic baseline is Bennett's law, which states that food composition shifts away from starchy staples as income rises, with increasing demand for fruits, vegetables, vegetable oils, and animal-source foods (Bennett 1941). The global expansion of multinational retail and foodservice chains has shaped consumer preferences and standardised the ways in which food is produced, delivered and consumed around the world (Unnevehr

2004; Martin 2018). Also, social globalisation contributes to convergence in tastes (Aizenman and Brooks 2008).

A great amount of previous research efforts is spent on relating the convergence in food consumption patterns to the integration within the European Union (EU). For the EU, economic and institutional integration has long been considered as fundamental goals. Hence, the 'catching-up' process among its member countries is of paramount relevance and any evidence of convergence across national borders would support the implementation of common policies across countries. In the existing literature, the issue of convergence within the EU is usually examined from a macroeconomic perspective, for instance using GDP per capita (Cabral and Castellanos-Sosa 2019), productivity (Sondermann 2014), and inflation rate (Brož and Kočenda 2018). Little attention has been paid to living standards across countries. As pointed out by Wan (2005), consumption convergence directly implies declining inequality in accessibility to food that is an indication of integrated living standards across regions/countries. Analyses of food consumption would assist governmental and non-governmental agencies in designing policies, infrastructure development plans and marketing strategies. Besides, convergence in food consumption could indicate that globalisation is having a homogenising impact on cultural identity (Erbe Healy 2014). These motivations lead to the blossoming of several research papers on this topic.

Proxying shares in food expenditure for the living standard of an average household in a given country, Dudek (2014) discovers a 'catching-up' process taking place in EU27 as well as EU15. However, the author does not find evidence for sigma convergence in EU15. A large number of researchers voice the same opinion that the dietary structure of EU countries is becoming more alike (Gil *et al.* 1995; Traill 1997; Elsner and Hartmann 1998; Grigg 1998; Sojková and Matejková 2001; Schmidhuber and Traill 2006; Sengul and Sengul 2006). Unlike the abovementioned researchers, Erbe Healy (2014) reports an increasing coefficient of variation in the food expenditure patterns, suggesting (sigma) divergence in four Western European countries (the United Kingdom, France, Ireland, and Italy) over the period 1985-2005. However, when Italy – the outlier with a strong tradition in preparing food at home as compared to the modern 'eating out' lifestyle in other countries – is removed from the study sample, converging trends are predominant. Despite the mounting evidence for converging patterns of food consumption in the EU, dietary differences still exist, for example between Mediterranean and non-Mediterranean countries, Southern Europe and Northern Europe. Against the backdrop of the significant convergence in socio-economic factors or even consumption at regional level (Nowak and Kochkova 2011; Otoiu and Titan 2015; Michail 2020), why should the food consumption not be completely converged?

Some researchers point to cultural and individual differences (Traill 1997; Gracia and Albisu 2001; Sengul and Sengul 2006). Clearly, culture affects consumer behaviour, and cultural diversity can resist pressures from foreign travel, media, telecommunications – the factors that are closely linked with globalisation and are believed to bolster the process of homogenising global food systems (Oberlander *et al.* 2017; De Sousa *et al.* 2018). On the other hand, individuals hold different values, resulting in diversified food preferences and habits. In this manner, country groupings themselves represent just an average of the national population whilst the market is rather fragmented. Considering the bewildering array of demographic, economic, psychometric, attitudinal, cultural, and lifestyle characteristics, the process of convergence is thus best viewed as the growing importance of homogeneous segments of consumers in the food markets across national boundaries.

Realising the relatively significant influence of price, income, consumer preferences and other factors stated in demand theory, some authors raise their concern over the validity of the absolute convergence results when these factors are not considered (Herrmann and Röder 1995; Srivastava *et al.* 2016). They argue that statistics tests such as coefficient of variation or time series regression indicate convergence just because income (as well as food price) converges across countries, and thereby it is necessary to control for income, price and other explanatory variables before associating convergence/divergence with long-run changes in consumer preferences. Other authors recommend considering other factors besides those mentioned in demand theory, for instance, socio-demographics (Erbe Healy 2014; De Sousa *et al.* 2018).

To summarise, the existing literature has identified: (i) rising similarity in total caloric supplies and dietary structure across national borders, (ii) convergence in the consumption of caloric intakes and certain food items. Despite the paramount importance of convergence in food consumption, research studies on this topic are rather dated and mainly focus on developed countries and the EU. Few studies have been conducted to examine food consumption patterns at the global level (Khoury *et al.* 2014; Azzam 2020; Bentham *et al.* 2020; Bell *et al.* 2021); yet, formal convergence testing has been largely missing. This fact reinforces the significance of the current research and the contributions it adds to the existing literature.

2.6 Chapter conclusion

This chapter aims to review the existing literature on the nutrition transition and the associated dietary changes. Although the nutrition transition model was originally developed in the early 1990s by Barry Popkin to describe five distinct patterns of diet and lifestyles, the term *nutrition transition* is commonly used to refer to the shift from traditional diets towards the ‘Western’ diet (Pattern 4 - Degenerative disease) that is rich in fat, sugars, meat and processed foods but low in fibre, and accompanied by increasingly sedentary lifestyles. Drivers for the nutrition transition are manifold and often involve a wide variety of economic, social and cultural factors with the complex interconnectedness. The impacts of underlying forces such as globalisation, rising income and urbanisation are well described in the literature. Nonetheless, the underlying mechanisms through which these take place remain ambiguous.

As there is mounting evidence for the dietary changes across the globe, the negative impacts of these shifts on human health and the environment have become a source of great concern. As pointed out by the nutrition transition model, the move towards healthier eating societies which is centred around Pattern 5 (Behavioural change) is necessary to induce large-scale changes to improve diets. But Pattern 5 so far remains hypothetical!

As a manifestation of the nutrition transition, global diets are predicted to be more similar as countries develop and become further globalised. There is a wealth of evidence for the increasing similarity in national food supplies that is indicative of convergence in consumption of caloric intakes and certain food items. Having said that, research studies on this topic are dated, mainly focus on developed countries and only few studies make connection to the nutrition transition literature. Few studies are conducted to examine food consumption patterns at the global level; however, formal convergence testing has been largely missing. Realising this gap in the extant literature, this research examines the dietary convergence in the light of beta and sigma convergence methods using global data, estimates the speed of convergence and probes into the role played by income. A challenge of employing global data lies in the nature of data varying across space (countries) and over time. In order to address this issue, this research utilises a statistical technique called *cluster analysis* to summarise and describe global diets on the basis of the historical trends. Potentially the results of cluster analysis will help to detect whether there is evidence for the existence of Pattern 5 in some countries and how the diet might be like at this stage.

Chapter 3

Cluster analysis

3.1 Chapter introduction

Researchers across various disciplines often face many tasks in which defining groups of homogeneous objects, whether they are individuals, firms, countries or even behaviours, might be useful. Overall, there is the need to search for a natural structure among observations based on a set of complex features. The most commonly used technique for this purpose is *cluster analysis*.

Fundamentally, cluster analysis involves sorting observations into groups so that members of a group are more similar to one another than they are to members of a different group. Each group is known as a *cluster* and the process of assigning observations to groups is referred to as *clustering*. If this task is done correctly, these clusters can be characterised by their *profile* – a summary of what members of a group are like in terms of the original variables used for clustering purpose. Observations in one group may have consistently high values on some variables but low values on others. As a cluster is more similar internally than it is to any other cluster, the analyst only needs to check for the profile of a cluster to have an insight into what all observations in that cluster are like instead of having to analyse all original variables at once.

This chapter aims to present a general review on cluster analysis and to highlight the suitability as well as novelty of this method for the current research. Section 3.2 lays down the conceptual definitions and main steps involved in cluster analysis. Section 3.3 introduces methods for clustering static data with a focus on both conventional methods (hierarchical and K-Means clustering) and the ever-growing literature on fuzzy clustering. Section 3.4 and 3.5 look at clustering methods for time series and spatial data respectively while Section 3.6 surveys methods for clustering space-time data. Section 3.7 examines the application of cluster analysis in food economics, focusing on food consumption studies. Limitations of earlier studies and the need for further research are pointed out. Section 3.8 concludes the chapter.

3.2 An overview of cluster analysis

3.2.1 Conceptual definitions

The aim of cluster analysis is to find groups (or clusters) of objects so that objects in the same cluster are more similar to one another whilst dissimilar from objects in other clusters. To put it differently, the main principle underpinning cluster analysis is to maximise the homogeneity of objects within the clusters while maximising the heterogeneity of objects between clusters (Figure 3.1). Cormack (1971) introduces the properties of internal cohesion (homogeneity) and external isolation (separation), and argues that clusters should exhibit high internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity. If the clustering task is done successfully, objects within clusters will be close to each other and objects belonging to different clusters will be far apart.

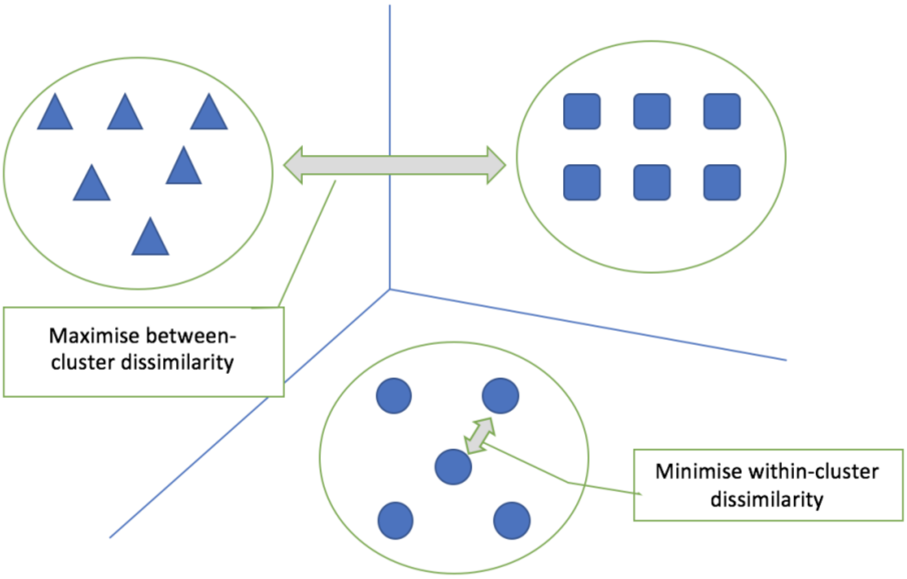


Figure 3.1 Visual illustration of cluster analysis.

Cluster analysis is an *unsupervised* technique to discover groups in data, and it should not be confused with *supervised* learning methods in which the groups are known *a priori* and the aim is to construct rules for classifying new individuals into one of the known groups. For supervised learning methods (such as regression or classification), one is given a set of p features $(x_1, x_2, \dots, x_p)$ measured on N observations and a response variable y also measured on the same N observations. The goal is then to predict y using $x_1, x_2, \dots, x_p$. An example of supervised learning is weather forecast apps which rely on some prior information (such as humidity level, and whether it is sunny or cloudy, etc) to predict the weather parameters for a given time. Another application of supervised learning in our daily life is the email filtering system which, based on past information about spams (such as the email wording or the sender's address), filters out an incoming email into Inbox (normal) or Junk folder (spam). Here, the goal is to try to predict a vector of continuous outcomes (the temperature in the first example) or a binary classification (normal email versus spam in the second example). In contrast, the *unsupervised* nature of cluster analysis implies that there is no rule for the initiation of classification, and the groups are not known *a priori* (Budayan *et al.* 2009). For an *unsupervised* task, the researcher only has information on a set of features $x_1, x_2, \dots, x_p$ measured on N observations, and he/she is not interested in predictions because there is not an associated response variable y. Instead, the goal is to discover interesting patterns about the measurements on $x_1, x_2, \dots, x_p$. Is there an informative way to visualise the data? Do meaningful subgroups among the variables or among the observations exist? Unsupervised learning techniques help to answer these questions. For instance, a biomedical researcher performs cluster analysis on a data set consisting of N observations (say tissue samples of patients with lung cancer) and the corresponding clinical measurements for each observation (say tumour stage or gene expressions) to find some *unknown* heterogenous subgroups of lung cancer. The goal here is not to make predictions, but to uncover structure, i.e. the distinct clusters on the basis of the data set.

Comparing cluster analysis with factor analysis – the two most common unsupervised learning methods, both techniques pursue the same objective of exploring structure in data. However, cluster analysis groups <u>objects</u> (which could be units, countries, firms, individuals, or behaviours) whilst factor analysis mainly concerns with grouping <u>variables</u>. Besides, the groupings in factor analysis are based on patterns of variation (correlation) in data whereas cluster analysis performs the groupings on the concept of distance (Hair *et al.* 2014). Some differences in these two procedures are highlighted in Figure 3.2. Assume that a set of p features $(X1, X2, \dots, Xp)$ measured on N objects is given. Cluster analysis seeks to assign N objects into K clusters $(K < N)$ based on the p features. The results of cluster analysis indicate the membership of each object to each cluster, for example, Object 1 belongs to Cluster 1, Object 2 belongs to Cluster 2, and so forth. Factor analysis, on the other hand, attempts to reduce the number of variables (p features in this case) into fewer number of factors H $(H < p)$ that are representatives of the original features. Factor analysis returns the factor loadings (a number ranging from -1 to 1), with the loadings close to -1 or 1 indicating that the factor strongly influences the variable. To illustrate, X1 has large positive loading (0.7) on Factor 1, so this Factor 1 best describes X1. Similarly, Factor 2 best represents X2 due to the large positive loading (0.8).

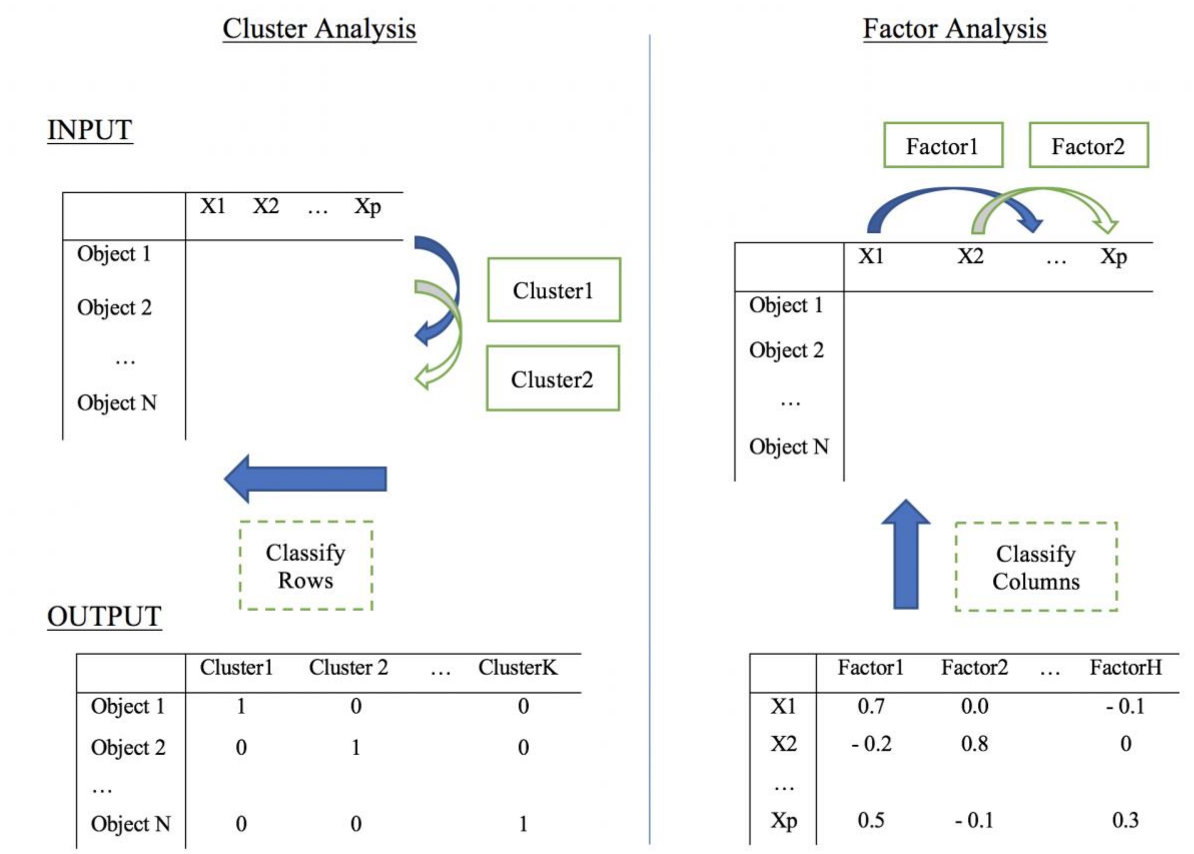


Figure 3.2 Cluster analysis versus factor analysis.

3.2.2 Basic stages in cluster analysis

In general, cluster analysis involves six steps as demonstrated in Figure 3.3.

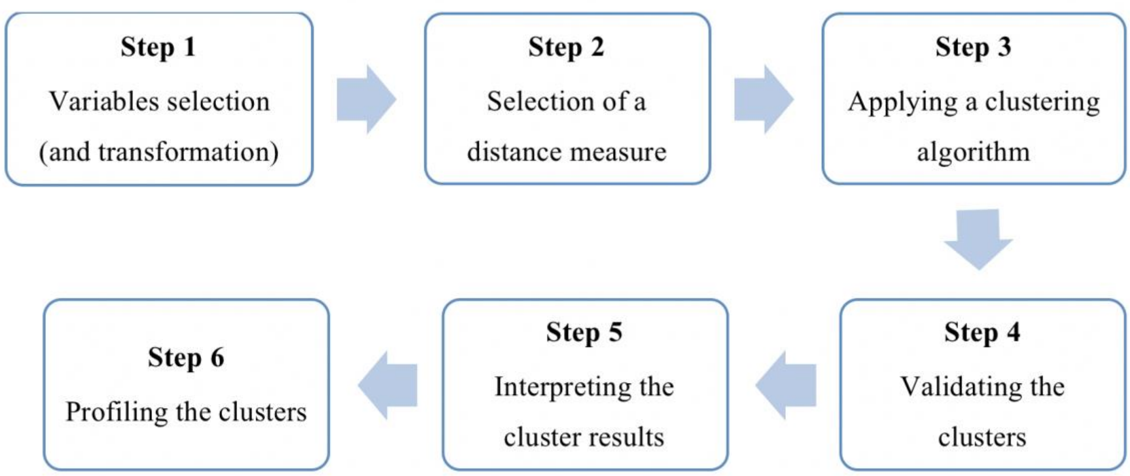


Figure 3.3 Stages in cluster analysis.

Step 1. Variables selection (and transformation)

The aim of cluster analysis is to partition a set of N objects into K groups based on the similarity of N objects for a set of specified characteristics. This aim should dictate the selection of variables used to characterise the objects being clustered because only variables that relate specifically to objectives of cluster analysis should be included (Hair *et al.* 2014). Cluster analysis has no means to differentiate relevant from irrelevant variables and yields the most consistent and distinct groups of objects across all variables. Theoretical, conceptual and practical considerations must be made when selecting input variables for cluster analysis. This step is critical as the derived clusters only reflect the structure of data as defined by those variables. To put it differently, two objects in the same cluster are only considered similar with regard to the selected input variables and might be dissimilar with respect to other variables which are not used for the clustering task.

To illustrate, assume that a marketing researcher wishes to understand patterns in customer spending on processed foods based on a survey conducted in a small community. The survey is divided into two parts: the first part contains basic questions about the respondent's socio-demographic profile (such as age, gender, marital status, accommodation postcode, job sector, etc), and the second part is related to the respondent's weekly expenditure (in pounds sterling) in a range of processed food products. In selecting the appropriate variables for a cluster analysis, the researcher should not utilise the socio-demographic questions in the first part but the questions in the second part instead because they are related to the customer's spending behaviour. If the food expenditure questions are used for a cluster analysis, the objective will be to see if there are groups of customers that exhibit distinctively different patterns of shopping for processed foods between the groups, but similar shopping patterns within each of the groups.

Despite the long history of cluster analysis, previous research has focused mainly on fine-tuning clustering algorithms and review articles mainly explore the wide range of variables that can be used for the basis of clustering (Dolnicar 2003; Tuma *et al.* 2011; Ernst and Dolnicar 2018). If survey data are utilised, it is common to contain variables that do not necessarily contribute to the clustering solution and they are referred to as *noisy variables*. It has been shown that irrelevant (noisy) clustering variables negatively affect the quality of the clustering task since they not only mask the cluster structure leading to less homogeneous clusters (Dolnicar *et al.* 2014) but also increase the dimensionality of the data making the clustering task substantially more complex (Hajibaba *et al.* 2019). As a result, prior researchers have proposed variable selection approaches to reduce the number of clustering variables (Carmone *et al.* 1999; Brusco and Cradit 2001) or methods that simultaneously select the most influential variables and group individuals (Dolnicar *et al.* 2011; Dolnicar *et al.* 2012; Legoherel and Wong 2012). It is worth mentioning that the inclusion of irrelevant variables could potentially lead to highly correlated variables, also known as multicollinearity. While multicollinearity is a serious issue in regression analysis as it is difficult to estimate the relationship between each explanatory variable and the response variable independently and consequently the beta coefficient estimates are not stable, the problem is different in cluster analysis because there is no response variable or beta coefficient. Normally, correlated variables are removed only if the number of variables is much larger than the sample size. Otherwise, all variables are kept, and the clustering result would be a very good partition. A certain number of observations measured on a specified number of variables are used to create clusters. Each observation belongs to one cluster and each cluster can be defined in terms of all variables used in the analysis. The aim is to maximise internal cohesion and external separation among the clusters. This aim is enforced if correlated variables are employed for the clustering task. Nevertheless, Dolnicar *et al.* (2016) point out that no matter high correlation occurs among all variables or among groups of variables it can decrease the performance of the clustering task.

The basic data input for most clustering methods is the $N \times p$ multivariate data matrix, X, containing the variable values for each object to be clustered, given by:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1a} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Na} \end{bmatrix} \tag{3.1}$$

in which the element x_{Na} denotes the value of the a-th variable on object N. X is normally called a *'two-mode'* data matrix since its rows and columns refer to two different things. The variables in X can be continuous, ordinal, categorical, or a mixture of these, and can include missing values. Sometimes, the rows of X may contain repeated measures of the same variable but under different conditions, at different times, or at different spatial locations.

It is usually the case that variables need to be *standardised* (or *scaled*) to have: (i) standard deviation of 1 and (ii) mean of 0, before any clustering algorithm can be applied. This is of paramount importance if variables are measured on different scales (for instance kilometres, metres, grams). Without this procedure, the dissimilarity measures (discussed in Step 2) will be distorted. Nonetheless, standardisation is criticised as it might remove some natural relationships in the scaling of variables (Aldenderfer and Blashfield 1984). Various approaches to standardisation have been proposed in the literature (see, *among others*, Milligan and Cooper 1988; Jajuga and Walesiak 2000), and often the formula for data standardisation is given as:

$$\frac{x_i - centre(x)}{scale\ (x)} \tag{3.2}$$

where $centre(x)$ is a function of mean (or median) of x values, and $scale(x)$ can be standard deviation, interquartile range, or median absolute deviation.

Step 2. Selection of a distance (dissimilarity) measure

The concept of similarity/distance is fundamental to cluster analysis as one should determine how 'close' or how 'far apart' objects are to each other before performing any groupings. Many clustering tasks start with an $N \times N$ one-mode matrix, which contains elements indicating the quantitative measure of closeness (also referred as *distance/similarity/dissimilarity*). Two objects are 'close' when the pairwise distance (dissimilarity) measure is small. Distance measures can be decided either directly (from a dissimilarity matrix) or indirectly (from a two-mode data matrix); however, the latter method is common in practice.

Many clustering techniques start by converting the two-mode data matrix X into an $N \times N$ matrix of inter-object dissimilarity matrix D, given by:

$$D = \begin{bmatrix} 0 & d(x_{11}, x_{21}) & d(x_{11}, x_{31}) & \cdots & d(x_{11}, x_{N1}) \\ d(x_{21}, x_{11}) & 0 & d(x_{21}, x_{31}) & \cdots & d(x_{21}, x_{N1}) \\ d(x_{31}, x_{11}) & d(x_{31}, x_{21}) & 0 & \cdots & d(x_{31}, x_{N1}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d(x_{N1}, x_{11}) & d(x_{N1}, x_{21}) & d(x_{N1}, x_{31}) & \cdots & 0 \end{bmatrix} \tag{3.3}$$

where $d(x_{21}, x_{11})$ denotes the quantitative distance measure between $x_{21}$and x_{11}. D is called *one-mode* matrix since its rows and columns indicate the same thing.

Broadly speaking, existing dissimilarity measures can be split into distance measures and correlation measures. Distance measures focus on the magnitude of the values and might indicate that two objects are close to each other even though they have different patterns across the variables. On the other hand, correlation measures put an emphasis on the patterns across the variables, not the magnitude of the differences between objects (Hair *et al.* 2014). The most common distance measures are listed in Table 3.1.

Table 3.1 Common dissimilarity measures for continuous data.

Distance	Rationale	Formula
Euclidean	A squared error distance	$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$ with x and y are two data vectors of length n
Manhattan	An absolute value distance	$d(x,y) = \sum_{i=1}^{n} \lvert (x_i - y_i) \rvert$
Minkowski	A generalisation of Euclidean distance (where $r = 2$) and Manhattan distance (where $r = 1$)	$d(x,y) = \sqrt[r]{\sum_{i=1}^{n} \lvert (x_i - y_i) \rvert^r}$ with $r \geq 1$
Pearson correlation	Measures the degree of a linear relationship between two objects	$d(x,y) = 1 - \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$ where $\bar{x}\ and\ \bar{y}$ denote the mean of x and y respectively
Spearman correlation	Calculates the correlation between the rank of x and the rank of y	$d(x,y) = 1 - \frac{\sum_{i=1}^{n}(x'_i - \overline{x'})(y'_i - \overline{y'})}{\sqrt{\sum_{i=1}^{n}(x'_i - \overline{x'})^2 \sum_{i=1}^{n}(y'_i - \overline{y'})^2}}$ where $x'_i = rank(x_i)$ and $y'_i = rank(y_i)$
Kendall correlation	Measures the correspondence between the ranking of x and y variables. Begin by ordering the pairs by x values. If x and y are correlated, they would have the same relative rank orders. For each y_i, count the number of $y_i > y_i$ (concordant pairs), and the number of $y_i < y_i$ (discordant pairs).	$d(x,y) = 1 - \frac{n_c - n_d}{\frac{1}{2}n(n-1)}$ where n_c , n_d are the total number of concordant pairs and discordant pairs respectively

Step 3. Applying a clustering algorithm

Cluster analysis is a data exploratory technique and different methods present different views of the same set of data (Leisch 2006). A range of clustering algorithms have been established in the literature. Details on clustering methods for static data, time series, spatial data, and time-space data are to be presented in the next sections.

In general, clustering approaches in the literature can be categorised into distance/dissimilarity-based methods and model-based methods (Everitt *et al.* 2011). The former technique groups objects into the same cluster based on the pairwise dissimilarity measure computed among objects. On the other hand, the latter assumes that data are generated by a finite mixture of probability distributions and that objects belong to the same j^{th} cluster if they have the same specific density function f_j (typically a Gaussian distribution function).

Step 4. Validating the clusters

This step typically involves determining the optimal number of clusters and evaluating the goodness of clustering results with the aid of clustering validation statistics. Following Charrad *et al.* (2014), clustering validation statistics can be grouped into three categories:

- Internal cluster validation measures, which use the internal information of the clustering process to assess the goodness of the cluster results without reference to external information. These statistics can be used to determine the number of clusters and the appropriate clustering algorithm for a given data set. The widely used internal cluster validation index is the *Dunn index*. If the derived clusters are compact and well-separated, Dunn index should be maximised (Dunn 1974).
- External cluster validation measures, where the objective is to quantify the agreement between the identified clusters and an external reference. This approach is often used for selecting the appropriate clustering algorithm for a specific data set. The most popular measure is the *Adjusted Rand index* (Hubert and Arabie 1985).
- Relative cluster validation measures, which assess the cluster structure by changing different parameters of the same algorithm and are mainly utilised to identify the optimal number of clusters. There are over 30 available indices or methods that have been proposed in the literature. The most common ones include *elbow method*, *silhouette method*, and *gap statistics*.

It is noted that the aforementioned methods for cluster validation are mainly applied for distance/dissimilarity-based clustering methods. For model-based clustering methods, the best model is usually selected according to the BIC criterion (Bayesian Information Criterion).

Step 5 + 6. Cluster interpretation and cluster profiling

In cluster interpretation step, each cluster is examined in terms of the clustering variables and other additional variables. Average/weighted average values of these variables are computed for each cluster and visual tools like boxplot are often used. A name or a label could be assigned to each cluster to reflect distinct nature of the identified clusters.

The profiling step involves describing the characteristics of each cluster to explain how they may differ on various features. Variables, which are not previously used in the clustering procedure and usually contain information on demographics, psychographics, consumption patterns are utilised. The emphasis is on the characteristics that differ significantly across clusters and could predict the membership of an object belonging to a cluster. ANOVA (Analysis of Variance), MANOVA (Multivariate Analysis of Variance), multinomial logit model or other econometric analysis could be employed to compare the average score profiles across clusters. As a rule of thumb, clusters should be distinct, consistent with theory and can be explained.

3.3 Clustering methods for static data

Static data are data that their feature values neither change with time nor change negligibly. According to how clusters are formed, clustering methods can be categorised into four types, as summarised in Table 3.2. The most popular techniques are non-hierarchical and hierarchical.

Table 3.2 Main clustering methods for static data: rationales and algorithm examples.

Clustering methods	Rationale	Example
Centroid-based	Clusters are represented by a cluster centre, which is not necessarily an object member to be clustered. Data points are grouped based on how close they are to the cluster centre.	Non-hierarchical clustering
Connectivity-based	Data points closer to each other in data space are more similar than data points farther away. Clusters are formed based on their distance.	Hierarchical clustering
Distribution-based	Data points belong to the same cluster if their observed values come from the same probability distribution, whose parameters are unknown and need to be estimated.	Gaussian mixture models
Density-based	Clusters are represented by areas of higher density within the data space as compared to other regions.	DBSCAN (Density-based spatial clustering and application with noise)

3.3.1 Non-hierarchical versus hierarchical clustering

Non-hierarchical clustering

Non-hierarchical methods aim to group observations around a centre and observation units are organised into K clusters, where K is pre-selected by the researcher. The two most popular algorithms in this category are *K-Means* and *K-Medoids*, both of which are partitional (i.e. breaking the data set up into groups).

K-Means algorithm attempts to minimise the distance from data points to the cluster centre, also known as *centroid* (which is simply the average of data points in the cluster) (MacQueen 1967). To illustrate using formula, K-Means algorithm aims to minimise the following function:

$$J = \sum_{k=1}^{K} \sum_{x_i \in C_k} (x_i - \mu_k)^2 \tag{3.4}$$

where J is the total within-cluster variation, K is the pre-specified number of clusters, x_i is the data point (object) in cluster C_k, and μ_k is the cluster centroid (or the mean value of points included in cluster C_k). The rationale behind this formula is that each object is assigned to a cluster so that the sum of squared distance of the observation (x_i) to the assigned cluster centroid (μ_k) is minimum. J, measuring the compactness of the clustering partition, should be as small as possible.

K-Means is an iterative algorithm and consists of the following steps:

Step 1: Specify the number of clusters (K) to be identified.

Step 2: Select randomly K (not actual) objects being the initial cluster centroids.

Step 3: Assign each object to the group with the closest centroid. This can be determined on the basis of the (Euclidean) distance between the object and the centroid.

Step 4: When all objects have been assigned, calculate the new mean value of each cluster. Once the cluster centroids are recalculated, each object is checked again if it might be closer to a different cluster. All the objects are reassigned using the updated cluster centroids.

Step 5: Iteratively repeat Steps 3 and 4 until the algorithm converges (when the assignments no longer change, and the centroids no longer move).

Figure 3.4 conveniently visualises the abovementioned steps. First, three cluster centroids (C1, C2, C3) are randomly initialised. The K-Means algorithm then goes through all the data points and depending on which cluster is closer, it assigns data points to one of the three cluster centroids (Figure 3.4a). Next, the algorithm re-calculates the averages of all the data points in a cluster and move the cluster centroid to the new average position (Figure 3.4b). This process is repeated until there is no further change in the cluster centroids.

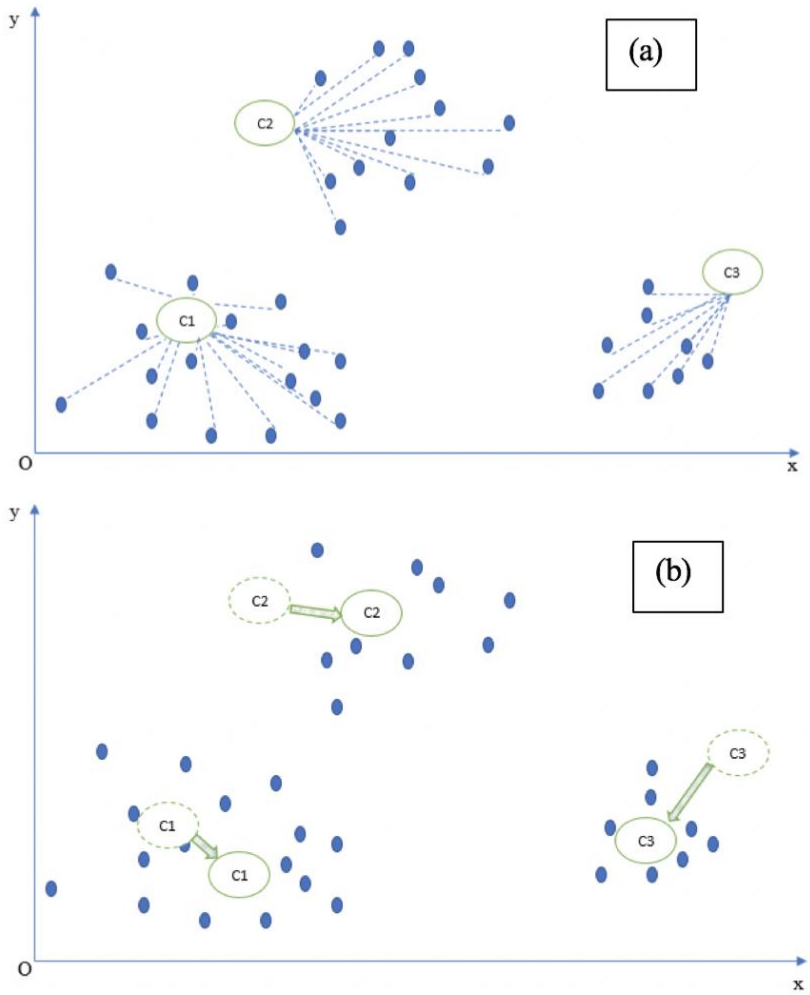


Figure 3.4 Visualisation of K-Means algorithm.

For K-Means, pre-specifying the number of clusters is a challenging task, and in practice it is usually the case that the researcher runs K-Means algorithm on different values of K, and then chooses the optimal K on the basis of cluster validation measures. Elbow, silhouette and GAP statistics are the most common methods. *Elbow method* depicts a line graph of the total within-cluster sum of square over a range of K values. This graph would typically look like an arm, and the 'elbow' on the arm indicates the optimal K. This method receives a wealth of criticism for its ambiguity, and an alternative is the *average silhouette* (Rousseeuw 1987). A silhouette, constructed for each data point, measures the clustering quality for that data point. The idea is to compute the average silhouette (for the whole data set) for varying values of K, and the optimal number of clusters is the one that maximises the average silhouette.

Specifically, the silhouette measure (Kaufman and Rousseeuw 1990b) of data point i is computed as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{3.5}$$

where $a(i)$ is the average distance between i and all other data points within the same cluster. It measures how well *i* is fitted into its cluster. $b(i)$ is the minimum average distance of *i* to all data points of other clusters. Generally speaking, $s(i)$ lies in the range [-1, 1]. If the value of $s(i)$ is close to 1, the heterogeneity of i's cluster is smaller than its separation, and *i* is considered to be well classified. If the value of $s(i)$ is close to -1, the heterogeneity of i's cluster is larger than its separation, and i is considered to be poorly classified. When $s(i) = 0$, the point *i* is on the border, meaning that it is not clear if *i* should have been assigned to the current or a neighbouring cluster.

The average silhouette width – the average of the $s(i)$ over the entire data set – is given by:

$$ASW = \frac{1}{N}\sum_i s(i) \tag{3.6}$$

Unlike the two above methods, GAP statistic (Tibshirani *et al.* 2001) is a statistical method that has been proposed in the literature as a measure for estimating the optimal number of clusters. The main goal is to formalise the idea of finding the elbow-type behaviour in the plot of the optimised cluster criterion against the number of clusters (K). Say the within-cluster measure W_K is the cluster criterion. The idea is to standardise the graph of $\log[W_K]$ against the number of clusters, by comparing it with its expectation under an appropriate null reference distribution. The null distribution is the one that produces a clearly unclustered data set. For this purpose, let E_N^* denote the expectation under a sample size of N from the reference distribution, the optimal number of clusters is the value of K that maximises the following statistics:

$$GAP_N(K) = E_N^*\{\log[W_K]\} - \log\,[W_K] \tag{3.7a}$$

$$\text{where } W_K = \sum_K \frac{1}{2N_K}\sum_{(i,j)\in\mathbb{C}_K} d_{ij}^2 \tag{3.7b}$$

An example is given in Figure 3.5 to illustrate how to select the optimal number of clusters using the three methods discussed so far. The K-Means algorithm is run with the number of clusters (K) varying from 2 to 10. The elbow method with the 'elbow' shape at $K = 4$ suggests that four-cluster solution is the best partition. The GAP statistic which is maximised at $K = 4$ also suggests four-cluster solution. The silhouette method, however, recommends three-cluster solution because the average silhouette reaches the maximum when $K = 3$.

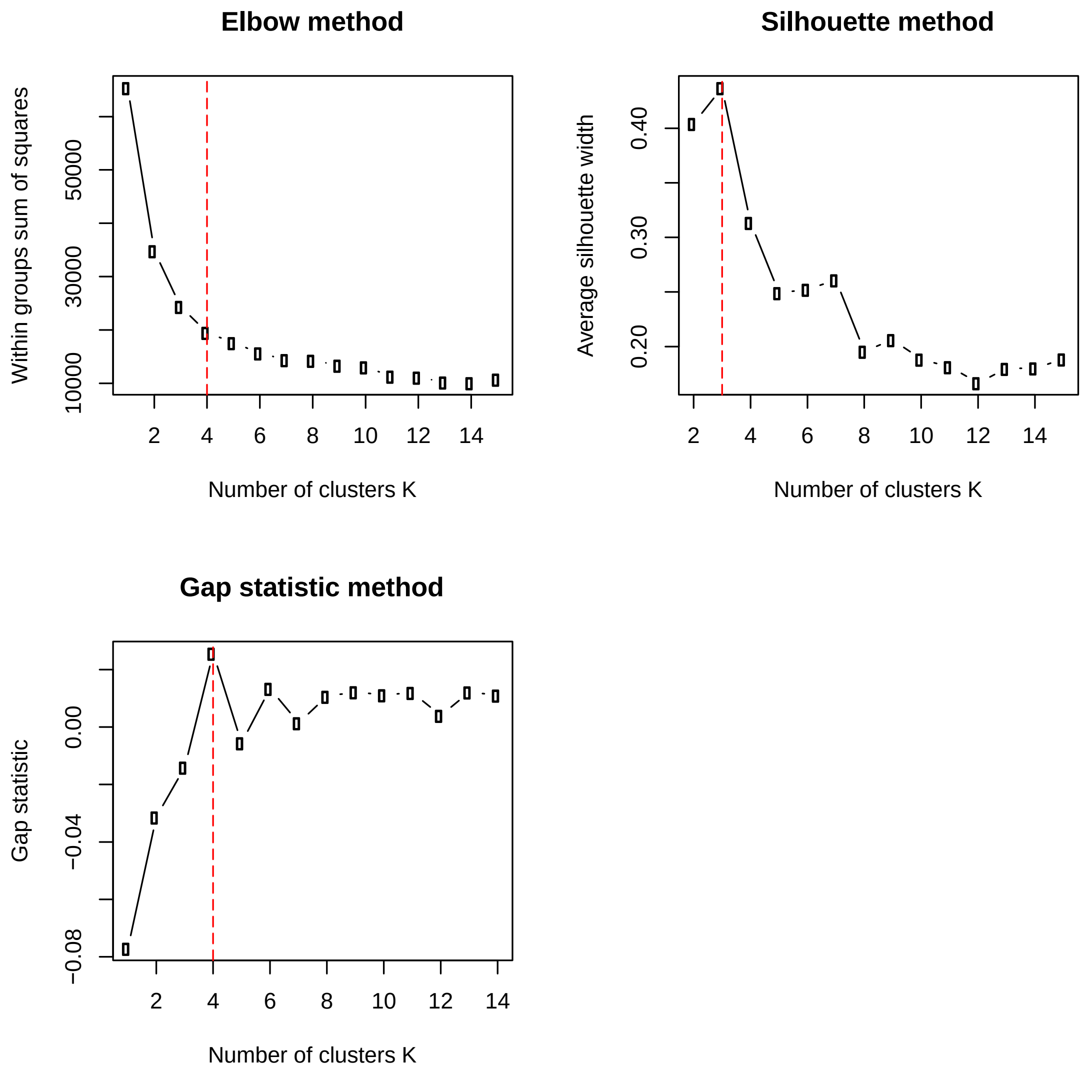


Figure 3.5 Illustration of elbow, silhouette and gap statistic methods.

Finally, it is worth mentioning that the cluster centre in K-Means algorithm, i.e. the centroid, is a fictitious notion, which is just the average of data points in the cluster. However, in some applications it is desirable to have the cluster centre to be one of the data points itself. This is where the *K-Medoids* (also known as *PAM – Partitioning Around Medoids*) comes in the picture (Kaufman and Rousseeuw 1990a).

PAM is a clustering algorithm related to K-Means and aims to partition a data set of N objects into K clusters, and each cluster is represented by one of the data points in the cluster. These representative points are called *medoids* (or exemplars). The PAM method searches for K representative objects for which the average distance of each exemplar and all the other members of the cluster is minimal. Thus, these representative objects are considered the most centrally located point in the cluster. PAM is proved to be less sensitive to noise and outliers than K-Means since it uses non-fictitious medoids as cluster centres instead of centroids (which are averages) (Kassambara 2017).

Similar to K-Means, K-Medoids algorithm requires the number of clusters (K) to be pre-defined. Various approaches mentioned earlier (elbow, silhouette, and GAP statistic methods) can help to determine the optimal number of clusters.

The procedure of the K-Medoids algorithm is as follows:

Step 1: Specify the number of clusters (K) to be created.

Step 2: Select randomly K of the N actual data objects being the initial cluster medoids.

Step 3: Assign each object to the cluster with the closest medoid, for example based on the (Euclidean) distance between the object and the medoid.

Step 4: For each medoid m and each data point o associated to that medoid, swap m and o, and recompute the average distance of o to all the data points belonging to m. Select the new medoid o with the lowest average distance. All the objects are reassigned using the updated cluster medoids.

Step 5: Iteratively repeat Steps 3 and 4 until there is no change in the assignments and the medoids no longer move.

Hierarchical clustering

Hierarchical method does not require pre-selecting the number of clusters, and involves the following steps:

Step 1: Given N objects to be clustered, assign each object to a cluster. This results in N clusters from N given objects. Let the distances between clusters be the same as the distance between data objects inside the clusters.

Step 2: Find the closest pair of clusters and merge them into one cluster.

Step 3: Compute the distances between the new cluster and each of the old clusters.

Step 4: Repeat Steps 2 and 3 until all objects are grouped into a single cluster of size N.

This procedure demonstrates *agglomerative hierarchical clustering* as clusters are merged iteratively. *Divisive hierarchical clustering* does the reverse process by starting with grouping all objects into a single cluster and then repeatedly splitting them into smaller clusters. In practice, divisive hierarchical clustering method is less common.

The result of hierarchical clustering method is a tree-based representation of the objects, known as the *dendrogram*. Observations can be partitioned into groups by cutting the dendrogram at a desirable height. An example of a dendrogram is plotted in Figure 3.6. Here, if the researcher decides to cut the dendrogram at the height of, say 180, the clustering algorithm derives two big clusters enclosed by the green border. If the dendrogram is cut at the height of 80, the clustering algorithm returns four clusters represented by four red rectangles. The lower the height at which the dendrogram is cut, the larger the number of clusters is identified. This example clearly demonstrates the fact that selecting the optimal number of clusters in hierarchical clustering method is somewhat subjective, and the interpretability of cluster results should be thoughtfully considered.

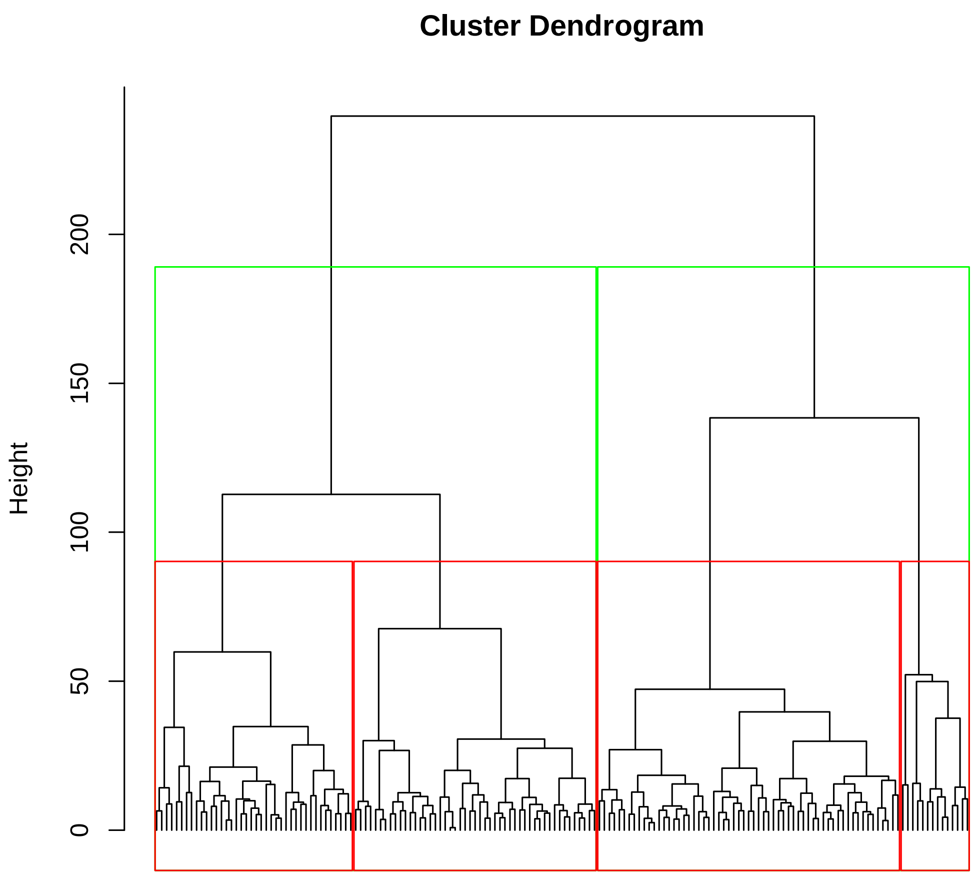
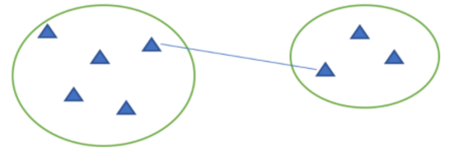
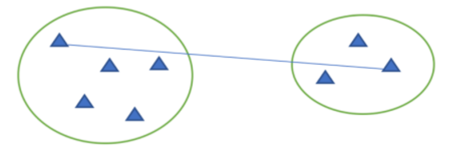
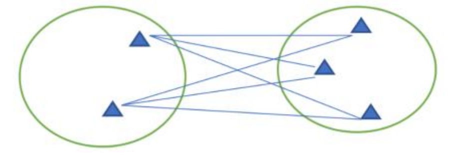
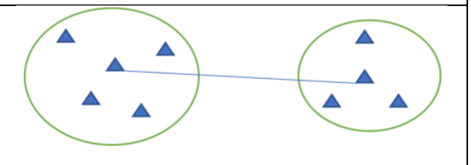


Figure 3.6 Example of a dendrogram in hierarchical clustering.

Given an inter-object dissimilarity matrix, hierarchical clustering methods can start and at each stage in the process merge individuals or groups of individuals formed earlier that are closest. Yet, a challenge arises in Step 3 where the distance between an object and a group of several objects or between two groups of objects needs to be computed. Various techniques (known as *linkage schemes*) propose different ways to define this distance. Table 3.3 lists the common linkage schemes along with their computations of the distance between the two groups. In the examples given in the last column of Table 3.3, the blue triangles represent the data objects, the circle around them denotes the cluster that they belong to, and the solid lines show how the distance between two clusters is calculated.

Table 3.3 Common linkage methods for hierarchical clustering algorithm.

Linkage methods	Distance between two clusters is	Visual illustration
Single linkage (or nearest neighbour)	The distance between the closest members of the two clusters	
Complete linkage (or furthest neighbour)	The distance between the members that are the furthest apart (most dissimilar)	
Average linkage	The average of all distances between all pairs	
Centroid linkage	The distance between mean vectors (centroids)	
Ward's method (or minimum variance)	The increase in sum of squares within clusters	

Pros and cons of non-hierarchical and hierarchical clustering techniques

For the purpose of comparison, only the most widely-used non-hierarchical algorithm, K-Means, is mentioned in the subsequent discussion. However, many characteristics of K-Means method to be discussed also apply for K-Medoids procedure.

It is shown that the K-Means method is more flexible and performs better with a large data set than hierarchical methods (Everitt *et al.* 2011). The latter has a main disadvantage that once objects are merged with others into a cluster, they cannot be removed from that cluster.

However, K-Means algorithm depends strongly on the starting selected centres because they are based on iterative procedures. Thus, running the K-Means algorithm twice on the same data set with different starting centres may result in two different solutions. The less clear the hidden data structure, the larger the difference between two solutions. From this point of view, K-Means is an unstable algorithm. The reason is related to the possibility of finding at each run only a local and unstable solution rather than a global one – 'local optima' problem (Jain *et al.* 1999). Another limitation with K-Means is that the number of clusters is required in advance, which is often unrealistic (Jain *et al.* 1999; Buttrey and Karo 2002).

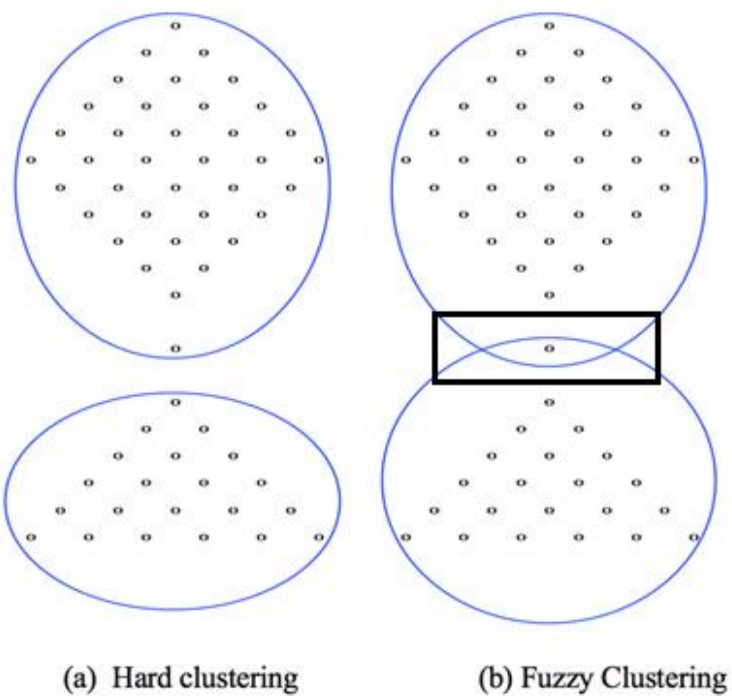
3.3.2 Fuzzy clustering

Hard clustering versus soft clustering

In terms of defining the cluster boundary, clustering techniques can be either hard or soft. In hard clustering (or *crisp clustering*), an object belongs to one and only one cluster (Figure 3.7a). In soft clustering (or *fuzzy clustering*), each object has a probability of belonging to each cluster, and thus, each object can belong to many clusters (as shown in Figure 3.7b) with different *membership degrees* ranging between 0 and 1. Points close to the cluster centre may belong to the cluster at a higher membership degree than points near the edge of a cluster.

K-Means and K-Medoids, which are crisp clustering methods meaning that the membership degree of an observation belonging to a cluster is one and of an observation not belonging to a cluster is zero, can be regarded as special cases of the Fuzzy K-Means and Fuzzy K-Medoids (also known as Fuzzy C-Means and Fuzzy C-Medoids).

Fuzzy clustering was first introduced by Bezdek (1974), Bezdek (1981) and Dunn (1973), and the literature on fuzzy clustering has been considerably expanded thence. Fuzzy K-Means is one of the most widely used fuzzy clustering algorithms, and its procedures are quite similar to K-Means method.



Note: each black circle represents an object, and the blue circles represent the clusters.

Figure 3.7 Hard clustering versus fuzzy clustering.

Fuzzy K-Means

Fuzzy K-Means aims to conduct a fuzzy partition of objects into a pre-defined number of clusters. Similar to K-Means method, Fuzzy K-Means also requires the number of clusters (K) to be pre-determined by the researcher. Unlike K-Means algorithm in which a data object can either wholly belong to a cluster or not, each data object in Fuzzy K-Means may belong to many clusters according to its membership degrees.

For a set of N objects and K clusters, Fuzzy K-Means algorithm attempts to minimise the following function:

$$\sum_{j=1}^{K}\sum_{i=1}^{N} u_{ij}^{m}\, d^2\left(x_i, c_j\right) \tag{3.8}$$

where x_i is the data vector; c_j is the centroid of the *j*-th cluster; $u_{ij} \geq 0$ for all $i = 1, 2, \dots, N$; and $\sum_{j=1}^{K} u_{ij} = 1$. The membership degrees u_{ij} are unknown; $d(x_i, c_j)$ denotes the distance between the data point and the cluster centre; m is the 'fuzzifier' (also known as *parameter of fuzziness*) and affects the final membership distribution. In the case of crisp clustering, $m = 1$.

The membership degree is inversely related to the distance to the cluster, and is defined by the following equation:

$$u_{ij} = \frac{1}{\sum_{p=1}^{K} \frac{d(x_i,c_j)}{d(x_i,c_p)}^{\frac{2}{m-1}}} \tag{3.9}$$

The cluster centroid in Fuzzy K-Means method is the mean of all points weighted by their degree of belonging to the cluster, as follows:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{m} x_i}{\sum_{i=1}^{N} u_{ij}^{m}} \tag{3.10}$$

Details of the algorithm are listed in the studies by Hathaway and Bezdek (1988), Yang (1993), and Baraldi and Blonda (1999).

The steps involved in Fuzzy K-Means method are summarised as below:

Step 1: Select the number of clusters (K).

Step 2: Start with a set of cluster centroids c_j $(j = 1, 2, \dots, K)$.

Step 3: For all data vectors x_i $(i = 1, 2, \dots, N)$ and all cluster centroids c_j $(j = 1, 2, \dots, K)$: calculate the distance between the data point and the cluster centroid: $d(x_i, c_j)$.

Step 4: Compute the membership degree u_{ij} using equation (3.9). Note that $u_{ij} \geq 0$ indicating the degree of association of data x_i with the centroid c_j, and u_{ij} is reversely related to the distance $d(x_i, c_j)$ so that: if $d\left(x_i, c_j\right) < d(x_i, c_p)$, then $u_{ij} > u_{ip}$ (with $p \in \{1, \dots, K\}$ and $p \neq j$).

Step 5: Recalculate the cluster centroids c_j using equation (3.10).

Step 6: Repeat Steps 3-5 until the membership degrees no longer change.

Fuzzy K-Medoids

Fuzzy K-Medoids (Krishnapuram *et al.* 1999; Krishnapuram *et al.* 2001) is another popular fuzzy clustering algorithm. For a set of N objects and K clusters, Fuzzy K-Medoids algorithm attempts to minimise the following objective function:

$$\sum_{j=1}^{K}\sum_{i=1}^{N} u_{ij}^{m}\, \tilde{d}^2(x_i, c_j)\ ;\ \sum_{j=1}^{K} u_{ij} = 1\ ;\ u_{ij} \geq 0 \tag{3.11}$$

where x_i is the data vector; c_j is the medoid (a representative data object) of the *j*-th cluster; u_{ij} is the membership degree of the *i*-th object to the *j*-th cluster; m is the fuzzifier. $\tilde{d}(x_i, c_j)$ denotes the distance between the data point and the cluster medoid.

The objective function (3.11) can be solved by the means of Lagrangian multiplier method, the solutions are as follows:

$$u_{ij} = \frac{1}{\sum_{p=1}^{K} \frac{d(x_i,c_j)}{d(x_i,c_p)}^{\frac{2}{m-1}}} \tag{3.12}$$

Comparing the two widely used fuzzy clustering methods (Fuzzy K-Means and Fuzzy K-Medoids), some comments can be made as below:

(i) Following Fuzzy K-Medoids, each cluster is represented by an observed exemplar object (*medoids*) instead of the fictitious representative (*centroids*, which are simply the average) in Fuzzy K-Means. The possibility of obtaining a non-fictitious representative offers more appeals in the interpretation of the final cluster results (Kaufman and Rousseeuw 2009).

(ii) Fuzzy K-Medoids algorithm is usually invariant with regard to the order in which the objects are presented. This feature does not apply to a whole host of other algorithms in the existing literature (Kaufman and Rousseeuw 2009).

(iii) Fuzzy K-Medoids method is slightly more robust than Fuzzy K-Means and is more resistant to the presence of noise in the data (Garcia-Escudero and Gordaliza 1999; Estivill-Castro and Yang 2004; Kaufman and Rousseeuw 2005; García-Escudero *et al.* 2010). This is because a medoid is less affected by the presence of noise or other extreme values than a centroid.

(iv) Solving equations (3.11) and (3.12) involves an exhaustive search for the medoids, which can be computationally intensive for a large data set (Maharaj *et al.* 2019).

(v) Since medoids always have the membership of 1 in the cluster, raising this membership value to the power of m is always equal to 1. Hence, the mobility of the medoids can be lost when

m is large. For this reason, Kamdar and Joshi (2000) suggest a value between 1 and 1.5 for m.

Advantages of fuzzy clustering

In the literature, many authors put forward different reasons for adopting fuzzy clustering approach (D'Urso 2004; Hwang *et al.* 2007). Major motivations include the following:

(i) Fuzzy clustering algorithm is a distribution-free procedure (Maharaj *et al.* 2019).

(ii) The overlapping classification of fuzzy clustering seems more realistic than the deterministic classification of hard clustering, reflecting the challenge of identifying a clear boundary between clusters in real world applications (McBratney and Moore 1985; Wedel and Kamakura 2012).

(iii) Fuzzy clustering models are computationally more efficient because dramatic changes in the value of cluster membership are less likely to occur during estimation procedures (McBratney and Moore 1985). Additionally, fuzzy clustering is shown to be less affected by local optima problems (Heiser and Goenen 1997; Menard and Eboueya 2002).

(iv) The memberships for any given set of observations indicate whether there is a second-best cluster almost as good as the best cluster – a feature that crisp clustering methods cannot provide (Everitt *et al.* 2011).

(v) Another advantage of fuzzy clustering over crisp clustering methods is that the cluster membership proportions can be conveniently combined with other information. For example, the memberships can serve as weights in calculating the weighted averages (Khoo-Lattimore *et al.* 2019), or they can be considered as probabilities in the formula related to Bayes' theorem (Everitt *et al.* 2011).

3.4 Clustering methods for time series

3.4.1 Clustering approaches

Time series clustering aims to identify similarities in the patterns across time. Due to the expansion of time series data across various assortments of disciplines, time series clustering attracts a growing interest of researchers and particularly the past two decades witness a considerable number of contributions to the topic.

Facing time series data, the researcher has two options. The first one is to divide the data set into discrete time points, to perform the conventional clustering algorithms on the derived static data sets, and then to compare the cluster results (see, *for example*, Di Lascio and Disegna 2017). Another approach is to perform the clustering on whole time sequences. Following the latter approach, the extensive literature examines the clustering task for both univariate and multivariate time series.

The algebraic representation of a univariate time series is formalised as:

$$X = \{x_{it}: i = 1, \dots, I; t = 1, \dots, T\} \tag{3.13}$$

For multivariate time series (or *multivariate time trajectories*), the data usually imply a three-way structure 'units × variables × times', and can be arranged in a data time array as follows:

$$X = \{x_{ijt}: i = 1, \dots, I; j = 1, \dots, J; t = 1, \dots, T\} \tag{3.14}$$

where i, j, and t denote the object, quantitative variable, and time respectively. The observation x_{ijt} represents the j-th quantitative variable observed for the i-th object at time t. The multivariate time series data matrix for the i-th object is as follows:

$$X_i = \begin{bmatrix} x_{i11} & \cdots & x_{ij1} & \cdots & x_{iJ1} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{i1t} & \cdots & x_{ijt} & \cdots & x_{iJt} \\ \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{i1T} & \cdots & x_{ijT} & \cdots & x_{iJT} \end{bmatrix} \tag{3.15}$$

The geometric representations of time data array for univariate and multivariate time series are provided in Figure 3.8.

Take the case of national food consumption. A researcher who wishes to explore different food consumption patterns around the world could exploit a univariate time series data set of the daily total calories per person for 200 countries over the past 20 years. Here, the data matrix becomes $X = \{x_{it}: i = 1, \dots, 200; t = 1, \dots, 20\}$, where i and t denote the country and time index respectively. Figure 3.8a shows 200 lines in a 2D plane. Each line represents the calorie consumption for each country and the horizontal axis shows the time (20-year period).

Otherwise, he/she could perform the same task but from the multivariate time series perspective. In this case, the 'units × variables × times' structure could be 'countries × energy supply from macronutrients (protein, fat, carbohydrate) × years'. Algebraically, the data matrix takes the form $X = \{x_{ijt}: i = 1, \dots, 200; j = 1, \dots, 3; t = 1, \dots, 20\}$ where i, j and t are indices for countries, energy sources, and time respectively. For illustration, different time series would be plotted on a 3D plane as in Figure 3.8b and three axes represent the time (20 years), units (200 countries) and variables (3 macronutrients).

Commentaries on historical developments and taxonomy of time series clustering methods can be found in studies by Corduas (2010), Fu (2011), Esling and Agon (2013), Liao (2005), Aghabozorgi *et al.* (2015), Caiado *et al.* (2015), D'Urso *et al.* (2016), and Bagnall *et al.* (2017). Broadly speaking, the clustering task of univariate and multivariate time series could be adopted by three main approaches shown in Figure 3.9.

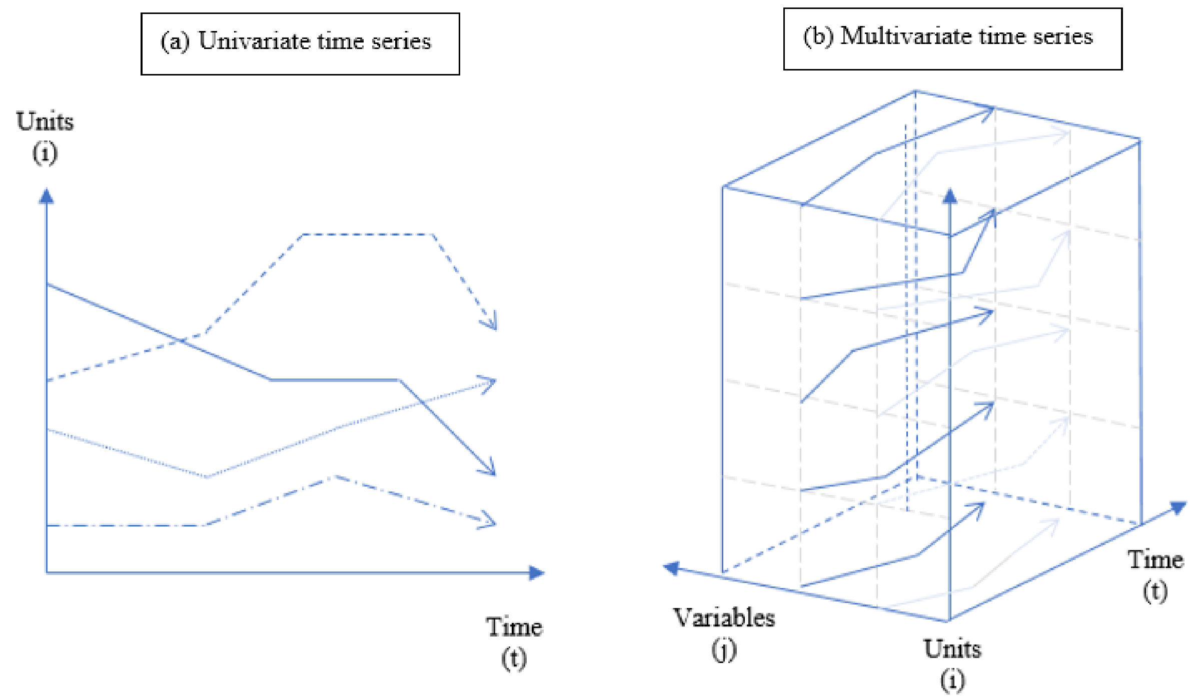


Figure 3.8 Geometric representation of data array for univariate and multivariate time series.

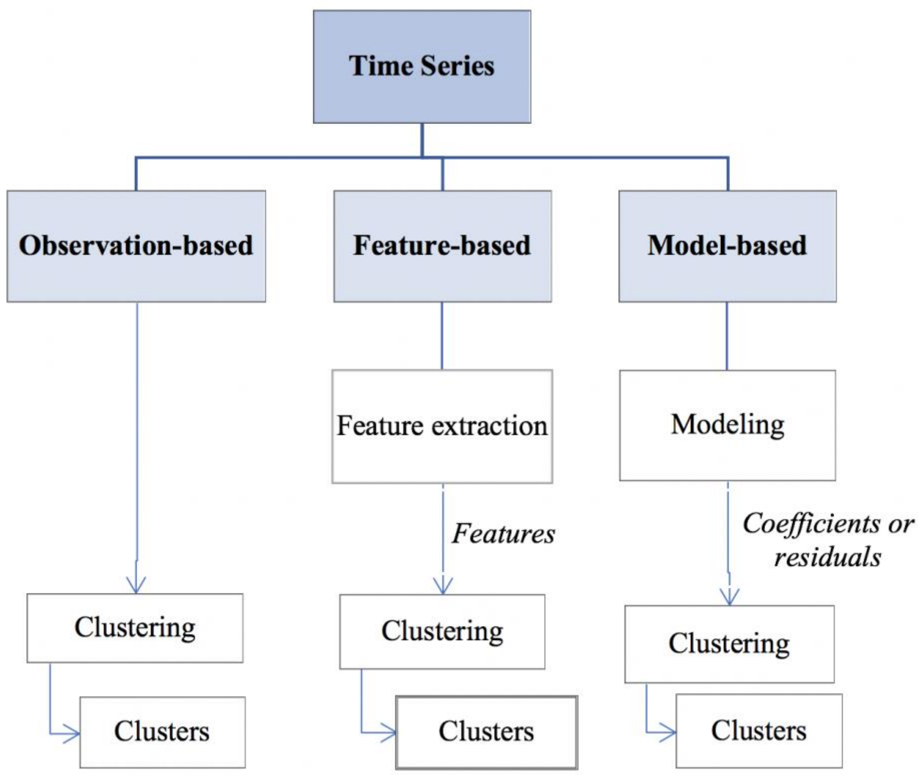


Figure 3.9 Time series clustering approaches.

Observation-based clustering

Following this approach, conventional clustering algorithms such as K-Means and hierarchical clustering are applied directly on the observed time series or their transformations, and clusters are identified via a suitable distance measure. This approach is particularly useful with short time series (D'Urso and Vichi 1998; Coppi and D'Urso 2003; D'Urso 2004; D'Urso 2005; Coppi and D'Urso 2006; Izakian *et al.* 2015).

Feature-based clustering

In this case, clustering algorithms are performed on suitable features derived from the original time series. Several features have been selected and can be categorised into:

- Time domain features including autocorrelation function (Alonso and Maharaj 2006; D'Urso and Maharaj 2009), inverse autocorrelation function (Caiado *et al.* 2006), quantile autocovariance function (Lafuente-Rego and Vilar 2016; Vilar *et al.* 2018);
- Frequency domain features such as periodogram (Caiado *et al.* 2009), coherence (Maharaj and D'Urso 2010), and cepstral (Maharaj and D'Urso 2011);
- Wavelet features (Wang *et al.* 2008a).

Model-based clustering

This approach assumes that time series generated from the same underlying statistical model exhibit similar patterns. Clustering algorithms are performed on the parameter estimates or on the residuals of the fitted models. Several models have been considered in the extensive literature, including:

- ARMA and ARIMA (Maharaj 1996; Kalpakis *et al.* 2001; D'Urso *et al.* 2013);
- GARCH (Caiado and Crato 2010; D'Urso *et al.* 2013; D'Urso *et al.* 2016);
- Extreme values (D'Urso *et al.* 2017);
- Kernel (Mikalsen *et al.* 2018);
- Tail dependence and copulas (De Luca and Zuccolotto 2011; Di Lascio and Giannerini 2012; Durante *et al.* 2014, 2015; De Luca and Zuccolotto 2017; Di Lascio and Disegna 2017; Zhang and An 2018).

3.4.2 Fuzzy approach in time series clustering

In time series clustering literature, growing attention has been paid to fuzzy clustering approach. Not only does it inherit the advantages of fuzzy clustering for static data, but it also captures the drifting or switching nature of time series (D'Urso *et al.* 2018). To illustrate this switching nature, a multivariate time series could have a dynamic pattern consistent with a cluster for a certain time period, and then a completely different dynamic pattern which is more similar to another cluster in the following time period.

Besides, D'Urso (2005) indicates another motivation for using fuzzy approach for time series data, which is greater adaptability in defining the *prototype* time series. A *'prototype'* is the object whose characteristics are exemplar of its cluster. Regarding static data, the prototype can simply be the centroid. The notion of a centroid multivariate time series is more complicated to comprehend and to compute. The fuzzy approach is more powerful when the observed time series do not show too much different patterns from each other. In this case, the fuzzy approach allows the possibility to single out the underlying structures, if these are likely to exist in the given set of time series. These motivations are justified in the studies by Coppi and D'Urso (2003), D'Urso (2004), D'Urso and Maharaj (2009), Maharaj and D'Urso (2010, 2011), and D'Urso *et al.* (2016).

3.4.3 Dissimilarity/distance measures

In the case of time series clustering, the dissimilarity concept is more complicated due to the dynamic nature of time dependent data. Conventional dissimilarity measures often fail to address the interdependence relationship between values, thus are not adequately effective (Montero and Vilar 2014; Di Lascio and Disegna 2017). Realising this shortcoming, earlier researchers have proposed different approaches to define an appropriate dissimilarity measure between time series. Following Montero and Vilar (2014), the distance measures in time series clustering methods can be classified into four classes: feature-based, model-based, complexity-based, and prediction-based. Table 3.4 compares these four types of distance measures.

Even though there are previous attempts to compare existing dissimilarity measures (see, *for instance*, Keogh and Kasetty 2003; Pertega and Vilar 2010), the most appropriate measure should be chosen according to the nature of the clustering and the specific purpose of the grouping (Montero and Vilar 2014). This is to ensure the cluster solution to be meaningfully interpreted in terms of the grouping target.

Table 3.4 Four types of distance measures in time series clustering.

Distance measures	Rationales	Studies	Advantages
Feature-based	Compares serial features extracted from the original time series	Galeano and Peña (2000), Caiado *et al.* (2006)	• Allows dimensionality reduction • Features can be obtained in a straightforward manner
Model-based	Assumes that time series are generated by specific underlying models and evaluates the dissimilarity between fitted models	Piccolo (1990), Maharaj (1996)	• Can detect clusters of different sizes • Cluster results are independent of variable scaling
Complexity-based	Compares the levels of complexity of time series	Keogh *et al.* (2007), Batista *et al.* (2011)	Assesses the level of shared information between time series
Prediction-based	Two time series are similar if their forecasts for a specific future time are close	Alonso *et al.* (2006), Vilar *et al.* (2010)	More practical if the real interest of clustering lies in the properties of predictions

3.4.4 Dynamic Time Warping distance

Dynamic Time Warping (DTW) algorithm was first proposed in 1970 in the context of voice recognition (Velichko and Zagoruyko 1970) but has been increasingly employed in time series clustering context as a more suitable alternative to the traditional Euclidean distance (Izakian *et al.* 2015). Given two time series, the DTW locally stretches or compresses them to make one resemble the other as much as possible (Berndt and Clifford 1994). After stretching, the distance between the two series is calculated by summing the distances of individual aligned elements.

Following this rationale, the function that allows to remap each time series needs to be specified. This function is called *warping function* and aims to realign the time indices of the time series.

Given a 'query' (or test) series $X = (x_1, x_2, \dots, x_Q)$ and a 'reference' series $Y = (y_1, y_2, \dots, y_S)$, with length $Q\ and\ S$ respectively, the total distance between the two time series is computed by the 'warping path'. The warping path allows to compare each element in X with the closest element in Y, and is defined as:

$$\phi(l) = (\phi_x(l),\ \phi_y(l)) \text{ with } \begin{cases} l = 1, 2, \dots, L \\ \phi_x(l)\ \in \{1, \dots, Q\} \\ \phi_y(l) \in \{1, \dots, S\} \end{cases} \tag{3.16}$$

under two constraints:

(i) Boundary condition: $\phi(1) = (1, 1)$ and $\phi(L) = (Q, S)$;

(ii) Monotonicity condition: $\phi_x(1) \leq \cdots \leq \phi_x(l) \leq \cdots \leq \phi_x(L)$,
and $\phi_y(1) \leq \cdots \leq \phi_y(l) \leq \cdots \leq \phi_y(L)$.

The warping functions ϕ_x and ϕ_y remap the time indices of X and Y respectively. Given ϕ, the average accumulated distortion between the two *warped* time series is computed as:

$$d_\phi(X, Y) = \sum_{l=1}^{L} d(\phi_x(l),\ \phi_y(l)) m_\phi(l) / M_\phi \tag{3.17}$$

where $m_\phi(l)$ is a per-step weighting coefficient, M_ϕ is the corresponding normalisation constant which ensures that the accumulated distortions are comparable along different paths, and $d(.,.)$ is usually the Euclidean distance (Giorgino 2009). As there are multiple warping paths, the DTW algorithm seeks to find the optimal warping function $\hat{\phi}(l) = \left(\hat{\phi}_x(l),\ \hat{\phi}_y(l)\right), (l = 1, \ldots, L)$ such that:

$$D(X, Y) = \min_\phi d_\phi(X, Y) \tag{3.18}$$

Or in other words, the equation (3.18) shows that one should pick the deformation of the time indices of X and Y which brings the two series as close to each other as possible. The distance $D(X, Y)$ in equation (3.18) is the DTW distance, or the ‘minimal global dissimilarity’. A simple example is given in Figure 3.10 to illustrate the idea of aligning two time series. A query series S is represented by the black solid line (left axis), a reference series T is represented by the red dashed line (right axis), and the blue dashed line shows the mapping between points of time series T and S.

DTW distance has been extensively used to compare the (dis)similarity between time series thanks to its various benefits over the traditional distance like Euclidean distance. Many of the advantages, which are mentioned by D’Urso *et al.* (2019a), are indicated subsequently. First, DTW distance preserves the time ordering of the sequence (due to the monotonicity condition in constructing the warping path), thus DTW distance goes beyond the instantaneous features of time series by incorporating both the instantaneous and variational characteristics of time trajectories. In this regard, not only the instantaneous position (i.e. the levels) of the time trajectories is considered, but also the varying rates at which the phenomena change over time. Second, DTW distance can be computed for time series of different lengths – a feature that Euclidean distance is not capable of (Wang *et al.* 2019). Third, no assumptions are required regarding the data distribution (for example, stationary or non-stationary), hence DTW distance can be directly calculated for raw data. Next, DTW distance compares time series based on a many-to-one (and vice versa) instead of a one-to-one manner adopted by Euclidean distance (Giorgino 2009). Finally, DTW distance can recognise time series of the same shapes even after they are transformed by shifting or scaling (Cassisi *et al.* 2012). This feature is illustrated in Figure 3.10: the warping paths harmoniously indicate time series S with the shape akin to that of the series T but being shifted to the right. Despite all these flexibilities, DTW is more computationally demanding than Euclidean distance. In general, DTW distance is often combined with K-Medoids, fuzzy K-Medoids and hierarchical methods in time series clustering (see, *for instance*, Izakian *et al.* 2015; D'Urso *et al.* 2018; Lee *et al.* 2018; Liu *et al.* 2018; D'Urso *et al.* 2019a; D'Urso *et al.* 2019b).

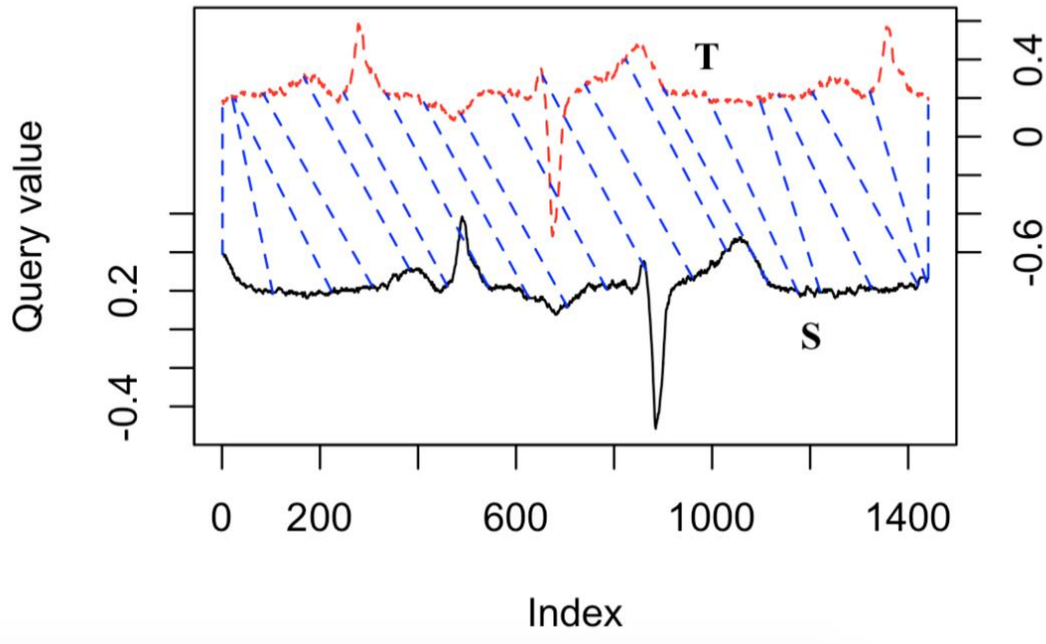


Figure 3.10 Aligning two time series.

3.4.5 Copula theory

For random variables X and Y with continuous marginal cumulative distribution functions $F_X(x)$ and $F_Y(y)$ respectively, let the joint cumulative distribution function be denoted by $F(x, y)$. Copula function is closely related to Sklar's theorem (Sklar 1959), according to which there is a unique copula function $C(u, v)$ that connects $F(x, y)$ with $F_X(x)$ and $F_Y(y)$ as:

$$F(x, y) = C(F_X(x), F_Y(y)) \tag{3.19a}$$

$$\text{or equivalently } C(u, v) = F(F_X^{-1}(u), F_Y^{-1}(v)) \tag{3.19b}$$

The copula function $C(u, v)$ is therefore said to be able to capture the dependence structure between the random variables X and Y.

In the multivariate context, the Sklar's theorem states that every joint distribution function $F(.)$ can be expressed in terms of K marginal distribution function F_k and the copula function C as:

$$F(x_1, \dots, x_k, \dots, x_K) = C(F_1(x_1), \dots, F_k(x_k), \dots, F_K(x_K)) \tag{3.20}$$

for all $(x_1, \dots, x_k, \dots, x_K) \in \bar{R}^K$, with $\bar{R}$ denoting the extended real line. Equation (3.16) suggests that any joint probability function $f(.)$ can be split into the margins and the copula, and that the copula represents the structure of association between random variables. Because of this unique feature, the copula has been widely adopted in various disciplines to capture the dynamic patterns of time series. There is a vast and growing literature on copulas that explores the dependence structure between different financial markets (Bedoui *et al.* 2019), the co-movements among stock prices (Fenech and Vosgha 2019; Xiao 2020) or the cross-market transmission of shocks in finance (Jin *et al.* 2020; Yang *et al.* 2020). In food economics, copulas have been adopted to investigate the co-movements of crop price during extreme market circumstances (Zimmer 2016) and to study the risk spill-overs between energy and agricultural commodity markets (Ji *et al.* 2018; Adhikari and Putnam 2020). At the micro level, copula-based decomposition methods are used to assess the determinants of nutrition transition in Vietnam regarding the consumption of macronutrients (Trinh Thi *et al.* 2018b).

Traditionally, the concept of *dependence* and *co-movement* is often quantified by a measure of correlation coefficient (often Pearson correlation). However, the Pearson correlation coefficient is only an indication of linear association and this method is inadequate in capturing non-linear dependence structures (Di Lascio and Disegna 2017). Furthermore, the traditional correlation coefficient relies on a stringent assumption that the joint distribution of two variables is multivariate normal, whilst this assumption can violate in practice, for example when assessing the association between financial asset returns which tend to have fat-tailed and skewed distributions (Adhikari and Putnam 2020). Another limitation of the traditional correlation coefficient is that zero correlation does not imply independence (Di Lascio and Giannerini 2012). For two random variables X and Y, zero correlation only means null covariance $cov[X, Y] = 0$, whilst the condition $cov[\varphi_1(X), \varphi_2(Y)] = 0$ for any functions $\varphi_1(.)$ and $\varphi_2(.)$ is needed for a conclusion of zero dependence. Copulas on the other hand offer a flexible and more robust measure of dependence and permit asymmetric or non-linear dependences. In addition, copulas are not affected by strict monotonic transformations (logarithmic and exponential) while the traditional correlation coefficient is not invariant under such transformations (Durante and Sempi 2016). The invariance property means that the random variables and their logarithms have the same copula.

Some copulas play a major role in measuring the co-movement or dependence between random variables: the Fréchet-Hoeffding copula bounds. For any copula $C(u, v)$, the following bounds hold:

$$W(u, v) \leq C(u, v) \leq M(u, v) \tag{3.21}$$

where $W(u, v) = max\{u + v - 1, 0\}$ and $M(u, v) = \min\{u, v\}$ are the lower and upper bound respectively, and these bounds correspond to the cases of extreme dependency. Specifically, the Fréchet-Hoeffding upper bound M is the co-monotonic copula and corresponds to the perfect positive dependence ('co-monotonicity'), whereas the lower bound W is the counter-monotonic copula and corresponds to the perfect negative dependence ('counter-comonotonicity').

In time series analysis, clustering methods have been employed to identify similarities among time series and can be either dissimilarity-based or model-based. Methods in the former category classify time series in the same cluster on the basis of (dis)similarity which is usually reported in a dissimilarity matrix and computed by a proper dissimilarity measure. A conventional dissimilarity measure is derived from the (Pearson) correlation coefficient. The main idea is to consider a sort of correlation coefficient between time series so that high positive correlation may be translated into some degree of similarity between the time series under consideration. However, this approach is subject to several abovementioned shortcomings of using the linear correlation coefficient as a dependence measure. Hence, there is a growing body of research that supports the application of a more meaningful dependence measure – the copula function (Durante *et al.* 2014; Disegna *et al.* 2017; Zhang and An 2018).

Following the copula-based framework, the dissimilarity matrix depends on the copula C_{ij} associated with two time series x_i and x_j. If two time series are perfectly positively dependent, or *co-monotonic*, $C_{ij} = M$ where M is the comonotonic copula (Equation 3.21) and the dissimilarity measure between x_i and x_j should be zero. Otherwise, the dissimilarity between two time series can be interpreted as the departure from the co-monotonic copula as follows: $\|C_{ij} - M\|$ where $\|.\|$ is a suitable norm. Once the dissimilarity matrix is obtained, a conventional clustering algorithm (for example, hierarchical clustering, K-Means, K-Medoids or the fuzzy variants) can be performed to identify the final clusters.

Unlike dissimilarity-based methods, model-based clustering techniques assume that each time series is generated by a specific underlying model or a mixture of underlying probability distributions. Time series are grouped together when the underlying models characterising them are similar. Various models have been adopted in the literature, including copulas (Di Lascio and Giannerini 2012; Di Lascio and Disegna 2017). Let $X_i = (X_{i1}, \dots, X_{iT_i})^T$ be the *i*-th time series ($1 \leq i \leq n$) where T_i is its length and n is the number of time series. Model-based clustering approach assumes that these time series are drawn from K clusters and the time series in each cluster share a common dependence pattern. For a positive integer l, $(X_{it}, X_{i(t+l)})$ and $(X_{jt}, X_{j(t+l)})$ have the same copula function if two time series X_i and X_j belong to the same cluster.

3.5 Spatial clustering

Cluster analysis aims to discover patterns in data by organising a set of N objects into K disjoint <u>unknown</u> clusters. However, in some cases, there is additional information about the types of clusters that are sought in the data and it is therefore relevant to impose constraints on the set of allowable solutions. This is usually referred as *constrained clustering*, where the membership of clusters is determined partly by external information (Everitt *et al.* 2011). A common type of constraints is contiguity constraint (in space). In this regard, objects in a cluster are not only similar to each other, but also required to be contiguous objects. To give an example, clustering technique can be applied to study the incidence of an infectious disease among municipalities within a country and the administrative boundaries between municipalities are used as the spatial contiguity constraint. The clustering results can help to uncover groups of counties which might correspond to disease 'hot spots' or 'cold spots'. Unless the clustering method is explicitly spatial, the geographic relevance might not be sufficiently accounted for (Grubesic *et al.* 2014).

With spatial constraints, one can be either strict or non-strict in clustering procedures. Strict procedures put objects, which are though very similar, into different clusters if they are spatially apart. A commonly used method involves the concept of *contiguity matrix* C which is an $N \times N$ matrix, consisting of elements c_{ij} that $c_{ij} = 1$ if objects *i* and *j* are contiguous, otherwise $c_{ij} = 0$. Hierarchical and non-hierarchical methods can be modified to take this constraint into account, allowing only contiguous objects to be clustered together (Murtagh 1985; Gordon 1996). On the other hand, clustering procedures are non-strict when the constraints are not necessarily neighbourhood. An example of a non-strict method considers a modified dissimilarity matrix which combines the dissimilarity matrix of geographical distances and dissimilarity matrix from non-geographical variables with different weights (Jain and Farrokhnia 1991; Coppi *et al.* 2010; D'Urso *et al.* 2019a).

Clustering with spatial constraints is also useful for finding homogeneous groups in *spatial units* – data that have a spatial component ranging from explicit geographical locations to implicit spatial neighbourhood relations such as topological, socio-economic proximity and direction relations (Bindiya *et al.* 2013). Several clustering models for spatial data have been proposed in the literature and in essence these methods take into account two distinguishing characteristics of spatial data: *spatial dependence* and *spatial heterogeneity* (Anselin 2013). The former refers to the existence of a functional relationship between what happens at one point in space and what happens elsewhere, whereas the latter refers to variation in this relationship over space (LeSage 1998; LeSage and Pace 2009). In addition, spatial data are often characterised by *positive spatial autocorrelation*, meaning neighbouring units tend to have similar features (Gordon 1996).

Following Fouedjio (2016), existing approaches can be grouped into four classes:

(i) Non-spatial clustering with geographical coordinates as additional variables;

(ii) Non-spatial clustering with a spatial dissimilarity measure (Oliver and Webster 1989);

(iii) Spatially constrained clustering (Pawitan and Huang 2003);

(iv) Model-based clustering (Ambroise *et al.* 1997).

3.6 Space-time clustering

As an extension of spatial clustering, space-time clustering technique involves classifying spatial units into clusters based on a set of quantitative features observed at several time occasions. To this aim, a three-way data array characterising 'same objects (spatial units) × same quantitative variables × different times' can be algebraically formalised as follows:

$$X = \{x_{ijt} : i = 1, \dots, I; j = 1, \dots, J; t = 1, \dots, T\} \tag{3.22}$$

where i, j, and t indicate the spatial unit, the quantitative variable, and the time respectively; x_{ijt} represents the j-th variable observed on the i-th spatial unit at time t. Such a three-way data array is usually known as the spatial-time array and can be illustrated geometrically on a three-dimensional space or sliced into different bi-dimensional matrices as in Figure 3.11. The bi-dimensional matrices are given by:

$$X_i = \{x_{ijt} : j = 1, \dots J; t = 1, \dots T\} \qquad \text{('space' slice)} \tag{3.23a}$$

$$X_j = \{x_{ijt} : i = 1, \dots I; t = 1, \dots T\} \qquad \text{('variable' slice)} \tag{3.23b}$$

$$X_t = \{x_{ijt} : i = 1, \dots I; j = 1, \dots J\} \qquad \text{('time' slice)} \tag{3.23c}$$

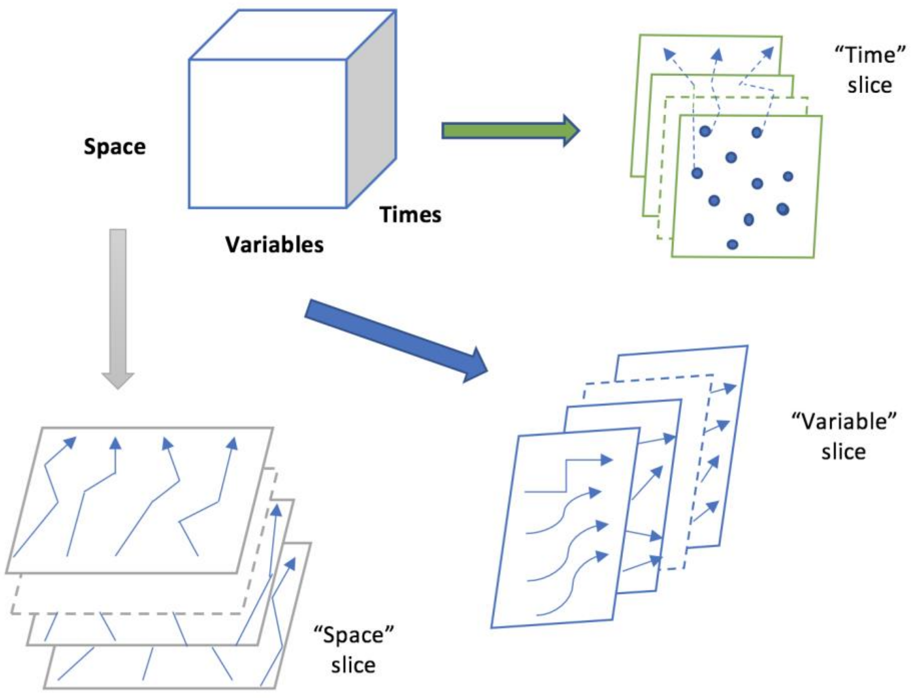


Figure 3.11 Representations of the spatial-time array.

The example of national food consumption again provides a good illustration. The researcher could explore the similarities in the evolution of total calorie intakes over time in the space-time clustering framework. The three-way data array can be arranged as 'countries × calorie intake × years'.

Algebraically, the data matrix takes the form $X = \{x_{ijt}: i = 1, \ldots, 200; j = 1; t = 1, \ldots, 20\}$ where i, j and t are indices for countries, caloric consumption, and time respectively. The form of the data matrix X looks identical to that from the time series clustering approach, however the i units here are spatial units which are related to each other by some kinds of geographical, economic, social, cultural, or institutional proximity.

Realising the complexity of space-time data, Coppi *et al.* (2010) advocate three main important considerations associated with space-time clustering: (i) the spatial nature of the objects, (ii) the characteristics of the multivariate time trajectories, (iii) the uncertainty of assigning a spatial unit to a specific cluster. The first two issues originate from the high dimensional aspect of space-time data and require to not only incorporate the spatial information into the clustering procedure but also define a suitable dissimilarity measure between time trajectories. The last issue can be resolved by the application of fuzzy clustering technique, which allows a certain degree of flexibility in assigning the spatial units to clusters.

In order to tackle the high-dimensional nature of spatial-temporal data set, data reduction techniques can first be employed and traditional clustering algorithms can then be applied on a two-mode matrix. There are several streams in the literature favouring this reduction approach. Krishnapuram and Freg (1992) and Shekhar *et al.* (2015) propose to add a new column containing the relationships between time and space to the traditional two-mode matrix. Athanasopoulos *et al.* (2009) introduce a hierarchical time series clustering method where the clustering task is performed at different spatial levels. Nonetheless, data reduction can lead to information loss, highlighting the need to develop ad-hoc clusteirng techniques which simultaneously incorporate spatial and temporal information.

In the same spirit as Disegna *et al.* (2017), existing space-time clustering techniques can be classified into four categories:

(i) Non-spatial time series clustering with a spatial dissimilarity measure (Izakian *et al.* 2013)
(ii) Time series clustering with spatial contiguity constraints (Hu and Sung 2006; Coppi *et al.* 2010; Qin *et al.* 2011; D'Urso *et al.* 2019a; D'Urso *et al.* 2019c);
(iii) Density-based clustering (Ester *et al.* 1996; Wang *et al.* 2016; Cheng *et al.* 2018);
(iv) Model-based clustering (Basford and McLachlan 1985; Viroli 2011; Torabi 2016; Disegna *et al.* 2017; Paci and Finazzi 2018).

3.7 Cluster analysis in food economics

The application of cluster analysis spans over a wide range of disciplines, and food economics is no exception. Traditionally, cluster analysis proves its usefulness for policymaking purposes. To illustrate, agriculture is a fundamental instrument for economic development and poverty reduction in many countries; however, the way agriculture works for development varies across countries depending on how they rely on agriculture as a source of economic growth. Designing policies most suited to each country's economic conditions is hence crucial to best implement agriculture-for-development agendas. Based on the contribution of agriculture to growth and the importance of rural poverty, the World Bank classifies developing countries as agriculture-based, transforming, or urbanised (World Bank 2007). This typology of countries provides a useful framework to formulate policy guidance. In agriculture-based economies where agriculture contributes significantly to growth, the key policy challenge is how to help agriculture play its role of the lead sector for overall growth. In transforming economies, as agriculture is no longer the major generator for economic development, growth in rural nonfarm economy is needed to eradicate rural poverty. In urbanised economies, agriculture, contributing only a little to growth, can help to reduce the rural poverty by including the rural poor as producers and by creating jobs for them.

There are earlier efforts in which identifying a number of country groups can provide some guidance about potential policies to address food and nutrition security. By categorising 167 countries into twelve distinct clusters on the basis of food security measures, Díaz-Bonilla *et al.* (2000) conclude that the WTO's classification of countries into developed, developing, least developed and net food importing developing does not adequately capture food security concerns. While the majority of least developed countries are food insecure, some food insecure countries do not belong to the least developed category. Also, 'developing countries' is not a useful label in terms of food security as these countries are scattered among all categories of food security. Recently, in a similar study 152 countries are classified into ten groups based on their food security profiles (Díaz-Bonilla and Thomas 2015). The authors demonstrate how cluster analysis helps to create typologies of food (in)security conditions and argue that the results depend on the variables selected.

In food economics, cluster analysis has been a useful tool to classify individuals/countries based on the homogeneity of their food consumption patterns. A summary of related studies is provided in Table 3.5. Although identifying population subgroups that share similar dietary patterns is not a new practice, the main focus has been placed on developed countries and the European Union (EU).

As put by Petrovici *et al.* (2005), the concept of dietary patterns comprises two underlying dimensions: nutrient intake (for instance the per capita daily supply of calories, protein, fat) and food consumption (the calorie consumption from specific foods such as cereals, meat, milk, vegetables, and fruits). For policymaking purposes, forming clusters based on the consumption of food groups is a better choice since individuals choose to consume (combinations of) foods, not nutrients (Coulston *et al.* 2017). This can explain why nearly all studies in Table 3.5 employ food groups as clustering variables.

In order to quantify food consumption, dietary intake data are needed, and they can be derived from food consumption surveys at national, household and individual level or similar dietary assessment methods. Further details on different sources of dietary data are described in Section 4.3.4 (Chapter 4). Among these various methods, the Food Balance Sheet (FBS) compiled by the Food and Agriculture Organisation of the United Nations (FAO) is a widely used source of dietary data measuring ‘apparent food consumption’ at national level. Despite some limitations of the FBS especially that it estimates food available for consumption rather than food consumption *per se*, the FBS is a cost-effective and standardised source of dietary data well suited for longitudinal comparison within and between countries (FAO 2018; Vilarnau *et al.* 2019). Thanks to this advantage, the FBS is preferred particularly when looking into dietary patterns among two or more countries. Having said that, a large number of studies in Table 3.5 rely on information on food consumption from surveys of household consumption or expenditure. Depending on the survey type (income/expenditure/budget survey or specialised food consumption survey), household surveys collect data on the quantities of foods available to, acquired or consumed by a representative sample of households from the population (FAO 2018). Although data for a given country from both the FBS and household survey refer to food availability, the former implies the total quantities of foods from both household and non-household sectors (for example restaurants, street food vendors, hospitals, etc) while the latter is only restricted to what flowing into the household sector (FAO 2020b). Also, the survey designs are not standardised but differ in key characteristics even within the same type of survey, both across countries and within countries over time (Zezza *et al.* 2017; Russell *et al.* 2018). Thus, “for the purpose of assessing the food availability or consumption of the population as a whole, it appears that the FBS is a more appropriate source than those of the household survey” (FAO 2020b).

Next, it is worth noting that food consumption data are usually high-dimensional, reflecting the plentiful food commodities as well as the multiple food groups making up the whole diet. One approach to deal with this type of data is to employ principal component analysis or factor analysis, and then run clustering algorithms on the derived components or factor scores. Almost half of the studies listed in Table 3.5 pursue this dimension reduction technique. Nevertheless, Chen and Hsu (1999) point out two main disadvantages of using factor scores instead of all variables for segmenting task, which are also applicable for clustering purpose. First, results of factor analysis might change if a different rotation method is used. Second, the original variables are more easily interpretable than the derived components with factor labels since the naming and interpretation of the components can involve personal judgements.

With regard to clustering methods, the 7th column of Table 3.5 lists the various methods employed by previous studies organised in chronological order. In the beginning, traditional clustering techniques for static data, such as hierarchical clustering and K-Means, are the most popular methods. Later on, it seems to be a common practice to combine both techniques in a two-step procedure: first a hierarchical cluster analysis gives an idea of an appropriate number of clusters and based on this information a non-hierarchical K-Means algorithm is then applied to define the clusters. Over the past decade, more advanced methods are adopted, particularly model-based clustering techniques such as finite mixture model, latent class analysis, or copula-based model. Among the previous literature, the study that is most methodologically aligned with this research is the study by Di Lascio and Disegna (2017), which also employs copula functions. Nevertheless, their study examines the diets of EU countries whereas this research widens the scope to investigate global diets. Even more, their study applies the copula-based clustering algorithm on discrete time periods (the first and last year of the period under examination) whereas this research aims to perform the clustering task on time series as a whole sequence. In doing so, the similarities among time series are interpreted in terms of their co-movements, which could be detected by a copula function. In principle, time series should be grouped together if they are positively concordant, or in other words large (small) values of one series at a given time tends to be associated with large (small) values of the other series at the same time.

The last column of Table 3.5 summarises key results from the previous studies. Albeit differences in the scope of study as well as the clustering method employed, these findings are shaped around two main themes. First, there is an increasingly similar dietary composition across countries particularly those belonging to the OECD or the EU. Across national borders, previous studies document clusters of countries that are close to each other both geographically and culturally: Northern Europe, Southern Europe, and the Balkans to name a few. This implies the existence of a common dietary structure in these regions making the 'Scandinavian diet', the 'Southern European diet', or the 'Balkan diet'. Nonetheless, the consumption level of single food items varies across countries in the same cluster. Thus, the label like 'Balkan diet' better describes an overall tendency rather than dictates rigid rules for a norm of diet. Within countries over time, clusters of populations with similar food consumption patterns exhibit an increase in total calorie availability, a rise in energy contribution from fats but a reduction in the energy proportion from carbohydrates – all of which represent different manifestations of the *nutrition transition* (Balanza *et al.* 2007).

Second, in many studies the application of cluster analysis helps to identify different subgroups in the population that are characterised by healthy versus unhealthy consumption patterns. Interestingly, there is evidence for switching behaviours, meaning that some segments of the population move to a (un)healthier cluster over time (Walthouwer *et al.* 2014; Di Lascio and Disegna 2017). Often, the consumption patterns and dietary healthiness are profiled in relation with socio-economic variables (say income, gender, marital status, family size, rural-urban residence, and education attainment). For example, Chinese consumers with food consumption behaviour towards a 'Western' diet tend to be richer, more highly educated and living in urban areas, whereas the traditional diet is associated with those living in rural areas and earning medium income (Zhang *et al.* 2008). From the viewpoint of healthiness, Casini *et al.* (2013) report that among young Italians those who want convenience are most likely to consume nutritionally poor diets due to the frequent recourse to 'easy-to-prepare' and 'ready-to-use' products. Furthermore, some authors attempt to make a connection between cluster's socio-economic characteristics and health status. Take overweight and obesity as an example. Zhang *et al.* (2008) conclude that higher BMI is found among urban consumers who have higher education attainment and higher socio-economic status. Although the clusters are established on the basis of food consumption patterns, the fact that their characteristics are so reflective of the nutrition transition framework highlights the merits of cluster analysis. It is possible to identify clusters with different levels of overweight and obesity by focusing solely on the consumption of various food groups. In spite of the manifold and complex factors associated with overweight and obesity, its dietary origin is obvious.

Overall, the findings summarised in Table 3.5 show that cluster analysis has long played an important role in understanding food consumption behaviours by dividing a set of countries, households, or individuals into groups that are meaningful and useful. Depending on the clustering variables, countries, households, or individuals within each group share common characteristics regarding various aspects of diets (for instance the quantities of food categories, food composition, or their evolution over time). However, in most cases, cluster analysis is only the starting point for further analysis, such as the examination of dietary patterns in relation to different socio-economic variables as well as health status.

After reviewing the existing literature, it is evident that despite the validity of cluster analysis in food economics there are a number of literature gaps which call for further investigation.

First, time series data related to food intake/dietary patterns are abundant, and previous researchers have attempted to make use of this type of data to explain the evolution of dietary patterns. Yet, either time series are merged into one large set of static data (see, *for example*, Blandford 1984; Staudigel and Schröck 2015; Sadowski 2019), or clustering algorithms are applied on discrete time periods and clustering results are compared between a baseline and a follow-up period (usually the first and last year) (for instance, Gil *et al.* 1995; Di Lascio and Disegna 2017). In both cases, the clustering task is not performed on the whole set of time sequences and the time dependent nature of the data is not appropriately addressed.

Second, previous studies extensively employ *crisp clustering* methods in which countries, households or individuals are grouped into non-overlapping clusters while scant attention is paid to *fuzzy clustering* which enables the possibility of overlapping clusters. In fact, fuzzy clustering is an attractive method offering considerable appeals to food/diet-related research. Allowing a country to simultaneously belong to more than one cluster with a membership proportion between 0 (absolutely does not belong) and 1 (absolutely belongs), fuzzy clustering facilitates the idea that multiple diets coexist within a single country. While geographical locations, weather conditions, religion and other cultural factors help to shape the so-called 'Chinese cuisine' distinctive from a typical meal in the USA, globalisation makes it possible to find McDonald's burgers across China or vice versa Chinese takeaways across the USA. Even within a country say China, the cuisine is so diverse that cooking styles, ingredients and flavours vary from region to region (Zhang and Ma 2020). That is not to mention the difference between diets of the poor and diets of the rich within each local municipality. Therefore, a diet observed in any country can be best considered as the 'average' approximation of several existing dietary patterns in the country. In that case, fuzzy clustering is able to disentangle the diet in China into say 90% of traditional eating patterns and 10% of Western-oriented food styles.

On the other hand, fuzzy clustering can deal with the uncertainty of assigning a time series into a specific cluster which was discussed in Section 3.4.2 when a unit (say a country) switches to a different cluster over time. Such a situation occurs if a time series exhibits a dynamic pattern consistent with a cluster for a certain time period and then a completely different dynamic pattern which is more similar to another cluster in the following time period. Indeed, the food economics literature documents the tendency that individuals/countries change to a (un)healthier cluster over time (Walthouwer *et al.* 2014; Di Lascio and Disegna 2017). Hence, the utilisation of membership degree between 0 and 1 would avoid the arbitrariness of assigning a unit (say a country) to only one cluster when it may be 'close' to several.

Third, the identified clusters on the basis of their food consumption patterns are often investigated in relation with income, overweight/obesity prevalence or other socio-economic characteristics (Zhang et al, 2008; Petrovici, 2005). Nonetheless, this is done as a post-cluster analysis. On the other hand, income and obesity prevalence can be considered as additional information, and hence can be incorporated in the clustering procedure as 'spatial' information constraining the defined clusters to be distinguished in 'space'. Importantly, the 'space' here is more than geography. The configuration of the spatial information is further explored in Section 7.2 (Chapter 7). To the best of the author's knowledge, space-time clustering algorithm has not been applied in the earlier food economics literature.

Acknowledging the above literature gaps, this research extends the previous literature in several ways. This is the first study using global data from the FAO to examine the convergence and similarity among the world's diets. Formal convergence methodologies will first be applied, and any evidence for convergence would indicate that diets across national borders have become more alike. To capture the similarity in the global diets, an innovative time series fuzzy clustering algorithm will be adopted. Unlike the prior studies in the existing literature, the clustering algorithm is performed on the whole sequence of time series instead of merging the longitudinal data into a set of static data. The main purpose is to ensure that the time dependent nature of the data is appropriately addressed. Specifically, the adoption of fuzzy clustering technique is attractive since it allows the coexistence of multiple diets within a single country and it can accommodate the switching behaviour that is often observed for time series data. This is an important contribution as fuzzy clustering has not been previously employed in the food economics literature. As cluster analysis divides countries into different clusters based on the historical food consumption trends, it is possible to find out the main dietary trends around the world, the direction (healthier/less healthier) have these trends become as well as their impacts on global health.

In addition, the second empirical analysis investigates the convergence and similarity among the global diets through the lens of spatial analysis techniques. Going beyond the geographical 'space', this study explores the spatial relationship among countries via the notion of *economic proximity*. Is it economic development that is driving the similarities among the global diets rather than geographical closeness? If so, economic proximity could act as spatial information and should be incorporated in the clustering procedure to assist the selection of final cluster solution.

Table 3.5 Applications of cluster analysis in food consumption studies.

Author(s)/Year	Country/Region	Clustering variables	Data source	Type of data	Factor analysis	Clustering algorithm	Key results
Blandford (1984)	OECD countries	Food groups	OECD	Time series	No	K-Means	Dietary composition becomes more similar
Gil *et al.* (1995)	16 EU countries	Food groups	FBS	Time series	Yes	Hierarchical	More similar dietary structures
Wirfält and Jeffery (1997)	The USA	Food groups	Survey	Cross-sectional	No	K-Means	• 2 clusters with high consumption of pastry and meat had higher fat intakes • Other 2 clusters with high intakes of skim milk had higher micronutrient levels
Bradatan (2003)	Balkan countries and Hungary	Food groups	FBS	Time series	No	Hierarchical	There is a common diet structure, but calorie intakes and meat consumption vary
Petrovici *et al.* (2005)	30 EU countries	Food groups and nutrients	FBS	Cross-sectional	Yes	Hierarchical	Strong similarities are found between Slovakia and Czech Republic; Finland and Norway; Estonia, Lithuania, and Latvia
Balanza *et al.* (2007)	20 EU countries	Food groups	FBS	Time series	Yes	K-Means	Similar trends include an increase in energy from fats and a decrease from carbohydrates
Bertail and Caillavet (2008)	France	Food groups	Survey	Cross-sectional	No	Finite mixture model	Among 6 segments with diverse demands for fruits and vegetables, one segment is income-independent
Zhang *et al.* (2008)	China	Food groups	Survey	Cross-sectional	Yes	Hierarchical and K-Means	Adherence to the traditional diet is negatively related to Body Mass Index
Honkanen (2010)	Russia	Food groups	Survey	Cross-sectional	Yes	Hierarchical and K-Means	• Strong preference for meat • Fish is the least consumed food

Casini *et al.* (2013)	Italy	Food groups	Survey	Cross-sectional	No	Latent class clustering	Out-of-homers and convenience seekers are at risk of consuming unhealthy diets
Erbe Healy (2014)	Italy, the UK, France, Ireland	Food groups	Survey	Cross-sectional	No	K-Means	• Italian diet is more keeping with the Mediterranean diet than before • Diets in other three countries are more reliant on food away from home
Walthouwer *et al.* (2014)	The Netherlands	Food groups	Survey	Time series	No	Hierarchical and K-Means	• 3 clusters associated with healthy, moderately healthy and unhealthy diets • 34% of the sample switched to a (un)healthier cluster
Staudigel and Schröck (2015)	Russia	Food groups	Survey	Cross-sectional	Yes	Hierarchical and K-Means	Affluent households drive the supply for food away from home
Di Lascio and Disegna (2017)	40 EU countries	Food groups	FBS	Time series	No	Copula-based	• Diets of EU countries become more similar • Some countries over the years moved to a (un)healthier diet
Heng and House (2018)	8 countries	Food groups	Survey	Cross-sectional	No	K-Means	Nearly 80% of the sample does not consume fruits frequently or only consumes common fruits
Scalvedi *et al.* (2018)	Italy	Food groups	Survey	Cross-sectional	Yes	Hierarchical and K-Means	34% of adult populations consume balanced diets, 30% consume unbalanced diets, 8% are protein overeaters, 28% are nibblers
Azzam (2020)	172 countries	Western Similarity index based on food groups	FBS	Cross-sectional	No	K-Means	A cluster consisting 16 countries whose dietary patterns resemble the American diet

3.8 Chapter conclusion

Cluster analysis is a data exploratory tool that organises data into homogeneous groups so that the within-cluster dissimilarity is minimised whilst the between-cluster dissimilarity is maximised. This chapter reviews the literature on cluster analysis covering a wide variety of clustering methods for static data and data that vary across time and/or space. In terms of defining cluster boundary, clustering technique can be either crisp or fuzzy. Crisp clustering methods assign data units into non-overlapping clusters, meaning that the membership of each unit associated with a cluster is either 0 (does not belong) or 1 (belongs). Fuzzy clustering on the other hand allows each unit to belong to more than one cluster with varying membership proportion between 0 (absolutely does not belong) and 1 (absolutely belongs), and therefore clusters are overlapping. Several advantages of fuzzy clustering are discussed particularly its ability to reflect the uncertainty of assigning a unit into a particular cluster when the data includes temporal (and spatial) attributes. In addition to traditional clustering techniques, a constraint in terms of geographical space can be introduced so that units in a cluster need to not only be similar to each other but also be geographically close.

In the final section of this chapter, the usefulness of cluster analysis in food economics is reviewed and a summary of related studies is provided. From this arise a number of literature gaps to be explored in the current research. Despite the abundance of time series data in food economics, the time dependent nature of the data is not appropriately addressed in earlier studies. Next, there is an overwhelming predominance of crisp clustering methods while fuzzy clustering has been largely neglected. Also, the clusters defined by differing food consumption patterns are examined in relation with income, overweight/obesity rates and other socio-economic characteristics. In general, this step is usually done separately from the cluster analysis. These factors however can play the role of spatial information and can assist the selection of the final cluster solution. To the best of the author’s knowledge, space-time clustering methods have not been employed in the previous studies related to food/diet. This further highlights the novelty of this research and reinforces the value it adds to the existing literature.

Chapter 4

Diet quality indices

4.1 Chapter introduction

Today, nearly one in three people worldwide suffers from at least one form of malnutrition: childhood stunting, wasting, vitamin and mineral deficiency, overweight or obesity and diet-related noncommunicable diseases (WHO 2017b). While these multiple forms of malnutrition undoubtedly pose serious threats to global health, the world is facing a nutrition crisis: approximately three billion people are consuming low-quality diets, which either contain insufficient calories, vitamins and minerals or contain too many calories from saturated fat, salt and sugars (Global Panel on Agriculture and Food Systems for Nutrition 2016). Given that the causes of malnutrition are manifold, poor diet is one of them. It is shown that low-quality diet is the leading risk factor for mortality and morbidity in the global burden of disease (Afshin *et al.* 2019). Thus, promoting high-quality diets is key to achieving the Sustainable Development Goal of "ending malnutrition in all forms by 2030" which is one of the largest challenges facing all countries in the 21st century (WHO 2019). Existing epidemiological studies have documented mounting empirical evidence for the link between the consumption of healthful diets and the promotion of normal growth and development in childhood and adolescence, as well as the mitigation of health problems in adults (Sares-Jäske *et al.* 2017). In general, there is the need to measure and monitor *diet quality* in public health policy agendas.

A number of methods have been proposed to assess overall diet quality and to find out healthy dietary patterns (Fransen and Ocké 2008). To this aim, numerous indices have been developed and validated to reflect different dimensions of diet quality. Ranging from indices simply measuring the compliance with certain dietary recommendations to more complex tools requiring consumption information of micro- and macro-nutrients, these indices are widely used in the existing nutritional epidemiology literature (Burggraf *et al.* 2018). However, previous systematic reviews all point to a main drawback of these indices regarding their limited predictive capacity for health outcomes, questioning their validity (Waijers *et al.* 2007; Arvaniti and Panagiotakos 2008). Indeed, assessing diet quality is a challenging task not only given the many dietary factors that can affect health but also because there is no official definition of diet quality in the related literature (Alkerwi 2014; Berdanier *et al.* 2014). This chapter aims to review the concept of diet quality and debated issues in constructing diet quality indices.

The remainder of this chapter is laid out as follows. Section 4.2 looks at the concept of *diet quality* and the nuances in quantifying it. Section 4.3 gives a brief overview on the construction methods of pre-defined diet quality indices, the myriad of existing indices, different potential dietary data sources and the predictive ability of these indices for health outcomes. Section 4.4 discusses motivations for the use of a diet quality index and introduces the most appropriate one for this research. Section 4.5 presents concluding remarks.

4.2 What is diet quality and how to measure it?

Conventionally, the notions of *healthy nutrition* and *adequate nutrition* tend to be treated as interchangeable and the focus has long been put on the promotion of diets that provide a sufficient amount of calories and essential nutrients (Ruel 2003). While nutrient quantity is an important indicator to gauge the progress towards eliminating hunger and undernutrition, it matters not only 'how many calories' but also 'what type of calories' bearing in mind the escalating overweight and obesity prevalence worldwide (Echouffo-Tcheugui and Ahima 2019). These days, the "double burden of malnutrition" requires promoting *high-quality diets* that on the one hand meet the minimal dietary energy requirement for day-to-day survival and on the other hand have a positive impact on overall well-being. So, what exactly is a high-quality diet? To answer this question, it is crucial to first define *diet quality*.

In nutritional epidemiology literature, *diet quality* is described in various ways, including a healthy diet, a balanced diet, nutritious foods, functional foods, and a nutrient-rich diet (Alkerwi 2014). Even though these terms all emphasise the importance of achieving an optimal level of health via balanced nutrient intakes, the existence of heterogeneous terminologies creates a huge confusion on what constitutes a high-quality diet. While it is fairly agreed upon certain nutrients or food components that should make up a high-quality diet, translating this knowledge into a standardised diet configuration is a challenging task due to dietary customs, cultural context and locally available foods. Individual factors (for example age, gender, disease status, and physical activity level) lead to different needs for macro- and micro-nutrients (FAO and WHO 2001).

A way to think about diet quality is to categorise food items into healthy and unhealthy components and on the basis of which individuals should maintain an *adequate* consumption of healthy food components (for instance fruits, vegetables, whole grains, fibre) and a *moderate* (or very limited) consumption of unhealthy ones (such as saturated fat, sugars, sodium) (Guenther *et al.* 2013). Many dietary recommendations including the Dietary Guidelines for Americans (USDA and HHS 2020) are established based on these two key dimensions (*adequacy* and *moderation*). Others, including the UK food-based dietary guidelines (the Eatwell Guide) (Public Health England 2016), consider *variety/diversity* as another important dimension of diet quality since a diverse diet should consist of sufficient nutrients (adequacy) and necessarily limit specific nutrients (moderation) (Kim *et al.* 2003; Gémez *et al.* 2019). So far, a universally agreed set of indicators to measure diet quality does not exist (Schwingshackl and Hoffmann 2015). Yet, current national food-based dietary guidelines often offer some guidance on how a healthful diet with respect to a particular national population should look like. The recently emerged concern around the topic of sustainability adds more nuance to the notion of a high-quality diet. In order to encapsulate the multidimensional nature of diet quality, the definition of high-quality diets can be as broad as those "eliminate hunger, are safe, reduce all forms of malnutrition, promote health and are produced sustainably without undermining the environmental basis to generate high-quality diets for future generations" (Béné *et al.* 2019).

How to measure the quality of a particular diet? Historically, nutritional epidemiology has predominantly emphasised the influence of single dietary components on disease development, for example the role of vitamin C against the risk of cancer or the role of fat intake against cardiovascular diseases (Maynard *et al.* 2005). Nonetheless, this so-called 'reductionist' approach has received a wealth of criticism. First, individuals do not consume single foods or nutrients but a complex combination of foods containing numerous nutrients and non-nutrients (Mertz 1984). Second, the existence of food synergy, meaning the interaction and interrelation of nutrients, might hinder their bioavailability and absorption (Jacobs and Steffen 2003; Tapsell *et al.* 2016). To further complicate the issue, entering several highly correlated variables (for instance, the interrelated nutrients) in a model predicting the risk of a disease is likely to cause multicollinearity, making the estimations less robust and the predictions less accurate (Hu 2002). In the end, foods are complex whilst our knowledge is limited, so the nutrition-health nexus is unlikely to be reduced entirely to molecular interactions (Katan *et al.* 2009). Thus, any research aiming to find an association between food intakes and disease prevention should follow a more *holistic* approach instead of focusing on specific nutrients or individual foods. Diet quality should be viewed via the lens of the totality of what one eats and drinks as a whole (Schwingshackl and Hoffmann 2015; Hiza *et al.* 2018). This leads to the emergence of an alternative approach (called *dietary pattern analysis*) as opposed to the *reductionist* approach.

Following dietary pattern analysis, there are two methods to measure and quantify dietary patterns (Waijers *et al.* 2007). One method is based on current nutritional knowledge on foods and nutrients important to health which are quantified and summed to arrive at an overall measure of dietary quality (Trijsburg *et al.* 2019). This is often coined *theoretically defined dietary patterns* or *pre-defined diet quality indices/scores*. The aim is to evaluate the dietary healthiness and categorise individuals according to the extent to which their eating behaviour is healthy or how their dietary patterns comply with the reference intake values recommended in dietary guidelines. A great advantage of pre-defined diet quality indices is generalisability, meaning that they can be applied to different populations (Román-Vinas *et al.* 2009; Vandevijvere *et al.* 2013). In addition, diet quality scores are useful for comparing diet quality of sub-groups within the same population, tracking diet quality over time, assessing impacts of an intervention on diet quality and examining the relationship between diet and disease risks (Lanham-New *et al.* 2019). However, the accuracy of *a priori* indices is limited by the current level of dietary knowledge regarding diet-health relationship, and uncertainties for the index construction process (Fransen and Ocké 2008).

The other method, known as *empirically defined dietary patterns*, examines diet quality *a posteriori* by using statistical techniques such as factor or/and cluster analysis. Such techniques aggregate intake variables into factors to investigate common underlying food consumption patterns within a population (Newby and Tucker 2004). Because the patterns are derived specifically for the population, they are often not reproducible across populations (Kant 1996). Besides, they do not necessarily define the healthiest patterns because they are not based on current nutritional knowledge or evidence on the diet-health relationships (Hu 2002).

In this research, the world's diets are first defined by cluster analysis, and a pre-defined diet quality score is then exploited to evaluate the healthiness of overall diet. The next section presents an overview of pre-defined diet quality indices.

4.3 Pre-defined diet quality indices

4.3.1 Construction framework

Being composite indices, diet quality indices follow the same methodology for index development which consists of five steps as follows:

Step 1: Select index variables (components);
Step 2: Select the number of partitions for each component;
Step 3: Design a scoring system and cut-off values for each component;
Step 4: Attribute a weighting to each component;
Step 5: Sum up the scores assigned to components to find the index value.

Essentially, all indices are combined measures of individual components, each of which represents a different dimension of the index and is based on some particular dietary recommendations/guidelines. Once the components to be included in the index are determined, they need to be quantified in terms of the number of partitions (or categories) for each component as well as the score assigned to each partition. A common technique is to use a *cut-off* value for each component and to assign 0 if the consumption of that component is lower (or higher) than the cut-off value, and 1 if the consumption is higher (or lower) than this threshold. The index components are scored using arbitrary weights and then summed up to arrive at a final score.

To demonstrate, Table 4.1 shows the composition of the Healthy Diet Indicator 2013 (HDI-2013) (Berentzen *et al.* 2013), which is the updated version of the original Healthy Diet Indicator (Huijbregts *et al.* 1997). HDI-2013 comprises seven components, namely saturated fatty acids, polyunsaturated fatty acids, cholesterol, protein, dietary fibre, fruits and vegetables (excluding potatoes), and free sugar. These components are chosen from the World Health Organisation (WHO)'s recommendations for the prevention of chronic diseases (WHO 2003). A dichotomous variable is generated for each component. If a person's intake is within the recommended range following the WHO's guidelines, this variable takes the value 1, otherwise, it is coded 0. All seven components have equal weightings towards the total score. Therefore, the HDI-2013 is the sum of all these dichotomous variables and takes a value in the range of 0-7 points. The higher the HDI-2013 value, the higher the diet quality.

Table 4.1 Composition of the Healthy Diet Indicator 2013 (HDI-2013).

Components (daily intake)	Cut-offs	Score	Weighting	The HDI-2013 value
1. Saturated fatty acids (% E)	< 10	1	Equal	$= \sum_{i=1}^{7} X_i$ where X denotes the component scores
	≥ 10	0		
2. Polyunsaturated fatty acids (% E)	6 - 10	1		
	< 6 or > 10	0		
3. Cholesterol (mg)	< 300	1		
	≥ 300	0		
4. Protein (% E)	10 - 15	1		
	< 10 or > 15	0		
5. Dietary fibre (g)	> 25	1		
	≤ 25	0		
6. Fruits and vegetables (excluding potatoes) (g)	≥ 400	1		
	< 400	0		
7. Free sugar (% E)	< 10	1		
	≥ 10	0		

Note: % E refers to the percentage of total energy intake excluding alcohol;
mg and *g* are abbreviations of milligrams and grams.

4.3.2 Existing diet quality indices

Numerous pre-defined diet quality indices have been proposed in the existing literature and the majority of them aim to evaluate diets of adults. Some specific indices are developed for children and adolescents, for example the KIDMED index (Da Rocha *et al.* 2020), or for elderly people for instance the Elderly Dietary Index (Kourlaba *et al.* 2009).

Overall, the four most widely used indices are the *Healthy Eating Index* (HEI-1995) (Kennedy *et al.* 1995), the *Diet Quality Index* (DQI-1994) (Patterson *et al.* 1994), the *Healthy Diet Indicator* (HDI-1997) (Huijbregts *et al.* 1997), and the *Mediterranean Diet Score* (MDS-1995) (Trichopoulou *et al.* 1995). Several modifications and updates of these indices have been proposed. For example, AHEI-2002, HEI-2005, HEI-2010, HEI-2015 are variants of the original HEI.

Primarily, indices are constructed following either nutritional guidelines or Mediterranean dietary patterns. The *Mediterranean diet* has received a lot of attention due to the fact that adults dwelling near the Mediterranean Sea have one of the lowest incidences in chronic diseases and one of the highest life expectancies in the world (Trichopoulou and Benetou 2019). Even though the positive association between adherence to the Mediterranean diet and risks of coronary disease and certain types of cancer is confirmed in prior studies (Panico *et al.* 2014; Anand *et al.* 2015; Rosato *et al.* 2019), a precise and quantified definition of the Mediterranean diet does not exist. Having said that, a traditional Mediterranean diet refers to the eating pattern of the Mediterranean basin which spans from Southern Europe to Northeast Africa, and is usually characterised by a high monounsaturated/saturated fat ratio, a moderate consumption of wine and dairy products, a low consumption of meat while a high consumption of vegetables, grains and fruit (Trichopoulos and Lagiou 2004; Mastorakou *et al.* 2019). The abundance and wide variety of healthful foods in the Mediterranean region might be partly attributed to the strategic location along the north 40th parallel, which stretches across the Mediterranean Sea, Asia, Japan, North America and the Iberian Peninsula (Vilarnau *et al.* 2019). However, Mediterranean eating patterns are not homogeneous for the Mediterranean basin but are spatially heterogeneous, shown by for example varied patterns in north versus south Mediterranean countries, due to the inherent socio-economic factors within each region (El Kinany *et al.* 2020).

Tables B.1 and B.2 in Appendix B summarise 61 indices, of which 41 indices are tailored to the national dietary guideline and 20 indices based on Mediterranean diet. It can be seen that national dietary recommendations mostly stem from the US, Canada and other high-income countries whereas there is an absence of indices developed for low- and middle-income countries. A recent report shows that only 2 out of 31 low-income countries and 12 out of 51 middle-income countries currently have some sorts of dietary guidelines (Global Panel on Agriculture and Food Systems for Nutrition 2016). Among 41 reported indices, a handful of them are created for cross-regional comparison and only two are globally applicable. The myriad of existing indices on the one hand demonstrates the incoherence of national dietary recommendations and on the other hand is evident for the constant update of these guidelines. To recap from previous sections, indices are proposed to suit the context and food culture of a country or geographical area, and thereby indices should ideally be updated when new nutritional knowledge and recommendations unfold (Guenther *et al.* 2013; Moraeus *et al.* 2020). To illustrate, the *Healthy Eating Index* (HEI-1995) was originally developed to evaluate the extent to which Americans are conforming to the nutritional recommendations outlined in the *Dietary Guidelines for Americans*. This index however has been revised since 2005 and the updated versions including HEI-2005, HEI-2010, HEI-2015 are proposed as new guidelines (the most recent one is the 2020-2025 Dietary Guidelines) emerge.

That said, the question remains whether these country-specific and cross-regional indices precisely capture all dimensions of diet quality. The majority of indices listed in Table B.1 and B.2 are based on both nutrients and food groups, and often the recommendations based on food groups predominate. This is somewhat expected as it is obviously easier for individuals to follow a guideline such as 600 grams of vegetables per day than 1,000 kcal from carbohydrates. Comparing indices listed in two tables, most of them are featured with an equal weighting scheme, implying that the relative importance of different food groups and/or nutrients is usually not taken into account. Besides, indices based on the Mediterranean diet tend to have a stronger association with health outcomes.

4.3.3 Similarities and differences among existing indices

Many similarities and differences among existing diet quality indices are engendered by the methodology for index development, which is not black-and-white and there is plenty of room for arbitrary choices made by the researcher (see, *among others*, Waijers *et al.* 2007; Arvaniti and Panagiotakos 2008; Kourlaba and Panagiotakos 2009; Wirt and Collins 2009; Burggraf *et al.* 2018; Aljuraiban *et al.* 2019; Trijsburg *et al.* 2019). Specifically, these choices are related to the selection of variables/components to be included in the index, the cut-off values that should be used for each component, the weights that should be assigned to each component, and the dietary assessment method to measure food intakes. As a result, the index development is susceptible to a high degree of subjectivity. Moreover, differences in the construction method make the comparison between indices impossible. Hence, an understanding of the rationale of the diet quality index as well as the components making up the score is crucial for the index to be correctly interpreted (Lanham-New *et al.* 2019). Key issues in the construction of a diet quality index are discussed as below.

Index components

According to Kant (1996), there are three approaches in constructing a diet quality index: based on nutrients (for example total calcium intake), food/food groups (for example intakes of meat, vegetables, cereals), or a combination of both. Some indices comprise of only nutrients (DQINB-1999), some contain only food groups (DQS-2007, MAI-1999, HFI-2001), and others contain both (DQI-1994, HEI-

2005, AHEI-2002). Overall, nutrient-based indices tend to consider consumption as a percentage of one of the nutrient-based reference values of the dietary reference intakes as a marker for diet quality whereas indices based on food groups examine the dietary patterns of foods to identify patterns associated with adequacy and positive health outcomes (Kant 1996). In the existing literature, the most commonly applied indices are based on the combination of nutrients and foods (Coulston *et al.* 2017).

Even though the nutrients or food groups exploited in each index vary case by case, some nutrients (such as total fat, ratio of saturated fat to mono- or polyunsaturated fat, and cholesterol) or food groups (such as fruits, vegetables, cereals, meat, and dairy) are often included due to their long-known positive impacts on health. The number of components used in each index is not fixed but ranges between 4 (HFI-2001) and 20 (DGAI-2005).

As a diversified diet comprising of a greater variety of foods is more beneficial than a monotonous diet, prior researchers recommend taking into account *diet variety* (reflected by the number of different foods consumed over a given period of time) in the index construction (Arimond and Ruel 2004). This has led to the development of indices that exclusively measure diet diversity, for example the Dietary Variety Score (Bernstein *et al.* 2002) or the Riksmaten Adolescent Diet Diversity Score (RADDS) (Moraeus *et al.* 2020). However, the necessity to incorporate variety as an index component is questioned by Waijers *et al.* (2007) as diet quality indices are generally made up of several adequacy indicators, automatically implying a varied diet.

Cut-off values

Cut-off values should be specific to not only country or region but also age, sex, weight, and physical activity level in order to take full advantage of the scientific knowledge available for the population under examination (Burggraf *et al.* 2018). In practice, cut-offs can be either *normative* or *percentile*. Normative cut-offs are derived from current evidence for diet-health relationships that reflect dietary requirements of healthy individuals, and thereby are chosen as a healthy level of intake usually in accordance to dietary recommendations (Drake *et al.* 2011). HEI-1995, HEI-2005, HEI-2010, HDI-1997 and HDI-2013 are examples of those indices being constructed in this way. This kind of cut-off prevents overestimating scores and allows the comparison of studies in meta-analyses (Aljuraiban *et al.* 2019). Although the rationale seems attractive, this approach is subject to a major drawback. If the intake of a component is below the cut-off value for the majority of subjects in a population, this index component does not contribute any extra discriminating power and could become redundant in index construction (Waijers *et al.* 2007).

On the other hand, percentile cut-offs (median or quartile) simply indicate the intake values below which a given percentage of observations in a population sample fall. Many Mediterranean diet-based indices use sex-specific median cut-off value (for instance MDS-1995 and MDS-2003). The main advantage of this method is that each component is well distributed with half of the subjects scoring positively and the other half scoring negatively. Nevertheless, it has certain disadvantages including that the median value does not necessarily reflect a healthy level of intake, and different population samples produce different median values (Kourlaba and Panagiotakos 2009). For example, a high MDS score in a southern European population can differ greatly from a high score in a northern European population.

Bearing in mind the pros and cons of normative and percentile cut-offs, researchers should take into account the intake levels of the included components in the population when choosing the cut-off value. Some indices (say MDS-1995, HDI-1997, HFI-2001) exploit only one cut-off value, some (say DQI-1994 and DGI-2002) use more than one cut-off value (with a lower bound, an intermediate range, an upper bound), and others (say AHEI-2002 and DQI-I-2003) assign the score of each component to be proportional to the extent with which the guideline is complied.

Weighting and aggregation

For the majority of existing indices, all items have equal contributions to the total score. However, it is not plausible that all components have the same health impact and hence, equal weighting can lead to an overestimated score (Aljuraiban *et al.* 2019). Only few indices are constructed using specific weights to some components (for example HEI-2005, RCI-2008, DQI-I-2003). However, the weights are decided arbitrarily with no sound rationales and the oft-quoted reason is that certain components are more important to diet quality based on dietary guidelines (Kim *et al.* 2003). The remaining indices apply linear aggregations with equal weights designated to the index components without giving further explanations.

Underlying dietary intake data

Diet quality indices are constructed based on food consumption data, which can be derived from food consumption surveys at national, household or individual level and can be expressed in terms of nutrients and/or foods. Main dietary assessment methods include Food Frequency Questionnaire (FFQ), 24-hour recall, diet history, food record and Food Balance Sheet. The validity of these methods as well as their strengths and limitations are discussed in the next section.

4.3.4 Dietary assessment methods for measuring dietary intake

According to Gibson (2005), there are four approaches in nutrition assessment to comprehensively evaluate the nutritional status of individuals, namely anthropometrics, biochemical parameters, clinical examination and dietary assessment. *Dietary assessment* refers to methods that estimate the consumption of food and nutrients at national, household and individual level. Figure 4.1 presents the most common dietary assessment methods, which are categorised depending on the nature of the method. *Indirect* methods assess diets by employing secondary data whilst *direct* methods gather primary data from individuals. Depending on the time that food consumption is recorded, direct methods can be either *prospective* or *retrospective*. The former refers to recording diets as foods are being consumed while the latter relies on a recall of foods that were consumed. Each of the methods shown in Figure 4.1 is to be discussed subsequently.

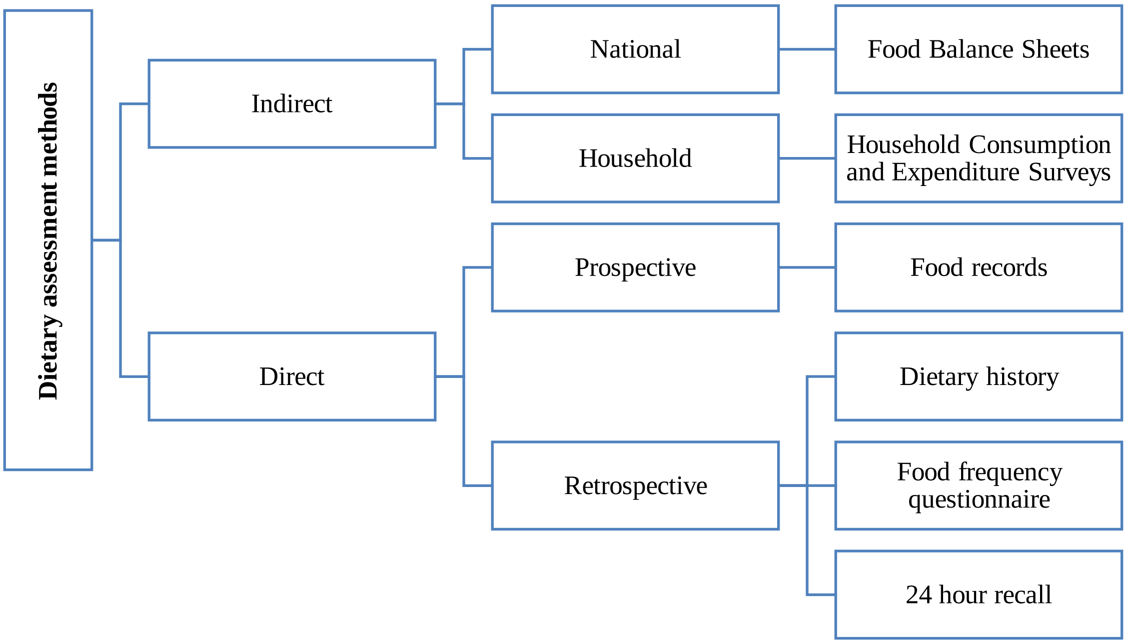
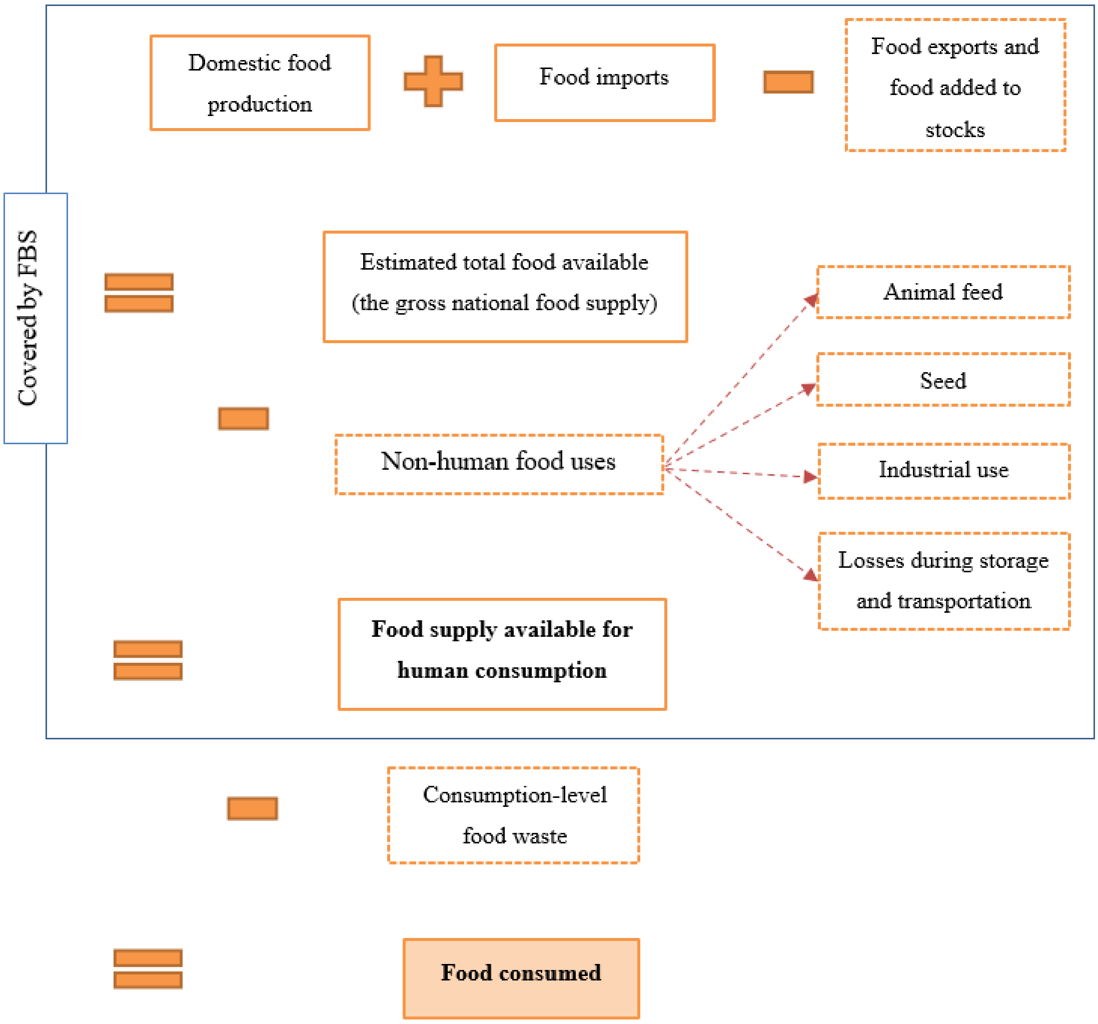


Figure 4.1 Overview of dietary assessment methods.

Measuring food consumption at national level

The most widely used method for estimating food availability for consumption at national level is the Food Balance Sheet (FBS) from the Food and Agriculture Organisation of the United Nations (FAO). By definition, the FBS is an aggregated data set that "presents a comprehensive picture of the pattern of a country's food supply during a specified reference period" (FAO 2001, p.1). The FBS compilation follows an accounting framework such that all potential sources of both food supply and food utilisation of a given food item are specified. Figure 4.2 shows how the food available for human consumption is calculated. Domestic food production plus the quantity imported and then deducted by the quantity exported and food added to stocks gives the total food available for supply. The *food available for consumption* is derived after subtracting other utilisations for non-human food use (including animal feed, seed, industrial use for food use and non-food use) and food losses during storage and transportation. This process is repeated for every food product (on the basis of primary commodity equivalent) consumed within a country, and all the primary commodity equivalent balances are combined into a single comprehensive data set – the FBS. Conventionally, the data are presented per capita after adjusting for the population size, in terms of kilocalories per person per day (kcal/capita/day) for energy intake or grams per day for protein and fat content. Figure 4.2 also illustrates that the FBS can reasonably proxy the national food availability for consumption ('apparent food consumption'), but not food consumption *per se* because food waste is not incorporated into the figures. However, the FBS framework allows viewing the domestic food supply and demand situation in a holistic approach, offering appeals to food supply analysis and food policy formulation. Its applications are numerous.



Note: the dotted boxes represent food utilisations.

Figure 4.2 The derivation of food supply from Food Balance Sheet.

Being one of the most extensive food consumption databases providing annual data on approximately 100 food commodities for over 180 countries worldwide, the FBS is a useful tool to monitor global food patterns and dietary habits (FAO 2018). To illustrate, analysing the FBS data helps to uncover trends in the prevalence of undernourishment in developing countries (Ambagna *et al.* 2019), to examine the evidence of the nutrition transition (Sheehy and Sharma 2010; Sheehy and Sharma 2013; Sheehy *et al.* 2019), and to predict future food demand (Gouel and Guimbard 2019). Even though the FBS covers per capita consumption of energy, protein and fat but lacks information on micronutrients, some indicators of diet quality such as the quantity of available fruits and vegetables in a given country, the share of energy from non-staple foods, and the share of energy from animal-source protein are easily derived. Petrova *et al.* (2011) estimate the national *n*-3 fatty acid intake based on the availability of fish and vegetable oils. When combined with food composition table, the FBS data can assist in evaluating the adequacy of zinc (Wuehler *et al.* 2005; Wessells and Brown 2012), magnesium (Joy *et al.* 2013), and other micronutrient contents (Arsenault *et al.* 2015) in national food supplies. Other studies employ the FBS to identify variations in adherence to the Mediterranean diet among Mediterranean countries (Balanza *et al.* 2007) or worldwide (Da Silva *et al.* 2009; Vareiro *et al.* 2009). These indicators give a relative picture of the availability as well as diversity of foods in a country but are not able to provide specific insights related to other dimensions of diet quality.

The FBS has been extensively utilised for diet monitoring purposes due to its various advantages. First, the FBS is a standardised and freely accessible source of indirect nutrition data which is relatively simple to analyse (FAO 2018). Despite the existence of direct dietary assessment methods at individual level, these are rarely standardised or comparable across countries or over time owing to several methodological differences (Micha *et al.* 2018). Second, FBS data are perhaps the only available dietary data for nearly all countries and territories worldwide, especially low-income countries (Grünberger 2014; Leclercq *et al.* 2019). Next, FBS data are useful to compare food availability among countries or to monitor trends over time within a country (Popkin and Reardon 2018). As the FBS tracks food availability, it can determine whether the national food supply adequately meets the nutritional requirements, especially for underdeveloped and developing countries where undernutrition and hunger are likely to exist to a great extent. On the other hand, the rising obesity rates within a country could be tracked by the increasing overall per capita food availability (FAO 2017b). Such information can support the government in introducing prompt policy measures to halt the rise.

Yet, several drawbacks of the FBS hinder its usefulness in assessing dietary consumption. The most important caveat is that FBS data do not imply actual food intakes/consumed as the food waste at household, retail and restaurants is not incorporated into the figures. As pointed out by Anand *et al.* (2015), almost a third of all food produced for human consumption is wasted before it is actually consumed. Thus, the FBS likely overestimates the amount of food actually consumed. According to Popkin and Reardon (2018), the FBS data are 20% or more higher than the true dietary intake. Looking at a wide variety of food items, Grünberger (2014) suggests that the FAO overestimates consumption figures for most foods and the greatest overestimation is for whole grains whereas the FAO underestimates intakes of nuts, seeds, and legumes. Second, estimates are derived from primary and/or basic country statistics which may be subject to methodological errors (FAO 2018). The extent to which the FBS data appropriately reflect the reality needs to be cross-checked with factors such as food losses, food waste, and unrecorded trades across national borders. Third, FBS data do not allow the examination of seasonal variations in the food supply (Grünberger 2014). Next, the FBS does not take into account small-scale agricultural production and harvest of wild plants, heightening the uncertainty mostly in developing countries and rural areas where people tend to be dependent on home production and wild foods (Beal *et al.* 2017). Neither does it collect information on consumption by tourists (Da Silva *et al.* 2009). In addition, the data coverage and data quality (such as data on crops, production, storage, and losses) limit the data reliability particularly in low- and middle-income countries (Jacobs and Sumner 2002; Schmidhuber and Traill 2006; Desiere *et al.* 2018; Godenau *et al.* 2020). Some authors show that the FBS tends to underestimate food availability in less developed countries (Balanza *et al.* 2007; FAO 2018). Finally, the FBS does not distinguish food supplies by age, gender, education, or socio-economic levels (Del Gobbo *et al.* 2015; Muhammad *et al.* 2017). Even though analysing the FBS can provide a macro picture of shifting diets but with no insight into the distribution of foods within a country due to the lack of disaggregated information. Therefore, the FBS can potentially mask the coexistence of over-consumption and under-consumption within a single country (Beal *et al.* 2017). These limitations should be recognised for the FBS estimates to be accurately interpreted and researchers should be cautious in linking the trends in national food consumption to variations in disease or mortality due to the influence of many lifestyle factors (Gibson 2005).

Measuring food consumption at household level

Depending on the primary purpose, household surveys gathering information on food consumption or expenditure can take various forms, including the Household Budget Survey (HSB), the Household Income and Expenditure Survey (HIES), the Living Standards Measurement Study (LSMS), the Household Expenditure Survey (HES), the Living Costs and Food Survey (LCFS). All of these surveys, whether nationally or sub-nationally representative, are collectively referred to as *Household Consumption and Expenditure Survey* (HCES). Unlike the FBS in which food consumption information is estimated from the perspective of food supply, the HCES estimates food consumption at household level from the perspective of food demand. Household members are responsible for recalling/recording all expenses and types of foods consumed during a specified time period, say one week to one month. This information is collected and analysed by national statistical offices to calculate the ‘apparent food consumption’ at household level (FAO 2018).

The validation of calculating individual nutrient intakes/food consumption from the HCES is examined in studies by Engle-Stone and Brown (2015), Coates *et al.* (2017), Sununtnasuk and Fiedler (2017), and Karageorgou *et al.* (2018). The HCES has gained its popularity for several reasons. First, the HCES is regularly conducted in a large number of low- and middle-income countries, and might be the only source of dietary information in such countries where hunger and undernutrition are most pronounced (Russell *et al.* 2018). Second, since the HCES contains a wealth of information on the household's food consumption and acquisition and is routinely conducted once every 3-5 years on large samples (10,000 households on average) that are representative of national and subnational demographics, it can be used for tracking trends as well as changes in national patterns of food consumption (Sununtnasuk and Fiedler 2017). Third, the HCES is less expensive than other sources of nutrition data as the survey is conducted and paid for by government agencies (Fiedler *et al.* 2013).

However, the HCES is not without limitation. A major drawback is the lack of information on food away from home. By definition, the HCES measures "the total amount of food available for consumption in the household, generally excluding food eaten away from home unless taken from home" (Putnam and Allshouse 1994). As such, the consumption figures might be underestimated. In many countries, the rising consumption of foods outside home might hamper the survey's ability to accurately determine the overall diet quality. Second, the survey designs are not standardised but differ in key characteristics even within the same type of survey, both across countries and within countries over time (Fiedler *et al.* 2012; Russell *et al.* 2018; Rippin *et al.* 2020). For example, different member states in Europe use different methods to collect food consumption data, making it difficult for cross-country comparisons (Ioannidou *et al.* 2020). Though there are international guidelines for the design and implementation of each of the HCES types, they are specific to each type of survey and generally lack coherence and leave a significant amount of leeway for the national survey statisticians (Zezza *et al.* 2017). Third, the HCES does not provide information on individual food consumption and hence does not allow the investigation of food distribution among household members (FAO 2018). Finally, HCESs are costly and most countries do not conduct them on an annual basis (Godenau *et al.* 2020).

In addition to the collection efforts by the federal government, a number of rich datasets on food purchases and consumption are collected by private market research firms in order to analyse food retail markets. One of such data is scanner data, which record sales of food purchased at stores or used by consumers at home. Popular suppliers of scanner data include Kantar, IRI, and Nielsen. The collection of scanner data can be facilitated: (i) at *point-of-sale (retail)* by the use of the universal product code (UPC) of products sold at retail checkout counters; (ii) by *household scanner panels* or random samples of households whose task is to scan in the UPC of the items they purchased using scanners provided to them (Muth *et al.* 2019). Generally speaking, scanner data allow for more thorough analyses of food purchase behaviour since the data recorded at the scannable barcode level can be linked to detailed information on characteristics of products (such as the brand name, the price for which the product was bought, the quantity, the weight and whether it was purchased on promotion), households, and stores (Levin *et al.* 2018). It is this granularity of the data at the product level that makes scanner data attractive for some specific uses (Griffith and O'Connell 2009; Sweitzer *et al.* 2017). Other important advantages of these data are that they are produced in a timely manner (Muth *et al.* 2016) and the household scanner panel samples are much larger than those for the HCES (National Research Council 2005).

Yet, scanner data have some limitations. Foremost, scanner data reflect purchases rather than consumption, and this type of data only accounts for food-at-home purchases, not food away-from-home (Chen *et al.* 2016). Infrequent purchases and the frequency with which people shop are other key issues. In addition, there is a big burden on respondents regarding the household scanner surveys as a respondent has to scan in all the items purchased after each shopping and report the results to the collecting firm (Lusk and Brooks 2011). Finally, scanner data are not designed to include comprehensive information on the household characteristics. Despite some basic demographic information (for example employment status, household income, or the age of the head of household), there is a lack of information on health knowledge, physical activity, sources of income, and participation in food assistance programs (National Research Council 2005).

It is worth mentioning that neither of the indirect methods for dietary assessment that have been considered so far directly collects primary dietary data from individuals to calculate the dietary intake or food consumption. Importantly, the nutrition data obtained by the two methods are best described as 'apparent food consumption' rather than 'effective food consumption'. Therefore, these indirect methods are most useful in identifying trends in food availability across geographical regions and over time. However, direct methods for individual-based dietary assessment do exist and the information gathered from which can be employed to investigate trends in food consumption, food and nutrient intakes, dietary patterns, and to examine the diet-disease association.

Measuring food consumption at individual level

Methods in this category provide either quantitative daily consumption data (recalls and records) or retrospective information on food consumption over a longer time period (diet history and food frequency questionnaire). Each of the direct methods is described subsequently along with their applications, strengths and shortcomings.

24-hour recall

This method requires subjects or their caretakers to recall food intake of the previous 24 hours in an interview. Therefore, the actual intake (consumption) of individuals is assessed. However, a single 24-hour recall is not sufficient to calculate an individual's usual intake of food and nutrients. Often, multiple non-consecutive 24-hour recalls on the same individual are recommended to eliminate daily variability in dietary patterns (Knüppel *et al.* 2019). This method is useful in assessing average usual intakes for a large population under the condition that subjects are representatives of the population and the days of the week are equally represented (Micha *et al.* 2018). This method is quick, easy, inexpensive, and can be used for illiterate individuals (Schoeller and Westerterp 2017). Since the respondent burden for a single 24-hour recall is small, those who agree to participate in 24-hour recalls are more likely to be representative of the population than those who agree to keep food records.

Therefore, 24-hour recall is useful across a wide range of populations (Coulston *et al.* 2017). Nonetheless, this method relies on memory aids to quantify food consumption, hindering the accuracy when conducting for elderly (Gibson 2005).

Food frequency questionnaire (FFQ)

This method attempts to obtain the frequency with which foods are consumed during a specified period of time. As the usual food intakes are assessed over a relatively long period of time, this method can help identify food patterns associated with inadequate nutrient intakes. The simplest questionnaires contain a list of comprehensive or specific food items along with a set of frequency categories (daily, weekly, monthly, or yearly). This method is widely used in epidemiological studies to rank individuals into different categories of low, medium or high intakes of specific foods (Kanerva *et al.* 2014). The method could be in the form of an interview, a self-administered questionnaire, or a computer-administered questionnaire with less than 30 minutes to complete. There is less respondent burden, and it is quick and easy to collect the results (Kennedy *et al.* 2011). Hence, the FFQ is commonly used to estimate dietary intake in large epidemiological studies (Coulston *et al.* 2017). Nonetheless, the accuracy is lower than other methods (Dynesen *et al.* 2003). Besides, questionnaires are context-specific since diets vary from place to place and the food list needs to be updated raising the need to revalidate the questionnaires in each context (Schoeller and Westerterp 2017). In addition, the FFQ collects neither eating pattern information (for instance meals per day) nor detailed information on foods consumed (for example brand names) (Berdanier *et al.* 2014).

Food record

This method assesses actual or usual food intakes by requiring subjects to keep a record of all food and beverage (including snacks) at the time of consumption over periods from one to seven days. Despite the virtue of yielding accurate estimation, this method is time-consuming (Berdanier *et al.* 2014). To complete a food record, each respondent must be trained to adequately describe the foods and the amounts consumed with plentiful information including brand name of the food, preparation methods, portion sizes (Coulston *et al.* 2017). Thus, the respondent burden is higher and the accuracy highly depends on the conscientiousness of subjects (Gibson 2005).

Diet history method

This method aims to estimate the usual food intake of individuals over a longer period of time, usually a month. It often involves an interview for 24-hour recall plus information on usual eating pattern, followed by an FFQ to verify the initial data. As the food or nutrient intakes are recorded over a relatively long period, this method is useful for food policy development programmes in identifying food patterns associated with inadequate nutrient intakes (Drake *et al.* 2011). Nevertheless, the process is labour intensive, time-consuming, and results highly depend on the skills of the interviewer (Schoeller and Westerterp 2017).

To sum up, accurate estimations of food consumption and nutrient intake are of utmost importance in understanding dietary patterns, monitoring dietary quality, and informing nutrition policies. Still, the data gathered by existing dietary assessment methods are far from optimal. Moreover, no single method is superior in terms of measuring all components of diets. Hence, the choice of an appropriate dietary assessment method should be in accordance with a given purpose and in any case its strengths as well as limitations should be borne in mind. If one is interested in uncovering the overall trend in dietary patterns over time and/or across countries, national-level standardised data are recommended. On the other hand, individual-level dietary data are preferred if the prime purpose is to assess diet quality in its full form.

4.3.5 Diet quality indices and health outcomes

As mentioned earlier, the complexity of individual's diets and the interactions between nutrients undermine the traditional approach of discovering the influence of single nutrients (or food) on the risk of related diseases (Mertz 1984). Developed in a response to the need of a holistic approach, diet quality indices have become an increasingly popular tool to investigate epidemiological associations between dietary intakes and nutrition-related health outcomes (Wirt and Collins 2009; Murakami *et al.* 2020).

To recap, the majority of existing indices are constructed based on either national/international dietary guidelines or the Mediterranean diet pattern. The components included in these indices are chosen among the nutrients or food groups that are suggested by a particular nutritional guideline or characterise a certain dietary pattern. As a result, these indices represent the degree of adherence to a particular guideline or dietary pattern (Bach *et al.* 2006; Waijers *et al.* 2007; Burggraf *et al.* 2018). Yet, their predictive ability for several health outcomes needs to be confirmed since components of a particular nutritional guideline are not necessarily good predictors of health outcomes (Kourlaba and Panagiotakos 2009).

The related literature witnesses several research attempts to investigate the validity of existing diet quality indices in terms of nutrient adequacy and the risk of various health outcomes including biomarkers of disease, mortality and chronic diseases such as cardiovascular diseases (CVDs) and cancer. Tables B.1 and B.2 report an inverse relationship between indices and health outcomes in most studies. However, the association is generally modest and the predictive capacity of most indices seems to be in the same vicinity (Waijers *et al.* 2007; Arvaniti and Panagiotakos 2008). The relative risks for health outcomes are almost attenuated after controlling for confounding factors such as age, education, smoking status, BMI, physical activity (Kant 2004). It is acknowledged that differences in population size, dietary assessment methods, index scoring system, and the adjustment approaches for confounders make it more challenging to obtain consistent results across studies (Berdanier *et al.* 2014). Previous researchers report a weak relationship between diet quality indices and reduced risks of CVDs in men (McCullough *et al.* 2000a), but fail to uncover any association with reduced risks of chronic diseases in women (McCullough *et al.* 2000b).

Specific health outcomes such as cancer risks are not predicted as strongly or as consistently as all-cause mortality or CVDs (Wirt and Collins 2009; Schwingshackl and Hoffmann 2015; Neelakantan *et al.* 2018). On the one hand, many index components (for instance those of HEI-2010) are chosen from epidemiological associations with reduced risks of CVDs and its risk factors, explaining why indices tend to have a better prediction for the risk of CVDs (McCullough *et al.* 2000a; Nicklas *et al.* 2012). On the other hand, the link between index components and cancer risks could be less significant. Inconsistent results between diet quality scores and chronic diseases are also documented. To take obesity as an example, many researchers discover that lower diet quality scores are related to overweight and obesity (Nicklas *et al.* 2012; De Miguel-Etayo *et al.* 2019; Yang *et al.* 2014; El Kinany *et al.* 2020). By contrast, Villegas *et al.* (2004), Asghari et al (2012) and Moraeus *et al.* (2020) question the ability of diet quality indices to predict weight status.

Though diet quality indices can be adequate measures of overall diet quality, it is difficult to relate an index with the risk of a specific disease unless the index is particularly designed for this purpose (Radwan *et al.* 2015b). Reviewing the use of Mediterranean diet indices in epidemiological studies, Bach *et al.* (2006) conclude that these indices do not have the best predictive ability but they have a sufficient one. This could be attributed to the fact that elements of the Mediterranean diet are strongly associated with a reduced risk of coronary heart disease and several forms of cancer (Anand *et al.* 2015).

Several reasons are put forward to explain the modest association between diet quality indices and health outcomes, and most often quoted is the unresolved methodological issues during index construction (with regard to the arbitrary choices in choosing the index components, cut-off values, scoring system, and weighting) (Kourlaba and Panagiotakos 2009; Berdanier *et al.* 2014). In essence, diet quality indices may not completely fulfil the requirement of a holistic approach in which the correlation between intakes of various dietary groups needs to be resolved (Waijers *et al.* 2007). In addition, while very few diet quality scores are developed for use internationally, the vast majority of existing indices are country/region-specific and tailored to a country/region's distinguished food habits as well as disease profiles (Trijsburg *et al.* 2019). Thus, the applicability of such pre-defined scores for a wide range of health outcomes relevant for diverse populations across different cultures is often debated (Fung *et al.* 2018; Pereira *et al.* 2020). As argued by La Vecchia and Majem (2015), inter-country disparities in dietary components fundamentally reflect the availability of the food items included in the score, rather than individual choices based on health-related indications and individual consciousness and attention. As a matter of fact, some nutrients or foods could be key factors for specific health outcomes while irrelevant to other health outcomes (Radwan *et al.* 2015a). Finally, diet quality scores can assess only one aspect of food consumption whereas the confluence of several other factors such as socio-economic class, education, accessibility to healthy food stores, lifestyle, and behavioural aspects can impact the compliance to healthy eating patterns and mediate the effects of food intakes on health outcomes (Aljuraiban *et al.* 2019).

To sum up, given the large number of existing diet quality indices, no index is superlative in assessing the risk of poor health outcomes. Moreover, an index targeting to determine the general health status of a population may not be an accurate predictor of a specific disease. As a rule of thumb, a more specific index is required in order to measure more specific health outcomes.

4.4 The application of diet quality indices in this research

4.4.1 Motivations for the choice of a diet quality index

Even though validating the index's predictive capacity for health outcomes should be the main determinant for choosing a specific diet quality index, current empirical evidence is limited to facilitate this choice. Heterogeneities in sample populations, intended use, diet assessment methods, data sets and development methodologies make it difficult to arrive at a solid recommendation from the multiple existing indices. The decision of which index to use should also consider other criteria.

First, the intended purpose of the index should be defined: whether it aims to measure absolute diet quality, to assess the adherence to a particular dietary guideline, or to assist in health promotion programmes. For the latter purpose, the index should be food-based in its composition as individuals choose to eat (combinations of) foods, not nutrients (Waijers *et al.* 2007; Coulston *et al.* 2017). The use of food-based indices is readily translatable to dietary advice and public health policy applications (Kourlaba and Panagiotakos 2009; Tapsell *et al.* 2016).

Second, the index composition should be revised. Indices constructed by foods and food groups are straightforward and they can overcome several limitations of food composition tables (to name a few, being incomplete, being outdated and lacking nutrient content of processed, fortified and cooked foods) (Trijsburg *et al.* 2019). Another advantage is that interactions among nutrients contained in food groups are taken into account (Kant 1996). Looking at an index made up from vegetables, the health effects of vegetables are attributed to not only fibres but also the antioxidants, carbohydrates, proteins and other non-nutritive components. Yet, it is not easy to keep track of the large heterogeneity within food groups (Coulston *et al.* 2017). For instance, the protective properties against cardiovascular diseases vary among varieties of vegetables. Another practical drawback is the applicability of food-based indices to populations with different dietary practices without being modified. On the other hand, indices based on nutrients are more robust and can be easily adapted to different populations and countries (Verger *et al.* 2012). The *Mean Adequacy Ratio* and the *Mean Probability Adequacy Index* for example can reflect nutritional quality regarding the adequate intakes of several nutrients; however, these scores do not consider the upper limits of intake and hence are poor indicators of the overall diet quality. Another strength of nutrient-based indices is that nutrient dosages and their effects on health are directly captured (Tapsell *et al.* 2016). However, the construction process is more data-demanding and the conversion of food intakes into nutrient intakes could lead to additional measurement errors (Burggraf *et al.* 2018).

The third consideration is the scoring system and cut-off values. Many foods show a U-shape correlation with health outcomes. For example, moderate consumption of meat is beneficial but high consumption is detrimental to health. As a result, it is preferred to design scoring ranges or let the score be proportional to intake, instead of using simple cut-off values. Furthermore, scores should depend on or adjusted for total energy intake to avoid confounding by energy intake (Waijers *et al.* 2007).

Regarding all the aforementioned considerations, the *Mediterranean Adequacy Index* (MAI) seems to be the appropriate choice and will be adopted in the empirical analysis of this research because of the following reasons:

- Mediterranean diet quality indices show an overall better predictive capacity for health outcomes than indices based on dietary recommendations (Fernadi *et al.* 2018).
- This research aims to provide new evidence to assist public health policymakers with dietary monitoring and health promotion purposes, thus the use of an index based on foods and food groups like the MAI may lead to a more easily applied tool.
- The MAI is a quotient measure instead of an add/subtract score, eliminating the necessity of cut-off values. Using cut-off points based on the distribution of selected food groups in the population being evaluated would hinder the appraisal of time trends and the comparison between groups or studies (Bach *et al.* 2006).
- This research examines trends in food consumption patterns around the world and the Food Balance Sheet deems to be the most widely and extensively used dietary assessment method for this purpose. The utilisation of the Food Balance Sheet dataset makes all data needed for the MAI's calculation readily available.

4.4.2 The Mediterranean Adequacy Index (MAI)

The MAI measures the adherence to the Italian Mediterranean dietary pattern reference (Nicotera in 1960) in two Italian cohorts of the Seven Countries Study (Alberti-Fidanza *et al.* 1999; Fidanza *et al.* 2004). It is created as a quotient between the sum of energy from typical Mediterranean diet and the sum of energy from non-typical Mediterranean diet products. The higher the index value, the greater the adherence to Mediterranean diet patterns, the better the diet quality. Some typical Mediterranean foods include bread, cereals, legumes, potatoes, vegetables, fruit, fish, red wine, vegetable oils while non-typical Mediterranean foods include milk, cheese, meat, eggs, sugars, animal fats and margarines, sweet beverages, cakes, pies and cookies. The formula for MAI's calculation is given by:

$$MAI = \frac{\sum energy\ intake\ from\ Mediterranean\ components}{\sum energy\ intake\ from\ non\text{-}Mediterranean\ components} = \frac{Calories\ (good)}{Calories\ (bad)} \tag{4.1}$$

An example of MAI values corresponding to the world regions is shown in Figure 4.3. Over the past half a century the global MAI (denoted by the red dashed line) has remained quite stable at around 3, meaning that the quantity of "good" calories in an average diet is three times larger than the quantity of "bad" calories. Given the burgeoning obesity crisis worldwide over the past four decades, the fact that the healthiness of the average diet has not changed much over the same period is surprising. By definition, the MAI is just a quotient itself, therefore it is unable to capture the increasing amounts of extra calories that inevitably packed on the pounds, not least the increased intake from cheap, convenient, ultra-processed foods high in fat, sugars and salt. That is not to mention the other side of the coin – the shift towards sedentary behaviours and physical inactivity which also plays important roles in creating the obesogenic environment. As the MAI is simply a ratio, the rising MAI observed for African diets has little to imply about the better food supply that has helped reduce under-nutrition in the region but indicates improved composition of their diets. Asian diet was once the healthiest, indicated by the historical highest MAI value which however has declined gradually and approaches the global figure in 2013. While this declining pattern reflects that the composition of Asian diets has varied dramatically, it does not account for millions of undernourished people in Asia that have been lifted out of hunger. Without significant changes in the composition, diets of Europeans and Americans are assigned with consistently lowest MAI values of slightly above 1 suggesting that their diets are made up of roughly equal amount of "good" and "bad" calories.

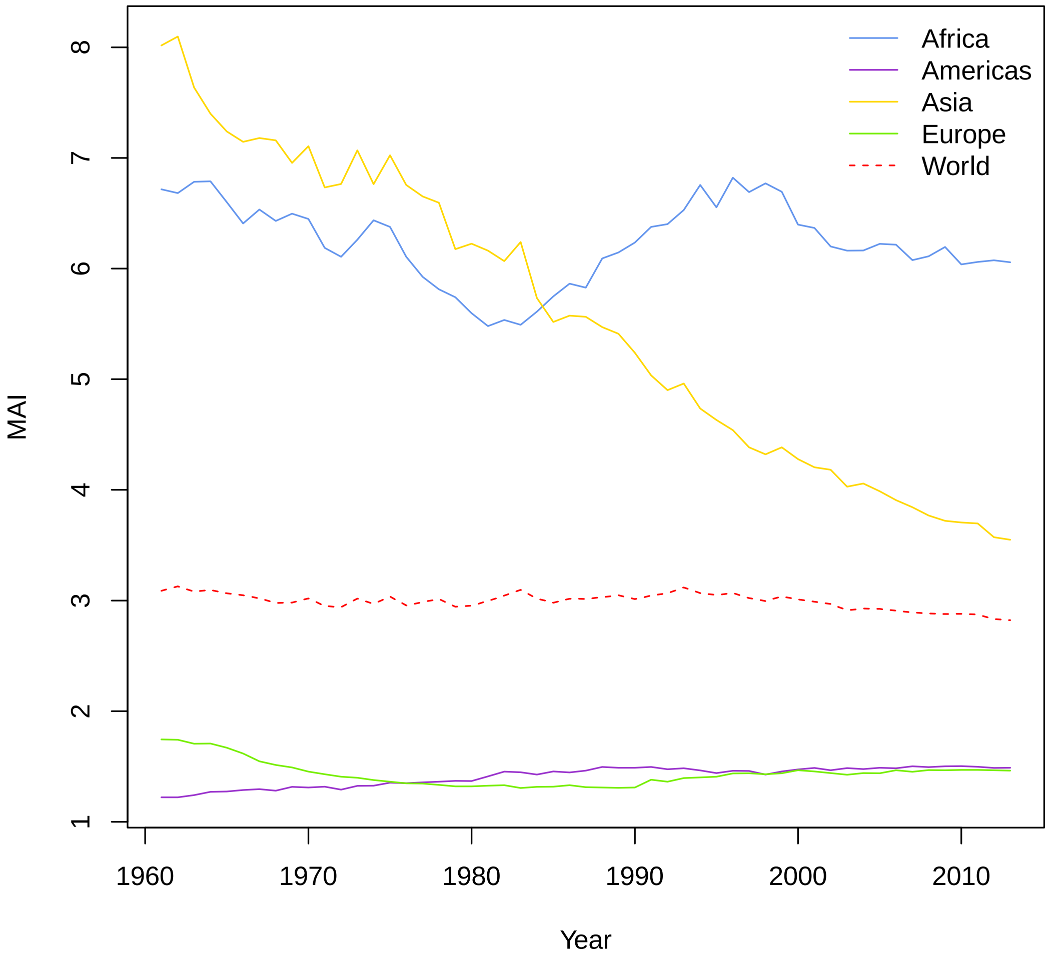


Figure 4.3 Mediterranean Adequacy Index by region, 1961-2013.

The MAI has been widely adopted by previous researchers to approximate diet quality in studies across different disciplines (Knoops *et al.* 2006; Filomeno *et al.* 2014; Di Lascio and Disegna 2017; Menotti *et al.* 2017; Montero *et al.* 2017; Finardi *et al.* 2018; Kromhout *et al.* 2018; Metro *et al.* 2018; Vilarnau *et al.* 2019). In fact, its usefulness has been extensively tested in the earlier literature (Alberti-Fidanza and Fidanza 2004; Balanza *et al.* 2007; Rodrigues *et al.* 2008; Alberti *et al.* 2009; Da Silva *et al.* 2009; Bach-Faig *et al.* 2011; Chang *et al.* 2017; Finardi *et al.* 2018). Table 4.2 provides a summary of the MAI's pros and cons as a diet quality indicator.

Table 4.2 Strengths and limitations of the MAI.

In terms of	Strengths	Limitations
Ease of calculation	• Simple and easy to calculate • Does not need to convert food intakes to nutrients using energy densities	None
National coverage	Identifies dietary characteristics of different populations	Its validity at household level needs to be confirmed
Representativeness of diets	• It is one of the most predictive indicators of a Mediterranean diet • Shows a strong inverse association with coronary heart disease mortality	• It is debatable if potatoes and seed oils are typical Mediterranean diets or not • Does not consider the overall calorie intake • Can be misleading for children, pregnant and lactating women, and those following special diets
Flexibility	It is a quotient measure instead of an add/subtract score, removing the necessity of cut-off points	Can be used with dietary data obtained in large-scale studies only if dietary assessment methods are reliable and valid
Efficiency	• Captures major shifts in food availability trends over time and across countries • Has satisfactory discriminating power for longitudinal data	• Red meat and poultry are not separated • Same importance is given to all food groups

4.5 Chapter conclusion

There has been a shift from the traditional approach focusing on the role of single nutrients or foods on disease development due to the complexity of human diets and the interaction among nutrients contained within food groups. Following a more holistic approach, diets should instead be considered as a whole when establishing any epidemiological association between overall diet and health outcomes. As a result, methods for quantifying diet quality have evolved and a widely used method in epidemiological studies is pre-defined diet quality indices. A large number of indices are proposed, and almost all of them aim to quantify the degree of adherence to a particular dietary guideline or dietary patterns. Overall, diet quality indices can be classified into two categories: indices based on national/international nutritional recommendations and indices based on the Mediterranean diet.

Several studies in the literature attempt to compare the great multitude of existing diet quality indices on the basis of construction criteria and associations with health outcomes. Many similarities and differences among available diet quality indices can be explained by the methodology for index development, which is highly subjective due to the arbitrary choices made by the researcher. These choices involve the selection of components that should be included in the index, the cut-off values that should be used for each component, and the weights that should be assigned to each component.

Diet quality indices are increasingly being used to measure associations with biomarkers and health outcomes. It is found that diet indices are associated with a reduced risk of all-cause mortality and/or mortality and selected diseases. However, this association is attenuated when confounding variables are adjusted. The proposed indices are adequate tools to evaluate the overall diet quality, but they have moderate ability to predict chronic diseases and health determinants, casting doubts on the validity of these indices.

In this research, the world's diets are first empirically defined by cluster analysis, and a pre-defined diet quality index is then exploited to evaluate the diet healthiness. More specifically, the Mediterranean Adequacy Index (MAI) is selected based on the research objectives, data availability as well as the construction characteristics and numerous proposed advantages of this index in earlier studies. In this regard, it is helpful to investigate the evolution of MAIs over time as it would indicate the direction that global diets are heading towards and which diet exhibits the most worrying trend.

Chapter 5

Changes in food consumption patterns over time

5.1 Chapter introduction

There is a general consensus that food and nutrition information is critical in monitoring the well-being of any population and therefore should be incorporated into national information system (FAO 2018). Without such robust information, the progress towards the achievement of Sustainable Development Goals 1 and 2 (Ending poverty and Ending hunger respectively) could not be adequately measured (Zezza *et al.* 2017). The global demand for reliable data on what people eat and drink was acknowledged at the Second International Conference on Nutrition (ICN2) in 2014: "Nutrition data and indicators, as well as the capacity of, and support to all countries, especially developing countries … need to be improved in order to contribute to more effective nutrition surveillance, policy-making and accountability" (FAO/WHO 2014, p.3). Despite unprecedented progress in the production of household consumption and expenditure data over the last two decades, national statistical agencies are using different methods to collect data on food consumption of their population, not to mention the great disparity in the level of detail and the quality of data across countries (Micha *et al.* 2018; Ioannidou *et al.* 2020; Placzek 2021).

Against this backdrop, the Food Balance Sheet (FBS) compiled by the Food and Agriculture Organisation (FAO) of the United Nations serves as an inexpensive source of nutrition data which is highly standardised and well suited for both within- and between-country comparison (Popkin and Reardon 2018; Lopez Barrera and Hertel 2021). Being one of the most extensively used databases on food supply and consumption, the FBS gives a comprehensive picture of food consumption by keeping track of annual energy supplies of around 100 primary commodities and food aggregates for approximately 185 countries dating back to 1961 (FAO 2018). Notwithstanding its popularity in empirical studies, the FBS is subject to a critical shortcoming that the consumption-level waste (food waste at retail, restaurants, and household) is not accounted for and hence the data should be interpreted as *food available for consumption* rather than actual food intake (Vilarnau *et al.* 2019).

Utilising the FBS data, this chapter provides a description of changing patterns of food consumption over time. The objective is to identify noticeable dietary changes worldwide and to highlight evidence of the nutrition transition. The great level of detail and the longitudinal nature of the FBS data allow the detection of not only common trends but also geographical heterogeneity in the evolution of the world's diets.

The rest of the chapter is structured as follows. Section 5.2 assesses major changes in the composition of the global average diet by macronutrients and by food aggregates. Section 5.3 disentangles dietary changes by the world regions and examines some indication of convergence/divergence in the energy supply of main food groups. Section 5.4 offers some insight into the meaningful correlation between food supply and economic development whilst Section 5.5 depicts important trends in obesity prevalence. Section 5.6 concludes.

5.2 Evolution of the global diet

5.2.1 Diet composition by macronutrients

The energy value of food (quantified as kilocalories per capita per day, or kcal/capita/day) is a primary indicator of food security and one of the most frequently used measures of food supply. However, the adequate supply of other macronutrients is of crucial importance in maintaining proper nutrition. Protein, of whatever sources, serves as the major structural component of muscle and other tissues and is needed to produce hormones, enzymes and haemoglobin (Hoffman and Falvo 2004). Fat is another macronutrient that plays a vital role in human diet as it facilitates the absorption of fat-soluble vitamins and forms a structural component of cell walls (Schmid 2010).

The FBS provides data on total energy (in kcal/capita/day) as well as fat and protein supply (both in grams/capita/day). To derive the energy supply of carbohydrates, an energy density of 4 kcal per gram of protein and 9 kcal per gram of fat is assumed. These values are based on the FAO's established nutritional guidelines (FAO 2003). The quantity of calories (in kcal/capita/day) from protein and fat is therefore calculated by multiplying the daily supply of protein (in grams) by 4, and the daily supply of fat (in grams) by 9. The daily carbohydrate supply (in kcal/capita/day) can be given as:

$$Carbohydrate\ supply = Total\ calories - (protein\ supply \times 4 + fat\ supply \times 9) \quad (5.1)$$

Figure 5.1 shows changes in total energy as well as composition of the global diet by three macronutrients (carbohydrate, fat, and protein) over the period 1961-2013. Overall, the daily per capita calories grew relentlessly from 2,196 kcal/capita/day in 1961 to 2,884 kcal/capita/day in 2013 - an increase of 31%. Particularly, the rise in total calories seems to be mostly driven by fat - the energy supply of which nearly doubled in 2013. Despite a modest increase of only 19% over the period, carbohydrate has always been the largest source of energy, followed by fat and protein.

Another feature standing out from Figure 5.1 is that not only the quantity of protein but also the composition from different sources of protein has changed over the past half century. In fact, protein can come from either animal or plant origin, and the quality of two protein varieties is not the same. Animal-derived protein, also referred as complete protein, contains all essential amino acids. Plant-derived protein, except from pulses, some nuts and seeds, is incomplete due to the absence of one or two essential amino acids, and therefore is usually considered of lower quality (Hoffman and Falvo 2004). Figure 5.1 reveals a clear dominance of plant protein in the global diet, yet the more marked growth of animal protein (63% compared with 18%) signals that the plant-to-animal ratio in protein supply is approaching one.

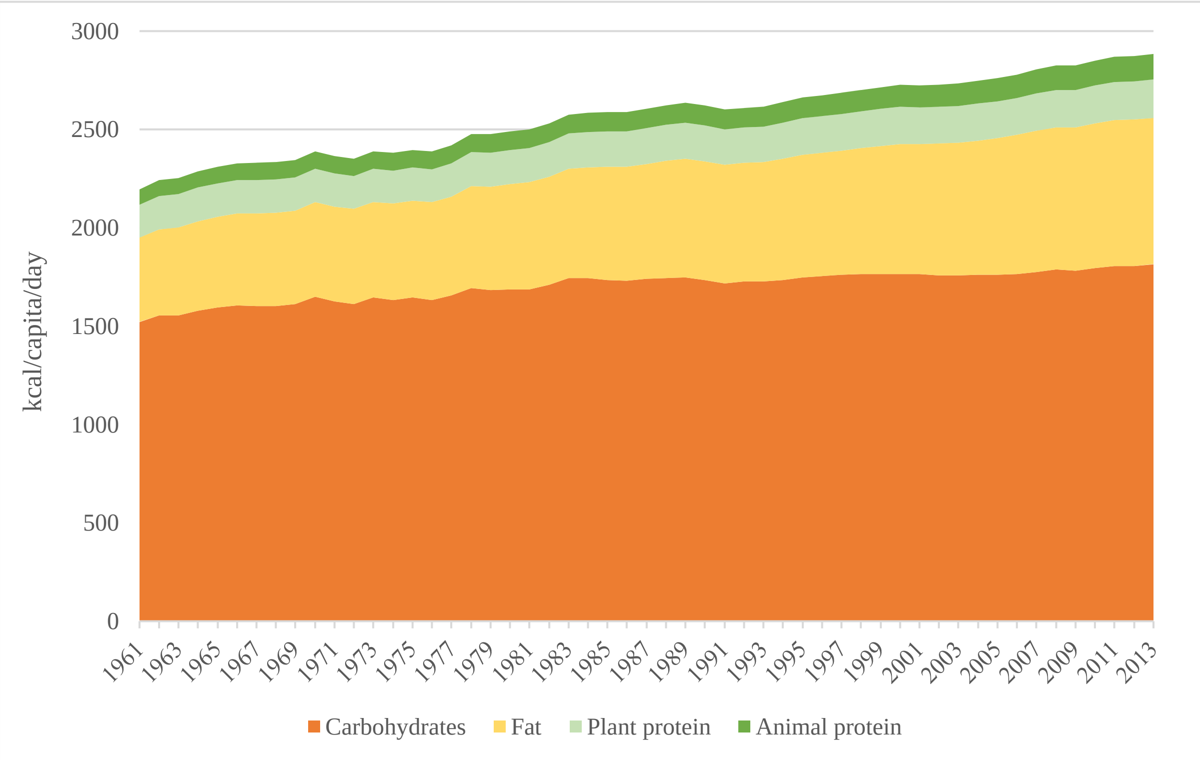


Figure 5.1 Macronutrient composition of the global diet, 1961-2013.

Figure 5.2 illustrates the contribution of energy supply from three macronutrients to the total caloric figure as well as the proportions of protein from two sources in 1961 and 2013. While the share of energy from carbohydrates dropped from 69% in 1961 to 63% in 2013, the energy share of protein remained unchanged (11%) and the contribution of fat jumped from 19% to 26%. Even though carbohydrates account for the majority energy of the global diet, its relative importance has declined, giving way to fat. In terms of protein composition, plants have been the main source of protein over the past half a century; however, individuals are collecting a greater amount of protein from animal sources, shown by a rise from 32% in 1961 to 40% in 2013. Evidence of the nutrition transition is therefore obvious in the replacement of carbohydrate with fat and the increased consumption of animal-sourced foods.

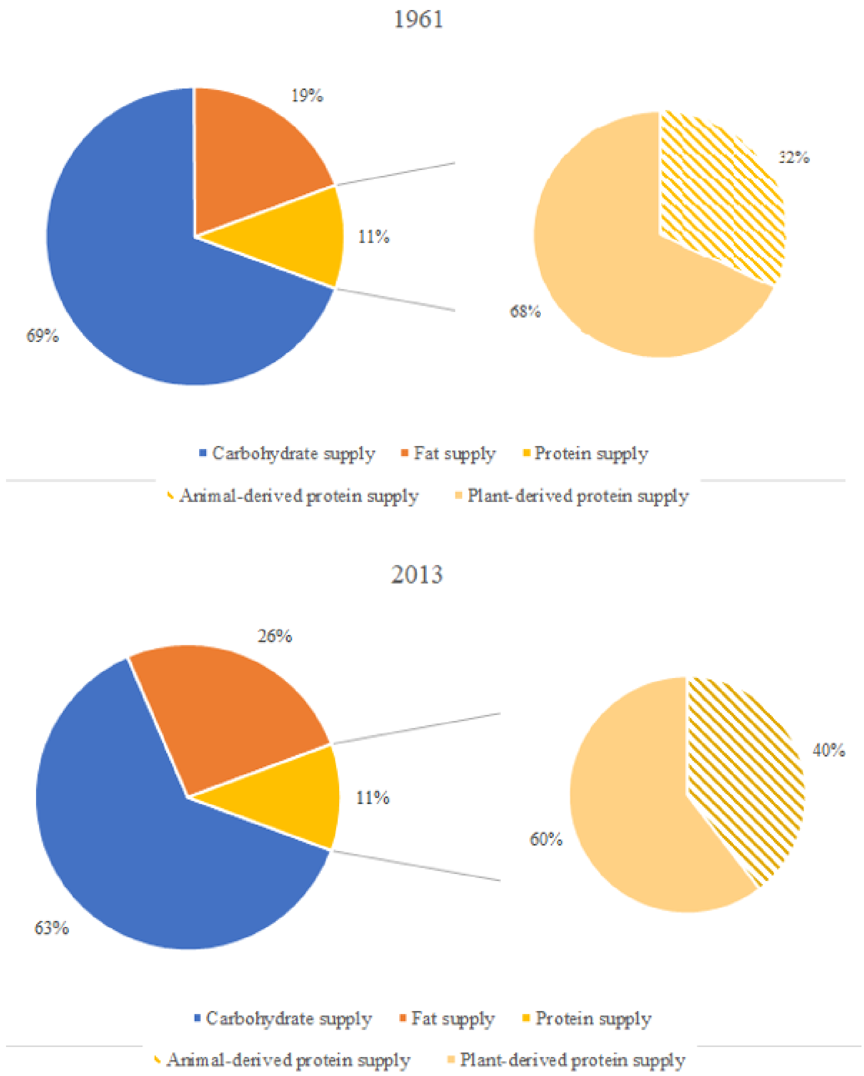


Figure 5.2 Energy share from macronutrients in the global diet, 1961 and 2013.

5.2.2 Diet composition by main food aggregates

As individuals consume (combinations) of foods rather than single nutrients, it would be more meaningful to look at diet composition from the perspective of individual food aggregates. Table 5.1 reports changes in the structure of the global diet regarding twelve main food groups: (1) Cereals – excluding beer, (2) Vegetable oils, (3) Vegetables, (4) Pulses, (5) Sugar and Sweeteners, (6) Starchy roots, (7) Alcoholic beverage, (8) Fruits – excluding wine, (9) Eggs & Milk (excluding butter), (10) Meat, (11) Animal fats, (12) Fish and seafood. These food groups represent a comprehensive picture of an overall diet in accordance with the FAO's aggregation and they can be separated into plant-sourced foods (1-8) versus animal-sourced foods (9-12). It is clear from the inspection of Table 5.1 that energy intakes from most food aggregates (except starchy roots, pulses, and animal fats) increased over the period 1961-2013, contributing to the rising calorie content of the global diet. Specifically, the growth rate is largest for vegetable oils. This result resonates with Drewnowski and Popkin (1997) who argue that the nutrition transition starts with a surge in the consumption of cheap cooking oils and is characterised by the substitution of carbohydrates for fat. As Section 5.2.1 mentions the increasingly important role of animal- versus plant-derived protein, it is worth examining the various specific sources of protein presented in Table 5.1. While the supply of pulses – a food commodity commonly found in diets of many developing countries declined by almost one fourth, the apparent consumption of meat as well as other animal products (such as fish and seafood) more than doubled. Over the past five decades, the growth rate of energy from animal-sourced foods is almost two times higher than vegetable-sourced foods even though the latter makes up approximately 80% of the total energy supply. So, not only we are collecting more calories, but we are consuming more from animal origin.

Table 5.1 Composition of the global diet by commodity group, 1961 and 2013.

	Food aggregates	**1961 (kcal/capita/day)**	**2013 (kcal/capita/day)**	**Change (%)**
Vegetable-sourced	Cereals - Excluding Beer	1,086	1,292	19
	Vegetable Oils	113	271	140
	Vegetables	44	95	116
	Pulses	89	68	-24
	Sugar & Sweeteners	193	236	22
	Starchy Roots	175	141	-19
	Alcoholic Beverages	53	69	30
	Fruits - Excluding Wine	51	97	90
	Total	*1,804*	*2,269*	*26*
Animal-sourced	Eggs & Milk (Excluding Butter)	136	174	28
	Meat	110	237	115
	Animal fats	71	61	-14
	Fish, Seafood	17	34	100
	Total	*334*	*506*	*51*
	Grand Total	**2,196**	**2,884**	**31**

Table 5.2 shows changes in the share of energy from twelve main food aggregates to the global diet in 1961 and 2013. Cereals remain to be the largest energy provider, however, their energy share dropped from 49% to 45%. In 1961, sugar & sweeteners and starchy roots made up a large proportion of the total energy intake (9% and 8% respectively). In 2013, sugar and sweeteners still represented a

considerable proportion (8%) whilst the energy share from starchy roots dropped to 5%. On the other hand, the proportion of energy from vegetable oils and meat increased significantly, each accounts for approximately one-tenth of the total dietary energy. At the same time, energy contributions from vegetables and fruits, despite a marginal increase, remained relatively low.

Table 5.2 Energy share from main food groups to the global diet, 1961 and 2013.

	Food aggregates	1961 (%)	2013 (%)	Change (%)
Vegetable-sourced	Cereals - Excluding Beer	49	45	-4
	Vegetable Oils	5	9	4
	Vegetables	2	3	1
	Pulses	4	2	-2
	Sugar & Sweeteners	9	8	-1
	Starchy Roots	8	5	-3
	Alcoholic Beverages	2	2	0
	Fruits - Excluding Wine	2	3	1
Animal-sourced	Eggs & Milk (Excluding Butter)	6	6	0
	Meat	5	8	3
	Animal fats	3	2	-1
	Fish, Seafood	1	1	0

Broadly speaking, two major modifications in the composition of the global diet are observed. Foremost is the decline in the contribution of staple foods such as cereals and starchy roots which were predominant in the 'traditional diet', in substitution for energy-dense foods such as vegetable oils and sugar & sweeteners. Second, the share of animal-sourced foods in the total energy has increased, fuelled by the shift from pulses to meat. Not only the global diet has become more calorific, but it starts to pick up elements of the 'Western' diet – a trend that is predicted by the nutrition transition model. Other evidence of the nutrition transition is available if one delves deeper into the changes regarding individual food aggregate.

Take the case of cereals as an example. According to Table 5.1, data from the FBS indicates a rise in the availability of cereals during the past half century. However, changes in food consumption entail both quantity and quality dimensions. As illustrated in Figure 5.3, the supply of traditional cereals such as maize, millet and sorghum declines whereas an increase is reported for more widely used grains (such as rice and wheat). This variation reflects the move from coarse to more polished grains in countries that are experiencing the nutrition transition (WHO 2003).

Another example comes from starchy roots. Represented in various forms such as cassava, yam and sweet potatoes, starchy roots appear in commonly consumed dishes in many parts of the world. As indicated in Table 5.1, the supply of starchy roots reduced significantly by 19% from 1961 to 2013. Nonetheless, the specific type of roots also altered. Figure 5.4 displays a marked decrease in the supply of sweet potatoes over the past 50 years but a rise in the supply of potatoes – a staple in the 'Western' diet since the 1980s. This signals the adoption of the 'Western' dietary pattern in the global diet for the last few decades.

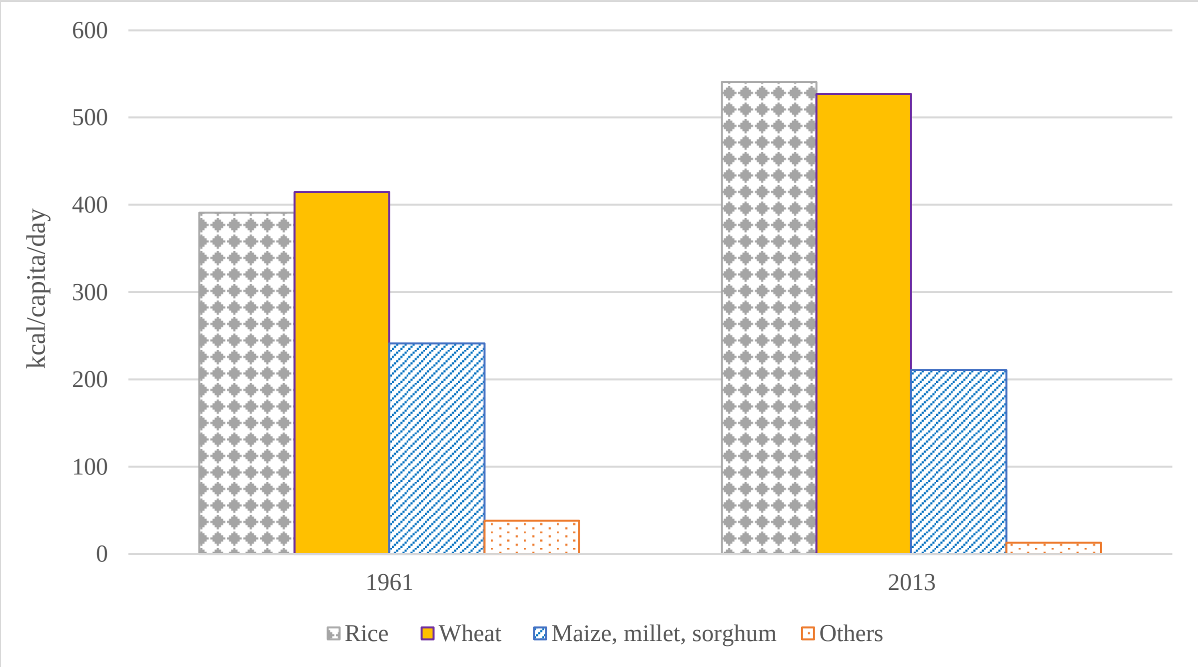


Figure 5.3 Daily per capita calories of cereals by commodity type, 1961 and 2013.

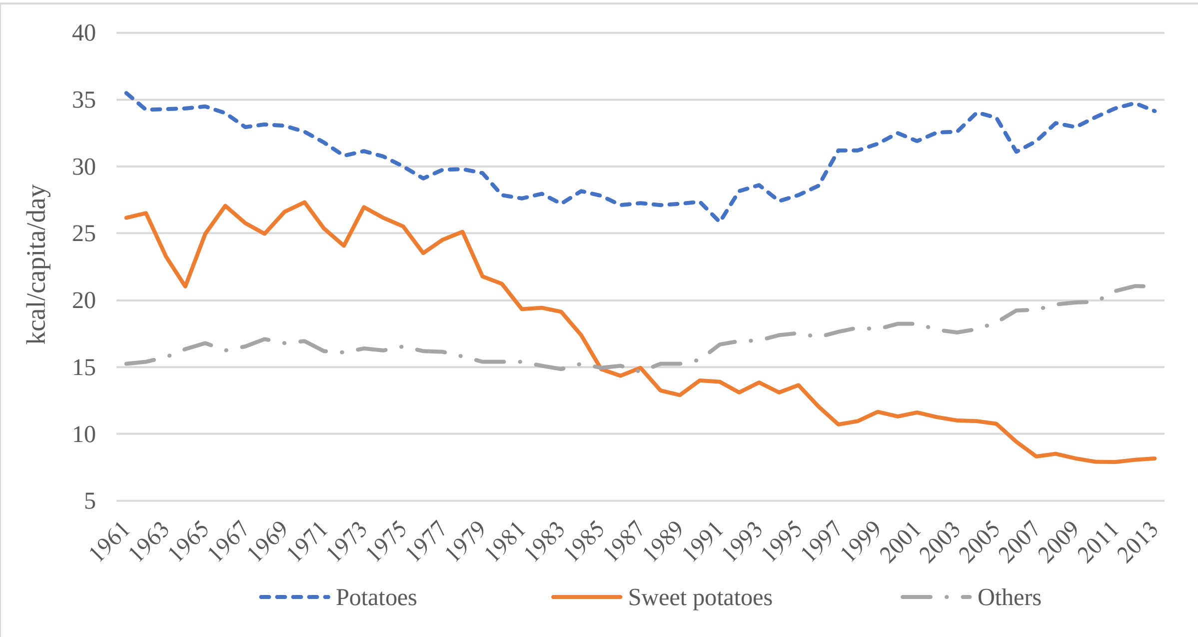


Figure 5.4 Daily per capita calories of starchy roots by commodity type, 1961-2013.

So far, the examination of composition of the global diet by main food groups has revealed evidence for the nutrition transition that has happened over the past half a century as the global diet becomes more calorific and 'Westernised'. But is it a homogeneous transition across the globe or does it better denote a common trend with regional/national nuances? The next section examines changing patterns of food consumption regarding main regions of the world.

5.3 Evolution in diets of the world regions

5.3.1 An overview on food consumption among the world regions

Using the FBS data for regional groupings, this section presents changes in regional consumption patterns for Asia, Africa, Europe, Oceania, Northern America and South America. Figure 5.5 depicts the average daily caloric consumption from 1961 to 2013. Two features seem noteworthy. First, a rising trend is observed for all regions; however, spatial variations do exist. While South America, Africa and Asia witnessed a remarkable growth in calorie availability from as low as 1,805 to somewhat 2,779 kcal/capita/day, the calorie figure in Oceania increased slightly but with a considerable delay. On the contrary, the total caloric supply in Northern America has plateaued since 1995. Surprisingly, the energy supply in Europe, despite a rapid escalation in the first three decades of the period, declined slightly and remained stagnant during 1990s. This sudden drop could be attributed by the falling calories in the countries experiencing economic transition after the disintegration of the former Soviet Union/block and the former Yugoslavia (WHO 2003).

Second, total calories have increased at varying speeds across world regions. The rising pace in Asia was most substantial (54%), followed by Africa and South America (32% and 31% respectively). On the other hand, the rises in Europe and Oceania – the two most calorific regions in 1961 were much smaller than the average figure. Overall, a steeper rise pertains in poorer regions, signalling that national food supplies are converging globally. The fact that the least calorific regions experienced the most remarked growths implies the 'catching-up' effect in the beta convergence literature.

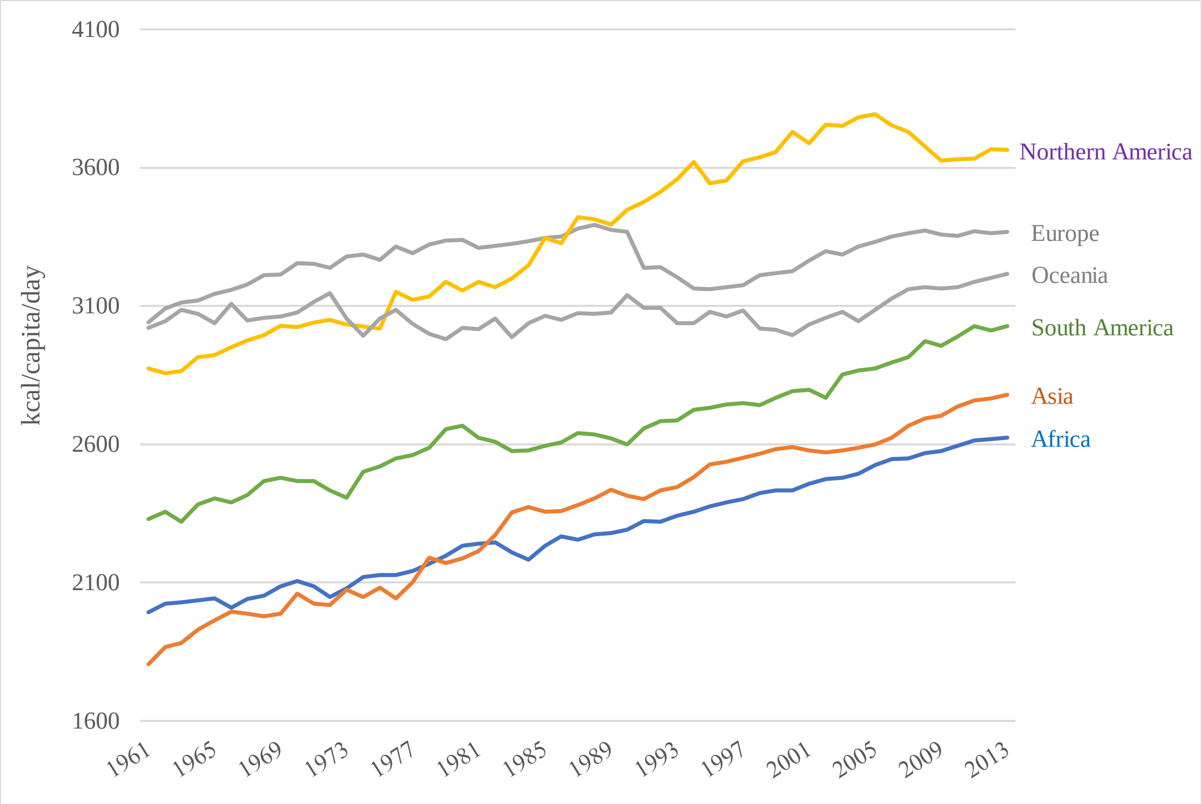


Figure 5.5 Daily per capita calories by region, 1961-2013.

For a better illustration of the caloric distribution over the world, Figure 5.6 maps the calorie content (in kcal/capita/day) by countries in 1961 and 2013. The level of calories is represented by a shade of orange: the darker the shade, the higher the calories. It is clear from the darker colour that individuals around the globe have accessed to an increased quantity of calories over the past half a century. In 2013, the energy supply in countries across Europe, Oceania and North America was mostly larger than 3,250 kcal/capita/day whereas the figure fell in the range between 2,750 and 3,100 kcal/capita/day in 1961. In poorer countries across South Asia, Sub-Saharan Africa and South America, the caloric consumption ranged from 2,300 to 3,000 kcal/capita/day in 2013 whilst the figure was well below 2,300 kcal/capita/day in 1961.

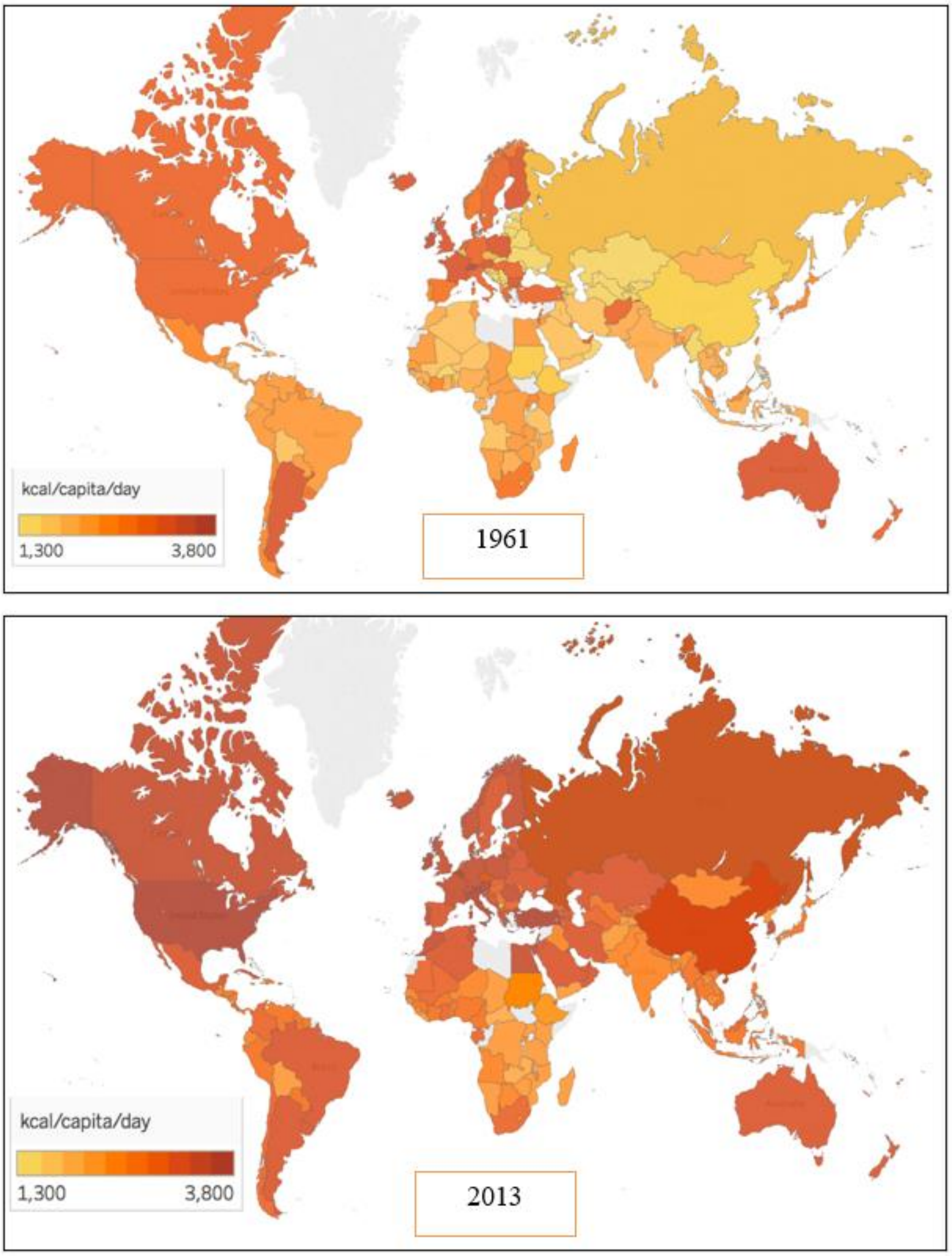


Figure 5.6 World map of daily per capita calories, 1961 and 2013.

As shown in the earlier section, the global diet experienced a surge in the energy contribution from animal products over the last five decades – an indication of the nutrition transition. Figure 5.7 compares changes in the share of animal-sourced foods among the world regions. Overall, a rising trend is observed in most regions; however, the rising speed differs greatly. While the increase in South America and Europe seems to be in line with the global average figure, the animal-derived energy supply nearly tripled in Asia from 6% to 16% but remained constant in Africa (at around 8%). Interestingly, Northern America and Oceania are the only two regions that witnessed a decrease in the energy share from animal-sourced foods between 1961 and 2013. Nonetheless, dwellers in these two regions are taking almost one third of their daily calories from animal products.

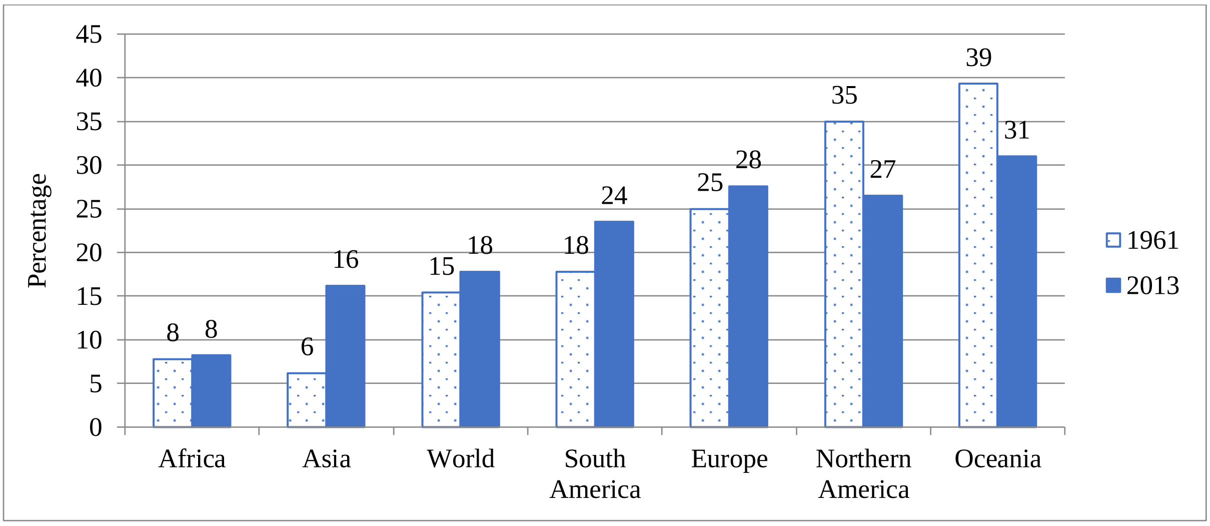


Figure 5.7 Energy share from animal-sourced products by region, 1961 and 2013.

5.3.2 Convergence versus divergence

The extant literature has documented mounting evidence for the dietary convergence that is happening on a global scale as a result of the nutrition transition (see, *inter alia*, Popkin 1993; Hawkes 2006; Kearney 2010; Popkin *et al.* 2012; Khoury *et al.* 2014; Bentham *et al.* 2020). Section 5.3.1 presents evidence for the convergence in the caloric consumption of the world regions. This section further examines the convergence/divergence among regional food consumption patterns by looking at the evolution in calories from select food aggregates: starchy roots, vegetables, meat, and vegetable oils. These food groups are chosen since their historical changes in the structure of the global diet are most remarkable among the twelve main food groups reported in Table 5.1. Changes in the caloric supply of the four food groups are plotted in Figure 5.8. In each line graph, the horizontal axis runs from 1961 to 2013, the vertical axis shows the number of calories in terms of kcal/capita/day, and each colour denotes a world region. In order to determine convergence in each graph, the concepts of sigma convergence and beta convergence are utilised. The former occurs if the gap between the lines reduces over time; otherwise, divergence is detected. On the other hand, beta convergence happens when regions with the initially lower levels of calories exhibit the most robust growth rate and thereby "catch up" with regions having the historical higher calorie levels.

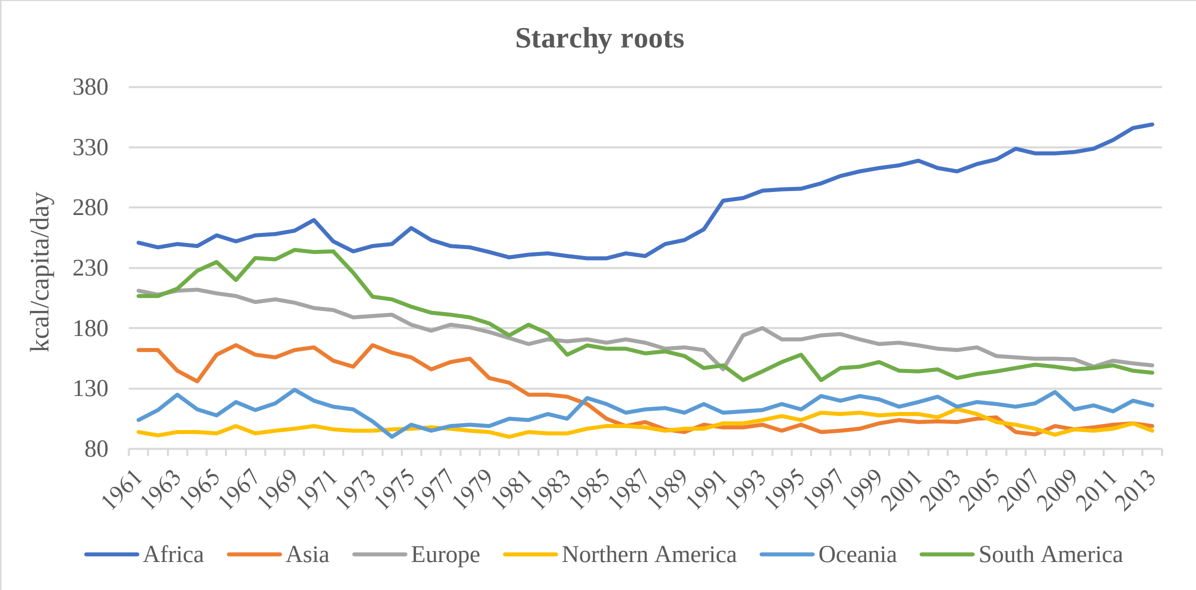
Confining attention first to starchy roots, Figure 5.8 clearly shows a declining trend for most regions. In spite of differing initial levels in 1961, the calories from starchy roots in Asia, Oceania, Northern America, South America, and Europe were all approaching 130 kcal/person/day in 2013. Moreover, the gaps between the lines representing these regions have lessened, indicating (sigma) convergence. At the same time, Africa exhibited a diverging pattern from the rest of the world in the sense that its calories from starchy roots have increased strongly, especially over the past 25 years.

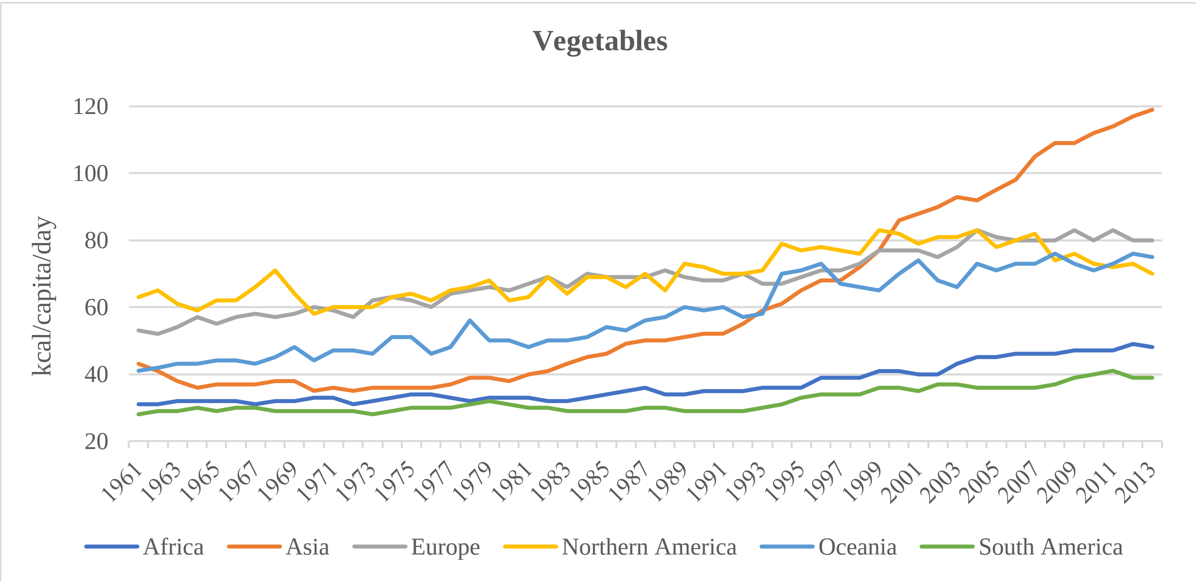
Turning to the calories from vegetables, an upward trend is observed for all six regions, yet the rising speeds differ greatly. The figure grew slightly in Africa and South America, and these two regions could represent a group of the lowest consumption level. Being associated with a low level of calories in the beginning two decades, Asia however witnessed the most remarkable growth where the calories from vegetables almost tripled between 1961 and 2013. Particularly, the fact that the calorie figure has been denoted by an almost straight line since 1991 suggests that the vegetable consumption in Asia would likely continue its rising tendency and approach a level that is distinctively higher than the remaining regions. Turning attention to Northern America, Europe and Oceania, Figure 5.8 reveals (sigma) convergence similar to the energy supply of starchy roots. Over the past half a century, the gaps between the lines have become narrower. Collectively these regions could represent a group of middle level of vegetable consumption.

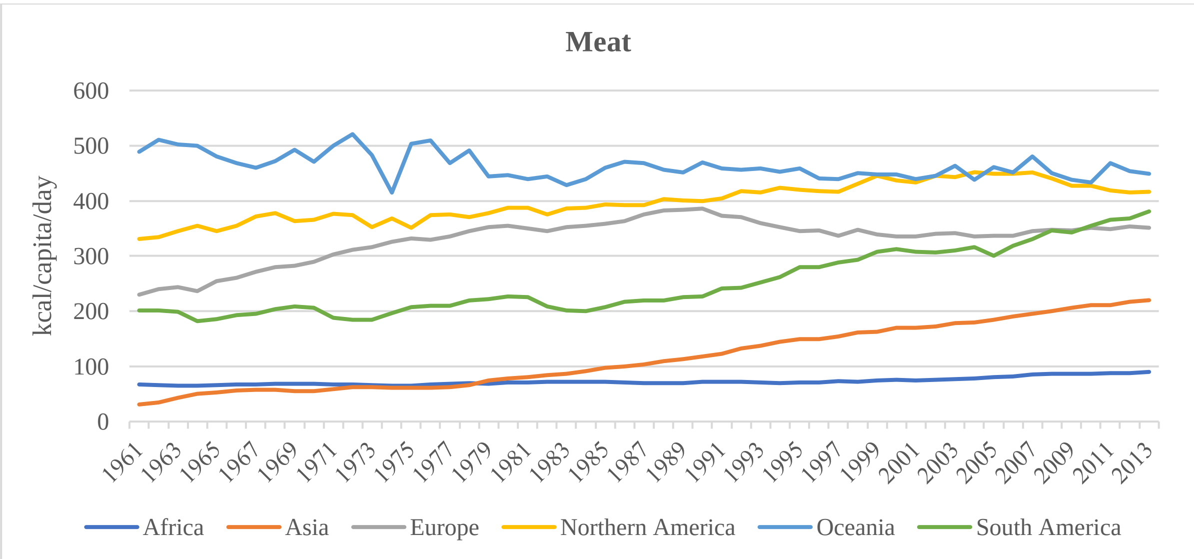
Regarding meat, an increasing trend is observed for all regions except Oceania. Over the past 50 years, the meat supply in Northern America, South America, and Europe rose dramatically and seemed to be converging on the level of Oceania. Both started out at the lowest level of about 50 kcal in 1961, the daily per capita supply in Asia has grown relentlessly and appeared to be on the trajectory path to converge with the West, whereas the supply remained stagnant in Africa at below 100 kcal and showed no signs of ‘catching up’ with the remaining regions.

In terms of vegetable oils, the supply was initially highest in Northern America, middle in Europe and low in the remaining regions. Throughout the 50-year-period, the supply remained highest in Northern America; nonetheless, the calorie figure increased dramatically and at least doubled everywhere between 1961 and 2013. Indeed, the rising speeds vary across regions. Notably, Oceania exhibited the most significant growth with a fivefold increase and was heading into the territory of middle consumption level with Europe and South America in 2013. This therefore provides evidence for the ‘catching up’ phenomenon in the beta convergence literature. In addition, Figure 5.8 illustrates a reduction in the distance between the lines representing Asia and Africa, which is indicative of (sigma) convergence. Diets of both regions contained the smallest number of calories from vegetable oils, at approximately 200 kcal in 2013.

Overall, the examination of changes in the energy supply of select food groups reveals several converging patterns as different groups representing different consumption levels can be observed in each line graph. As a result, diets of the world regions have become more similar for example in reducing calories from starchy roots and gaining calories from meat or vegetable oils. Nevertheless, with regard to each food group, it is not hard to find region(s) that seemed to buck the general upward/downward trend and diverged from the rest of the world. Hence, it can be too soon to conclude that diets around the world are converging on a single international norm and convergence better describes a tendency rather than a universal dietary type.


Starchy roots
kcal/capita/day
380
330
280
230
180
130
80
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
Africa
Asia
Europe
Northern America
Oceania
South America


Vegetables
kcal/capita/day
120
100
80
60
40
20
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
Africa
Asia
Europe
Northern America
Oceania
South America


Meat
kcal/capita/day
600
500
400
300
200
100
0
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
Africa
Asia
Europe
Northern America
Oceania
South America

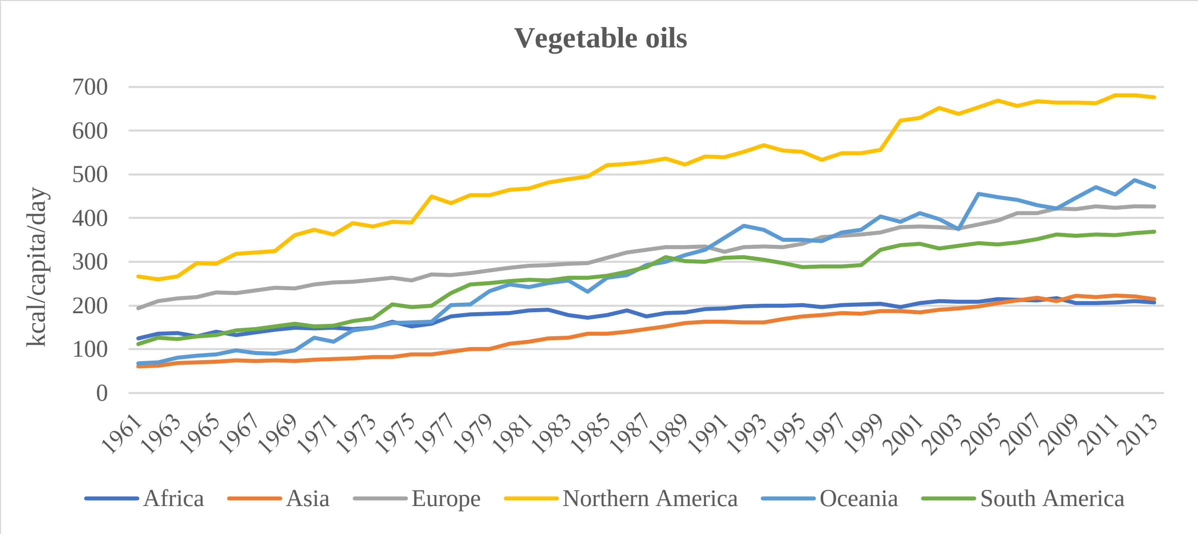


Figure 5.8 Daily per capita calories of select food groups by region, 1961-2013.

5.4 General trends in global obesity

Major changes in the global patterns of food consumption were demonstrated in the earlier sections. Setting aside regional/local heterogeneities, there is a common theme among the global diets: rising calorie content and the shift towards fat and animal-based foods (typical of the nutrition transition). These dietary changes occur in tandem with the growing worldwide epidemic of obesity which is now recognised as one of the most important public health problems facing the world today (OECD 2019b). This section aims to depict the important trends regarding the obesity prevalence worldwide.

Before showing the data relative to the phenomenon in different geographical areas/country groups, it is useful to clarify what is meant by obesity. At a basic level, obesity or weight gain is determined by an imbalance of energy. When energy intake (often measured in kilocalories) exceeds the energy burnt off through daily activities, the excess energy (also known as *energy surplus*) is stored by the body as fat and weight gain is expected. Conversely, an *energy deficit* is caused by consuming less energy than what is expended in physical activity. From this arise two drivers of obesity: increased kilocalorie intakes (particularly via consumption of fatty and sugary foods) and reduced energy expenditure (by leading a sedentary lifestyle). It was shown earlier in this chapter that the calorie supply went up in most countries over the past half a century. If this increase was not met with an adequate level of energy expenditure, weight gain and a rise in obesity rates is inevitable.

WHO deems overweight and obesity as a state of excessive accumulation of fats ("adiposity") (WHO 2017a). An individual's degree of adiposity cannot be immediately measured; however, there are some proxies based on anthropometric features that can be easily measured. *Body Mass Index* (BMI), defined as the ratio between weight (in kilograms) and the square of the height in metres (m^2), is the most used indicator (BCFN 2012). For example, an adult who weighs 70kg and is 1.7m tall will have a BMI of $70/1.70^2 = 24.2$. Based on the measured BMI values, an individual can be considered as underweight, overweight, or obese. The WHO defines these different conditions using the cut-off points in Table 5.3. An adult with a BMI greater than or equal to 25 kg/m^2 is classified as being overweight, and a BMI greater than or equal to 30 kg/m^2 is defined as obese.

Table 5.3 WHO classification of underweight, overweight and obesity.

Classification	BMI (kg/m^2)
Underweight	<18.50
Normal range	18.50 – 24.99
Overweight	≥ 25.00
Pre-obese	25.00 – 29.99
Obese	≥ 30.00
Obese class I	30.00 – 34.99
Obese class II (severe obesity)	35.00 – 39.99
Obese class III (morbid obesity)	≥ 40.00

Source: WHO (2004).

Being recognised internationally, BMI is a quick and effective method for estimating body fat and monitoring obesity trends at population level (Lau *et al.* 2020). However, it is subject to several criticism. First, BMI is a poor indicator of percentage of body fat as it does not distinguish between mass due to body fat and mass due to muscular physique, nor the distribution of fat (Nuttall 2015). Other factors, such as the waist-to-hip ratio, waist-to-height ratio and the amount as well as distribution of fat on the body are important in assessing the metabolic as well as mortality consequences of excessive fat accumulation. As explained by Yusuf *et al.* (2005), waist-to-hip ratio is a better predictor of coronary disease among BMI, waist-to-hip ratio and waist circumference. Second, the cut-off values for BMI need adjusting when implementing for a specific ethnic. For example, Lau et al (2020) recommend BMI cut-offs of 23 and 27.5 for Asian adults since they have a higher body fat percentage and greater cardiovascular risks compared with non-Asian with the same BMI due to variations in muscularity and body frame.

Data on obesity prevalence, i.e. percentage of defined population with a BMI ≥ 30 kg/m^2 (%, age-standardised) are available from the WHO Global Health Observatory data repository[1]. The data

[1] https://apps.who.int/gho/data/node

spans from 1975 to 2016, and can be disaggregated by geographical regions, by income groups and by gender. Using this data set, major trends in obesity prevalence are explored subsequently.

Figure 5.9 shows the obesity prevalence by WHO regions. On average, the worldwide prevalence of obesity almost tripled from 4.7% in 1975 to 13% in 2016. A striking increase is observed for all regions. Back in 1975, the obesity rate was largest in Europe and Americas (around 9%) and these two regions continue to lead with the highest rates in 2016 (23% and 29% respectively). By contrast, the figure is smallest in South-East Asia and Western Pacific; yet, the rises are staggering, from less than 1% of the total population being obese in 1975 to 4.7% and 6.4% respectively in 2016. These results are in accord with the sobering fact that after 30 years of trying, no country has been able to significantly reverse its rising obesity trend (Swinburn *et al.* 2011; Swinburn *et al.* 2019).

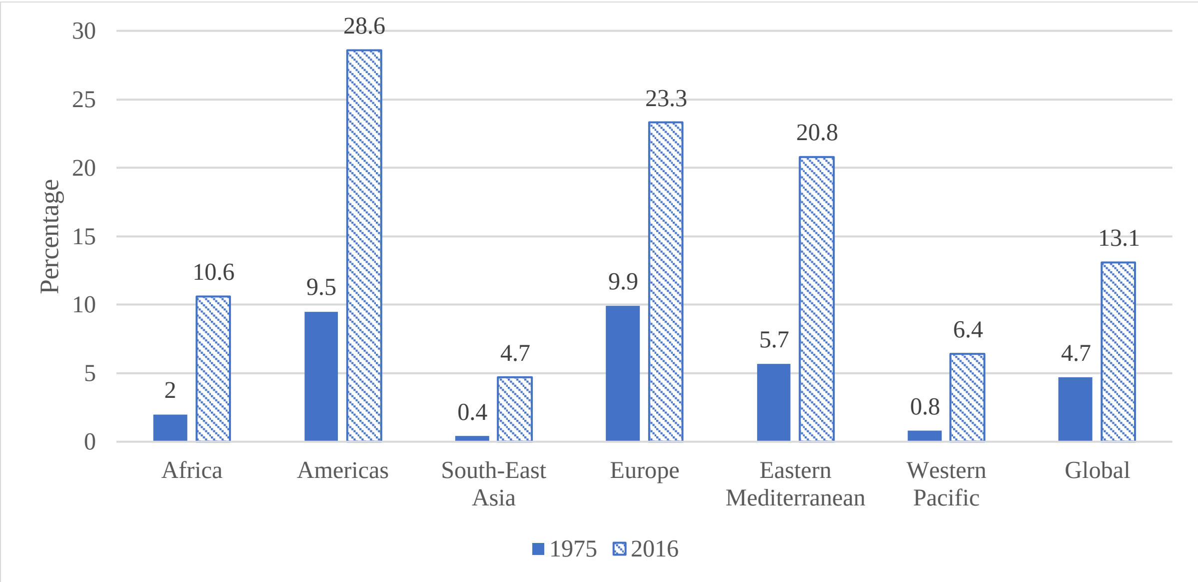


Figure 5.9 Prevalence of obesity in adults by region, 1975 and 2016.

Figure 5.10 reports the obesity prevalence by World Bank income groups. Broadly speaking, the richer the country the higher the obesity prevalence. The obesity rate has always been highest in high-income countries where approximately a quarter of the population is obese in 2016. Once being the "disease of the affluent", obesity is inflicting all income levels. Particular attention should be paid to lower-income countries in which the proportion of obese population more than tripled over the past 40 years, outpacing the speed associated with the wealthiest countries.

Figure 5.10 Prevalence of obesity in adults by country-income level, 1975-2016.

Figure 5.11 plots the prevalence of obesity in adults for males and females. The obesity prevalence has increased substantially for both sexes but with varying speeds. Over the past four decades, the proportion of obese males almost quadrupled from below 3% in 1975 to 11% in 2016 whereas the obesity prevalence among females doubled from 6.4% to 15%. Also, it is worth noting that obesity has always been more prevalent in females than males; nonetheless, the differences have waned over time.

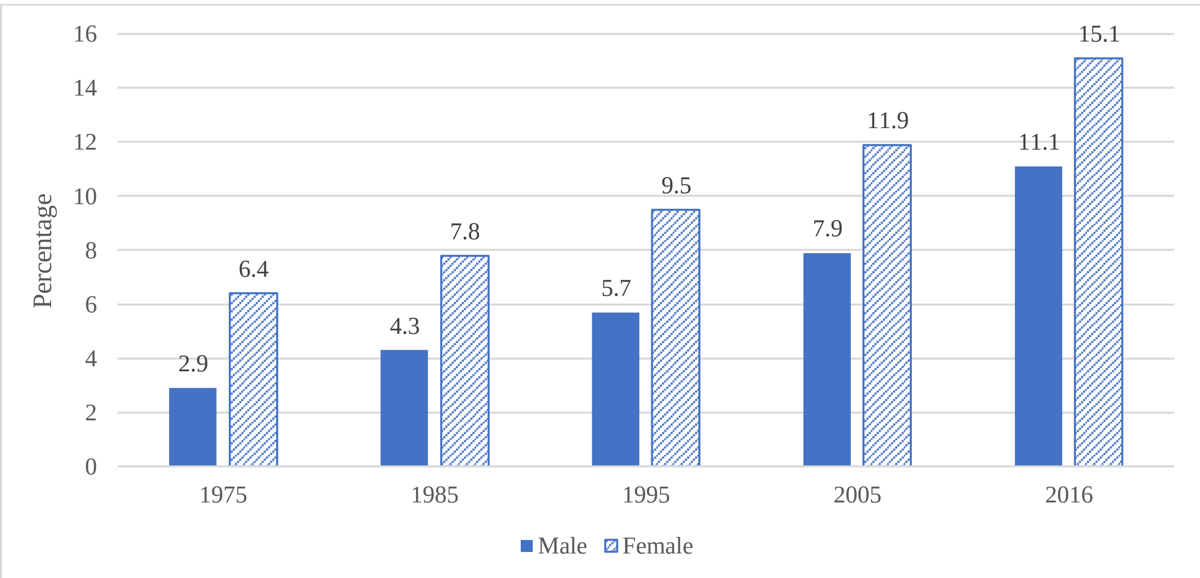


Figure 5.11 Prevalence of obesity in adults by sex, 1975-2016.

5.5 The link between food, health, and economic prosperity

The relationship between food and health is complex since everyone needs food to eat, live, and survive but a poor diet (in terms of either quantity or quality) could result in negative consequences on health (Bleich *et al.*, 2015). On the other hand, Figure 5.6 indicates that the total food supply is found to be larger in developed countries than developing countries, in richer regions (Northern America, Europe, Oceania) than poorer regions (Asia, South America, and Africa). This seems to suggest a positive correlation between food supply and economic status. If such a relationship is valid, it is of crucial importance to monitor and mitigate any adverse effects of the highly calorific diets that will loom on the horizon as countries further develop and "climb up the economic ladder".

In order to answer the question "How does food supply relate to income?", an illustration is given in Figures 5.12. The interaction between food supply, health and income over time is depicted by two bubble charts for 1975 and 2013. Each chart can be considered as a world map for food and wealth. The horizontal axis represents the GDP per capita (in current US$) and the vertical axis represents various levels of food supply (in kcal/capita/day). Each bubble is a country, the colour corresponds to a geographical region, and the size denotes the obesity prevalence.

The overall message from Figure 5.12 is that food supply tends to increase as national income rises. In 1975, on average an individual earned 1,000$ per year and consumed about 2,200 kilocalories a day. In 2013, the average income increased fivefold, and the energy contained in an average diet rose by almost a third to 2,800 kilocalories. Over the past four decades, not only did national economies grow and the bubbles shift uniformly to the right, but they also moved upwards. As a result, a number of Asian and African countries (purple and blue bubbles) eventually caught up and arrived at the same level of calories with their European peers (grey bubbles).

Broadly speaking, individuals consume more calories in richer countries, but not the other way around. For example, in 1975 the calorie availability in Saudi Arabia was significantly lower than in the majority of South American countries despite its remarkably higher GDP per capita. In either bubble chart, the evidence for high-income countries with low-calorie diets is not thin. In fact, a wide spectrum of calories is observed between countries in the same range of income. Therefore, economic development is a better predictor of the *trend* in food availability than the *level* of caloric consumption.

Another striking feature from Figure 5.12 is that over the past 40 years all bubbles increased in sizes – an indication of rising obesity prevalence. Even though the causes of obesity are manifold and often involve an array of genetic, metabolic, social, environmental, and behavioural factors, its dietary origin is obvious. Whilst individuals are increasingly exposed to a larger quantity of calories especially in the forms of energy-dense, ultra-processed foods, the reduced level of physical activity can easily make the excess of calories consumed end up being stored in their body as fat and consequently weight gain is expected.

Figure 5.12 The food – income – health relationship, 1975 and 2013.

It should be noted that the food availability data represented in the vertical axis of the two graphs in Figure 5.12 do not account for food waste which actually varies over time and across countries. According to recent FAO estimates, 14% of food is lost after harvesting and before reaching the retail level (FAO *et al.* 2019). While the global figure for food waste is rising, food waste by households and/or retailers is concentrated in industrialised countries where the figure is more than 40% (FAO 2015). However, little is known about how much food is wasted by consumers at household and retail level mostly due to the lack of data at the national and international level (Hall *et al.* 2009; FAO *et al.*

2019). Despite some existing efforts in quantifying food waste (see, for example, Lopez Barrera and Hertel 2021), the methodology is greatly heterogenous from country to country (Xue *et al.* 2017). The bottom line is the absence of one unique definition of food waste (Bellemare *et al.* 2017). Without taking into account food waste and how it varies temporally and spatially, one cannot make accurate inference about food consumption in Figure 5.12.

Another important limitation of the FBS data is that the food availability figures only present an average picture, ignoring the heterogeneity inherent in diets within a country. Although food availability has increased over the past 50 years to a level that is on average higher than the amount of calories required (Lopez Barrera and Hertel 2021), it does not necessarily mean that everyone is obtaining sufficient calories without considering the distribution of calories within each country. For example, poorer groups in a country may consume a particular food item less than the recommended level whereas richer individuals may consume well above the threshold, leading to a 'cancelling out' effect when the average food supplies are calculated. The use of average data in this case would not reflect the true consumption level of either group. In fact, the distributional concern is of greater importance for developing countries who are struggling with the 'double burden of malnutrition', i.e. the co-existence of undernutrition and overnutrition (Srinivasan *et al.* 2006). Thus, when discussing the effect of income on food, it would be insufficient to not mention the within-country distribution as previous studies have pointed out that obesity rates depend on income distribution (Doorslaer and Koolman 2004; Costa-Font and Gil 2008; Drewnowski 2009; Clément *et al.* 2021) and perhaps so does food consumption. In the literature, the extent to which income is distributed unevenly across a population is defined as income inequality, and a widely used measure for income inequality is the Gini index (Gini 1936). As shown in Section 2.5.1, this index lies between 0 and 1. A low value indicates more equal distribution and a high value indicates more unequal distribution. Value of zero corresponds to perfect equality while value of one corresponds to perfect inequality where income is concentrated in the hands of one person. Existing evidence suggests that high inequalities (GINI index of over 40) are associated with undernourishment rates above 10% in Africa and South America and moreover, countries experiencing rising income inequality are most vulnerable to overnutrition problems (Traill *et al.* 2014).

In order to illustrate the relationship of income inequality and food consumption, a graph similar to Figure 5.12 is reproduced replacing income with income inequality index (GINI) and examining homogenous groups of countries of low-, middle- and high-income levels. It can be seen from Figure 5.13 that high-income countries are characterised by consistently low inequality (GINI index below 40) and high food availability in both 1993 and 2013. Regarding middle-income countries, mixed trends in income inequality are found. For some countries (such as China and Indonesia), the substantial shift towards the right (i.e. rising inequality) is accompanied by the strong shift upwards to a much higher level of food availability, indicating warning signs for obesity problem. Within these countries, the increase in food consumption (and consequently obesity rate) is mostly driven by wealthy individuals while the poorer segment of the population might still be struggling with hunger and undernutrition. For other middle-income countries, the slight decrease in income inequality is coupled with a slight increase in food consumption. The rise in food availability in such countries is thus more evenly distributed among the rich and poor groups of the population. Finally, the pattern for low-income countries is not clear. Overall, Figure 5.13 describes a somewhat converging trend towards inequality index of approximately 40 and food availability of 3,000 kcal/person/day among countries of different income levels.

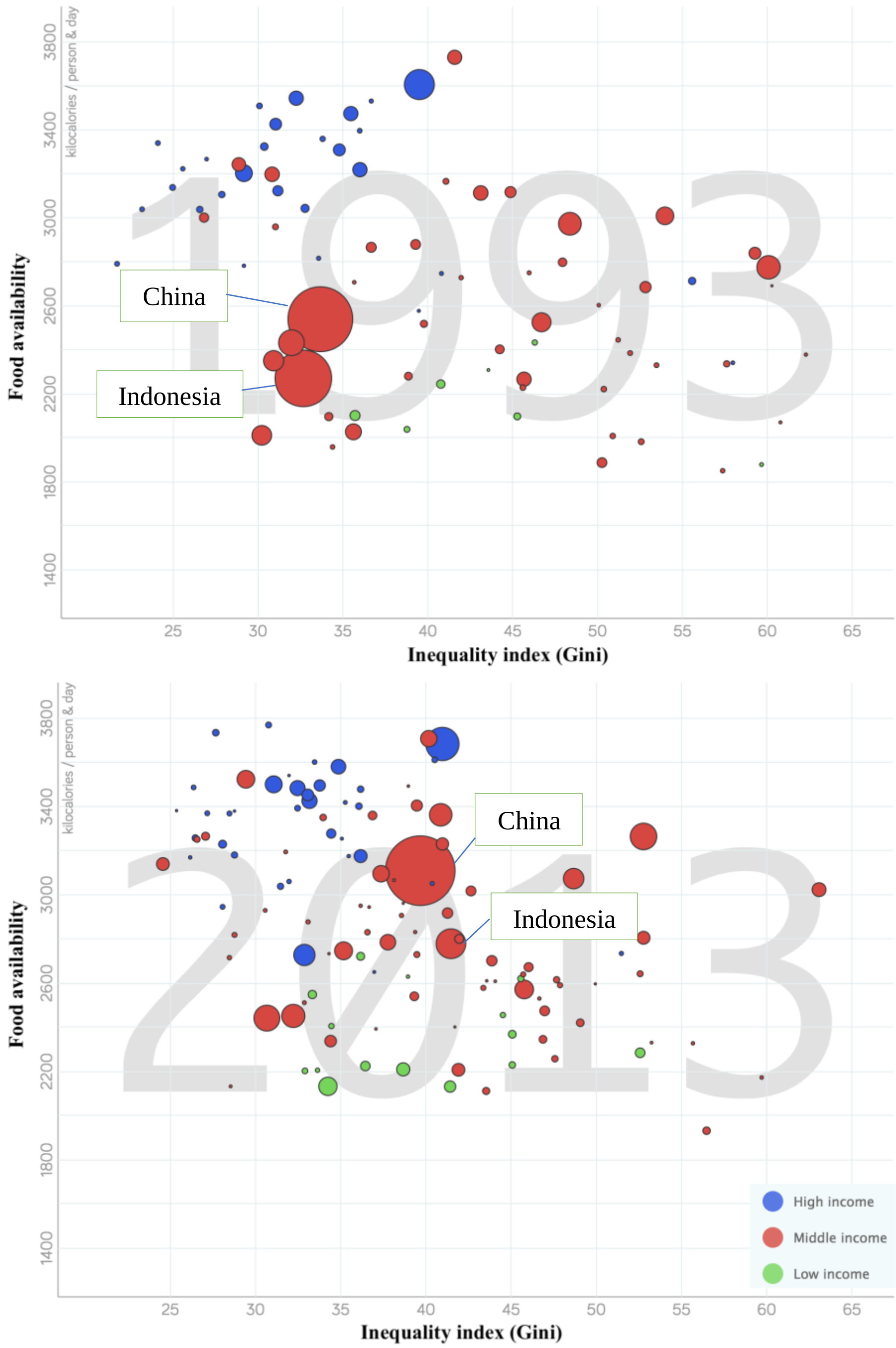


Figure 5.13 The food – income inequality relationship.

5.6 Chapter conclusion

To sum up, the past 50 years have witnessed a steady rise in food availability worldwide, coupled with significant changes in the structure of the global diet in terms of both macronutrients and individual food groups. These changes, such as a robust increase in fat supply, a surge in the energy contribution of vegetable oils and meat whereas a plummet in the contribution of cereals and starchy roots, are once again characterised by the *nutrition transition* that is happening worldwide.

In addition, there is evidence for the converging consumption among the world regions when looking at total caloric supply as well as the calorie figure of starchy roots, vegetables, meat, and vegetable oils. However, the convergence concept can be nuanced as sometimes both converging and diverging patterns coexist. Even though the examination of global diets through the lens of geographical regions reveals important evolutions in the quantities of calories as well as the dietary composition, both diets and dietary shifts are greatly heterogeneous even among countries in the same region. Moving beyond aggregate diets by geographical regions, the next chapter will employ a country-based analysis. Using a global data set varying across countries and time, such an analysis will help to better decodify the world's diets by identifying the common dietary trends and dietary patterns.

Chapter 6

Revisiting the convergence in global patterns of food consumption

6.1 Chapter introduction

In the previous chapter, different data visualisation techniques are applied to the Food Balance Sheet data, revealing critical changes in both macronutrients and food composition of national diets. Two key messages include: (i) rapid rising food availability over the past 50 years; (ii) a robust increase in fat supply and in energy contribution of vegetable oils and meat, as well as a move from coarse grains to more polished grains. But is it a universal shift and are transitional speeds uniform across countries?

In the extant literature, the abovementioned dietary changes and their nutrition impacts are referred to as the *nutrition transition* – one of the five stages describing predictable shifts in dietary and physical activity patterns (Popkin 1993; Kearney 2010; Popkin *et al.* 2012). While nutritional shifts are occurring at varying speeds in different regions, the most rapid changes are observed in the developing world (Hawkes *et al.* 2017). It is shown that the transition that occurred in the West over a couple of centuries took place in developing countries within just a few decades (Popkin 2002b; Popkin and Gordon-Larsen 2016). This seems to suggest converging diets at the global level. Various drivers from both supply- and demand-side of the food system have been proposed for the dietary changes. Rising incomes and urbanisation around the world are blamed for promoting diets rich in animal-source foods, fat, salt and sugars (Martin 2019). Globalisation and its influences, including mass marketing and the rise of supermarkets, are oft-quoted contributors to a tendency towards the consumption of energy-dense and processed foods (Drewnowski and Popkin 1997; Kennedy *et al.* 2004; Unnevehr 2004; Oberlander *et al.* 2017; Martin 2018).

Despite mounting evidence for the nutrition transition that is occurring worldwide, the literature on dietary convergence remains limited and mainly focuses on developed countries and the European Union. Prior studies have quantified the convergence in food consumption patterns across countries by comparing a descriptive statistic (often the coefficient of variation) over time (Blandford 1984; Angulo *et al.* 2001; Regmi and Unnevehr 2006; Regmi *et al.* 2008b; Nowak and Kochkova 2011) or via cross-sectional/time series regression (Herrmann and Röder 1995; Elsner and Hartmann 1998; Ravallion 2012; Ogundari and Ito 2015). The majority document convergence in the composition of diets across Western countries and that convergence speed has slowed down. Some authors recommend taking into consideration other explanatory variables including those mentioned in demand theory (such as price, income, consumer preference) as well as those not mentioned (such as socio-demographics) (Herrmann and Röder 1995; Erbe Healy 2014; Srivastava *et al.* 2016; Gouel and Guimbard 2019). So far, very few studies have examined patterns of food consumption at the global level (Khoury *et al.* 2014; Azzam 2020; Bentham *et al.* 2020; Bell *et al.* 2021); yet, the emphasis has been mostly put on evaluating the similarity across national diets and to date formal convergence tests have been largely missing.

This chapter examines the evolution of global diets using the daily per capita total calories available for consumption for 118 countries over the period 1961-2013 and revisit the problem of global dietary convergence in light of econometric tests. More specifically, this research first assesses *sigma convergence* for the global average calories. Any decrease/increase in the coefficient of variation over time indicates convergence/divergence. The concept of *beta convergence* is then applied to the calorie availability data. Beta convergence refers to the phenomenon in economics whereby poor countries tend to grow faster and 'catch up' with rich countries (Barro and Sala-i-Martin 1992). To test this formally, an unconditional beta convergence model is introduced to describe the 'catching-up' effects in economic development which can be measured by regressing the economic growth rates over a period of time on the initial level of development so that *beta* refers to the regression coefficient. In the beta convergence model of food consumption, a significantly negative beta coefficient indicates that low-calorie countries of the historical past exhibit higher growth rates over time and are approaching the calorie levels of high-calorie countries. Additionally, the structural parameters considered driving forces behind dietary changes (Section 2.3.2, Chapter 2) are added in a conditional beta convergence specification. Unlike the unconditional model which assumes that the growth rate of calories only depends on the initial level of calories and structural conditions such as the population growth rate, the rising speed of income, or the land use for food production are homogeneous across countries, the conditional beta convergence model accommodates heterogeneous nature of countries differing in these parameters. In essence, both diets and dietary changes are greatly heterogeneous.

The results of conditional beta convergence model could highlight that the changes in calorie consumption are not uniform across countries and neither is the convergence process. This brings up the question of to what extent the territorial aspects of countries are relevant in convergence analysis. Does the country's location matter in convergence modelling of the global diets? That the convergence process takes place over time and across space is a legitimate reason for taking into account the spatial component. As it has been proven in economics literature that rich regions tend to be surrounded by rich regions (Annoni *et al.* 2019), perhaps countries with high-calorie diets might likely be located near countries with similarly high-calorie diets, implying similar convergence. While a large body of work has emerged on the nexus between regional income convergence and the location of the region (Chocholatá and Furková 2017; Lolayekar and Mukhopadhyay 2019), very few studies have investigated spatial patterns in explaining health-related outcomes such as obesity prevalence (for example, Hajizadeh *et al.* 2016; Stańczyk 2016). In fact, little has been done in food economics.

Conceptually, the spatial dimension should be added in a model if it plays an important role in explaining the process of interest or it can act as a reasonable proxy for other factors not included in the model. To illustrate the former, rice is predominantly produced and consumed in Asia due to the climate ideal for growing rice. As an example of the latter, the neighbourhood food environment impacts food choice: limited access to supermarkets and living in proximity of fast-food retailers could lead to frequent purchases of less healthy food products. However, it is hard to identify and quantify the neighbourhood characteristics but easier to look at the spatial variation, thus the space becomes an alternative proxy. There are three practical reasons for integrating spatial dimension into beta convergence analysis of food consumption. First, this allows to account for the fact that national diets are *spatially dependent*, meaning that changes in a country's food consumption are determined by factors belonging to not only the country but also the neighbours. In the real world, technological communications, knowledge diffusions, cross-border trades, migration, contagious diseases, and global economic crises are examples showing how the socio-economic environments that facilitate food demand and food supply of a country can extend beyond international boundaries and interact with those of the neighbours – a phenomenon known as 'spatial effects' in the spatial literature. Often spatial interdependence captures the impact of all variables that are omitted in the unconditional specification of beta convergence model. Therefore, including spatial effects into the beta convergence model could eliminate omitted variable bias and enable us to obtain a better estimation of the convergence pattern. To further elaborate on the concept of countries being spatially dependent, the second reason stems from econometric perspective: ignoring the spatial interconnectedness causes a misspecification of the convergence model. The standard OLS approach assumes independence among the error terms whilst this assumption can be violated if spatial interactions are found among countries (LeSage and Pace 2009; Anselin 2013). As a result, any statistical inference based on estimates from the OLS method will not provide reliable results. Third, incorporating the spatial dimension into beta convergence analysis allows to document and quantify spill-over effects. Knowing that public policies can interact (or "spill") across international borders is tremendously useful to policymakers when formulating their policies, and this is recognised as the first important step in ensuring policy coherence (OECD 2021). Despite these abovementioned justifications, the influence of the space on dietary convergence has been under-studied in the earlier literature. This research fills this void. In the second part of the empirical analysis, a spatial beta convergence model is adopted. The novelty of this method is in the quantification of spatial effects, the identification of the source of spatial interactions, and the treatment of such effects in the beta convergence analysis.

On another note, in spatial analysis literature, the spatial interdependence between countries, regions or counties is commonly quantified by a geographical measure such as physical distance between locations or contiguity based on administrative boundaries. Nevertheless, an increasingly large number of studies have extended the notion of spatial closeness beyond geography to capture network of interaction, colonial links, and shared characteristics in for example language spoken, religion, or origins of the legal system (Ahmad and Hall 2017). Building on this development, spatial analysis of food consumption can also benefit from adopting a non-geographical measure of spatial proximity. Historically, civilisations emerged primarily in regions where the natural conditions were more favourable for stable agricultural production (Landes 1998). While weather conditions shape a country's productive capacity determining a person's immediate food sources, eating practices are conditioned by deeply rooted cultural and religious beliefs, perceptions and values which are naturally tied to the geographical area in which one inhabits (Dekker *et al.* 2017). Nonetheless, that paradigm has shifted since trade and other economic manifestations of globalisation have replaced food production and as a result, food availability, accessibility and consumption tend to depend more on a nation's wealth than geographical location itself (Pawlak 2016; Sadowski 2019). Nowadays, "the root cause of most food insecurity is poverty" (OECD 2020). Thus, economic factors have become an important spatial dimension of food security. Against that backdrop, a great novelty of the spatial analysis in this chapter is the examination of different kinds of spatial relationship among countries proxied by both traditional and non-traditional proximity measures. Importantly, a proximity measure in terms of economics rather than geography is proposed. This specification points to income level (rather than geographical closeness) that is driving the similarities in diets observed worldwide.

The remainder of this chapter proceeds as follows. Section 6.2 presents the methods, Section 6.3 introduces the data. Section 6.4 discusses the empirical results and Section 6.5 concludes.

6.2 Methods

6.2.1 Quantifying convergence

Convergence theories

Convergence in the most general sense refers to a process of gradual reduction in differences among observed countries during a certain period of time. For economists, the convergence of income between countries/regions has long been an intriguing topic. The term 'income' is a general one and depending on the research topic and objectives, it can be GDP per capita, wage per worker, consumption level, etc. Different convergence concepts are described in Section 2.5 (Chapter 2). Briefly, there are two main convergence theories in convergence empirics: sigma and beta convergence.

Sigma convergence refers to the reduction in cross-sectional dispersion of income over time (Quah 1993), while *beta convergence* occurs when economies with initially lower levels of income tend to grow faster than and 'catch up' with those with initially higher levels of income (Baumol 1986; Barro and Sala-i-Martin 1992).

The concept of beta convergence is supported by the neo-classical growth model (Solow 1956) which argues that the source of convergence is the diminishing return to capital. In simple terms, it means that national incomes converge with one another in the long-term regardless of the initial conditions – a hypothesis commonly known as *unconditional* (or *absolute*) *beta convergence*. The absolute beta convergence is usually tested through a cross-sectional equation regressing the average growth rates on the initial income levels. Absolute convergence is detected by a negative association between average growth rates and initial income levels even if no other explanatory variables are included in the regression model (Barro and Sala-i-Martin 1992). The assumption is that countries eventually converge to the same global steady state equilibrium.

If national incomes converge with one another in the long-term only providing that their structural conditions (such as technologies, human capital, population growth rates, legal institutions) are identical, that implies *conditional beta convergence* (Mankiw *et al.* 1992). The equilibrium differs by economy, and each country approaches its own unique equilibrium. In the case of conditional convergence, the negative relationship between initial incomes and the average growth rates holds only after controlling for the structural characteristics. Therefore, the cross-sectional regression equation testing conditional beta convergence includes other controls as explanatory variables.

However, some researchers argue that countries sharing similar structural characteristics and initial factors (for example, GDP per capita, human capita, preferences, public infrastructure) converge with one another in the long-term but need not converge on the same equilibrium path (Galor 1996). This gave rise to the *club convergence* hypothesis according to which countries belonging to the same 'club' move toward a club-specific steady-state equilibrium, and there is no convergence across different sets of equilibria. The empirical testing of club convergence usually involves the regression-based technique developed by Phillips and Sul (2007) to endogenously classify countries with similar characteristics into unique groups (or clubs).

Approaches to testing convergence in patterns of national food consumption

The concept of convergence has been applied in food economics for the past few decades. A summary of the related literature is provided in Section 2.5.2 (Chapter 2). In this chapter, sigma and beta convergence theories are examined for food availability data to determine whether or not national diets are becoming more alike.

An initial step is to investigate how the mean deviation of the global average figure for total calories changes over time. The dispersion in per capita daily calories, measured by the coefficient of variation (CV), is calculated as:

$$CV_t = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_{it}-\bar{y}_t)^2}}{\bar{y}_t} \tag{6.1}$$

where y_{it} is the per capita daily calories observed in the *t*-th year for the *i*-th country, $\bar{y}_t$ is the average per capita daily calories in the *t*-th year for all countries. Any reduction in the CV indicates a decline in the dispersion among countries, thus suggesting sigma convergence.

In order to quantify this converging pattern, absolute beta convergence will be then examined. This specification involves regressing the cross-sectional growth rates on the initial levels. In this analysis, a simplified version of the growth equation shown in Barro and Sala-i-Martin (1992) is utilised. Specifically, the following linear regression is estimated:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \beta\log\left(y_{i,t_0}\right) + \varepsilon_i \tag{6.2}$$

where y_{i,t_0} and y_{i,t_0+T} are respectively the per capita daily calories of country *i* at the initial and final periods; T is the number of years; ε_i is the standard error being independently and identically distributed with mean 0 and variance σ_ε^2. The left-hand side of equation (6.2) represents the average annual growth rate of per capita daily calories of the *i*-th country.

If β is significantly negative, there is an inverse relationship between the initial calorie level and the calorie growth rate that is indicative of convergence; a significantly positive β implies divergence. The estimated β also indicates the rate at which regions approach their steady state, i.e. the *speed of convergence*. From the estimation of β, the annual speed of convergence can be computed as:

$$\theta = -\frac{\ln(1+T\beta)}{T} \tag{6.3}$$

and the *half-life* (the number of years required for progress halfway towards the steady-state level when convergence is assumed to have been achieved) is given by:

$$\tau = -\frac{ln2}{\ln(1+\beta)} \tag{6.4}$$

However, the intercept in equation (6.2) may be influenced by structural factors that vary among groups of countries, directing them on a path to different steady-state levels. For example, Barro and Sala-i-Martin (1992) cite the between-country differences in technology or consumer preferences. Since the rate of convergence might be influenced by such structural differences, the intercept could vary among countries at different levels of development. Therefore, the beta convergence model represented in equation (6.2) will be further investigated among four country groupings: low-income, lower-middle-income, upper-middle-income, and high-income. Furthermore, a conditional beta convergence specification can be introduced by incorporating the structural factors affecting dietary changes into the right-hand side of equation (6.2) as additional explanatory variables.

6.2.2 Incorporating spatial effects into regression models

In the second part of this chapter, the global dietary convergence is examined by the means of spatial data analysis. *Spatial dependence*, or *spatial autocorrelation*, refers to the existence of a functional relationship between a phenomenon happening in a location and what happens in other locations (Anselin 1988). Spatial autocorrelation is positive when similar values for a variable are clustered together and negative when dissimilar values are clustered in space.

An overview of spatial dependence

The First Law of Geography, according to Tobler (1970, p.236), states that "Everything is related to everything else, but near things are more related than distant things". This has been the fundamental foundation for the concept of spatial dependence in spatial analysis which deals with interaction effects among geographical units (for example cities, municipalities, regions or countries). As spatial units (say countries) have closer and further neighbours, it is reasonable to assume that the proximity in space (say geographical closeness) can have an influence on different characteristics (say income).

Since the seminal papers by Cliff and Ord (1973, 1981), much of the spatial analysis literature has elaborated to handle spatial interactions. In the past, models that explicitly incorporated space or geography were developed in regional studies (Anselin 1992; Anselin and Rey 1997). However, the notion of space is not restricted to geographical meaning and "there is more to space than geography" (Beck *et al.* 2006). This idea is conveyed in Tobler's Second Law (Tobler 2004) which argues that "proximity and near can take on many meanings in different situations". Recently, many spatial studies handle cross-unit interactions in situations such as social network (Kelejian and Prucha 2010; Drukker *et al.* 2013), the economic diffusion of local labour markets, or electoral systems in political science (Manski 2000). In-depth reviews on spatial techniques and their usages are referred to Anselin (1988), Griffith (1988), Haining (1990), Cressie (1993), Haining (2003), Anselin (2006), Arbia (2006), LeSage and Pace (2009) and Anselin (2010). Though the adjective *spatial* is consistently used throughout this thesis to avoid confusion, the wider applicability of spatial models still remains.

In general, there are sound rationales for the need of spatial models. First, spatial, organisational or social interactions between economic agents are common phenomena in economics. The decision of an economic agent might depend on the decision of other agents, and the decision made by an economic agent could be reliant on a scarce resource (Anselin 2002). An example of the former is given in the context of competition where a firm wants to maximise its profits by considering the production levels of its competitors and its own characteristics (such as availability of raw materials, number of workers, labour productivity, etc). In the latter case, the quantity of the scarce resource consumed by a production firm depends on the quantity consumed by competing firms. These two examples conveniently highlight that the notion of the 'space' is not strictly geographical.

In addition to the above economic reasons, spatial interaction models are attractive from an econometric point of view. Often in economics one needs to explicitly account for the interaction of an economic agent with heterogeneous agents due to social norms, neighbourhood effects, copycatting. Such models seek to determine how the magnitude of a variable at a given location is influenced by values of the same variable at other locations ("spatial autocorrelation"). Besides, the growing popularity of spatial data thanks to the incorporation of geographic information system (GIS) as well as the availability of geocoded data reinforces the urge to account for spatial autocorrelation of residuals within spatial data. This type of dependence, if it exists, will violate a key assumption of the OLS estimation (independence of residual terms), thus OLS estimators will be less accurate or even biased, and inconsistent (LeSage and Pace 2009; Anselin 2013). Because of spatial interactions, in the linear model $y_i = \alpha + \beta_0 x_i + \varepsilon_i$ it does not hold that $\varepsilon_i \sim i.i.d$ (independently and identically distributed).

In the simplest terms, autocorrelation means the correlation of a variable with itself. For time series where observations are aligned in a linear order and the frequency of the series is set, the *temporal dependence* (or *temporal autocorrelation*) refers to the phenomenon that values of a variable depend on past values of the same variable, and the subscripts $"t-1", "t-2", \ldots, "t-l"$ are used to denote observations of the same variable at different time lags. In spatial analysis, the primary focus lies in the correlation between values of a variable at different locations, and the dependence structure is usually arranged in the so-called *spatial weight matrix* W, which will be discussed in the next section. Similar to the case of temporal autocorrelation, the OLS approach is inadequate and the estimations can be inconsistent and biased (Anselin and Bera 1998; LeSage and Pace 2009). Spatial dependence however is not a straightforward extension of temporal dependence. To illustrate, two geographical units can influence each other mutually, representing multidirectional nature (region A ↔ its neighbours ↔ other regions ↔ region A) whereas two observations in time exhibit unidirectional interaction (past → present). In addition, the existence of numerous measurements that could quantify spatial relationship (distance, neighbours, connections to name a few) presents another challenge in modelling spatial dependence as compared to capturing the only dimension (time) in temporal dependence (Getis 2007). These complexities lead to an expansion of the recent literature to developing theoretical and methodological frameworks on this issue (Griffith 2005; LeSage and Pace 2009; Anselin 2010). Even though the exploration of spatial dependence dates back over 40 years (Cliff and Ord 1969), many basic questions regarding the construction method and the testing procedure remain unsolved (Chen 2013).

Part of the disagreement involves the specification of spatial weight matrix W, which describes the spatial relationship among spatial units. Any modification of this matrix presents a new spatial dependence structure and will potentially change the results of a spatial model.

Spatial weight matrix: construction and specification

If there are N cross-sectional units, the non-negative $(N \times N)$ square matrix $W = (w_{ij}: i, j = 1, \ldots, N)$ is called the *spatial weight matrix* summarising the spatial relationship. Each spatial weight w_{ij} represents the spatial influence of the j-th unit on the i-th unit. It is assumed that $w_{ii} = 0$ for all $i = 1, \ldots, N$ (a unit does not influence itself directly); and hence the spatial weight matrix has zero diagonal. Two units i and j are neighbours if $w_{ij} > 0$, and not neighbours when $w_{ij} = 0$. The neighbourhood of the i-th unit is the set of observations with which it has a certain spatial linkage (i.e the spatial weight greater than 0). Despite being relatively simple and intuitive, the construction of the spatial weight matrix W is highly debatable and does not involve any standard protocol. The foremost challenge is how to define the space metric adequately for the problem at hand. Space here could be geography or something else (proximity, remoteness or a function of economic, social, cultural factors). Once an appropriate definition of space is chosen, a measure for the strength of spatial relationship is required.

Various ways to construct W have been proposed in previous studies depending on the definition of spatial relationship, and two most popular approaches include: (i) boundary-based and (ii) distance-based. The boundaries (due to either geographical or administrative nature) shared between spatial units play an important role in determining the spatial influence. Often, it involves the concept of *contiguity* which means that two spatial units share a common border of non-zero length. For instance, France is contiguous to Spain, France is contiguous to Germany, and Spain is not contiguous to Germany. Boundary-based matrices are easy to comprehend, yet it can be problematic in the case of isolated regions (for example, Sicily island in Italy, or Malta). One solution is to assign the nearest region as the only neighbour or to modify the method accordingly. Also, the concept of boundary is adequate for polygon regions but not for points.

On the contrary, distance-based matrices translate the spatial weights as a function of how far two spatial units are from each other. A distance measure d_{ij} between the i-th and j-th units indicates the intensity of the spatial linkage between two units. W is then established to reflect the distance decay, or in other words further units exert weaker influence than closer ones. It is easy to think of d_{ij} as a distance between points but it can represent a distance between polygon centroids or central points. Unless the distance d_{ij} is given explicitly, it can be computed from a wide range of distance metrics that are introduced in Section 3.2.2 (Chapter 3). This method works well even in the presence of isolated regions, when the spatial units are no longer characterised by polygons but points, and when the logic of border-sharing is invalid. Another advantage is that the spatial distance can be measured in a non-geographical way (Corrado and Fingleton 2012).

Based on these two approaches, the main types of spatial weight matrices are as follows.

Contiguity weight matrix

The entries of a spatial contiguity weight matrix indicate whether the spatial units share a boundary or not. The spatial weights are given by:

$$w_{ij} = \begin{cases} 1 \; if \; units \; i \; and \; j \; are \; contiguous \\ 0 \; otherwise \end{cases} \tag{6.5}$$

Despite being simple in its representation, the definition of contiguity is not obviously determined in particular for spatial units in polygon/grid shapes. Different contiguity types have been considered, among which the three most common ones are Rook, Bishop and Queen, in analogy to the moves for such pieces on a chess board. Two areas are neighbours if they share at least two common boundary points according to Rook’s criterion and at least one common boundary point following Queen’s contiguity. In the sense of Bishop’s definition, two polygons are adjacent if they share a common node. Figure 6.1 demonstrates how Rook, Bishop and Queen’s contiguity are defined. The neighbours of the unit i take the colour shades while the non-neighbours are not coloured.

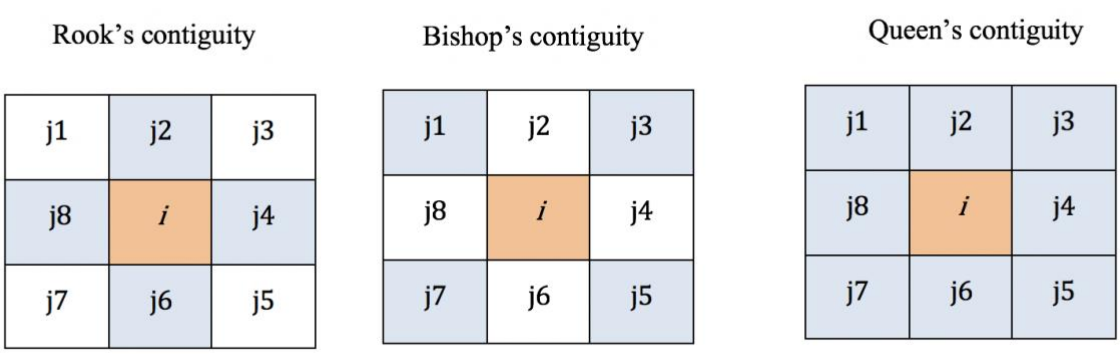


Figure 6.1 Different contiguity criteria.

K-nearest neighbour distance weight matrix

Following this approach, the distances between a given observation and the rest of the set are computed, ranked, and on the basis of which the k-closest observations are considered as neighbours of the given observation. Let the distances from each spatial unit i to all units $j \neq i$ be ranked as $d_{ij(1)} \leq d_{ij(2)} \leq \cdots \leq d_{ij(N-1)}$. Then for each $k = 1, \dots, N-1$, the set $N_k(i) = \{j(1), j(2), \dots, j(k)\}$ contains the k-nearest units to i. For each given (non-negative) k, the k-nearest neighbour weight matrix W, has spatial weights of the following form:

$$w_{ij} = \begin{cases} 1 \text{ } if \text{ } j \in N_k(i) \\ 0 \text{ } otherwise \end{cases} \tag{6.6}$$

Radial distance weight matrix

An alternative way to define neighbourhood is to draw a circle of a pre-defined radius with the *i*-th unit being the centre and consider all observations inside the circle neighbours of i. Let d_θ be the pre-defined threshold distance between two units beyond which there is no direct spatial effect between spatial units. There is an assumption of non-diminishing effects in distance up to the threshold d_θ. Figure 6.2 illustrates the k-nearest neighbour distance and radial distance. The spatial weights of the radial distance weight matrix are given by:

$$w_{ij} = \begin{cases} 1 \text{ } if \text{ } 0 \le d_{ij} \le d_\theta \\ 0 \text{ } if \text{ } d_{ij} > d_\theta \end{cases} \tag{6.7}$$

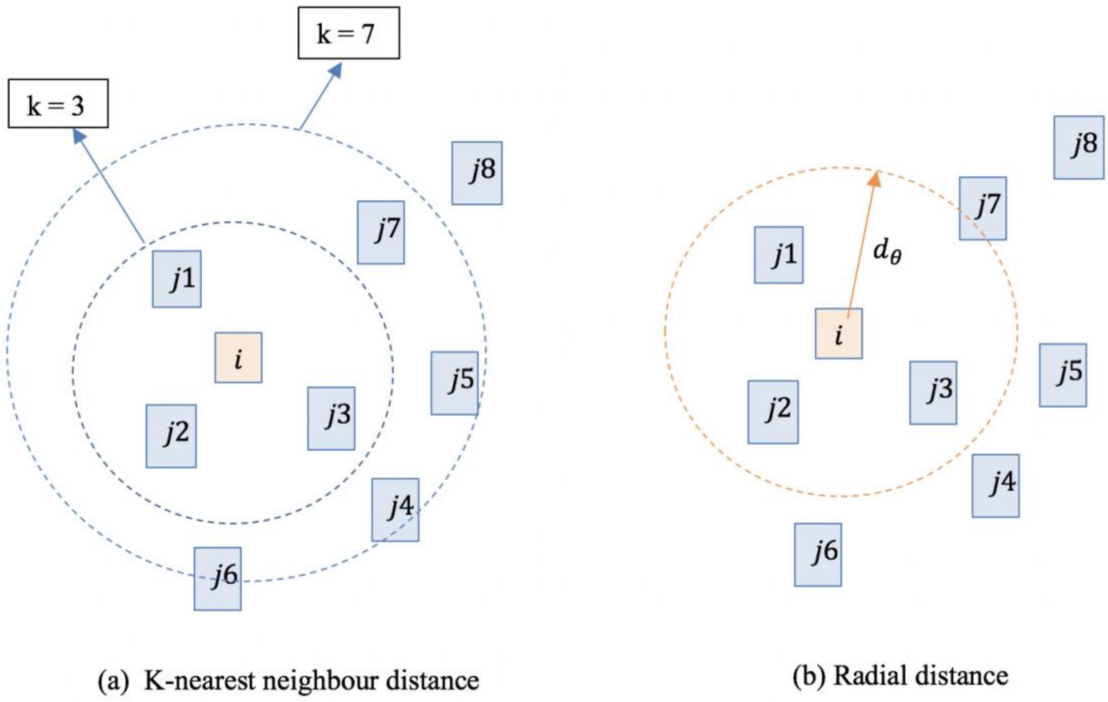


Figure 6.2 K-nearest neighbour distance versus radial distance.

Negative power distance weight matrix

Unlike the previous matrices which are binary, negative power distance weight matrix has continuous weights. The spatial weights take the following form:

$$w_{ij} = d_{ij}^{-\alpha} \tag{6.8}$$

where α is a positive integer (usually 1 or 2). If $\alpha = 1$, W is referred to as inverse distance matrix.

Exponential distance weight matrix

This is an alternative to the negative power distance. The spatial weights are determined by:

$$w_{ij} = e^{-\alpha * d_{ij}} \tag{6.9}$$

where α is a positive integer.

Once the spatial weight matrix W is developed, it is a common practice to normalise the spatial weights to remove any dependence on the scales (for example the measurement unit of distance in exponential or negative power distance weight matrix). Assuming that the normalised matrix $\widetilde{W} = (\widetilde{w}_{ij})$ is derived from the original spatial weight matrix $W = (w_{ij})$. Various ways to obtain $\widetilde{W}$ have been proposed, and the most popular techniques include *row normalisation* and *scalar normalisation*.

Regarding row normalisation, the *(i, j)*-th element of $\widetilde{W}$ becomes $\widetilde{w}_{ij} = {}^{w_{ij}}/_{r_i}$, where r_i is the sum of the i-th row of W. After row normalisation, each row of $\widetilde{W}$ will sum up to one: $\sum_{j=1}^{N} \widetilde{w}_{ij} = 1,\ i = 1, \dots, N$. The symmetric W matrix will produce an asymmetric $\widetilde{W}$ matrix. Since W is non-negative, all weights $\widetilde{w}_{ij}$ are between 0 and 1. In practice, W can also be column normalised so that elements of each column sum up to one. The column elements of a spatial weight matrix show the impact of a particular unit on all other units, while the row elements of a spatial weight matrix display the impact on a particular unit by all other units. Due to row normalisation, the impact on each unit by all other units is equalised whereas column normalisation makes the impact of each unit on all other units equalised (Elhorst 2014). In spite of being a common approach, row normalisation is not without criticism. Kelejian and Prucha (2010) warn that row normalisation might lead to misspecification problem. This is particularly true when the inverse distance matrix is row normalised since the economic interpretation reflected by distance decay will no longer be valid (Anselin 1988; Elhorst 2001). Row normalisation also alters the internal structure of W and hence the comparison between rows becomes troublesome (Elhorst 2014).

An alternative to row normalisation is scalar normalisation in which $\widetilde{W}$ is derived by multiplying W with a scalar ψ . The main benefit of scalar normalisation is that the symmetry and the internal structure of spatial weight matrix are preserved. Often:

$$\psi = {}^{1}/_{w_{max}} \tag{6.10}$$

where w_{max} represents the largest element of W (Elhorst 2001; Kelejian and Prucha 2010). This ensures that the resulting spatial weights $\widetilde{w}_{ij}$ are between 0 and 1.

Another option is to derive the min-max normalised matrix $\widetilde{W}$ by selecting:

$$\psi = {}^{1}/_{m} \tag{6.11}$$

where $m = \min\{max_i(r_i), max_i(c_i)\}$, with $max_i(r_i)$ and $max_i(c_i)$ being the largest row sum and column sum of W (Kelejian and Prucha 2010).

Alternatively, to obtain a spectral-normalised matrix the scalar is set as:

$$\psi = {}^{1}/_{\gamma} \tag{6.12}$$

where γ is the largest eigenvalue of W (Ord 1975).

The existence of numerous methods to develop the spatial weight matrix poses a question of selecting the optimum W. Previous researchers advise that any choice should be backed up by a solid theoretical standpoint (Kelejian and Prucha 2010; Corrado and Fingleton 2012; Neumayer and Plümper 2016). In reality, the construction procedure is subjective (Chen 2012) and economic theory underpinning spatial econometric applications often has little to guide the specification of W (Anselin 2002; Leenders 2002). Therefore, examining the robustness of the results derived from a spatial model subject to the specification of the spatial weight matrix is usually recommended (Elhorst 2014). Despite these criticisms, the ultimate concern should be: 'Does W adequately represent the spatial linkages between units regarding the variable of interest?'.

Having specified an appropriate spatial weight matrix, the next step before identifying a spatial model is to ensure that a spatial phenomenon does exist. Both graphical visualisation and statistical tests can characterise spatial dependence if it exists in the data. The next section introduces the most popular statistical procedures to measure spatial autocorrelation.

Detecting spatial dependence

To review, spatial autocorrelation measures the correlation of a variable of interest (say calorie intake) with itself across space (say European countries). Positive spatial autocorrelation suggests that observations closer together have more similar values, for example the number of calories in an average German diet is more comparable to an Austrian diet than a Greek. Conversely, negative spatial autocorrelation suggests that observations close together have dissimilar values than observations further away. This occurs when for instance the calorie content of a German diet is more different from an Austrian diet than a Greek. The task of a spatial autocorrelation test is to determine if a variable is more positively or negatively spatially autocorrelated than one would expect in a random distribution.

It should be made clear that the test of spatial autocorrelation can be run for: (i) cross-unit observations of a variable of interest, (ii) residuals from a linear regression model estimated by OLS technique (to evaluate if a spatial interaction model is needed), (iii) residuals from a spatial regression model. Thus, the identification of spatial autocorrelation can be done before or after estimating the regression model. In any case, the spatial dependence structure, represented by W, should be provided externally and results of the spatial autocorrelation tests are conditioned to that information. Broadly speaking, spatial autocorrelation testing procedures can be categorised into two classes: (i) general tests without a specified alternative hypothesis (the most popular ones include Global and Local Moran's I statistics), and (ii) specific tests with a specified alternative hypothesis (for instance LM test and its variants).

Among the proposed measurements, *Global Moran's I statistics* (Moran 1950a, 1950b) – a generalisation of Pearson's correlation coefficient is unarguably one of the most widely used indices to calculate spatial autocorrelation. The intuition behind this index is pretty straightforward. Returning to the food consumption example above, assume that a researcher wants to measure how similar or different the calorie consumption in Germany is from other countries in Europe. One way to do this is to compare how much the intake in Germany differs from the average (of all European countries) versus how much the intakes from the remaining countries (its neighbours) differ from the average. If German intake is much higher than the regional average and its neighbours have much higher consumption levels than the average, there exists positive spatial autocorrelation. Likewise, if the intakes of Germany and its neighbours are much lower than the average, there also exists positive spatial autocorrelation. Nonetheless, if the calorie consumption is much higher in Germany but much lower in its neighbours than the European average (or vice versa), the spatial autocorrelation would be negative.

The formula to compute the Global Moran's I index is given by:

$$I = \frac{N}{S_0} \times \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \times z_i \times z_j}{\sum_{i=1}^{N} {z_i}^2} \tag{6.13}$$

where z_i is the deviation from the mean $(x_i - \bar{x})$ of the value of variable at the location *i*; w_{ij} is the spatial weight determining the relationship between locations *i* and *j*; *N* is the total number of observations; S_0 is the sum of all the spatial weights: $S_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$.

The Global Moran's I index ranges between -1 and 1. Positive (or negative) value indicates positive (or negative) spatial autocorrelation, i.e. observations move in same (or opposite) direction as their neighbours. A value of zero indicates that observations follow a random pattern in space. Regarding its advantages, Global Moran's I statistics is shown to be locally best invariant (King 1981) and is proved by Anselin and Florax (1995) to consistently outperform the proposed alternatives such as the tests of Burridge (1980) and Anselin (1994), or the robust tests of Kelejian and Robinson (1992).

The formula in equation (6.13) shares some similarities with the Durbin-Watson test for temporal autocorrelation in time series analysis (Anselin and Bera 1998), and can be applied to a variable of interest or residuals from a linear regression model. Similar to other analyses, it is important to know if the computed Moran's I value is significantly different than the "ideal" value representing a random distribution. This would be possible if the variance of *I* is known. This variance can be computed in three different ways: Monte Carlo simulation, normality of x_i, and randomisation of x_i. The second approach assumes that the random variable x_i follows a normal distribution whereas this normality assumption is not necessary following the third approach, but it instead considers the observed values of x_i to be repeatedly randomly permuted. A spatial autocorrelation test based on the Global Moran's I statistic can be formulated with the following hypotheses:

H₀: there is no spatial autocorrelation ($I = 0$);

Hₐ: there are spatial effects ($I \neq 0$).

The alternative hypothesis could be one-sided or two-sided, however, the specific process that generates spatial effects (i.e. the source of spatial effects) is not stated.

In the above discussion, the Global Moran's I statistic is explored to assess the similarity or spatial dependence across locations with respect to a variable of interest (say calorie consumption). In other words, are countries with similar calorie intakes located close together or are consumption levels randomly distributed across Europe? To recap, the significance of the Global Moran's I gives evidence for the existence of spatial autocorrelation (i.e. observations are not spatially distributed in a random manner), the sign of the index (negative/positive) suggests whether observations near each other tend to be like/unlike each other, and its value indicates the strength of the association. It is worth noting that the Global Moran's I index is a global statistic that describes the spatial distribution of the entire dataset with a single value. In doing so, the Global Moran's I statistic compares how similar every observation is to its neighbours, and then averages out all these comparisons to give an overall idea about the spatial pattern of the variable. While the global statistic is useful in summarising the whole map, it tends to average local variations in the strength of spatial autocorrelation and thus does not allow to have further insights into interesting geographical subsets of the data. Particularly, in addition to global trend in the entire sample some localities exhibit values that are extreme, geographically homogeneous, and not in line with the global trend. In this case, it is interesting to identify *hot spots* or *cold spots* regions where the value of the variable under consideration is extremely pronounced across localities. This however can be deducted by the *Local Moran's I statistic*, also known as *Local Indicators of Spatial Association* (LISA) (Anselin 1995). Unlike the Global Moran's I statistic, each location in space, *i*, has its own unique spatial autocorrelation value as well as its own variance.

The Local Moran's I statistic can be calculated by:

$$LI_i = \frac{x_i - \bar{x}}{S_i^2} \times \sum_{j=1, j\neq i}^{N} w_{ij}(x_j - \bar{x}) \tag{6.14}$$

where x_i is the value at location *i*; $\bar{x}$ is the mean value of all observations; w_{ij} is the spatial weight between locations *i* and *j*; N is the total number of observations, and S_i is the variance defined as: $S_i^2 = \frac{\sum_{j=1, j\neq i}^{N}(x_j - \bar{x})^2}{N-1}$.

Similar to the Global Moran's I index, a spatial autocorrelation test can be conducted for the Local Moran's I index. To derive *p*-values of the Local Moran's I index, the Bonferroni adjustment method is utilised and the number of neighbours for each unit is required. In short, the Bonferroni correction divides the level of significance α by the average number of neighbours in each test (Anseline

1995). If the adjusted p-value is lower than 0.05, the Local Moran's I statistic is said to be significant at 5% significance level.

A significantly positive value of LI_i indicates that location i has neighbours with similarly high or low values, and the location is called a "spatial cluster". A negative value of LI_i indicates that the value at location i is different from its surrounding locations, and location i is a "spatial outlier". Classifying each observation in the data set depending on the value of itself and its neighbours, the Local Moran's I can detect concentrations of high values ("hot spots"), concentrations of low values ("cold spots") as well as spatial outliers. Are there countries in the sample with anomalous calorie intake? Where can we find the unexpectedly high calorie consumption across the sample? The local statistic helps to answer these questions.

In addition, the values of the Local Moran's I can be plotted in a scatterplot to show the relationship between each location and the average of its neighbours. An example is given in Figure 6.3. The horizontal axis represents value at location i, whereas the vertical axis represents values in the neighbourhood of i. Since the graph is centred on the mean (of zero), considering the x-axis all points to the right of the vertical line $x = 0$ represent values higher than average whereas all points to the left represent values lower than average. These values are respectively referred to as "high" and "low". Likewise, considering the y-axis points above and below the horizontal line $y = 0$ can be regarded as "high" and "low" respectively.

An important feature of the Moran scatterplot is the connection between Global and Local Moran's I statistics by classifying the nature of the spatial autocorrelation into four categories. Points in the top right corner of the scatterplot denote locations in which values at location i and its neighbours are "high" (above the mean), signalling positive spatial autocorrelation. Points in the lower left corner of the plot indicate locations in which values at location i and its neighbours are "low" (below the mean), also suggesting positive spatial autocorrelation. Points in the top left corner of the plot represent locations in which the value at location i is "low" while values at its neighbours are "high". In contrast, points in the lower right corner represent locations where the value at location i is "high" but values at its neighbours are "low". Both situations describe negative spatial autocorrelation. Therefore, based on the values of the Local Moran's I statistic, a location can be categorised into four groups: High-High (cluster of high values, or hot spots), Low-Low (cluster of low values, or cold spots), High-Low (outlier with a high value and is surrounded primarily by neighbours with low values), or Low-High (outlier with a low value and is surrounded primarily by neighbours with high values).

Finally, the slope of the linear fitted line in the Moran scatterplot equals value of the Global Moran's I statistic. In Figure 6.3, the best fit line is coloured in red, indicating the presence of positive spatial autocorrelation in the entire data set.

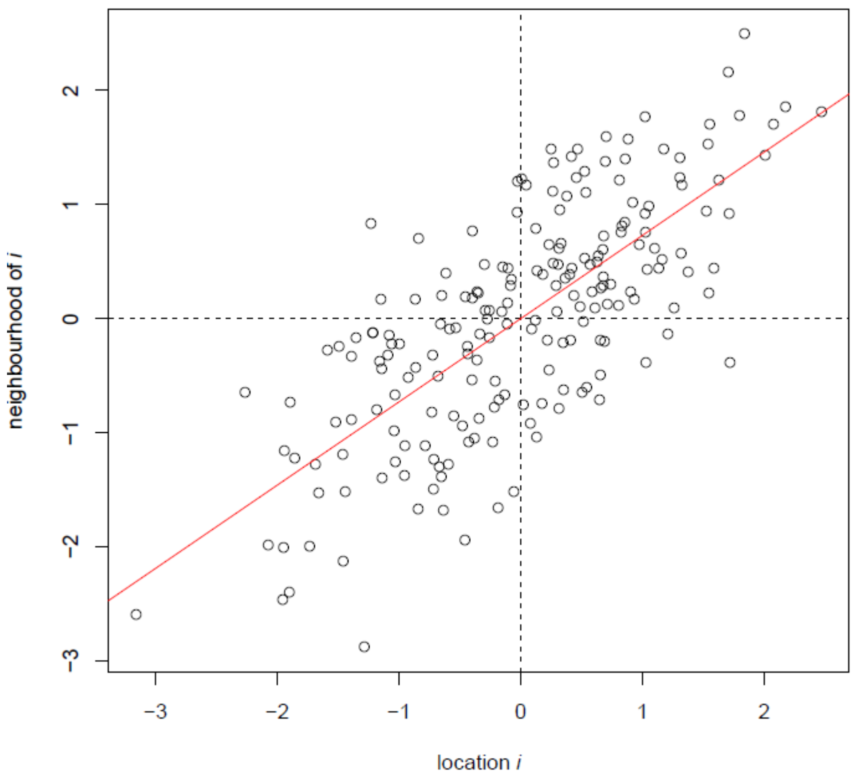


Figure 6.3 Moran scatterplot.

It is worth noting that the Global and Local Moran's I statistics are general tests for spatial autocorrelation, meaning that they provide a measure of how similar locations are to their neighbours without addressing the source of spatial dependence. Unlike these statistics, classic LM (Larange Multiplier) tests (Anselin 1988) and robust LM tests (Anselin *et al.* 1996) are specific tests. Therefore, the LM test and its variants can discriminate the source of spatial dependence and identify a more appropriate spatial model alternative. Both classic and robust LM tests are run for the residuals of the OLS model and follow a chi-squared distribution with 1 degree of freedom. The purpose is to justify the need of an alternative spatial model. The null hypothesis of no spatial autocorrelation is tested against a specific alternative hypothesis based on the type of LM tests. Below is the list of hypotheses according to the LM test and its variants (Kelejian and Prucha 2001), the name of the test is given inside the parentheses. Each alternative hypothesis suggests an alternative spatial model, the details of these models are discussed in the next section.

H_0: there is no spatial autocorrelation;

H_a *(LM-err):* spatial error model;

H_a *(LM-lag):* spatial lag model;

H_a *(RLMerr):* robust version of LM-err;

H_a *(RLMlag):* robust version of LM-lag;

H_a *(SARMA):* a combination of RLMerr and RLMlag.

Specifying a spatial regression model

Spatial regression deals with the specification, estimation and diagnostic tests of regression models that include spatial effects. The starting point is the non-spatial linear regression model given as:

$$y_i = \alpha + \sum_{j=1}^{K} \beta_{ij}\, x_{ij} + \varepsilon_i, \qquad i = 1, \dots, N \tag{6.15}$$

where y_i is the dependent variable corresponding to the i-th observation, α is the intercept, x_{ij} ($j = 1, \dots, K$) denotes the j-th independent variable of the i-th observation, K is the number of exogeneous independent variables, β_{ij} represents the coefficients to be estimated, and ε_i is the error term. It is convenient to rewrite the model in (6.15) in matrix notation as:

$$Y = \alpha\iota_N + X\beta + \varepsilon \tag{6.16}$$

where Y denotes an $N \times 1$ vector of the dependent variable for each observation in the sample ($i = 1, \dots, N$), ι_N is an $N \times 1$ vector of ones associated the constant α to be estimated, X is an $N \times K$ matrix of the independent variables (K is the number of exogeneous independent variables), β is a $K \times 1$ vector of the parameters to be estimated, and ε is an $N \times 1$ vector of the error terms with ε_i assumed to be independently and identically distributed with zero mean and constant variance σ^2. To be consistent with conventional models introduced in major spatial econometric textbooks, the use of matrix notation will be carried out throughout this chapter.

In a linear regression model, three types of interaction effects can be broadly categorised into: (T1) endogenous interaction effects among the dependent variable, (T2) exogenous interaction effects among the independent variables, and (T3) interaction effects among the error terms. These interaction effects can be extended to a spatial regression model where spatial interaction effects could be found among dependent variable, independent variables, or error terms. Figure 6.4 demonstrates interaction effects and the corresponding appropriate econometric model.

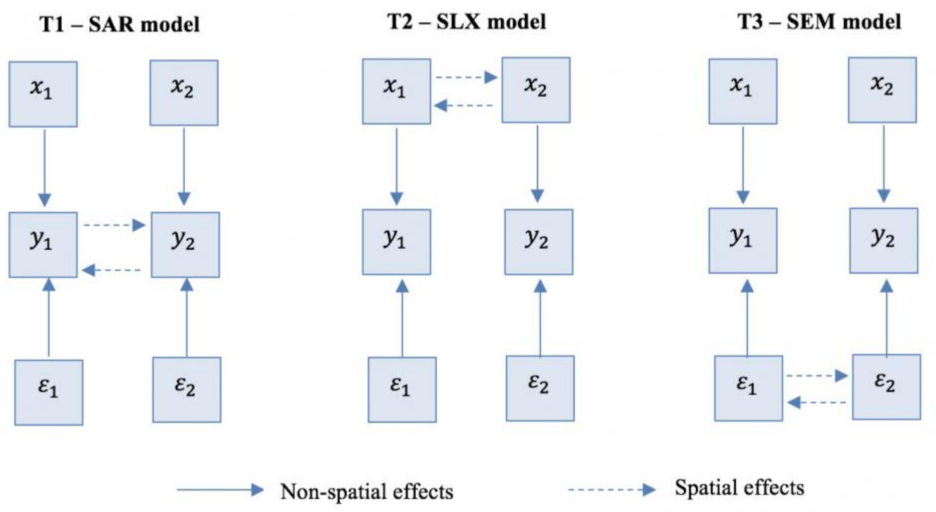


Figure 6.4 Spatial effects and the corresponding spatial interaction models.

According to the endogenous interaction (T1), the dependent variable of a spatial unit, say y_1, depends on the dependent variable of other units, say y_2. For example, y_1 can be thought of as the price of a house so that when agreeing on a sale price, a house buyer considers both the characteristics of the house (the number of rooms, the quality of amenities, its location, transport access, proximity to shops, etc) and the current selling price of nearby houses (y_2). To account for this endogenous interaction effect, *Spatial Lag Model*, or also known as *Spatial Autoregressive Model* (SAR) includes spatial lag (dependent) variables to explain the spatial linkage. A *spatial lag* of a variable is defined as the weighted average of observations on that variable over the neighbouring units. The SAR model (Cliff and Ord 1973; Ord 1975; Bivand 1984; Anselin 1988; LeSage and Pace 2009) generally takes the following form:

$$Y = \alpha\iota_N + \rho WY + X\beta + \varepsilon \tag{6.17}$$

where ρ is the spatial autoregressive coefficient and WY represents the endogenous interaction effects among the dependent variable. Of importance is the spatial autoregressive parameter ρ which measures the intensity of the spatial dependence. Positive (negative) ρ yields positive (negative) spatial dependence, and the value of 0 equates model (6.17) with the traditional OLS model in (6.16).

The exogenous interaction effect (T2) refers to the situation where the dependent variable of a particular unit, say y_1, depends on independent variables of other units, say x_2. Continuing with the previous example, the selling price of a house (y_1) may be influenced by the characteristics of nearby houses (for instance their size, location, the availability of garage, proximity to shops) which are denoted by x_2. If this phenomenon occurs, the specification of '*Spatial Lag of X Model'* (SLX) (Gibbons and Overman 2012) is required:

$$Y = \alpha\iota_N + X\beta + WX\theta + \varepsilon \tag{6.18}$$

where WX represents the exogenous interaction effects among the independent variable, and θ measures the magnitude of this interdependence.

Regarding the other interaction effect (T3) in which spatial autocorrelation is present among the error terms, this phenomenon is explained when two or more determinants of the dependent variable omitted from the model are spatially autocorrelated. Following the housing example, the opening of a major economic centre can represent a shock influencing the selling price of houses in that neighbourhood as well as nearby houses. *Spatial Error Model* (SEM) (Cliff and Ord 1973; Ord 1975; Anselin 1988; LeSage and Pace 2009) aims to correct the model's errors and takes the general form:

$$Y = \alpha\iota_N + X\beta + \varepsilon \tag{6.19a}$$
$$\varepsilon = \lambda W\varepsilon + u \tag{6.19b}$$

where λ is the spatial error coefficient and $W\varepsilon$ denotes the interaction effect among the errors of different spatial units.

Originally, the central focus of spatial regression was put on spatial lag model and spatial error model, both of which deal with one type of interaction effect. Anselin (1988) and Anselin *et al.* (1996) devise the testing procedure for a spatial lag or spatial error model based on the LM tests. Nevertheless, the past decade witnesses the blossoming interest for model that includes two or more types of interaction effects. Kelejian and Prucha (1998) and Elhorst (2010) advocate *SARAR* model that contains both endogenous interaction effects T1 and interaction effects among error terms T3. Models that include endogenous as well as exogenous interaction effects (T1 and T2) are labelled *Spatial Durbin Model* (SDM) (Anselin 1988), whereas the combination of T2 and T3 effects requires the use of *Spatial Durbin Error model* (SDEM).

Ultimately, the most general model, which includes three types of interaction effects, takes the following form:

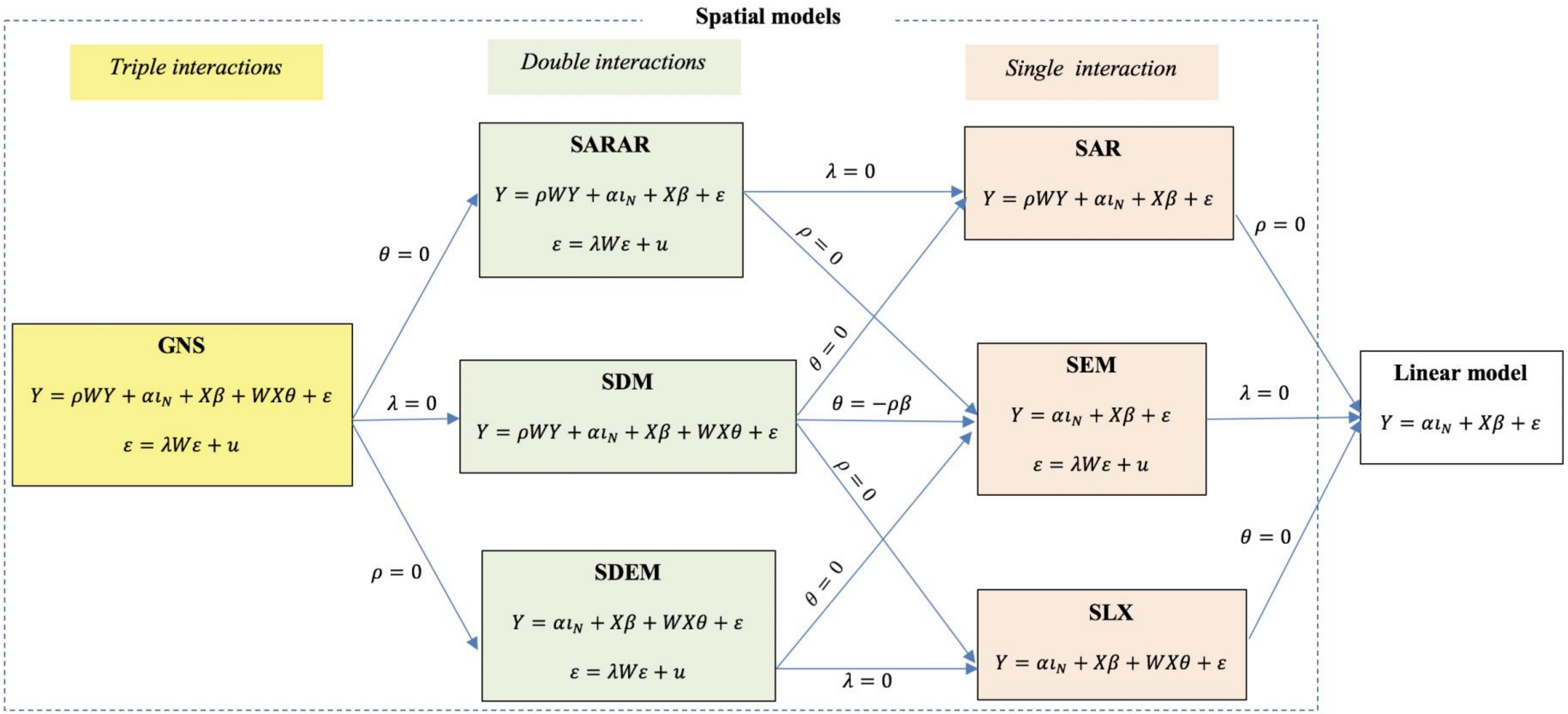
$$Y = \rho WY + \alpha \iota_N + X\beta + WX\theta + \varepsilon \tag{6.20a}$$

$$\varepsilon = \lambda W\varepsilon + u \tag{6.20b}$$

where WY denotes the interaction among dependent variable, WX the interaction among exogenous independent variables, and $W\varepsilon$ the interaction among error terms. This model is usually referred to as the *General Nesting Spatial model* (GNS). Important parameters include: ρ – the spatial autoregressive coefficient, and λ – the spatial autocorrelation coefficient. θ, just like β, is a $K \times 1$ vector of fixed but unknown parameters to be estimated.

Figure 6.5 shows a spectrum of linear spatial regression models with linear OLS model on the right side and GNS model on the left side. Models with one source of interaction, two sources and three sources are highlighted in different colours. The interrelationships among these models are indicated by arrows. From the left- to right-hand side, more specific models can be attained by imposing one or more parameters of the general model GNS (ρ, λ, θ). These restrictions are demonstrated right next to the arrows. It is worth noting that among the family of spatial models in Figure 6.5 models including exogenous interaction effects such as SDEM and SLX generally do not intrigue theoreticians and econometricians since the estimation of these models does not require any special consideration and standard estimation techniques could perform well. On the other hand, SAR, SEM and SARAR that incorporate endogenous interaction effects and interaction effects among errors are the most widely explored frameworks due to econometric problems accompanied in estimating these models. In the failure of standard econometric techniques, these spatial models could be estimated by Maximum Likelihood (Ord 1975), quasi-Maximum Likelihood (Lee 2004), instrumental variables (Anseline 1988), or generalised method of moments (Kelejian and Prucha 1998, 1999). An overview of these methods is included in Fischer and Nijkamp (2013).

The bewildering array of spatial models displayed in Figure 6.5 can create confusion in selecting the most appropriate model. Overall, two approaches are exploited. The first approach is *general-to-specific* or *top-down*, departing from the more general model GNS to a more specific model to explain the spatial phenomena. The selection criterion is based on likelihood ratio test (Hendry 1995). By testing models in Figure 6.5 from left- to right-hand side and gradually searching for a more parsimonious model, the top-down approach ensures correct inference as long as no variables are omitted in the more general model (Elhorst 2010). Nonetheless, GNS is by no means the most general model. Further spatial lags could be added to the model specified in equations (6.20a, b). On the other hand, three or more sources of spatial interaction effects usually imply weak statistical identification of the model (Cook *et al.* 2015). In practice, it is difficult to distinguish between ρ, λ, and θ; thus, GNS generally leads to an over-parameterised model. Wrong decision made at the beginning of the specification procedure can make the whole inference process go astray (Elhorst 2014).

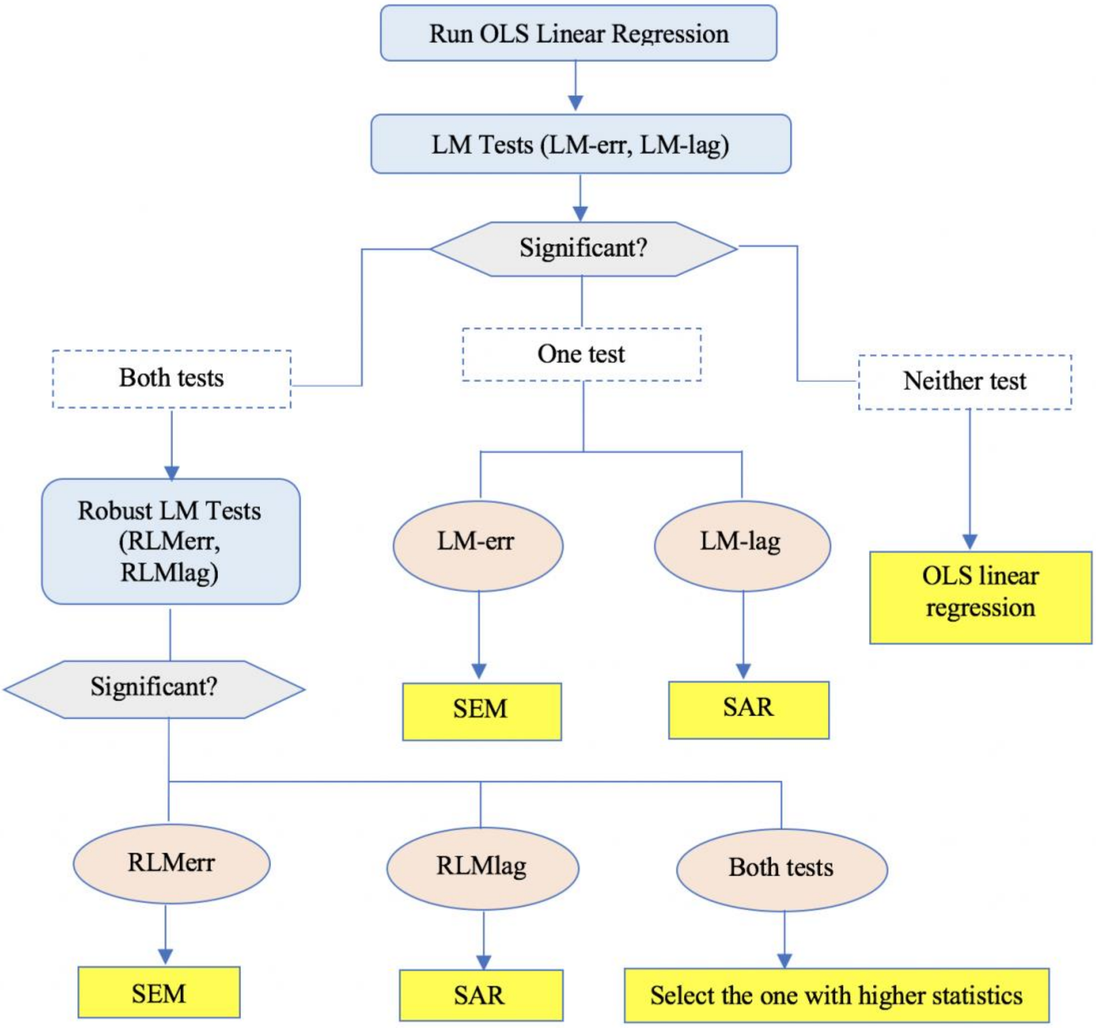


Note: SAR = Spatial Autoregressive Model (or Spatial Lag Model), SEM = Spatial Error Model, SLX = Spatial Lag of X Model, SDM = Spatial Durbin Model, SARAR = Spatial Autoregressive Combined Model, SDEM = Spatial Durbin Error Model, GNS = General Nesting Spatial Model.

Source: Author’s modification based on Halleck Vega and Elhorst (2014).

Figure 6.5 Taxonomy of spatial linear models.

The second approach in specifying a spatial model is *specific-to-general* or *bottom-up*. The testing procedure is illustrated in Figure 6.6. In a nutshell, the researcher starts with a non-spatial linear regression model and tests whether it is necessary to include spatial interaction effects into the benchmark non-spatial model. LM tests and the robust variants, as introduced in earlier discussion, make it possible to discriminate between SAR, SEM or non-spatial linear regression model. Florax *et al.* (2003) show that this procedure is most effective when the real underlying model is SAR or SEM. However, the authors argue that the likely presence of omitted variable bias might lead to false partial conclusions, and overestimating β coefficient(s) under the omission of spatial interaction terms may hinder the detection of spatial autocorrelation among residual terms of the OLS linear model.



Note: SAR = Spatial Autoregressive Model (or Spatial Lag Model), SEM = Spatial Error Model.

Figure 6.6 The bottom-up approach.

6.3 Data

In this chapter, the Food Balance Sheet (FBS) data from the Food and Agricultural Organisation of the United Nations (FAO) are retrieved. The FBS is one of the most extensive databases on national food consumption, describing the annual availability of total calories as well as primary food commodities (measured in kcal/person/day) for over 200 countries and territories (FAO 2019c). Details about the compilation of the FBS are mentioned in Section 4.3.4 (Chapter 4) and elsewhere (FAO 2017b). Despite being emphasised in previous chapters, it is important to re-stress that the consumption-level waste (i.e. food that is wasted at retail, restaurants, and household) is not incorporated in the FBS, therefore the data are not equivalent to average food intake or average consumption but best described as 'apparent consumption'. In the subsequent empirical analysis, the terms *caloric consumption*, *food consumption* and *diet* should thence be interpreted as *food available for consumption*. Another key limitation of the FBS is the reliability of data coverage and data quality for less developed countries (FAO 2018). In spite of these caveats, the FBS has been greatly utilised in the existing literature as it is the only source of standardised information on food consumption that enables longitudinal comparison between a large set of countries (Grünberger 2014; Leclercq *et al.* 2019).

The daily per capita food availability (in kcal/capita/day) over the period 1961-2013 is used for the empirical analysis. Countries with missing values are excluded and so are those with the population of less than one million in 2013 as these tend to be small islands with peculiar diets (Gouel and Guimbard 2019). This leads to a sample of 118 countries.

For the conditional beta convergence analysis, a range of economic, demographic, social and agroecological indicators are included in the model regressing the average growth rate of calories on the initial levels of calories. The purpose is to control for country-specific structural factors that can influence the changes of calorie consumption as well as the rate of convergence/divergence. A thorough discussion on the impact of these factors on the dietary changes is referred to Section 2.3.2 (Chapter 2). Such variables have been previously employed in several studies, as in Du *et al.* (2004), Hawkes *et al.* (2012), Choudhury and Headey (2017), Oddo *et al.* (2017), and Azzam (2020) among others. In this study, four of such variables are included and the details are as follows:

- Arable land (% of land area) – average for the period 1961-2013.
- GDP per capita growth (annual %) – average for the period 1970-2013.
- Urban population (% of total population) – average for the period 1961-2013.
- Labour force participation rate, female (% of female population aged 15 and over) – average for the period 1990-2013.

Data on arable land, urban population and female labour force participation rate are retrieved from the World Bank (World Bank 2020e) and data on GDP per capita are collected from the FAOSTAT (FAO 2019b). Arable land is used as a proxy for the capacity of agricultural production.

Higher share of land use for arable agriculture means larger potential for growing crops and hence higher food supplies which will have a positive effect on calorie growth. The average growth rate of GDP per capita is used to proxy economic development. Higher earnings not only allow individuals to afford a wider range of foods, but also generate demand for non-staple foods usually from animal source and in processed forms (Law *et al.* 2019; Umberger *et al.* 2020). Consequently, rising incomes are related with higher calorie consumption and more robust growth of calories due to the increased consumption of animal-source foods, sugars, fat and oils. Next, share of urban population is used as an indicator of urbanisation transition and subsequent changes to lifestyle and food acquisition strategies. To illustrate, increased urban growth is accompanied by increased access to processed foods, changing consumption preference towards convenient, 'ready-to-eat' and 'ready-to-heat' products, and increased connections with people from different cultures facilitating the acceptance of 'Western' dietary styles (Pingali 2007; Hawkes *et al.* 2017). All of these will exert a positive influence on the calorie consumption. Finally, female participation in the labour force is expected to affect the calorie growth positively. On the one hand, growing economic participation of women translates into more income for purchasing foods especially energy-dense foods as they become more affordable (Oddo *et al.* 2017). On the other hand, improved female employment would raise the opportunity cost of food preparation, decrease the amount of time women spend on cooking, and boost the demand for processed foods (Popkin and Reardon 2018). To the extent that arable land, economic growth, urbanisation and female participation in the labour force increase/decrease proportionately more in poor countries than in rich countries, these factors contribute to the dietary convergence. Rising income is more prevalent in poorer economies since richer economies have already had high levels of GDP per capita. Likewise, urbanisation and female employment are expected to increase faster in poorer countries than in richer ones. Therefore, these three variables are likely to have a positive impact on the rate of convergence.

For the exploratory spatial analysis, a fundamental task is to define a 'space' metric where the relationship among countries in the 'space' is drawn upon. Acknowledging that the definition of 'space' is more than geography, different possibilities are proposed in order to ensure that the adequate spatial relationship is not ignored. The starting point is to examine the spatial relationship through geographical lens if nearby countries tend to have similar diets. To quantify the geographical closeness, a distance-based matrix W^a is constructed and normalised as such $W_{ij}^a = {d_{ij}}/{\max(d_{ij})}$, where d_{ij} represents the geographical distance (in kilometres) between the capital cities of countries i and j. Also dictating the geographical proximity, the second option is to use a contiguity matrix W^b in such a way that $W_{ij}^b = 1$ if countries *i* and *j* are geographically contiguous and $W_{ij}^b = 0$ otherwise. Third, a contiguity matrix W^c is proposed to consider the economic proximity measure among countries. The generic element W_{ij}^c equals 1 if the average GDPs per capita 1970-2013 of countries *i* and *j* fall in the same quantile, and 0 otherwise. Such a contiguity matrix supports the argument that countries with similar income levels tend to have similar diets. The information of geographical distance and geographical contiguity is retrieved from the CEPII database (CEPII 2019), and the GDP per capita data are obtained from the FAOSTAT (FAO 2019b).

6.4 Empirical results

6.4.1 Results of convergence tests in non-spatial context

Sigma convergence

The convergence in patterns of national food consumption is first investigated by sigma convergence. Changes in the coefficient variation (CV) of caloric availability are plotted in Figure 6.7. Overall, the dispersion across countries shows a downward path over the whole period, indicating sigma convergence. However, the reduction in CV is uneven with a significant downturn in the early 1990s.

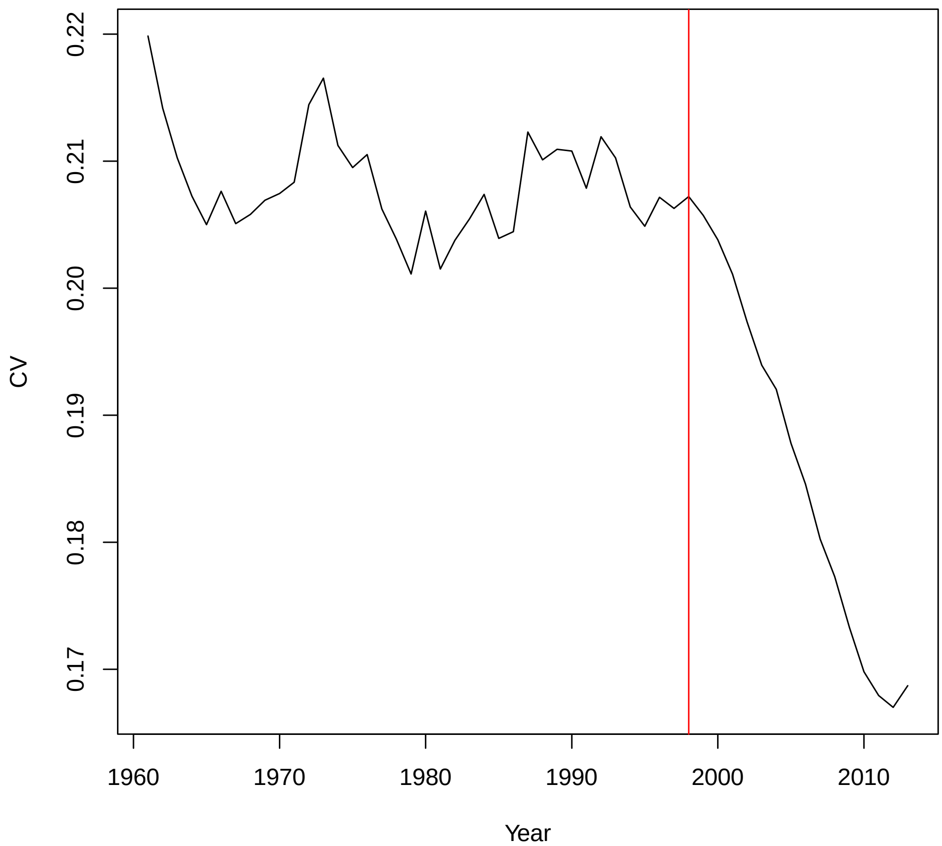


Figure 6.7 Coefficient of variation in daily per capita calories, 1961-2013.

While the FAO has made steady improvements in the methodology to compile the FBS, it would be fair to believe this structural break reflects a real change in the data rather than a methodological issue. In a FAO/WHO joint publication (WHO 2003, pp.14-15), they acknowledge: “This change [in the calorie availability] has not, however, been equal across regions. The per capita supply of calories has remained almost stagnant in sub-Saharan Africa and has recently fallen in the countries in economic transition ...”. Such variant patterns in the calorie availability across countries likely result in divergence as indicated by the strong peaks over the period from 1961 to 1998. Nonetheless, the period 1999-2012 is characterised by sigma convergence consistently over a long period of time. In 2013, the CV increases slightly, signalling that the trend is about to reverse towards divergence across countries.

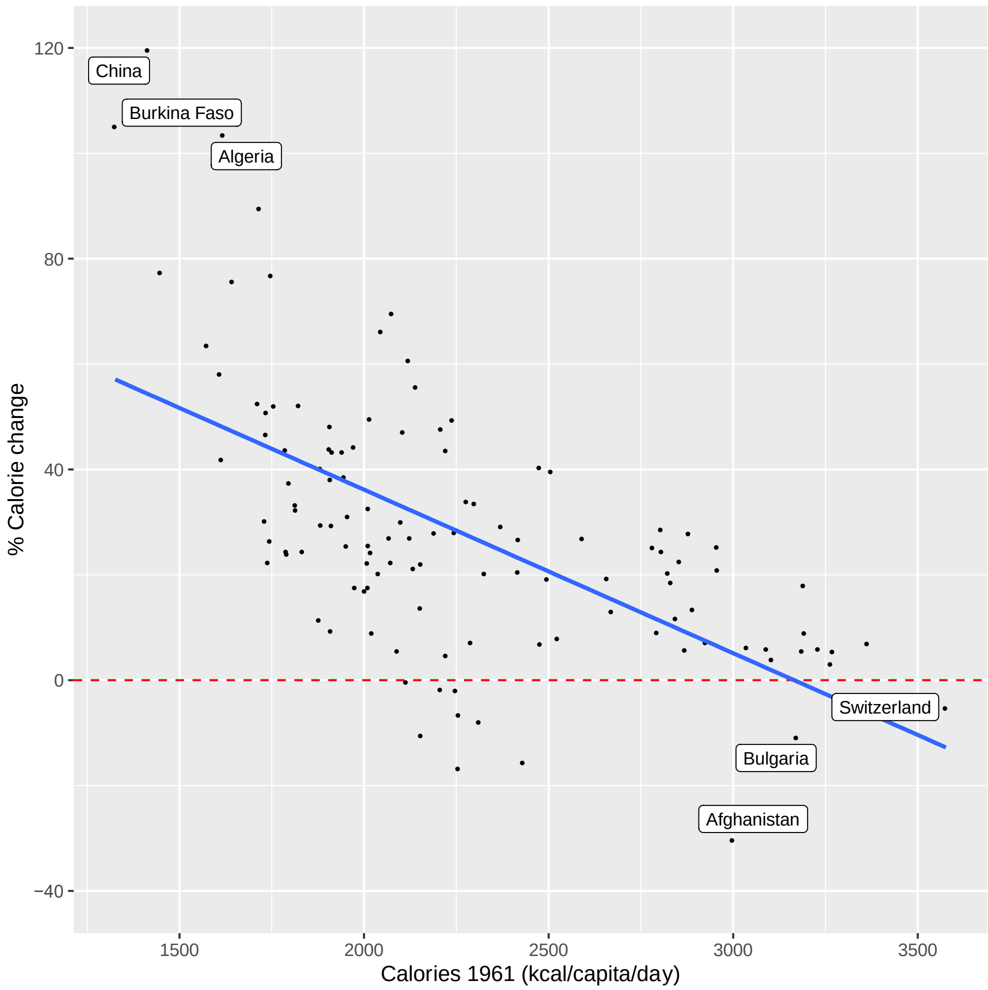
Unconditional beta convergence

To better understand the variations of calorie availability across countries, changes in calories are plotted against the initial calorie level in 1961. Two main features are highlighted in Figure 6.8. First, calorie changes are not even across countries. Over the past half a century, the calorie availability of most countries in the sample has increased by between 10 and 50% and doubled in China, Burkina Faso and Algeria. Yet, roughly 10% of the sample have witnessed declining calories – a pattern that contradicts the predominant trend in the existing literature. Further, except Switzerland and Bulgaria, these countries, which lie below the red dashed line, are listed by FAO as in need of external assistance for food due to civil conflict and population displacement (Afghanistan, Central African Republic, Chad and Uganda) or unfavourable climatic conditions (Kenya, Madagascar, Namibia and Zimbabwe) (FAO 2020a).

As highlighted in Figure 6.8, Bulgaria and Afghanistan are associated with an average of 3,000 kcal/person/day in 1961 – just right after Switzerland and among countries with the initially highest calorie consumption. This figure might seem surprising at first; however, there are some reasons to believe this is not due to measurement errors. The food economics literature has mentioned the rapidly rising trends in food consumption for Bulgarians in the 1960s. As documented by Ghodsee (2004), from 1960 to 1979, Bulgarians saw significant increases in the per capita consumption of all food categories (except potatoes). In 1977, Bulgarians ate 143% of the daily caloric requirement – the highest level among both industrialised market economies of the West and Eastern bloc economies and had a higher per capita daily caloric intake than the US. Such a substantial rise in food availability has been attrributed to strong economic growth and the transformation of Bulgaria into a modernised insutrial country following the communist government (Dimitrov and Atanasova 1964). Since the economic slowdown in the 1980s, the calorie consumption however has reduced, especially after the economic crisis in the mid-1990s. Economic contraction and hyperinflation caused food prices to rise at a higher speed than real income (Ivanova *et al.* 2006). Regarding Afghanistan, the literature is relatively silent about its food consumption in the past. Nonetheless, the high level of calorie availability could be due to a strong increase in cereal production (especially wheat) as the country benefited from the Green

Revolution (Borlaug 1971). The consumption has decreased possibly owing to the severe economic hardship during the 1979 Soviet invasion and ensuing civil war destroyed much of the country's limited infrastructure, and disrupted normal patterns of economic activity (Roy 2020).

Another notable feature from Figure 6.8 is the general downward tendency, or an inverse relationship between calorie changes and initial level of calories that is illustrated by the downward-sloping blue line. On average, countries with lower levels of calories in 1961 experience higher growth rates, signalling the 'catching-up' process of beta convergence.



Note: The blue line represents the trend, the red dashed line represents the zero-growth level, and the grey area denotes the 95% confidence level.

Figure 6.8 Relationship between calorie growth rate 1961-2013 and initial calorie level.

Next, beta convergence in the per capita daily calories is estimated for the entire period 1961-2013. Due to the structural break observed in the sigma convergence plot (Figure 6.7), beta convergence tests are also run for two sub-periods: 1961-1998 and 1999-2013. Even though convergence process is a long-run phenomenon, it is interesting to split the period into two sub-periods to see whether this process has been homogeneous during the period under consideration. The OLS estimations for the regression model (6.2) are presented in Table 6.1.

The significantly negative beta coefficient of -0.011 for $\log(y_{i,t_0})$ suggests that for a 10% increase in the initial level of calories, the average annual growth rate of calories would decrease by $0.011 \times \log(1.1) = 0.001\%$. The negative sign of the beta coefficient is indicative of convergence, meaning that initially high-calorie countries tend to exhibit lower growth rates of calories. Following the formula (6.3), the convergence speed is computed at 1.6% per year. In other words, countries bridge the gap between the current levels of calories and the steady-state levels by, on average, 1.6% annually. This convergence speed implies a half-life measure of 63 years, that is to say it takes almost 63 years for follower countries to eliminate 50% of the initial calorie gap regardless of country-specific characteristics.

When the data is split into two sub-periods, the significantly negative beta coefficients (-0.0009 and -0.016) prove a consistent trend of convergence and support the finding from Figure 6.8 that countries with lower initial levels of calories show a more robust convergence process. On the other hand, the convergence speed increases in the last two decades, and the half-life measure drops by almost one third from 73 years during 1961-1998 to approximately 42 years during 1999-2013.

Table 6.1 Absolute (unconditional) beta convergence: Estimation results.

Variables	Whole period	Divided into two sub-periods	
	1961-2013	1961-1998	1999-2013
Intercept	0.089*** (9.565)	0.076*** (5.587)	0.133*** (7.926)
$\log(y_{i,t_0})$	-0.011*** (-9.092)	-0.0009*** (-5.291)	-0.016*** (-7.627)
Adjusted R^2	0.411	0.187	0.328
Number of observations	118	118	118
Convergence speed per year (%)	1.6	1.2	1.8
Half-life (years)	62.8	73.4	42.3

Note: All test results are not significant unless indicated otherwise. *t*-statistics are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Absolute beta convergence is useful in identifying the 'catching-up' effect in the whole sample of 118 countries. However, there are good reasons to believe that the convergence process is not always linear, and the conditional beta convergence theory posits that countries can converge on a path to different steady-state levels ("convergence clubs"). To illustrate, countries at different income levels are represented by distinguished colours in Figure 6.9a. The country's classification into high-income, upper-middle-income, lower-middle-income and low-income is in line with the World Bank's categorisation (World Bank 2020d). It can be seen that the wealthiest countries pertain the largest per capita daily calories in 1961 and the lowest growth rate of calories. In general, total calories in middle-income countries tend to grow most robustly while the pattern for low-income countries is less conclusive. These are indications that the convergence process has not been homogenous across countries at different development levels.

One way to accommodate the nonlinearity is to allow for some flexibilities in equation (6.2) by letting the intercept vary among different groups of countries. Dummy variables with high-income countries being the reference category are added in the model to denote country groupings: d_L for low-income countries other than the high-income peers, d_{LM} for lower-middle-income countries, and d_{UM} for upper-middle-income countries. The regression model is represented by:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \beta \log\left(y_{i,t_0}\right) + \beta_L d_L + \beta_{LM} d_{LM} + \beta_{UM} d_{UM} + \varepsilon_i \qquad (6.21)$$

The significance of the dummy coefficients $\beta_L, \beta_{LM}, and\ \beta_{UM}$ would testify if there are equal changes in the average annual growth rates (the left-hand side of equation 6.21) due to a fixed change in the initial level of calories $\log\left(y_{i,t_0}\right)$ across four groups. The first three columns of Table 6.2 report the OLS estimation results for model (6.21).

Adding dummy variables improves the model fit, reflected in the considerably higher adjusted R-squared (0.628 as compared to 0.411). However, a large portion of the model remains unexplained. The beta coefficient for $\log\left(y_{i,t_0}\right)$ is significant and more robust (-0.018), leading to a higher annual convergence speed (5.3%) and a smaller half-life measure (38 years).

Column (1) reveals highly significant dummy coefficients for d_L, d_{LM}, and d_{UM}, proving that a fixed increase/decrease in the initial calorie level would lead to an unequal decrease/increase in the average annual growth rate for each group of countries. The negative sign of all dummy variables highlights that the changes are larger for the wealthiest countries as compared to others. For example, the coefficient -0.006 of d_L implies that for any increase in initial calorie level the annual growth rate of calories for low-income countries would on average decrease by 0.006% lower than the reference group (high-income countries). Similarly, the magnitude of the change in annual growth rate for lower-middle- and upper-middle-income countries is averagely 0.005% and 0.002% smaller than high-income peers. These results suggest that lower-income countries are converging to a steady state of calorie growth rate that is lower than for high-income countries. This might reflect the larger contribution of energy-dense foods (mostly from animal origin and in processed forms) in diets of rich countries.

Columns (2) and (3) of Table 6.2 show how this relationship changes across time. When the regression equation is run for two sub-periods, the coefficients for d_L, d_{LM}, and d_{UM} are still significantly negative in the former period but become insignificant in the latter period. The historical gap in responses of annual growth rate of calories to a change in the initial level of calories across different groups of countries is no longer perceivable during the last 15 years. Alternatively, income plays an important role in the early stage of the convergence in national food consumption. Nonetheless, when national incomes reach a certain threshold, income alone cannot explain the path countries are converging to different steady-state levels.

Table 6.2 Beta convergence with country groupings: Estimation results.

Variables	Adding dummy variables			Adding interaction terms		
	1961-2013 (1)	1961-1998 (2)	1999-2013 (3)	1961-2013 (4)	1961-1998 (5)	1999-2013 (6)
Intercept	0.147*** (13.653)	0.163*** (10.242)	0.144*** (5.137)	0.125*** (6.082)	0.149*** (4.844)	0.151** (2.047)
$\log(y_{i,t_0})$	-0.018*** (-13.365)	-0.020*** (-10.002)	-0.018*** (-5.099)	-0.015*** (-5.927)	-0.018*** (-4.717)	-0.019** (-2.032)
d_L	-0.006*** (-7.983)	-0.009*** (-7.551)	-0.002 (-1.108)	0.083*** (2.879)	0.092** (2.126)	-0.12 (-1.050)
d_{LM}	-0.005*** (-6.225)	-0.007*** (-6.714)	0.001 (0.379)	-0.001 (-0.018)	-0.053 (-1.065)	0.007 (0.086)
d_{UM}	-0.002*** (-3.648)	-0.004*** (-3.589)	0.0003 (0.186)	-0.001 (-0.041)	-0.011 (-0.284)	0.006 (0.070)
$\log(y_{i,t_0}) \times d_L$				-0.012*** (-3.137)	-0.013** (-2.371)	0.015 (1.051)
$\log(y_{i,t_0}) \times d_{LM}$				-0.0004 (-0.090)	0.006 (0.949)	-0.001 (-0.084)
$\log(y_{i,t_0}) \times d_{UM}$				-0.0001 (-0.024)	0.001 (0.210)	-0.001 (-0.070)
Adjusted R^2	0.628	0.481	0.339	0.662	0.517	0.332
Number of observations	118	118	118	118	118	118
Convergence speed per year (%)						
Low-income countries	5.3	3.7	2.0	4.1	15.0	2.1
Others				3.1	3.1	
Half-life (years)						
Low-income countries	38.2	34.3	38.2	40.5	25.4	36.1
Others				45.9	38.2	

Note: All test results are not significant unless indicated otherwise. *t*-statistics are reported in parentheses. *, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

Column (4) reports the coefficients for d_L, d_{LM}, and d_{UM}, which specify differences in intercepts, and the interaction coefficients for $\log(y_{i,t_0}) \times d_L$, $\log(y_{i,t_0}) \times d_{LM}$, and $\log(y_{i,t_0}) \times d_{UM}$ which indicate differences in slopes. While the former may not be of much interest, the latter is definitely important since the significance of which would reveal if there is a common rate of convergence. Two comments can be made here. First, the coefficient of the interaction term for low-income countries $\log(y_{i,t_0}) \times d_L$ is statistically significant, suggesting that the convergence rate is statistically different between the poorest and the richest group. Second, the negative sign of the interaction coefficient for low-income countries (-0.012) implies that the trajectory for low-income countries is 0.012 lower than high-income countries. This feature is clearly observed in Figure 6.9b: the regression line for the poorest countries substantially deviates from the middle- and high-income peers, being the steepest. Thus, low-income countries converge at the fastest pace and convergence rate reduces as income rises.

According to the average marginal effect, it turns out that for low-income countries the annual growth rate of calories would on average decrease by 0.0027 (%) per 10 percent increase in the initial calorie level. This rate of change implies an annual convergence speed of 4.1% and a half-life of 40 years. To compare, the beta coefficient for high-income countries is derived from the coefficient for $\log(y_{i,t_0})$ in Column (4), which is -0.015. If the initial calories increase by 10%, the average annual growth rate of calories would decrease by 0.0015 (%). As a result, high-income countries converge at the pace of 3.1% per annum and the half-life measure of approximately 46 years. These statistics reaffirm that poor countries have converged at a faster speed than the richer ones.

When splitting the data into two sub-periods, the interaction coefficient for low-income countries is significant over the pre-1998 period but appears insignificant after 1998. The interaction coefficients for lower- and upper-middle-income countries remain insignificant in either period, confirming that the convergence rates of middle-income and high-income countries are not statistically different. The negative sign of the interaction coefficient (-0.013) suggests the faster convergence speed of the poorest group of countries as compared to the wealthier peers. Convergence statistics show that prior to 1998 low-income countries were converging at a rapid pace of 15% per annum – the figure that is three times larger than the average rate over the whole period (4.1%). However, there is no evidence for any gap in the convergence speeds among different groups of countries in the latter period.

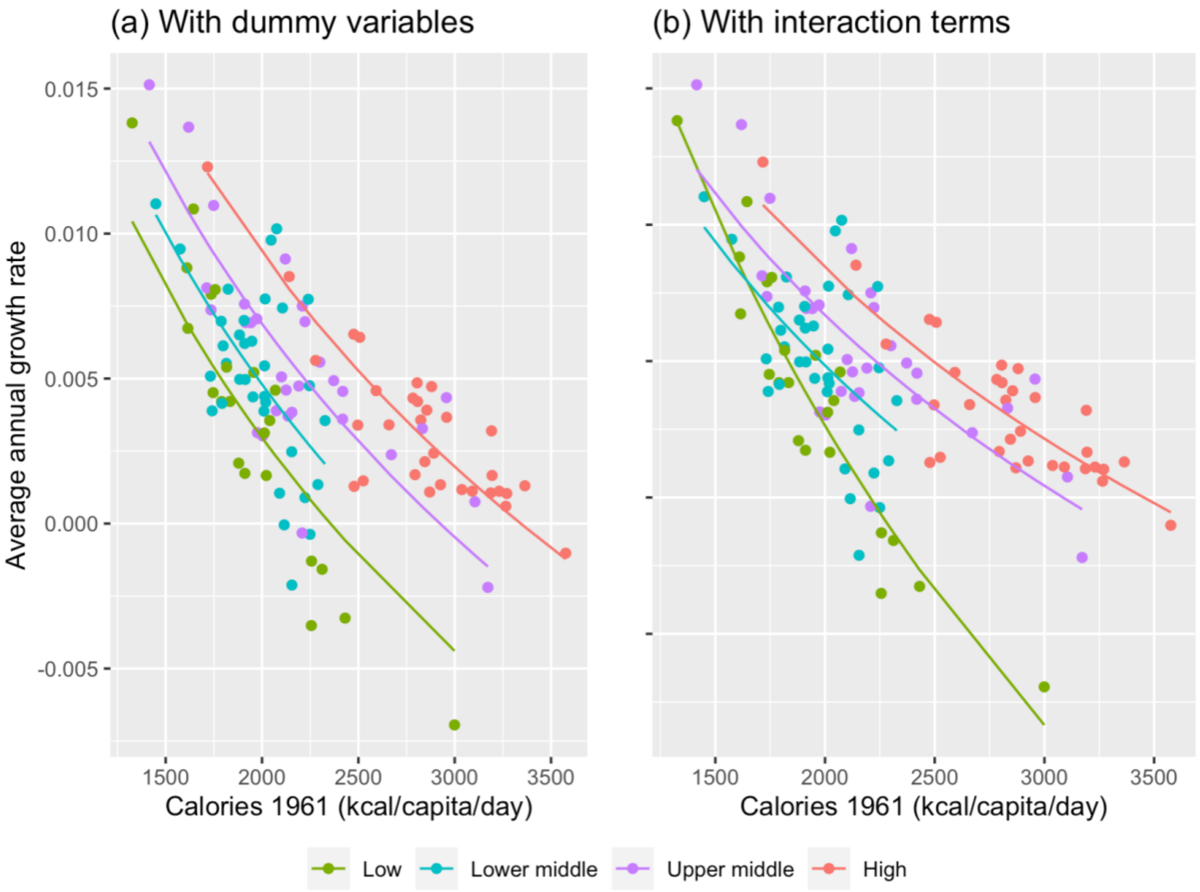


Figure 6.9 Fitted lines for regression model.

Conditional beta convergence

The analysis in this section examines a conditional beta convergence model with the inclusion of demographic, agroecological, and socio-economic variables. A description of these variables is provided in Section 6.3. Following Mankiw *et al.* (1992), a conditional growth model is defined as follows:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \beta\log\left(y_{i,t_0}\right) + \sum_{j=1}^{J}\pi_j Z_{ji} + \varepsilon_i \qquad (6.23)$$

where y_{i,t_0} and y_{i,t_0+T} are respectively the per capita daily calories of country *i* in the initial and final periods; T is the number of years; Z_{ji} includes the set of J additional control variables which vary over countries; and ε_i is the standard error term being independent and identically distributed with mean zero and variance σ_ε^2. More specifically, Z_{ji} comprises the average percentage of arable land in the total land area, the average growth rate of income, the average percentage of urban population, and the average female participation rate in the labour force. B, β and $\pi_j (j = 1, \ldots, 4)$ are the parameters to be estimated. Since the factors of dietary changes are often highly correlated, the variance inflation factor

(VIF) is computed for each independent variable in the OLS regression and any value exceeding 3 would indicate multicollinearity problem.

The OLS estimations of the model (6.23) are reported in the first column of Table 6.3. After controlling for a range of structural indicators, the beta coefficient is still significantly negative (-0.017), implying conditional convergence. The coefficients for arable land, income growth and urban population are significantly positive whereas the coefficient for female employment is significantly negative. In addition, the magnitude of the effect of GDP growth is the largest among structural indicators, suggesting that rising income has been the most influential determinant of dietary changes. The OLS estimates suggest that a one percent increase in the percentage of arable land between 1961 and 2013, the share of urban population and the growth rate of GDP is associated with a 0.4, 0.8 and 11.1 percent increase in the average annual growth rate of calories, respectively. Likewise, a one percent increase in the female participation rate in the labour force is associated with a 0.3 percent decrease in the average annual growth rate of calories. While the results for arable land, GDP growth and urban population are in line with the evidence suggested in the literature (as discussed in Section 6.3), the negative effect of female employment is surprising. The reason could be ascribed to the fact that higher proportion of female employment also means higher levels of education and training for women, which in turn lead to better nutritional knowledge. Having some extra amount of earnings enables households to access more calories from either nutrient-dense foods that contribute to higher-quality diets or calorie-dense foods rich in salt, fat and sugars that can undermine diet quality. The finding presented in this analysis supports the former, possibly thanks to the improved health awareness. Better employability for women also means less time for family and household activities (such as shopping for raw food materials and cooking). As a result, the family will not tend to eat home-cooked foods and instead opt for pre-cooked (junk) foods. For example, it is not uncommon to see family having sandwich instead of full-fledged meals at lunch or dinner. Since junk foods with their high convenience often cost more, this can cause a decline in the purchase of such foods and in calorie consumption.

The essence of conditional beta convergence is that in the long run countries need not converge to each other but instead converge to their own steady-state level. In this respect, countries with the initially low level of calories do not necessarily exhibit higher growth rates of calories, but countries that are further from their own steady-state level exhibit faster growth. With the inclusion of structural variables, results of the conditional beta convergence analysis indicate that countries are converging on average at an annual rate of 4% and it takes approximately 17 years for countries to eliminate half of the current disparity with the equilibrium. Compared with the unconditional beta convergence model, these convergence statistics imply a faster convergence process than when the structural parameters are not included.

Table 6.3 Conditional beta convergence regression: Estimation results.

Models	(1) OLS	(2) OLS	(3) OLS
	1961-2013	1961-1998	1999-2013
Intercept	0.128*** (0.010)	0.145*** (0.015)	0.070*** (0.026)
log(initial calories)	-0.017*** (0.001)	-0.019*** (0.002)	-0.016*** (0.003)
Arable land	0.004** (0.002)	0.005** (0.002)	8.7×10^{-5} (0.003)
Income growth	0.111* (0.057)	-0.004 (0.027)	0.770*** (0.139)
Urban population	0.008*** (0.001)	0.011*** (0.002)	0.006** (0.003)
Female employment	-0.003** (0.001)	-0.005*** (0.002)	-0.0001 (0.002)
Number of observations	118	118	118
Maximum VIF	1.98	2.09	2.29
Convergence speed per year (%)	3.99	3.18	1.82
Half-life (years)	17.37	21.82	38.09

Note: All test results are not significant unless indicated otherwise. Standard errors are inside the parentheses. *** Significant at $p \leq 0.01$; ** Significant at $p \leq 0.05$; * Significant at $p \leq 0.1$.

Similar to the earlier section, the analysis of conditional beta convergence is replicated for two sub-periods: 1961-1998 and 1999-2013. For the former period, the OLS estimations in Column (3) of Table 6.3 report a significantly negative beta coefficient which is however larger in magnitude than for the whole period. Turning to the structural parameters, arable land, urban population and female employment are significant predictors. The first two variables exert a positive effect on the annual growth rate of calories whereas the last variable exerts a negative effect. Noticeably, the impact of the structural indicators is larger in magnitude than the OLS estimates for the whole period shown in Column (1) of Table 6.3. Nonetheless, income growth does not affect the growth rate of calories.

The OLS estimations for the conditional beta convergence in the latter period 1999-2013 are presented in the last column of Table 6.3. The beta estimate is significantly negative and smaller in magnitude than in the former period (-0.016 versus -0.019) indicating a less robust convergence process. Indeed, the annual rate of convergence (1.82%) is lower than in the former period (3.18%) and is less than half of the average speed (3.99%) for the whole period. This result suggests that the structural parameters are closer to their steady-state values in the latter period, therefore countries are converging to their own equilibrium levels of calories at a slower speed. Investigating the effects of the structural parameters, rising income and urban population are the only two significant predictors and rising income has been the major driver for the global convergence in this period. A one percent increase in the growth rate of GDP is related to a 77 percent increase in the calorie growth whereas a one percent increase in the share of urban population is related to a 0.6% increase. The remaining variables (proportion of arable land and female employment) do not influence the growth of calories. That is to say agroecological, social and demographic indicators exert a stronger impact in the initial period; yet, economic factors have become a more important determinant of the dietary convergence over the past 15 years.

To sum up, the results of sigma and beta convergence point to convergence in calorie availability across countries over the past 50 years. The beta coefficient is significantly negative, implying that countries with lower levels of initial calories tend to exhibit higher growth rates of caloric consumption. Results of absolute beta convergence model show that countries bridge the gap between the current levels of calories and the steady-state levels by on average 1.6% per year. Additionally, the convergence process is proved to be non-homogenous across countries and how income affects the path countries are converging towards different steady states. The results indicate that income plays an important role in the early stage of convergence in national food consumption. Further evidence confirms the fastest speed of convergence for low-income countries. Thus, these findings call for scrupulous attention in monitoring and supporting low-income countries at the early stage of development as they exhibit the most worrisome trend. Results of the conditional beta convergence analysis complement the results from the unconditional beta convergence analysis and further elucidate the role of income. The structural conditions of low-income countries are further from their steady-state level than those of high-income countries. While agroecological, social, economic and demographic factors are proved to influence the growth rate of calories, rising income has been the main driver for rising calorie consumption over the past 15 years.

6.4.2 Detecting spatial dependence

Before embarking on specifying a spatial model, it is necessary to justify the presence of spatial phenomena in the data. A simple but useful approach is to visualise the data. Figure 6.10a illustrates the distribution of average calorie consumption for the period 1961-2013. The caloric consumption level of each country is represented by a shade of blue: the darker the shade, the larger the calories. Countries shown in grey are excluded from the analysis. Three comments can be made by inspecting Figure 6.10a. First, nearby countries tend to have similar level of calories, for example Australia and New Zealand at roughly 3,000 kcal/person/day, or China and India ranging from 2,200 to 2,500 kcal/person/day. Second, countries located within the same geographical region are likely to have the same level of calories. This phenomenon is most apparent in Asia, Northern America and Europe, but less pronounced in Africa and South America where almost a full spectrum of calories is observed. In these two continents, the average per capita daily calories of contiguous countries tend to fall in the same range, for instance Morocco and Algeria, or Brazil and Colombia. Hence, the influence of geographical closeness on the average per capita calories is evident; yet, it is a nuanced process. While geographical proximity is important in explaining some similarities in the calories of, say the United States and Canada, France and Italy, New Zealand and Australia, it is less so between each of these groups.

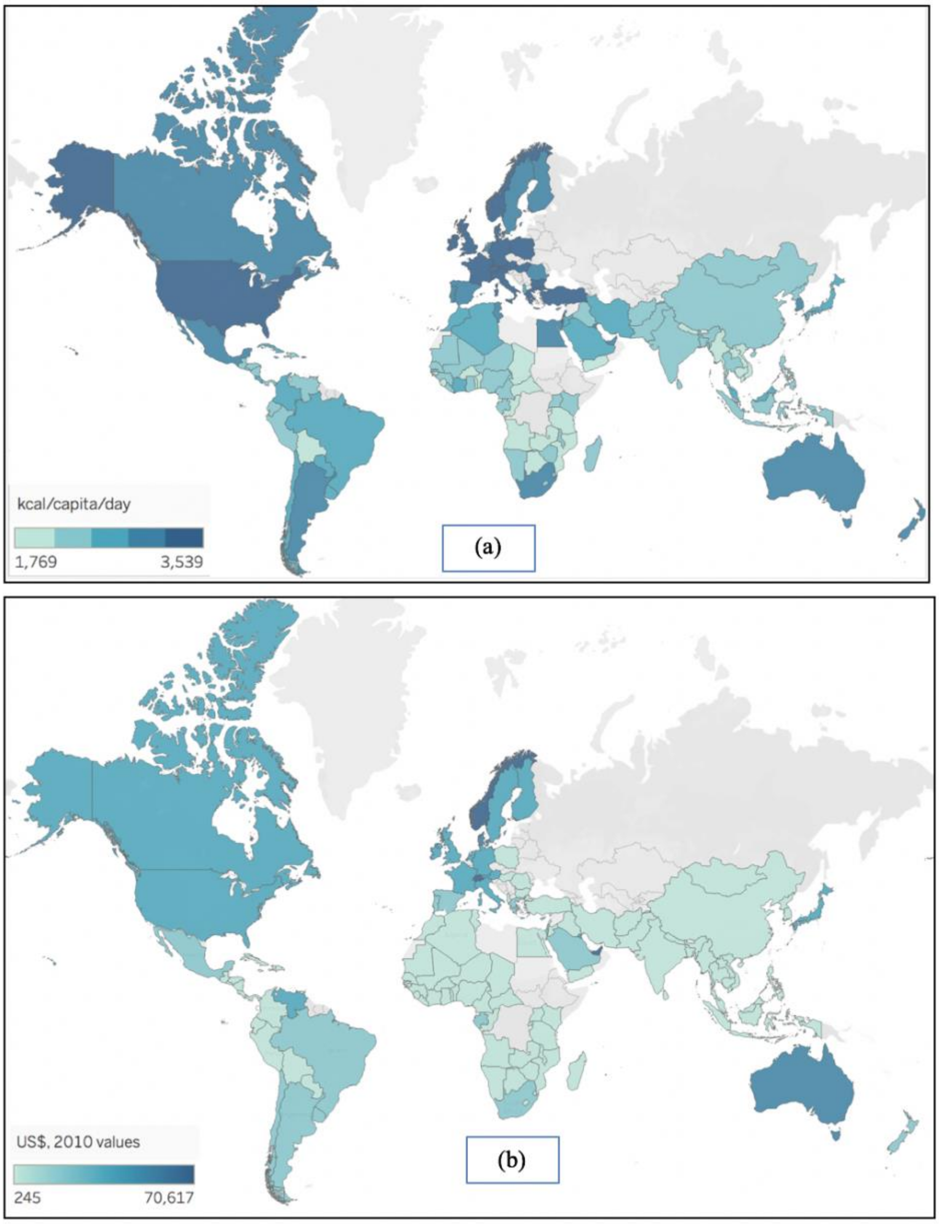


Figure 6.10 World map of (a) average daily per capita calories 1961-2013, (b) average GDP per capita 1970-2013.

Third, there is a concentration of highly calorific countries in Northern America, Europe, and Oceania while low-calorie countries are mainly located in East Asia and Africa. Zooming into continents with a clear polarisation in the distribution of total calories, richer countries are exposed to higher calories: South Africa and Argentina are obvious examples for Africa and South America respectively. For the purpose of comparison, a world map showing the geographical distribution of average income (GDP per capita) is plotted in Figure 6.10b. The level of income is represented by different gradations of blue: the darker the shade, the higher the income. Overall, a large number of blue shades overlap between two maps, suggesting that the richer a country the higher its calorie consumption. If such an association exists in the data, it signals a positive spatial autocorrelation, where the 'space' here is in terms of economic development rather than geography. Thus, the inspection of Figure 6.10 reveals possibilities of spatial autocorrelation among countries due to their proximity in either geographical closeness or economic development. Next, spatial autocorrelation tests can verify whether spatial autocorrelation is statistically significant as well as the strength and sign of the autocorrelation.

Results of the Global Moran's I statistics are reported in Table 6.4. Three spatial weight matrices are chosen to proxy different spatial relationships among countries regarding geographical distance (W^a), geographical contiguity (W^b), and economic contiguity (W^c). Given the three comments made by the inspection of Figure 6.10, one would expect significant Global Moran's I statistics for all three spatial weight matrices. However, Table 6.4 indicates little statistical evidence for spatial autocorrelation when the proximity among countries is measured in geographical distance (W^a). The negative sign is unsurprising when looking at the comparable levels of calorie consumption between distant countries such as Australia and Canada, or the United Kingdom and New Zealand in Figure 6.10a. On the other hand, there is no statistical evidence for spatial autocorrelation when the geographical boundaries between countries are considered (W^b). This result, though unexpected, makes sense taking into account the similar calorific levels, say between the United States and the United Kingdom, even though no actual boundary is shared between these countries. Hence, geographical contiguity, although probably being a driver for similarities in national diets within a continental region, is not a valid determinant at the global level. When the third spatial weighting matrix (W^c) is applied, the Global Moran's I index suggests a significant positive spatial autocorrelation, supporting that countries with similar income level tend to have similar levels of calories. To sum up, the examination of Global Moran's I statistics points to income level rather than geographical closeness that is driving similarities in national patterns of food consumption. This result corroborates the idea that spatial information should not be neglected in the study of dietary convergence. The specification of W^c will be applied for the rest of the analysis.

Table 6.4 Global Moran's I statistics for average daily per capita calories 1961-2013.

Spatial weight matrix	W^a	W^b	W^c
Global Moran's I	-0.001 *	0.002	0.665 ***

Note: Calculated by randomisation approach and two-sided alternative hypothesis;

*** Significant at $p \leq 0.01$; ** Significant at $p \leq 0.05$; * Significant at $p \leq 0.1$.

Using the economic proximity W^c, the Global Moran's I statistic detects positive spatial autocorrelation in calorie consumption for 118 countries. Overall, countries with highly calorific diets tend to be high-income nations and vice versa, countries with low calorie consumption are often low-income nations. While the global statistic provides a simple summary value for the entire map meaning that it informs if calorie consumption is randomly distributed across the space, it does not indicate where specific patterns (for example unexpectedly high/low values of calorie consumption) occur. In order to identify local variations in the strength of spatial autocorrelation across the sample, the local measure of spatial autocorrelation is needed.

In this section, the local Moran's I statistic is computed for each country using the spatial weight matrix W^c. For each country, the local statistic allows the computation of its similarity with its neighbours and to test its significance. It should be emphasised that the term 'neighbour' here refers to neither countries sharing the actual geographical boundaries nor countries located in the immediate vicinity but should be interpreted as countries with similar development level. Based on the value and significance of the local Moran's I index, a country can be classified to one of the following five groups:

(i) Countries with high calorie consumption associated with highly calorific *neighbour*s. These are labelled as "High-High" countries (also known as "hot spots").
(ii) Countries with low calorie consumption associated with low-calorie *neighbour*s. These are labelled as "Low-Low" countries (also known as "cold spots").
(iii) Countries with low calorie consumption associated with highly calorific *neighbours*. These are potential outliers and labelled as "Low-High" countries.
(iv) Countries with high calorie consumption associated with low-calorie *neighbour*s. These are potential outliers and labelled as "High-Low" countries.
(v) Countries with no significant local Moran's I statistic.

These five categories can be first identified from a scatter plot showing the observed calorie value against the averaged value of the neighbours (commonly referred to as *Moran scatterplot*). Once a significance level is set, local spatial autocorrelation values can be shown on a map (known as *LISA map*) to display the specific locations of hot spots, cold spots and other spatial phenomena. Figure 6.11 presents the Moran scatterplot and LISA map with regard to calorie consumption 1961-2013.

The Moran scatterplot shows the variable of interest (calories) for each country on the horizontal axis with the average values of its *neighbouring* countries on the vertical axis. Values on both axes are scaled to have mean zero and standard deviation of one. In this respect, values above 0 indicates the consumption level higher than the global average while values below 0 indicates the consumption level lower than the average. To recap, the label 'neighbouring' here reflects proximity in terms of economic development rather than geography, and two countries are said to be neighbours if their incomes are in the same quartile. To demonstrate different income levels, each country in the Moran scatterplot is represented by a point with the shape and the colour denoting the quartile that country belongs to. In ascending order, the colour red, green, blue and purple respectively corresponds to country in the 1st, 2nd, 3rd and 4th income quartile.

In general, the Moran scatterplot shows the impression that the higher the income the higher the calorie consumption. To illustrate, all countries in the 1st income quartile (shown by red circles) lie on the left side of the vertical line $x = 0$, meaning that the calorie consumption of the poorest countries is well below the global average. On the other hand, the richest countries (shown by purple plus signs) mostly lie on the right side of the vertical line $x = 0$, indicating a higher consumption level than the global average. The pattern for middle-income countries (shown by green triangles and blue squares) is less pronounced as a wider range of calorie figures is observed. If the positive relationship between income and diet is perfectly linear, the calorie content would progressively increase as a country proceeds to the higher income quartile. In that scenario, all countries in the 2nd income quartile (represented by green triangles) would be associated with higher values and therefore be positioned more towards the right side of the graph than countries in the 1st income quartile (represented by red circles). Nonetheless, this is hardly the case given the plot in Figure 6.11. Indeed, there is a wide spectrum of calorie consumption for each income level. For instance, a handful of richest countries shown in purple colour seem to divert themselves from the rest of the highest income group and are associated with a consumption level lower than the global average. So, in addition to the global tendency that higher income higher calorie consumption, some exceptions exist and countries within the same income quartile do not necessarily follow the same diet. Another noteworthy feature from the Moran scatterplot is that points of the same shape and colour tend to be arranged in a straight imaginary line and the four imaginary lines are not parallel to the horizontal axis but angle at a slope. This pattern implies that on average the calorie content of a country is not always in line with its *neighbours*. Again, two countries with similar level of development need not always exhibit similar level of calorie consumption. Overall, income is an important predictor for the (dis)similarities in diets globally, but other factors such as migration, heterogeneous consumer tastes and socio-cultural influences might play an underestimated role in explaining local deviations from the global patterns.

Returning to the Moran scatterplot, the horizontal and vertical axes naturally divide the graph into four quadrants. The North East quadrant belongs to countries which have similarly high level of calorie consumption as their neighbours (“High-High” countries). Countries in the South West quadrant have low level of calories and are surrounded by neighbours that also have below the average level of calorie consumption; hence, these are “Low-Low” countries. Both “High-High” and “Low-Low” quadrants correspond to positive spatial autocorrelation, and the former is referred to as *hot spots* while the latter represents *cold spots*. The North West quadrant belongs to “Low-High” countries whose levels of calories are below average but are associated with highly calorific neighbours. Finally, the South East quadrant includes “High-Low” countries whose consumption levels are above the average while being surrounded by neighbours with low levels of calories. Both quadrants correspond to negative spatial autocorrelation and represent potential *outliers* (in the sense that they are surrounded by neighbouring countries that are very dissimilar to them). In Figure 6.11, all countries from the 1st income quartile and the majority from the 2nd quartile belong to the "Low-Low" quadrant while most richer countries from the upper quartiles lie in the "High-High" category. The remaining countries occupy the atypical locations of the "Low-High" and "High-Low" quadrants. Overall, the Moran scatterplot reveals the predominance of "High-High" and "Low-Low" countries in the data, and as a result, the Global Moran's I statistic, being sort of an average of local values of spatial autocorrelation, detects positive spatial autocorrelation for the whole data set. Nevertheless, one should not jump into any conclusion too soon as the significance of the local spatial autocorrelation measure is not shown in the Moran scatterplot. In fact, the LISA map depicts these types of spatial relationship whilst distinguishing the significant autocorrelation values from the insignificant ones.

A LISA map illustrating different categories of countries based on the local Moran's I values is given in Figure 6.11. Hot spots ("High-High" countries) are shown in red and cold spots ("Low-Low" countries) in green. "High-Low" and "Low-High" countries are respectively shown in orange and yellow. Countries shown in grey are devoid of any spatial autocorrelation. Broadly speaking, the LISA map highlights 66 countries, equivalent to 56% of the sample, falling into the "High-High" and "Low-Low" quadrants. Agglomerations of hot spots exist across Europe, Northern America, and Oceania with wide stretches of the same red colour. While the notion of neighbourhood in this analysis is determined by the level of economic development, the fact that these countries aggregate geographically into actual regional territories reflects a relatively equal economic distribution in Europe, Northern America and Ocean. For these regions, more than likely nearby countries experience the same degree of economic development. Besides, other "High-High" countries are scattered in East Asia (South Korea, Japan), West Asia (Israel, Kuwait), and South America (Argentina). These countries are classified in the same group (High-High) even though they are not nearby each other in the usual geographical sense. The reason is that these are considered rich countries and their levels of calorie consumption are similar to those of other wealthy countries across Europe and Northern America. This grouping is therefore in line with the economic proximity measure adopted in this analysis (instead of the conventional geographical distance). On the other hand, there are some cold spots located in Africa and South Asia which are coloured in green. The LISA map also shows some atypical cases of calorific countries without spatial dependence with their neighbouring countries. Africa is an intriguing continent which is home to the only two significant "High-Low" (Morocco and Egypt) and "Low-High" countries (Botswana and Gabon). The message conveyed by these labels is that the calorie levels in Morocco and Egypt are on average higher than in other countries of the same income level (say India). By contrast, individuals in Botswana and Gabon – the two upper-middle-income countries – consume on average lower calories than their peers with comparable incomes in say Turkey. The existence of these "High-Low" and "Low-High" cases demonstrates that the association between income and calorie consumption is not always positive but negative for a small number of countries. This nuance highlights that the influence of income on diets is non-linear and that even among countries with similar level of economic development, socio-cultural differences can lead to heterogenous food consumption patterns.

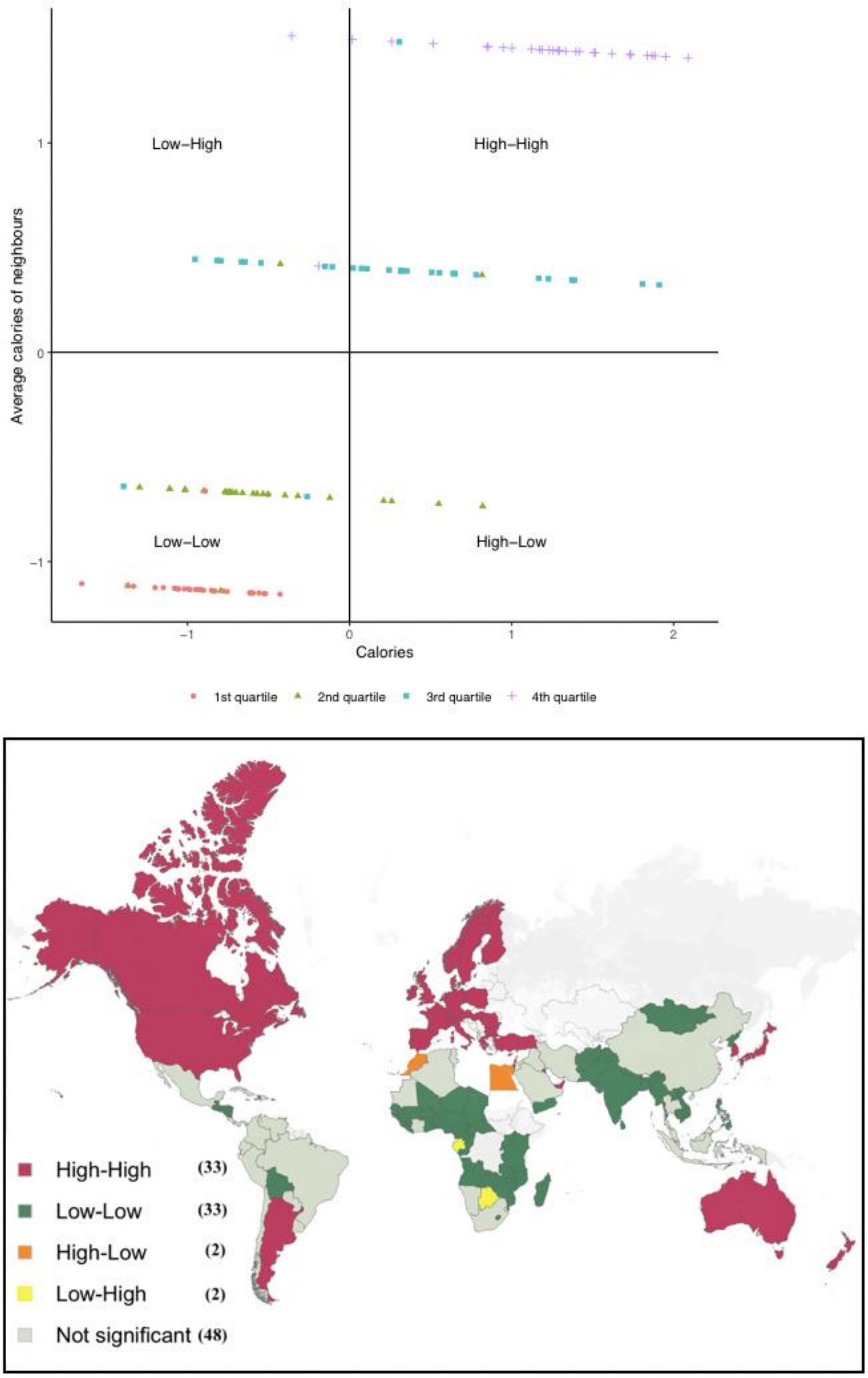


Figure 6.11 Moran Scatterplot (top) and LISA map (bottom) for average daily per capita calories 1961-2013.

Another notable feature from the LISA map is that the local Moran's I values for 48 countries (approximately 41% of the sample) are not statistically significant at 5% significance level. Income, even though can account for on average 56% of the similarities among national diets, is not the only predictor. For example, as shown in the LISA map the equator somewhat divides the globe into two halves so that low-calorie countries tend to be clustered around the tropics whereas highly calorific countries in places further towards the poles. Clearly, even after accounting for economic dependence the impact of climatic conditions remains somewhat profound. While it seems reasonable for countries further away from the tropics to have higher need for calories due to the colder temperature and countries near the tropics lower need due to the warmer temperature, this may not capture the whole picture. In fact, most of countries located around the tropics are developing countries where the availability of foods and fruits that are wild or home grown (not going through the marketplace) is larger than in Western/temperate countries. Such foods are not accounted for in the Food Balance Sheet data, and this could potentially explain why diets around the tropics are less calorific.

To sum up, as countries further develop and become richer, they tend to increase their calorie consumption. Broadly speaking, the degree of economic development determines the overall trend in diets (high- or low-calorie). The heterogeneity of country typologies based on the local Moran's I values implies that if convergence process exists it would be different between regions ("clubs") made up by say "High-High" versus "Low-Low" countries. The fact that a large portion of the sample is associated with insignificant spatial autocorrelation questions the role played by income. Nevertheless, it should be mentioned that the spatial metric used in this analysis is merely average income over the period 1970-2013. While the majority of countries have always belonged to a specific income category for example Canada in the richest group and Cambodia in the poorest group, for others the progress/regress in GDP per capita could be profound. For instance, China was associated with the 1st quantile group (low-income) in 1970 but the 3rd quantile group (upper-middle-income) in 2013. Consequently, its neighbourhood matrix could vary greatly between 1970 and 2013, and so does its local Moran's I value. Since the switching behaviour in the income quantile group is documented for 33 countries (28% of the sample), more than likely the insignificant autocorrelation figures observed in the LISA map are attributed to the time-varying nature of the associated neighbourhood matrix. Another important factor to consider is the distribution of income within each country. The use of an aggregate measure like national GDP per capita in this analysis may mask detailed patterns of differences in income levels across population subgroups. This distinction can be crucial for assessing within-country patterns of food consumption. As an illustration, though Vietnam is classified as a "Low-Low" country indicating the low-calorie characteristic of an average Vietnamese diet, such a label does not apply to some wealthy individuals with the adoption of the 'Western' dietary patterns. That said, even after taking into account income, individual attributes including demographic, psychometric, attitudinal and lifestyle factors have created persistent heterogeneity of consumption patterns within and between countries (Dubois *et al.* 2014). A discussion on these factors is referred to Section 2.2.2 and 2.5.2 (Chapter 2). In addition, migration (Atkin 2016) and cultural distance (De Sousa *et al.* 2018) might play an ineluctable role.

6.4.3 Results from the spatial beta convergence model

Unconditional spatial beta convergence analysis

The previous section reveals significant spatial dependence among countries in terms of national food consumption: the richer the country, the higher the calorie consumption. As such, the positive spatial autocorrelation in the average calorie data is likely to signal a similar kind of dependence when estimating the convergence rate and convergence level from a cross-sectional beta convergence model. As discussed in Section 6.2.2, ignoring the 'space', i.e. disregarding such a correlation in the regression analysis could hinder the OLS estimation results of the beta convergence model due to omitted variable bias. In this section, in order to identify the presence and type of spatial effects in the regression model the bottom-up approach described in Figure 6.6 is adopted. First, a non-spatial model of beta convergence in the following form is estimated by the OLS estimation method:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \beta\log\left(y_{i,t_0}\right) + \varepsilon_i \qquad (6.2)$$

where y_{i,t_0} and y_{i,t_0+T} are respectively the per capita daily calories of country *i* in the initial and final periods; T is the number of years; ε_i is the standard error term being independent and identically distributed with mean zero and variance σ_ε^2. A significantly negative β is indicative of beta convergence whereas a significantly positive β indicates divergence.

The next step is to detect the presence of spatial autocorrelation in the residuals of the non-spatial model (6.2) using the Global Moran's I statistics. If spatial autocorrelation exists, it is necessary to determine the source of the spatial autocorrelation as well as the alternative spatial model. To reiterate, there are two main sources of spatial dependence: (i) measurement errors due to omitted variables that are otherwise not crucial to the model; and (ii) the interaction of localities. The former, also known as 'nuisance dependence', is likely to occur and evident in most data sets of empirical studies while the latter is referred to as 'substantive form' of spatial autocorrelation. If the data shows spatial dependence of the nuisance form, the detected spatial dependence stems from country-specific factors (or *national effects*) rather than interaction effects between localities (or *spill-overs*). In that case, the Spatial Error Model (SEM) is more appropriate to describe the data, and the error terms in equation (6.2) are modelled using spatial moving average or spatial autoregressive process. If the spatial dependence is of substantive form, it means that spill-overs not only exist but are also an important determinant of the convergence process: the growth rate in one country is affected by the growth rate of its neighbours. This problem is resolved by applying the Spatial Lag Model (SAR) which adds a spatially lagged variable (i.e. the calorie level of neighbouring countries) on the right-hand side of equation (6.24).

How do we discriminate between these two spatial dependence effects? The LM procedures illustrated in Figure 6.6 can provide some guidance on model specification. If the LM-lag statistic is significant whilst the LM-error is not, then the SAR model is recommended. On the contrary, if the LM-error is significant while the LM-lag is not, then the SEM model should be selected. When both LM test statistics are highly significant, then the one with the higher robust LM test statistic is likely to be the correct specification.

Results of the regression model and relevant tests are displayed in Table 6.5. The Global Moran's I statistic is significantly positive (0.211), implying positive spatial autocorrelation. Being a global measure of spatial dependence, it gives neither conclusions about the source of spatial dependence nor guidance on which alternative spatial model is more appropriate, which are tasks of the LM tests. The significant LM-err statistic strongly points to the spatial dependence of nuisance form and advises the employment of the SEM model as the alternative to the original OLS model. Hence, a SEM model of the following form is to be estimated:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \beta\log\left(y_{i,t_0}\right) + \varepsilon_i \tag{6.24a}$$

$$\varepsilon_i = \lambda W^c \varepsilon_i + u_i \tag{6.24b}$$

where λ is the spatial autoregressive parameter quantifying the intensity of spatial autocorrelation between regression residuals, and W^c denotes the spatial weight matrix representing the economic contiguity.

To recap, an SEM model in equations (6.24a, b) can be estimated by Maximum Likelihood (ML), quasi-Maximum Likelihood, instrumental variables, or Generalised Moments (GM) methods. In this analysis, the GM procedure (Kelejian and Prucha 1999) is adopted and the results are presented in Column (2) of Table 6.5.

The GM estimator was originally motivated by the computational difficulties of the ML estimator which may not be computationally feasible if the sample size is moderate or large. A critical advantage of the GM estimator is simplicity of computation and that it ignores the Jacobian term, thus avoiding many problems related to matrix multiplication, matrix inversion, the computation of characteristic roots and/or Cholesky decomposition which are often involved in an ML procedure (Elhorst 2014). Another advantage of the GM estimator is that it does not rely on the assumption of normality of the disturbances u (Bell and Bockstael 2000). Nonetheless, it assumes that the disturbances u_i are independently and identically distributed for all i with zero mean and variance σ^2. The rationale behind the GM estimator is that Kelejian and Prucha (1999) use nonlinear least square to obtain a consistent generalised moment estimator for lambda λ so that the consistency of the resulting spatially weighted estimator is assured. That is to say the authors were not necessarily interested in the inference about lambda λ per se, but in its estimate as a way to obtain consistent estimators for beta β. Lambda therefore is considered a nuisance parameter whose only role is to provide a consistent estimator for the regression coefficients (Anselin 2003). Its significance cannot be assessed and in fact the authors do not provide an asymptotic variance for lambda[2].

Table 6.5 Unconditional beta convergence regression: Estimation results.

Models	(1) OLS	(2) SEM	(3) OLS	(4) SEM	(5) OLS	(6) SEM
	1961-2013	1961-2013	1961-1998	1961-1998	1999-2013	1999-2013
Intercept	0.089*** (0.009)	0.134*** (0.011)	0.076*** (0.014)	0.148*** (0.016)	0.133*** (0.017)	0.133*** (0.017)
log(initial calories)	-0.011*** (0.001)	-0.017*** (0.001)	-0.009*** (0.002)	-0.019*** (0.002)	-0.016*** (0.002)	-0.016*** (0.002)
Lambda λ	-	0.047 (0.262)	-	0.048 (0.384)	-	0.0009 (0.966)
Number of observations	118	118	118	118	118	118
Spatial diagnosis						
Moran's I (errors)	0.21***	-	0.19***	-	0.002 [0.37]	-
LM-lag	1.04 [0.31]	-	1.04 [0.31]	-	0.10 [0.75]	-
LM-error	73.90***	-	61.86***	-	0.01 [0.94]	-
Robust LM-lag	41.26***	-	45.26***	-	0.21 [0.65]	-
Robust LM-error	114.12***	-	106.09***	-	0.12 [0.73]	-
Convergence statistics						
Convergence speed per year (%)	1.63	4.00	1.15	3.18	1.84	1.85
Half-life (years)	42.64	17.30	60.05	21.81	37.59	37.52

Note: All test results are not significant unless indicated otherwise. Standard errors are inside the parentheses. [p-values] are reported for spatial diagnosis tests if the test statistic is not significant.

*** Significant at $p \leq 0.01$; ** Significant at $p \leq 0.05$; * Significant at $p \leq 0.1$.

[2] The output presented in Table 6.5 however reports standard error for lambda which is derived from the analytical method outlined in http://econweb.umd.edu/~prucha/STATPROG/OLS/desols. pdf.

As can be seen in Table 6.5, the lambda parameter λ is positive (0.047), confirming positive spatial autocorrelation among the error terms. More specifically, the detected spatial dependence is caused by country-specific factors (national effects) rather than spill-overs. These results suggest that the spatial effects among neighbouring countries are limited to error terms or unmodeled factors. In this respect, a random shock to a specific country will not only affect the growth rate of calorie consumption in the respective country but will also diffuse throughout the entire neighbourhood because of the spatial dependence of the error terms. Movements away from a steady state equilibrium induced by a shock are not restricted to the corresponding country but apply to a set of economically adjacent countries.

Turning to the coefficients estimated from the model (6.24a, b), the beta coefficient is significantly negative (-0.017), implying that countries with lower initial levels of calories tend to exhibit faster growth rates (the ‘catching-up’ effects). The associated annual convergence speed over the period is 4% and the half-life measure is 17 years. As compared to the OLS model, the consideration of the spatial error autocorrelation leads to a more robust beta convergence process. This finding thus shows that economically close countries tend to converge more rapidly to similar levels of calories and ignoring spatial interaction leads to biased estimates of the speed of convergence. While it has been shown that theoretically a spatial model would always yield a greater convergence speed as being augmented with neighbouring effects (Ahmad and Hall 2017), in practice such neighbouring effects can stem from the interaction between agricultural policies. For example, developing countries have tended to pursuit anti-agricultural and anti-trade policies by taxing their agricultural sectors (rather than subsidising them). Conversely, government policies in developed countries have been characterised by high levels of support and protection for agricultural sectors. This disarray in agricultural policies has exerted uniform impacts of over-production in high-income countries and under-production in more-needy developing countries (Anderson 2010; OECD 2019a).

In Figure 6.7, the coefficient variation shows a break in year 1998 and that the period 1999-2013 is characterised by a sharp sigma convergence process. The beta convergence was examined by splitting the data into two sub-periods. The overall results suggest a faster convergence process over the past decade, and when income dummies are included in the beta convergence equation neither the dummy variables nor their interaction terms with the initial level of calories prove to be significant in the latter period. These findings signal the less likelihood of a spatial dependence process among national diets since the Millennium. In order to test this formally, the spatial analysis is replicated for two sub-periods: 1961-1998 and 1999-2013.

Columns (3) of Table 6.5 reports the OLS estimation results for the former period. The Global Moran’s I statistic of 0.19 is highly significant and rejects the null hypothesis of uncorrelated errors. Next, the strongly significant LM-error statistic favours the SEM as the appropriate model to handle the spatial effects. Therefore, the detected spatial dependence is caused by omitted variables that otherwise are not crucial for the model. A SEM model as specified in equations (6.24a, b) is estimated by the GM estimation method and the coefficients are displayed in Column (4) of Table 6.5. The spatial autoregressive parameter λ is positive (0.048), suggesting positive spatial autocorrelation among error terms. Similar to the SEM estimation results for the whole period, the inclusion of spatial effects in the unconditional beta convergence model during the period 1961-1998 leads to a significant and more negative beta coefficient (-0.019 for SEM versus -0.009 for OLS). Again, a faster convergence process is revealed: the associated speed of convergence is 3.18% per year and the half-life is 22 years.

When the beta convergence model is investigated for the latter period 1999-2013 as shown in Column (5), however, the Global Moran's I statistic becomes insignificant, eliminating the possibility of spatial dependence. In addition, none of the LM tests or the robust variants appears to be significant. For the sake of comparison, estimations from a SEM model is provided in the last column of Table 6.5. As expected, the beta coefficient remains unchanged and is significantly negative (-0.016). Nonetheless, the lambda becomes very close to zero (0.0009), offering little evidence for the spatial interaction among the error terms. In this case, the rate of convergence obtained from a spatial model (SEM) is very similar to what obtained from a non-spatial model (OLS). Both results indicate that national diets are converging at the pace of approximately 1.8% per annum. Here, the less apparent role of spatial effects implies that the influence of other countries might still be strong in the latter period but not confined to countries within the proximate 'neighbourhood'. The acceleration of globalisation process accompanied by an increased number of bilateral and international trade agreements as well as the rise of supranational organisations and transnational companies has allowed individuals in developing countries to increasingly be exposed with and to adopt the eating styles from richer countries. This signals the 'Westernisation' transition of diets around the world, and as a result, dietary patterns that are once characterised by wealthier countries are no longer limited to the West.

So far, unconditional beta convergence analysis (with and without spatial effects) points to the 'catching-up' effects, i.e. countries with the initially low levels of calories exhibit higher growth rates of calories and therefore the calorie consumption across national borders would converge in the long run. However, a more difficult question is to decide whether countries converge to the same steady-state level of calories or if by virtue of the existence of distinct equilibria, convergence is rather relative than absolute. Given the slow absolute convergence speed (in the absence of spatial interaction) and the high relevance of spatial effects (due to omitted variables), it is unlikely that countries would converge to the same steady-state level. In that case, an analysis of conditional beta convergence is recommended.

Conditional spatial beta convergence analysis

Similar to the previous section, the bottom-up approach described in Figure 6.6 is adopted to identify the presence and type of spatial effects in the conditional beta convergence regression model. First, a non-spatial conditional growth model is defined as follows:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B + \beta\log\left(y_{i,t_0}\right) + \sum_{j=1}^{J}\pi_j Z_{ji} + \varepsilon_i \tag{6.23}$$

where y_{i,t_0} and y_{i,t_0+T} are respectively the per capita daily calories of country i in the initial and final periods; T is the number of years; Z_{ji} includes the set of J additional control variables which vary over countries; and ε_i is the standard error term being independent and identically distributed with mean zero and variance σ_ε^2. More specifically, Z_{ji} comprises the average percentage of arable land in the total land area, the average growth rate of income, the average percentage of urban population, and the average female participation rate in the labour force. B, β and $\pi_j (j = 1, \dots, 4)$ are the parameters to be estimated.

As can be seen from the results of the spatial diagnosis tests in Table 6.6, the Global Moran's I statistic is significantly positive, rejecting the null hypothesis of uncorrelated errors. Further, the significant LM-lag statistic points to the spatial lag model (SAR) as the appropriate model to handle the spatial effects. Thus, a spatial lag model in the case of conditional beta convergence is specified as follows:

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = \rho W^c \frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) + B + \beta\log\left(y_{i,t_0}\right) + \sum_{j=1}^{J}\pi_j Z_{ji} + \varepsilon_i \tag{6.25}$$

where ρ is the spatial autoregressive parameter quantifying the intensity of spatial autocorrelation between the dependent variable, and W^c is the spatial weight matrix representing the economic contiguity.

The model (6.25) is estimated by the Maximum Likelihood (ML) method and the results are displayed in the second column of Table 6.6. Overall, the spatial lag regression results are comparable with the non-spatial OLS model with a few notable exceptions. Among the structural indicators, the effect of a one percent increase in the proportion of arable land between 1961 and 2013 is lessened from a 0.4 per cent decrease to a 0.3 per cent decrease. Conversely, the effect of a one percent increase in the growth rate of GDP is heightened from a 11.1 per cent increase to a 12.1 per cent increase. Economic conditions become more important whereas agroecological conditions become less influential when controlling for the spatial dependence. For the remaining independent variables, the direction of the relationship as well as the magnitude of the effect remains the same as with the OLS model.

The rho parameter (ρ) represents the average influence of the calorie growth in neighbouring countries on the dependent variable. This parameter is statistically significant and indicates the presence of spatial dependence among countries in terms of spill-overs. Particularly, the negative sign of rho shows a negative spatial feedback effect. In other words, if a country is surrounded by countries with high growth, this negatively affects its own growth rate of calories.

Table 6.6 Conditional beta convergence regression: Estimation results.

Models	(1) OLS	(2) SAR	(3) OLS	(4) SEM	(5) SAR	(6) OLS
	1961-2013	1961-2013	1961-1998	1961-1998	1961-1998	1999-2013
Intercept	0.128*** (0.010)	0.134*** (0.010)	0.145*** (0.015)	0.151*** (0.015)	0.145*** (0.015)	0.070*** (0.026)
log(initial calories)	-0.017*** (0.001)	-0.017*** (0.001)	-0.019*** (0.002)	-0.019*** (0.002)	-0.018*** (0.002)	-0.016*** (0.003)
Arable land	0.004** (0.002)	0.003** (0.002)	0.005** (0.002)	0.005** (0.002)	0.005** (0.002)	8.7×10^{-5} (0.003)
Income growth	0.111* (0.057)	0.121** (0.053)	-0.004 (0.027)	-0.006 (0.026)	-0.007 (0.026)	0.770*** (0.139)
Urban population	0.008*** (0.001)	0.008*** (0.001)	0.011*** (0.002)	0.010*** (0.002)	0.011*** (0.002)	0.006** (0.003)
Female employment	-0.003** (0.001)	-0.003*** (0.001)	-0.005*** (0.002)	-0.005*** (0.002)	-0.006*** (0.002)	-0.0001 (0.002)
Rho ρ	-	-0.026*** (0.010)	-	-	-0.012*** (0.012)	-
Lambda λ	-	-	-	0.008 (0.779)	-	-
Number of observations	118	118	118	118	118	118
Maximum VIF	1.98	-	2.09	-	-	2.29
Spatial diagnosis						
Moran's I (errors)	0.009 [0.12]	-	0.025**	-	-	-0.007 [0.53]
LM-lag	3.074*	-	0.252 [0.61]	-	-	0.071 [0.79]
LM-error	0.149 [0.69]	-	1.107 [0.29]	-	-	0.077 [0.78]
Robust LM-lag	5.549**	-	3.151*	-	-	0.326 [0.57]
Robust LM-error	2.63 [0.11]	-	4.010**	-	-	0.332 [0.56]
Convergence speed per year (%)	3.99	4.40	3.18	3.41	3.12	1.82
Half-life (years)	17.37	15.74	21.82	20.31	22.21	38.09

Note: All test results are not significant unless indicated otherwise. Standard errors are inside the parentheses. [p-values] are reported for spatial diagnosis tests if the test statistic is not significant.
*** Significant at $p \leq 0.01$; ** Significant at $p \leq 0.05$; * Significant at $p \leq 0.1$.

It should be noted that the inclusion of spatially lagged variables in the regression model complicates the interpretation of the coefficients and the coefficients for independent variables cannot be explained in the same manner as the OLS or the SEM. As a matter of fact, "past studies using spatial regression models frequently interpreted the model estimates incorrectly" (LeSage and Pace 2014, p.77). The reason is that a change in the initial level of calories in country i not only affects the calorie growth in country i ("direct impact"), but it also influences the growth rate of calories in the surrounding countries. Since the calorie consumption of the neighbours has changed, this in turn affects the growth rate of calories in country i ("indirect impact"). In order to appropriately understand the spatial relationships, the coefficient estimates of the spatial lag model are decomposed into direct and indirect impacts and reported in Table 6.7. The estimated coefficients of the direct effects are used to compute the speed of convergence and the half-life measure.

Table 6.7 Impact calculations for spatial lag regression estimates.

Variables	SAR (1961-2013)			SAR (1961-1998)		
	Direct impacts	Indirect impacts	Total impacts	Direct impacts	Indirect impacts	Total impacts
log(initial calories)	-0.017 ***	0.007 ***	-0.010 ***	-0.019 ***	0.005	-0.014 ***
Arable land	0.003 **	-0.001 *	0.002 *	0.005 **	-0.001	0.004
Income growth	0.122 **	-0.053 *	0.069 *	-0.007	0.002	-0.005
Urban population	0.007 ***	-0.003 ***	0.004 ***	0.011 ***	-0.003	0.008 ***
Female employment	-0.003 **	0.002 *	-0.001 **	-0.005 ***	0.001	-0.004 **

Note: Significance is established using a simulation approach (LeSage and Pace 2009).
*** Significant at $p \leq 0.01$; ** Significant at $p \leq 0.05$; * Significant at $p \leq 0.1$.

Direct impact expresses the marginal effect of a one percent change in the independent variable on the dependent variable in the same country. For example, if country i increases the share of arable land, what will be the influence on the growth rate of calorie in country i? Indirect impact represents the marginal effect on the dependent variable of the country itself due to a one percent change in the independent variable of all neighbouring countries. This measure indicates the effects of all other countries j raising the share of arable land on the growth rate of calories in country i. Total impact is the sum of direct and indirect impact. If all countries increase the share of arable land, what will be the average effect on the growth rate of calories in a typical country?

It can be seen from Table 6.7 that for each predictor, the indirect impact is smaller in magnitude than the direct impact suggesting little spill-over effects. Still, a couple of significant indirect effects are worth mentioning. Regarding the initial level of calories, a positive spill-over effect is detected. Neighbouring countries with more calorific diets can spread their eating habits for example via food trades, international marketing campaigns and increased cultural exchange, consequently exerting a positive effect on the calorie growth in one's own country. A negative spill-over effect is reported for the share of urban population, suggesting that being associated with more urbanised neighbouring countries reduces the calorie growth rate of one's own country. In nations with more rapid urbanisation, there are increasing demands for calorie-dense foods such as meat, dairy products and vegetable oils, and this leads to a rise in the price of these non-staple products, making them less affordable and less consumed in one's own country. Another feature from Table 6.7 is that the total impacts comprise mostly the direct impacts which are more or less comparable with the SAR coefficients reported in Table 6.6.

Next, the analysis of conditional spatial beta convergence is replicated for two sub-periods: 1961-1998 and 1999-2013. For the former period, the significant Moran's I index offers evidence for the spatial dependence while both of the robust LM statistics are significant. For the sake of comparison, both spatial lag and spatial error models are estimated, and the results are presented in Columns (4) and (5) of Table 6.6. Broadly speaking, the estimates from the SEM are not dramatically different from those reported for the OLS. Therefore, it's not surprising that the rate of convergence and the half-life measure do not vary much, being 3.2% and 22 years respectively. The lambda parameter is negligible (0.008) and statistically insignificant, indicating no spatial dependence among the error terms.

For the SAR model, the impacts of controlling for spatial dependence are apparent in the initial level of calories and the female participation in the labour force. The former exerts a smaller magnitude of effects than in the OLS model and the latter exerts a larger magnitude. The rho parameter is significantly negative (-0.012) and hence points to negative spill-over effects. To better understand the influence of spatial dependence, the SAR estimates are disentangled into direct and indirect effects. Table 6.7 shows that none of the indirect effects is statistically significant, suggesting that changes in the structural conditions in neighbouring countries do not affect the calorie growth of one's own country. Most of the direct impacts are actually comparable with the SAR estimates in Table 6.7. Two notable exceptions are the initial level of calories and the female employment. A one percent increase in the former is associated with a 0.19 percent decrease in the calorie growth rate, which represents a larger effect than the 0.18 percent decrease predicted by the SAR coefficient. A smaller impact is observed for the latter: there is a 0.5% reduction in the growth rate of calories per one percent increase in the female employment rate – which is less than the 0.6% decrease implied by the SAR coefficient. In terms of convergence statistics, the SAR indicates a slightly slower convergence process with the annual speed of 3.1%. This slowdown could certainly be attributed to the negative spill-over effects.

The OLS estimations for the conditional beta convergence in the latter period 1999-2013 are presented in the last column of Table 6.7. Spatial diagnosis tests are applied and the results of both Moran's I statistic and LM tests are not statistically significant. Similar to the result from the unconditional beta convergence model, spatial dependence is not detected in the latter period 1999-2013. The absence of spatial effects is somewhat predicted since the conditional beta convergence specification incorporates a range of structural indicators controlling the impact of omitted variables which otherwise could ignite interactions among countries.

In summary, the conditional beta convergence analysis of calorie growth reveals a more robust convergence process over the past half a century at an annual rate equivalent to that from the spatial model of unconditional beta convergence. These findings confirm that countries are open to a range of demographic and socio-economic flows and exhibit spatial proximity in dietary convergence. With the liberalisation of trade, the diffusion of technologies and the expansion of global mass media, individuals in less developed countries are continually exposed to the heavily advertised dietary habits and eating behaviours related to the 'Western' diet typical of those living in developed countries. Even after considering these structural conditions, there is some evidence for the negative spill-over effects which might imply behavioural changes due to the adverse health consequences of the highly calorific diet.

While agroecological, social, economic and demographic factors are proved to influence the growth rate of calories, rising income has been the main driver for rising calorie consumption over the past 15 years. This finding has crucial policymaking implications especially when a measure of economic proximity is considered. During the process of economic development, low-income countries may direct their attention and efforts towards improving their infrastructure characteristics to a level similar to that of high-income countries. Since the convergence speed is augmented with spatial effects, two countries with similar characteristics are expected to have greater spatial interaction and would eventually converge to similar levels of calories and therefore worsening diets. Although the economic gap between poor and rich countries might seem too big to be bridged soon, the half-life signals that it will not be long until the world will be drowning in unhealthy diets.

6.4.4 Testing the exogeneity of income

An important concern in the estimation of calorie-income relationship is the potential endogeneity of income and this issue has been raised in many empirical studies (see, among others, Bouis and Haddad 1992; Abdulai and Aubert 2004; Ogundari and Abdulai 2013; Zhou and Yu 2015; De Sousa *et al.* 2018; Trinh Thi *et al.* 2018a). Particular attention is paid to the endogeneity due to simultaneity bias as the direction of causation between calorie consumption/availability and income occurs in both directions. On the one hand, it has been long established in the literature that nutritional status is determined by income, and therefore undernutrition and hunger which are assumed to be related to economic underdevelopment in developing countries could be alleviated by the means of economic growth (Abdulai and Aubert 2004). This idea is conveyed in the Engel's law, which states that individuals tend to increase calorie consumption as income rises, but the marginal growth rate would reduce when the calorie intake reaches the saturation point (Skoufias *et al.* 2011). On the other hand, several researchers argue that unemployment and poverty could be the result of not having enough to eat or poor diet (Subramanian and Deaton 1996). According to the efficiency wage hypothesis (Stiglitz 1976), higher intakes of calories/nutrients lead to higher productivity of workers thanks to improved health status and this can contribute to better wages and higher wealth. Such a reverse causality can be a source of endogeneity related to using income as the explanatory variable in the conditional beta convergence estimation and could cause biased estimates of beta coefficient. Therefore, it is of crucial importance to acknowledge the causal link between dietary intake (health status) and economic growth especially in poor countries (Well 2007) and that income growth is not a strictly exogenous contributor to better diets (Traill *et al.* 2014).

From the econometrics perspective, the error term must be unrelated to the regressors, or $E(\varepsilon|x) = 0$, in order for an OLS estimation to give consistent estimators. If this assumption is violated, the endogeneity issue arises. A common approach to deal with endogeneity in the empirical literature is using instrumental variables. The instrumental variable, φ, is chosen in such a way that it needs to be correlated with the endogenous variable x but uncorrelated with the error term, $E(\varepsilon|\varphi) = 0$ (Cameron and Trivedi 2009). In the context of estimating the relationship between income and calorie consumption, some instrumental variables have been proposed in previous studies, including non-food expenditure (Subramanian and Deaton 1996; Trinh Thi *et al.* 2018a) and rainfall variation (Mangyo 2008). In this research, the share of non-food expenditure is employed as an instrumental variable for income in the conditional beta convergence specification.

To this end, a conditional beta convergence model for a subset of 90 countries during the period 1990-2010 are examined. The smaller size of sample data as well as shorter time span is due to the availability of expenditure data. The share of non-food expenditure is derived as:

$$nf_{it} = 1 - f_{it} \tag{6.26}$$

where nf_{it} and f_{it} refer to the share of non-food expenditure and food expenditure respectively for country *i* in year *t*. Data on the share of food consumption expenditure in total consumption expenditure are retrieved from the FAO Statistics Household Survey Database, International Labour Organisation and country publications (FAO 2017a). Food consumption expenditure refers to the monetary value of acquired food, purchased and non-purchased, including non-alcoholic and alcoholic beverages as well as food expenses on away from home consumption such as in bars, restaurants, canteens, and street vendors. Total consumption expenditure refers to the monetary value of acquired goods for consumption, food and non-food items, consumed by members of household.

Results from using instrumental variable (IV) and OLS regressions are presented in Table 6.8. The sign of all explanatory variables is similar in both estimations; however, the significance of many variables differs. Notably, the beta coefficient for log(initial calories) remains the same, and as a result, the IV estimation leads to a similar conditional convergence process with more or less the same convergence speed and half-life measure as the convergence statistics obtained from the OLS approach. Nonetheless, instrumenting appears to attenuate and remove the effect of income growth, arable land and urbanisation on calorie growth.

The IV approach assumes that income growth is endogenous; yet, if income growth is in fact exogeneous, the OLS estimates would be more efficient. In order to test for the exogeneity of income growth, the Durbin-Wu-Hausman test with the null hypothesis that income growth is exogeneous is utilised and the test statistics (Durbin 1954; Wu 1974; Hausman 1978) are reported in Table 6.9. The difference between the Durbin and Wu-Hausman tests of endogeneity is that the former uses an estimate of the error term's variance based on the model assuming the variables being tested are exogeneous while the latter uses an estimate of the error variance based on the model assuming the variables being tested are endogenous. Under the null hypothesis that the variables being tested are exogenous, both estimates of the error variance are consistent (StataCorp 2019). As can be seen in Table 6.9, both test statistics are highly insignificant. The associated large *p*-values indicate that the hypothesis cannot be rejected, and one cannot reject the exogeneity of income growth in the conditional beta convergence model. Therefore, the OLS estimates are more efficient.

Table 6.8 Comparing regression results from OLS and IV estimations.

Models	(1) OLS	(2) IV
Intercept	0.173*** (0.023)	0.176*** (0.030)
log(initial calories)	-0.023*** (0.003)	-0.023*** (0.003)
Arable land	0.007** (0.003)	0.006 (0.004)
Income growth	0.096*** (0.024)	0.001 (0.597)
Urban population	0.011*** (0.003)	0.009 (0.012)
Female employment	0.003 (0.003)	0.003 (0.005)
Number of observations	90	90
Convergence speed per year (%)	3.06	3.04
Half-life (years)	22.69	22.81

Note: All test results are not significant unless indicated otherwise. Standard errors are inside the parentheses. *** Significant at $p \leq 0.01$; ** Significant at $p \leq 0.05$; * Significant at $p \leq 0.1$.

Table 6.9 Exogeneity test results.

	Test statistic	*p*-value
Durbin	0.030	0.862
Wu-Hausman	0.028	0.868

6.5 Chapter conclusion

This chapter aims to shed a light into the convergence in global diets using the per capita daily calories available for consumption for 118 countries over the period 1961-2013. To this aim, sigma and beta convergence methodologies are investigated.

The narrowing dispersion between national calorie availability suggests sigma convergence particularly over the last two decades. Unconditional beta convergence is confirmed and countries with lower levels of initial calories tend to exhibit higher growth rates (the 'catching-up' effect). Dummy variables representing different income levels are incorporated into the model and the results highlight income to be an important determinant at the early stage of convergence. In addition, low-income countries have been converging at the fastest pace, and the convergence rate reduces as income rises. In order to account for different structural conditions between countries, a range of agroecological, socio-economic, and demographic variables are included in the conditional beta convergence specification. While agroecological, social, economic and demographic factors are proved to influence the growth rate of calories, rising income has been the major driver for rising calorie consumption over the past 15 years.

A significant contribution of the convergence analysis in this study is the consideration of a spatial dimension in the traditional beta convergence model. This is an innovative approach to examine the role of space as a contextual factor for dietary behaviour. Three different proxies for spatial relationship are employed: (i) geographical distance, (ii) geographical contiguity, (iii) economic contiguity. The proposal of income (average GDP/capita) as the proximity measure emphasises income level (rather than geographical closeness) that is driving the similarities in diets observed worldwide. Results from the Global Moran's I statistics reveal that economic proximity is the only to yield significant (and positive) spatial autocorrelation, i.e. countries with similar income level tend to have similar diets. Economic proximity is hence considered in the spatial beta convergence testing.

In the unconditional beta convergence specification, there is evidence for the spatial dependence stemming from country-specific factors (national effects) rather than spill-overs. The spatial error model (SEM) is estimated and results point to a faster absolute beta convergence. Thus, ignoring the spatial relationship underestimates the convergence dynamics. In the conditional spatial beta convergence model, even after controlling for the structural indicators, evidence suggests the negative spill-over effects which might imply behavioural changes due to the adverse health consequences of the highly calorific diet.

This study is not without limitations. First, the Food Balance Sheet data should be interpreted as *food available for human consumption* rather than *food consumption* as food waste is not accounted for. This kind of apparent consumption data tends to mask other issues such as hunger and undernutrition that often coexist with overnutrition. Second, other controls could be added to the convergence model, and the club convergence analysis could be conducted with different country groupings (for example, based on population size or language spoken). Another limitation is that food prices are not considered in the analysis. While income is a significant determinant for calorie consumption, without the cost of food, it is impossible to know the extent to which individuals substitute one food for another. However, the absence of historical data on relative food prices corresponding to food categories listed in the FBS is a great barrier.

Chapter 7

What are the world's diets? Identifying common trends of food consumption around the world

7.1 Chapter introduction

Motivated by the convergence phenomenon in economics that poorer economies tend to grow faster than richer economies (the 'catching-up' effect), the empirical analysis in Chapter 6 applied the convergence testing frameworks in the topic of food consumption.

Dietary convergence implies that food consumption patterns across countries are becoming more similar. In assessing the similarity in diets across national borders, a large number of previous studies utilise cluster analysis to identify which countries naturally group together in terms of food budgets (Bertail and Caillavet 2008; Erbe Healy 2014; Staudigel and Schröck 2015) or food consumption behaviours (Gil *et al.* 1995; Balanza *et al.* 2007; Di Lascio and Disegna 2017), and how this grouping evolves over time (Walthouwer *et al.* 2014). In fact, time series data on food consumption/dietary intakes are abundant; nonetheless, previous researchers either merge time series into one large set of static data (for example, Blandford 1984; Staudigel and Schröck 2015; Sadowski 2019), or apply clustering algorithms on discrete time periods comparing the results between a baseline and a follow-up period (usually the first and last year) (for example, Gil *et al.* 1995; Di Lascio and Disegna 2017). In both cases, the clustering task is not performed on the whole set of time sequences and the time dependent nature of the data is not appropriately addressed. In order to fill in this gap in the literature, this research first employs an innovative copula-based time series fuzzy clustering algorithm. The copula function captures the dependence among time series and the fuzzy logic allows a country to belong to multiple clusters with different degrees of membership. Fuzzy clustering is an attractive method as it allows the possibility that individuals within a country do not eat the same diet and therefore several diets/dietary trends coexist within a single country. By grouping countries into clusters of relatively homogeneous trajectories of caloric consumption, it is possible to assess which dietary characteristics most resemble each cluster, how healthiness measures differ between clusters, and how these have changed over time.

Although such an analysis addresses the time dependent nature of the data, it neglects the spatial dimension. Since dietary data are often cross-sectional or longitudinal, the commonly used clustering methods in the earlier literature are those for static data and time series (Section 3.7, Chapter 3). In this research, the FBS data however are characterised by both temporal and spatial components. To a great extent, the countries under examination can be considered spatial units. When dealing with this type of data, *spatial clustering* techniques are required (refer to Section 3.5, Chapter 3). To recap, spatial clustering describes the situation in which cluster membership is constrained by some external information on the spatial relationship among units (often contiguity in space) so that units belonging to a cluster are not only similar to each other but also required to be contiguous. Unless the clustering method is explicitly spatial, the geographic relevance might not be sufficiently accounted for (Grubesic *et al.* 2014). Taking an example, while the dietary patterns of Germany, Austria, Canada and the United States are largely considered 'Western' style, several aspects of diets in Germany tend to be more comparable to Austria and likewise Canada to the United States due to the closer geographical distance and consequently the more similar corresponding environmental characteristics.

In food economics, spatial clustering has been employed to track health-related outcomes and previous studies mainly focus on clustering obesity prevalence (Gartner *et al.* 2016; Hughey *et al.* 2018; Qiu *et al.* 2020) or food insecurity (Kim *et al.* 2016; Tomita *et al.* 2020). Even though earlier authors realise that communities at risk of unhealthy food intakes could be spatially clustered (Austin *et al.* 2005), the literature on spatial clustering of dietary patterns is thin (Dekker *et al.* 2017; Tamura *et al.* 2017). In fact, adding a spatial dimension in the cluster analysis of food consumption patterns is an innovative way to examine the role of space as a contextual factor for dietary behaviour and the findings could lend support to the design and implementation of place-based policy interventions that target communities at risk of worsening diets (Leonard *et al.* 2018). Thus, the implications for intervention are profound since policy solutions for improving food insecurity and health are geographic in nature.

In principle, the relationships between environment, food consumption, and health are embedded in a spatial context. Nonetheless, the majority of previous studies have not taken spatial relationships into consideration. Thus motivated, this study significantly adds to the nascent literature of this research area. In the second part of the empirical analysis in this chapter, similarities in the evolution of global diets are captured in the light of an innovative Copula-based Fuzzy K-Medoids Space-Time clustering algorithm (Disegna *et al.* 2017). This cluster analysis has the merit of enabling the inclusion of spatial information into the clustering procedure dealing with both the spatial and temporal dimensions observed for data in reality. Returning to the earlier example of four countries, if the innovative copula-based fuzzy time series cluster algorithm classifies them in a cluster denoting the 'Western' dietary patterns, the space-time clustering algorithm would identify two clusters: one including Germany and Austria and the other Canada and the US. The purpose is to form clusters that are both data-coherent (in temporal dimension) and spatially coherent. In this respect, it would be possible to investigate the differing environments between two groups which otherwise would be masked in analysing the aggregate 'Western' diet cluster.

The rest of this chapter is organised as follows. Section 7.2 explains the innovative time series and space-time clustering algorithms. Section 7.3 presents the data, Section 7.4 discusses the empirical results and Section 7.5 concludes.

7.2 Methodology

7.2.1 The copula-based fuzzy time series clustering algorithm

The aim of this chapter is to derive groups of countries with similar patterns of food consumption. *Cluster analysis* is a data-driven technique that is well-suited for this purpose. Chapter 3 provides a taxonomy of clustering techniques and Section 3.4 focuses on clustering time series – the type of data employed in this empirical analysis. This section delves further into details of an innovative time series fuzzy clustering algorithm (Disegna *et al.* 2017) which belongs to the copula-based category. As pointed out in Section 3.4.5 (Chapter 3), the copula function offers more appeals than the traditional correlation coefficient in quantifying the dependence (or co-movement) between time series, particularly if the dependence structure is non-linear or asymmetric.

The starting point of the cluster analysis is represented by an $(N \times T)$ data matrix X defined as:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1T} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NT} \end{bmatrix} \tag{7.1}$$

where x_{it} is the value of the i-th unit $(i = 1, 2, \dots, N)$ at the t-th time period $(t = 1, 2, \dots, T)$. Hence, the i-th row of the data matrix X represents the univariate time series of the i-th unit. Here, cluster analysis aims to classify N rows of the data matrix X into K groups based on the behaviour of the time trajectories over T period so as to minimise the intra-cluster dissimilarity and maximise the inter-cluster dissimilarity. First, one needs to define an appropriate dissimilarity measure.

The copula-based dissimilarity measure between any pair of rows x_{it} and x_{jt} $(i, j = 1, \dots, N$ and $i \neq j)$ can be defined as:

$$d_{ij} = f(\|M - C_{ij}\|) \tag{7.2}$$

where $\|.\|$ is the L[2] norm, f is a convenient real-valued function, the copula C_{ij} expresses the dependence between time series of units i and j, M is the Frechet upper-bound copula: $M(u,v) = \min(u,v)$, which is the *maximal degree of similarity* (or *comonotonicity*) among time series. Based on equation (6.8), if time series of units i and j are exactly co-monotone, their dissimilarity (d_{ij}) is equal to zero. It is noted that the copula C_{ij} in equation (6.8) can be estimated either parametrically or non-parametrically.

Having established a suitable dissimilarity measure, the next step in cluster analysis is to apply a suitable clustering method. Overall, clustering procedures can be either hard or soft. In *hard* (or *crisp*) *clustering*, each unit can only belong to one cluster, and the cluster membership is either 0 or 1. In *soft* (or *fuzzy*) *clustering*, a unit, however, can belong to multiple clusters with varying degrees of membership between 0 and 1. In this analysis, the fuzzy approach is adopted to reflect the uncertainty that arises from assigning the (time series) units to different clusters. This kind of uncertainty is proxied by the membership proportion obtained from the clustering results. Other benefits of fuzzy clustering over crisp methods in the context of time series clustering can be referred to Section 3.4.2 (Chapter 3).

In the fuzzy clustering literature, there are two main clustering methods: Fuzzy K-Means and Fuzzy K-Medoids (see Section 3.3.2, Chapter 3). In this analysis, the latter is chosen due to its two major advantages. First, the prototypes obtained through the algorithm are actually observed time series (known as 'medoids' in the clustering literature) instead of fictitious average series ('centroids'). This allows to characterise the obtained clusters by the exemplar time trajectories - a feature that is appealing for policy targeting purposes (D'Urso *et al.* 2019a). Second, Fuzzy K-Medoids prove to be more robust to the presence of noise and outliers in the data than Fuzzy K-Means because a medoid is less affected by outliers or extreme values than a centroid (Kaufman and Rousseeuw 2009; García-Escudero *et al.* 2010; Maharaj *et al.* 2019).

The Fuzzy K-Medoids approach when combined with the copula-based dissimilarity measure leads to the novel Copula-based Fuzzy K-Medoids time series clustering algorithm. The objective function of this algorithm can be formalised as follows:

$$\begin{cases} min\ \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}^{m} d_{ik}(x_i, x_k) = \sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}^{m} f(\|M - C_{ik}\|) \\ \sum_{k=1}^{K} u_{ik} = 1,\ u_{ik} \geq 0 \end{cases} \quad (7.3)$$

where x_i and x_k are respectively the time series of the *i*-th unit and the medoid time series of the *k*-th cluster; u_{ik} denotes the membership degree of the *i*-th unit in the *k*-th cluster $(k = 1, 2, \ldots, K)$; $m > 1$ is the fuzziness parameter to be set by the user; $d_{ik}(.,.)$ is the dissimilarity measure between time series of the *i*-th unit and the *k*-th medoid.

In simple terms, the objective function (7.3) implies that the clustering algorithm aims to partition N time series into K clusters so that the dissimilarity between time series and the representative of the cluster they belong to is minimal.

Some remarks regarding the clustering algorithm

Remark 1: Data pre-filtering

Since time series data can exhibit a variety of patterns, it is helpful to split a time series into different components, each represents an underlying category of pattern. In general, an observed time series X_t can be decomposed into:

- T_t, the *trend* component, which reflects the slow and long-run evolution ("secular variation"), for example, a continued increasing or decreasing direction over time.
- C_t, the *cycle* component, which represents a regular repetition of the same pattern in the long run. Commonly, the trend and cycle components are known together as the trend-cycle (or simply trend) component.
- S_t, the *seasonality* component, which represents a repeated pattern that occurs every unit of time (no longer than a year) in the short run (for example quarterly, monthly or weekly).
- R_t, the *random* component (or "noise"), which reflects random, irregular, unpredicted influences, for example turning points and unexpected occurrences (maybe due to war or natural disaster). It is considered as the residuals of the time series after other components have been extracted.

Therefore, a time series can be thought of as a function of these components: $X_t = f(T_t, C_t, S_t, R_t)$. To help improve the understanding of time series, these components are often extracted and examined in isolation.

In Disegna *et al.* (2017), the authors recommend undertaking a data pre-filtering step before running the clustering algorithm. The observed time series are decomposed and the residual series (after removing trend and seasonality) serve as input for the clustering procedure. The main purpose of this pre-filtering step is to remove the effects of heteroscedasticity and autocorrelation in the copula estimation of the dependence structure (Durante *et al.* 2014, 2015). The authors illustrate the application of the clustering algorithm in detecting common behaviours of tourist flows. It is well-known that time series data in tourism are usually characterised by a strong common seasonality component (Gil-Alana *et al.* 2020). Unless the objective is to derive clusters of time series with similar seasonal patterns, performing the clustering algorithm on the original data can mislead the clustering results owing to the overwhelming influence of seasonality.

In this analysis, the clustering algorithm is applied on food consumption data (or more precisely food availability). The data under consideration are annual time series, hence there is no seasonality and the observed time series can be split into two components: trend (T_t) and random (R_t). So, a decision to be made here is whether to apply the clustering algorithm on the residual series R_t (after removing the trend) or the observed series X_t (including the trend). The existing literature does not offer much guidance on this and to the best of the author's knowledge to date no prior study has attempted to perform cluster analysis on the whole sequences of the FAO food availability. The interpretation of clustering results would differ depending on which clustering variable is employed.

In order to capture the twin nature of trend and fluctuation present in food consumption data, the clustering algorithm will be performed in two scenarios when the clustering variables are the observed series X_t and the detrended residual series R_t. It should be noted that due to differing inputs the number of clusters identified in two scenarios is not necessarily the same. Regarding the clustering results, some distinctions need to be recognised. Using the original time series, the clustering algorithm relies on information related to the level, trend and variability of calorie consumption. Since trend is the more influential component (than random), the obtained clusters would reflect different patterns of evolution in calorie consumption. This analysis is henceforth denoted as 'Trend analysis'. When the original series are detrended, the systematic information on the continued variability of the series, including direction (upward/downward) and speed of change, is removed. The remainder is the random patterns characterising the unpredicted and irregular variability of the series ("shocks" due to for example war, natural disaster, economic crisis or political instability). Utilising such information, the clustering algorithm would group together countries exhibiting similar deviations from the trend (sudden increase/decrease) in calorie consumption and is hereafter referred to as 'Fluctuation analysis'. Employing de-trended series, 'Fluctuation analysis' could help to uncover other patterns in data that might be masked by the trends.

To highlight this, the differences between two analyses are illustrated in Figure 7.1 using simulated data. Considering four annual time series X_{it} $(i = 1, \dots, 4)$, each time series can be split into two components: trend $(T_{it}, i = 1, \dots, 4)$ and random $(R_{it}, i = 1, \dots, 4)$ so that $X_{it} = T_{it} + R_{it}$. The time index *t* runs from 1 to 30. The trend lines are represented by quadratic trends as follows:

$$T_{1t} = -9.7 - 0.01t + 0.02t^2 \quad (7.4a)$$

$$T_{2t} = -9.95 - 2t + 0.06t^2 \quad (7.4b)$$

$$T_{3t} = -4.7 - 0.01t + 0.02t^2 \quad (7.4c)$$

$$T_{4t} = -14.95 - 2t + 0.06t^2 \quad (7.4d)$$

The random components of the first two series follow an ARMA model (1,1) and those of the other two series follow an ARMA model (0,1), with the parameters given as:

$$R_{1t} = 0.35R_{1t-1} + \varepsilon_t + 0.4\varepsilon_{t-1} \quad (7.5a)$$

$$R_{2t} = 0.01R_{1t-1} + \varepsilon_t + 0.8\varepsilon_{t-1} \quad (7.5b)$$

$$R_{3t} = \varepsilon'_t + 0.8\varepsilon'_{t-1} \quad (7.5c)$$

$$R_{4t} = \varepsilon'_t + 0.01\varepsilon'_{t-1} \quad (7.5d)$$

where ε_t and ε'_t are white noise processes being identically and independently distributed with normal distribution.

The copula-based time series fuzzy clustering algorithm is performed on the original data (X_{it}) and the pre-filtered data (R_{it}). The results are displayed in Figure 7.1. Time series of the same group are denoted by the same colour. 'Trend analysis' classifies the observed series into two groups (X_1, X_3) and (X_2, X_4) as shown in the top panel. Scrutinising further into separate components, such a classification is largely driven by the existence of common trends. By construction in equation (7.4a, c), the trend lines of the first and third series have the same shape with the latter being shifted up by 5. Likewise, the trend lines of the second and fourth series have an identical shape but with a gap of 5. The commonality of the trends is revealed in the middle panel of Figure 7.1 where T_1 and T_3 are represented by a somewhat almost monotonic increase while T_2 and T_4 by a parabolic U-shaped line. This describes how the 'Trend analysis' identifies different clusters – by searching for common trends in the historical series of calorie consumption, and as a result a cluster for example might be characterised by monotonic rising calories, a cluster by a quadratic rising trend whilst another by a decreasing trend. On the other hand, 'Fluctuation analysis' categories the four simulated series into two groups (X_1, X_2) and (X_3, X_4) according to the similar behaviour of random series (R_1, R_2) and (R_3, R_4). The lower panel of Figure 7.1 shows that large/small values of one (random) series at a given time tends to be associated with large/small values of the other (random) series at the same time. For example, at time 25 although all four series exhibit a strong rising trend there is a negative deviation from the trend observed for X_1 and X_2 (denoted by a trough in the corresponding random series) but a positive deviation from the trend observed for X_3 and X_4 (denoted by a peak in the corresponding random series). So, even though the 'Trend analysis' would group together two countries sharing a constant rising trend in the calorie consumption, the 'Fluctuation analysis' might classify them into two different clusters unless common deviations from the trend exist.

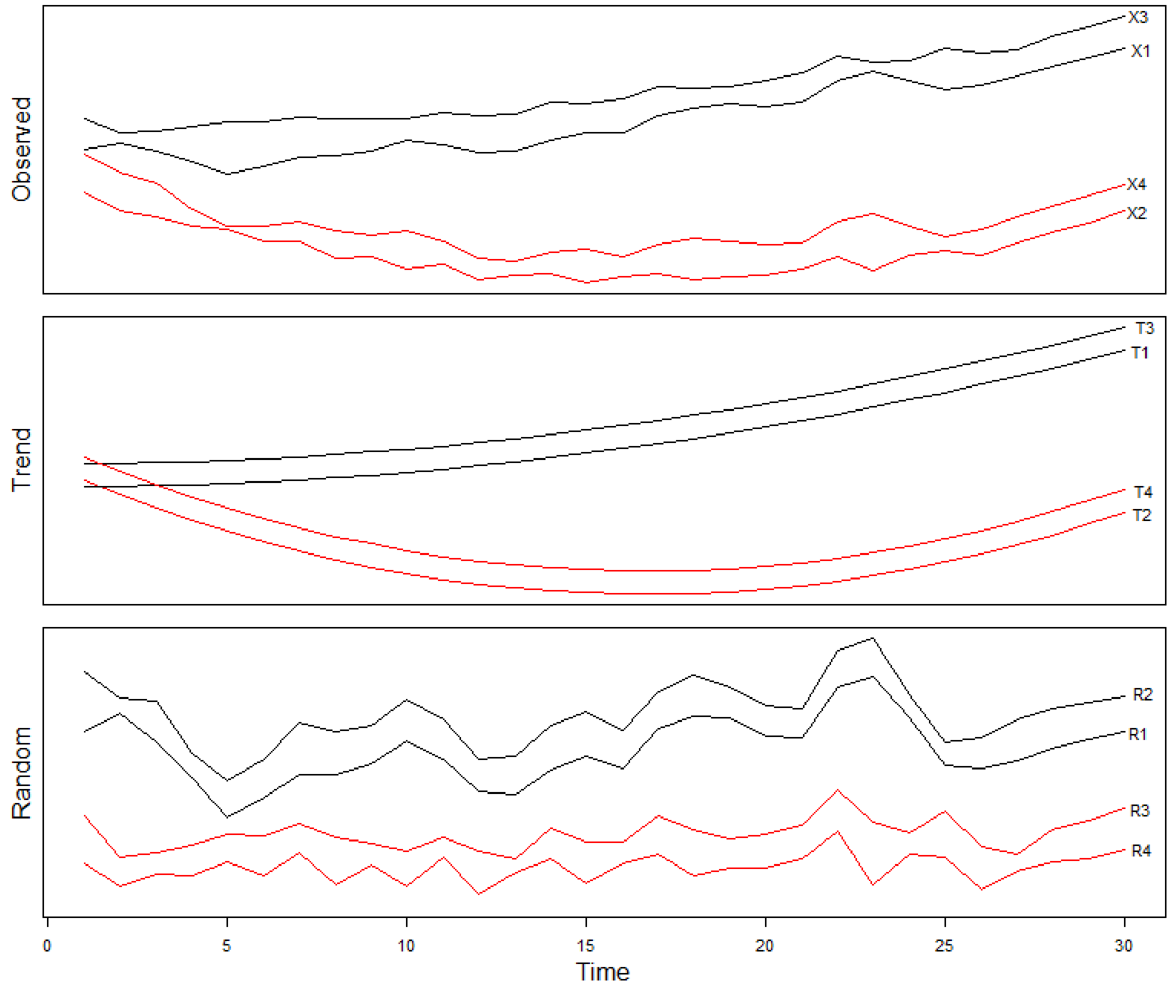


Figure 7.1 Clustering results: observed series versus detrended series.

Remark 2: The fuzziness parameter

The parameters to be fixed in equation (7.3) are the fuzziness parameter m and the number of clusters K. The fuzziness parameter $m > 1$ is usually chosen by the user and has an influence on the clustering results. If m is close to 1, the algorithm will produce the clustering output in which most of the membership degrees will be very close to 0 or 1. In other words, the results will be very similar to those of crisp clustering methods. In contrast, a large value of m will lead to all membership degrees being close to $^1/_K$. Thus, Kamdar and Joshi (2000) suggest selecting m in the range (1, 1.5].

Remark 3: Cluster validity

The number of clusters K in equation (7.3) is usually selected according to a cluster validity measure. A host of different cluster validity indices and their distinctive characteristics are explored in Section

3.2.2 (Chapter 3). In fuzzy clustering framework, the widely used validity measures for selecting K include *Xie-Beni* index (Xie and Beni 1991) and *Fuzzy Silhouette* index (Campello and Hruschka 2006).

Xie-Beni criterion (XB) is defined as the ratio between compactness and separation among clusters, and can be computed by:

$$XB = \frac{\sum_{i=1}^{N}\sum_{k=1}^{K} u_{ik}^{p} d(x_i, x_k)}{N min_{p \neq q} d(x_p, x_q)} \tag{7.6}$$

where $(p, q) \in \{1, \dots, K\}$. The numerator in equation (6.10) represents the *total within-cluster distance*, which is the objective function (called J) of the Fuzzy K-Medoids clustering procedure. The ratio $\frac{J}{N}$ is called the *compactness* of the fuzzy partition. The smaller this value, the more compact a partition with a fixed number of clusters is. The other element $min_{p \neq q} d(x_p, x_q)$ in the denominator is called the *separation* because the larger this value the more separate the clustering partition with a fixed number of clusters. Thus, for a fixed number of clusters, the <u>smaller</u> the XB value, the more compact and separate the clusters are, the <u>better</u> the assignment of the units to the clusters. Despite being intuitive, XB index has two major disadvantages: it decreases monotonically as K approaches N, and XB goes to infinity as the fuzziness parameter m becomes very large (Cebeci 2019).

Another criterion for selecting the number of clusters K in fuzzy clustering framework is the Fuzzy Silhouette index (FS). This is the fuzzy extension of the Silhouette criterion in crisp clustering (see Section 3.3.1, Chapter 3). FS criterion aims to determine how well the units are assigned into clusters in terms of simultaneously minimising the intra-cluster distance and maximising the inter-cluster distance. The FS index is defined as:

$$FS = \frac{\sum_{i=1}^{N}(u_{ik} - u_{ik'})^{\alpha}\lambda_i}{\sum_{i=1}^{N}(u_{ik} - u_{ik'})^{\alpha}}, \lambda_i = \frac{(b_i - a_i)}{max\{b_i, a_i\}} \tag{7.7}$$

where a_i is the average distance between the i-th unit and all units belonging to the k-th cluster ($k = 1, \dots, K$) with which i is associated with the highest membership degree; b_i is the minimum (over clusters) average distance of the i-th unit to all units belonging to the cluster k' with $k \neq k'$; $(u_{ik} - u_{ik'})^{\alpha}$ is the weight of each λ_i calculated upon the fuzzy partition matrix $U = \{u_{ik}; i = 1, \dots, N; k = 1, \dots, K\}$, where k and k' are respectively the first and second best clusters (accordingly to the membership degree) to which the i-th unit is associated; α is the user defined weighting coefficient. The <u>higher</u> the FS value, the <u>better</u> the assignment of the units to the clusters.

It is noted that the FS criterion explicitly considers the fuzzy membership matrix $U = \{u_{ik}: i = 1, \dots, N; k = 1, \dots, K\}$ in its calculation. Maharaj *et al.* (2019, p.40) further comment:

> "[Fuzzy Silhouette] may be able to discriminate between overlapped data clusters even if these clusters have their own distinct regions with higher data densities, since it [Fuzzy Silhouette]

considers the information contained in the fuzzy partition matrix U based on the degrees to which clusters overlap one another. This information can be used to reveal those regions with high data density by stressing the importance of time series data concentrated in the vicinity of the cluster prototypes while reducing the importance of objects in the overlapping areas".

Using artificial data sets, Cebeci (2019) compares the performance of these two popular indices and concludes that FS index tends to be more stable. In addition, Campello and Hruschka (2006) show that FS is less computationally intensive and performs similar or better than XB under a range of scenarios with different data sets and fuzzy clustering algorithms. Nonetheless, Rawashdeh and Ralescu (2012) argue that FS index tends to ignore the clustering of points in the overlapping regions because data points around cluster centres are assigned with higher weights and become more significant to the computation of the index than data points in overlapping regions.

Remark 4: Parameters of the time series clustering algorithm in the empirical analysis

The objective function in equation (7.3) is subject to the following specifications:

(1) The fuzziness parameter is set to $m = 1.5$ because of its better performance in earlier studies (Kamdar and Joshi 2000).

(2) The function f is set to: $$f(t) = \exp(t) - 1 \quad (7.8)$$ because this has been empirically tested and appears to be the most convenient to highlight small differences among dissimilar values.

(3) Following the empirical application provided in Disegna *et al.* (2017), the copula C_{ij} in equation (7.2) is estimated by:

$$C_{ij}(u, v) = \frac{1}{T}\sum_{t=1}^{T} 1\left(\frac{R_{it}}{T+1} \leq u, \frac{R_{jt}}{T+1} \leq v\right) \quad (7.9)$$

where R_{it} and R_{jt} are the ranks associated with the observed series for 'Trend analysis' and the detrended series for 'Fluctuation analysis'.

(4) The optimal number of clusters K is detected by Fuzzy Silhouette and Xie-Beni indices as discussed in Remark 3.

(5) For the 'Fluctuation analysis', the original time series are detrended by applying the Hodrick-Prescott filter. This detrending method is attractive since the detrended series can be obtained even when the relationship of the trend component on time, $f(t) = T_t$, is unknown.

7.2.2 The Copula-based Fuzzy K-Medoids Space-Time clustering algorithm (COFUST)

Embedding spatial information in cluster analysis: Proximity matrix

Cluster analysis discovers patterns in data by organising a set of N objects into K disjoint *unknown* clusters so that the within-cluster dissimilarity is minimised while the between-cluster dissimilarity is maximised. Sometimes there is additional information about the types of clusters that are sought in the data and it is therefore relevant to impose constraint(s) on the set of allowable solutions. As such, the membership of clusters is determined partly by external information (Everitt *et al.* 2011). A popular type of constraint in empirical research is spatial constraint. Simply speaking, not only should the within-group dispersion be minimised but also the spatial autocorrelation between units should be considered. One real-life scenario where the proximity between objects matters in the clustering task is for the mapping of virus during an epidemic spread (i.e. the identification of hot spots/cold spots).

In clustering spatial data, the spatial information can be incorporated into the clustering process by using a matrix S, which is a symmetric matrix with zero diagonal and the off-diagonal elements indicate the spatial relationship between two spatial units (Pham 2001; Coppi *et al.* 2010; D'Urso *et al.* 2019a). S is labelled *proximity matrix* to avoid confusion with the spatial weight matrix W that was discussed in Chapter 6. Again, the proximity here could be geographical, biological, or social. Numerous approaches in building W could be applied to construct S. Two widely used methods of developing the proximity matrix S involve the concepts of *contiguity* and *connectivity*. To illustrate, two territorial units are contiguous if they are neighbours or if they belong to the same macro-area (even when they are not adjacent). In this regard, the generic element of S is given as: $s_{ij} = 1$ if units i and j are contiguous ($i \neq j$), and 0 otherwise. Alternatively, the generic element s_{ij} could be the inverse of a distance measure between units i and j ($i \neq j$) and is normalised to be in the range [0, 1]. The more connected the two units, the lower the value in S.

The copula-based dissimilarity measure

The starting point is represented by an $(N \times T)$ data matrix X, defined as follows:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1T} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NT} \end{bmatrix} \tag{7.1}$$

where x_{it} is the value of the i-th unit ($i = 1, 2, \ldots, N$) at the t-th time ($t = 1, 2, \ldots, T$). In other words, the i-th row of matrix X represents time series of the i-th unit. The additional spatial information is represented by an $(N \times N)$ proximity matrix S given by:

$$S = \begin{bmatrix} 0 & s_{12} & \cdots & s_{1N} \\ s_{21} & 0 & \cdots & s_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ s_{N1} & s_{N2} & \cdots & 0 \end{bmatrix} \tag{7.10}$$

where the generic entry s_{ij} is the proximity measure between units i and j $(i, j = 1, 2, \dots, N)$: $s_{ii} = 0$ and $s_{ij} \geq 0$. Usually, S is a symmetric matrix: $s_{ij} = s_{ji}$ $(i \neq j)$.

In the case where spatial information is not considered, the copula-based dissimilarity d_{ij} between units *i* and *j* can be defined as:

$$d_{ij} = f(\|M - C_{ij}\|) \tag{7.2}$$

where f is an increasing and continuous real-valued function with $f(0) = 0$, $\|.\|$ is the suitable norm in the copula space, C_{ij} measures the rank-invariant similarity ("dependence") between time series of units i and j, and M is the Frechet upper-bound copula $M(u, v) = \min(u, v)$ representing the maximum degree of similarity among time series. If $C_{ij} = M$, the dissimilarity between x_i and x_j is 0.

In order to incorporate spatial information into the dissimilarity distance, one first needs to transform the proximity matrix S into a matrix whose entries are objects of the copula space:

$$S_{ij} = s_{ij}L + (1 - s_{ij})M \tag{7.11}$$

where L is the Frechet lower-bound copula $L(u, v) = \max(u + v - 1, 0)$ representing the maximum degree of dissimilarity among time series. When $s_{ij} \approx 0$ (i.e, units i and j are proximate), $S_{ij} \approx M$, whilst $S_{ij} \approx L$ when $s_{ij} \approx 1$ (i.e, units i and *j* are far away).

Thus, each pair $(i,\ j)$ of units is associated with a copula that combines the dependence and proximity information, shown by:

$$\tilde{C} = \beta C_{ij} + (1 - \beta)S_{ij} \tag{7.12}$$

where $\beta \in [0, 1]$ is a tuning parameter that reflects how much influence proximity information exerts on the clustering procedure. If $\beta = 1$, the clustering output only reflects the dependence among time series while the clustering solution is solely based on the proximity information when $\beta = 0$.

Therefore, with the presence of spatial information the copula-based dissimilarity measure between units *i* and *j* is defined as:

$$d_{ij} = f(\|M - \tilde{C}_{ij}\|) \tag{7.13}$$

Substituting (7.11) and (7.12) into (7.13) gives:

$$d_{ij} = f(\|\beta(M - C_{ij}) + s_{ij}(1 - \beta)(M - L)\|) \tag{7.14}$$

Applying a clustering algorithm

Clustering techniques could generally be grouped into two categories: crisp versus fuzzy clustering. Within a crisp clustering framework, each unit can only belong to one cluster and clusters are mutually exclusive. Such a crisp assignment of data to clusters can be restrictive in presence of data points that are equally distant from two or more clusters. *Crisp clustering* arbitrarily assigns those data points to one of the clusters although they should equally belong to all of them. By contrast, *fuzzy clustering* is an overlapping approach which allows a unit to belong to multiple clusters simultaneously with different degrees of membership (Bezdek 1981). This overlapping assignment reflects cluster structure in a more natural way especially when clusters overlap or when there is not a clear boundary between clusters (McBratney and Moore 1985). The membership degrees produced by fuzzy clustering methods also indicate whether there is a second-best cluster almost as good as the best cluster - a feature that crisp clustering methods cannot capture.

Two main fuzzy clustering methods in the empirical literature include Fuzzy K-Means and Fuzzy K-Medoids, details of which are discussed in Section 3.3.2, Chapter 3. Here, the latter is proposed for the space-time clustering task due to its two major advantages. First, the algorithm returns *medoids* that adequately synthesise the structural information of each cluster, and furthermore these medoids are actual units instead of 'imaginary' average series as in the case of the Fuzzy K-Means technique. These non-fictional medoids provide better interpretations of the final obtained clusters, especially in clustering territorial units. Kaufman and Rousseeuw (2009, p.71) argue:

> "In many clustering problems, one is particularly interested in a characterisation of the clusters by means of typical or representative objects [geographical territories]. These are objects [geographical territories] that represent the various structural aspects of the set of objects [geographical territories] being investigated. There can be many reasons for searching for representative objects [geographical territories]. Not only can these objects [geographical territories] provide a characterisation of the clusters, but they can often be used for further work or research, especially when it is more economical or convenient to use a small set of K objects [geographical territories] instead of the large set one started off with".

Second, the Fuzzy K-Medoids technique tends to be more robust to the presence of noise in the data than the Fuzzy K-Means (García-Escudero *et al.* 2010).

The general Fuzzy K-Medoids algorithm can be formalised as:

$$\begin{cases} min\colon \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik}^{m} d_{ik}(x_i, x_k) \\ \sum_{k=1}^{K} u_{ik} = 1,\ u_{ik} \geq 0 \end{cases} \tag{7.15}$$

where x_i and x_k are time series of the *i*-th unit and medoid time series of the k-th ($k = 1, 2, \dots, K$) cluster; u_{ik} denotes the membership degree of the *i*-th unit in the k-th cluster; $m > 1$ is the fuzziness parameter; $d_{ik}(x_i, x_k)$ is the dissimilarity measure between time series of the i-th unit and the k-th medoid, and is defined as in equation (7.14). According to the objective function (7.15), the COFUST algorithm aims to minimise the copula-based space-considered dissimilarity between data points and cluster centres (i.e. medoids in this case).

Data pre-processing

As discussed in Section 7.2.1, the choice of using actual data or pre-filtered data offers different interpretations, and hence is a merit. In the same spirit, the COFUST algorithm in the subsequent empirical analysis will be performed for actual data ('Spatial trend analysis') and pre-filtered data ('Spatial fluctuation analysis'). Some distinctions can be drawn here. The 'Spatial trend analysis' detects groups of countries with similar evolution in calorie consumption whilst simultaneously being close in spatial proximity. On the other hand, clusters derived from the 'Spatial fluctuation analysis' include countries whose calorie consumptions experience similar deviations from the trend while countries in the same cluster are required to be spatially close.

Some remarks on cluster validity

The parameters to be fixed in model (7.15) are the number of clusters K, the fuzziness parameter m and the tuning parameter β. Details on how to choose K and m are referred to Section 7.2.1.

Here, the selection of the optimal value of K and β is more complicated and cannot be determined simultaneously. The optimal number of clusters K can be suggested by the Fuzzy Silhouette Index (FS) (Campello and Hruschka 2006) or Xie-Beni Index (XB) (Xie and Beni 1991). However, there is no subjective criteria/test for choosing the tuning parameter β which can take any value between 0 (no time series information is used to formulate clusters) and 1 (no spatial information is incorporated into the clustering procedure). Indeed, Disegna *et al.* (2017) recommend setting β between 0.5 and 1 since a value of β lower than 0.5 would indicate an overpowering influence of spatial information.

A possible solution is represented by the following heuristic procedure: assuming K has already been chosen (in the scenario when spatial information is not considered in the cluster analysis, i.e. $\beta = 1$), for every specified value of β the obtained clusters are constructed in such a way that the within-cluster dispersion is minimised while the within-cluster spatial autocorrelation is maximised. To this purpose, for fixed values of K and m, the algorithm is run for varying values of β and the optimal value of β is chosen so that the post-cluster spatial autocorrelation is maximised. The rationale behind this approach is that the identification of clusters remains largely based on the dependence among time series and the spatial information only fine-tunes the final cluster obtainment, perhaps by some adjustments in the membership degrees. Obviously, a crucial task is to compute the spatial autocorrelation among post-cluster units – the basis to determine β.

7.2.3 Spatio-temporal autocorrelation

Measuring spatio-temporal autocorrelation in time series data

As explored in earlier sections, a commonly used tool to measure spatial autocorrelation in the literature is the Global Moran's I statistic (Moran 1950a; Gittleman and Kot 1990). Its computation and the significance test heavily rely on the exogenous spatial weight matrix W, which must be externally provided and appropriately specified. Conventionally, this spatial weight matrix is strictly based on spatial nature even when data are collected over time. Chasco and López (2008, p.102) argue:

> "spatial dependence has usually been defined as a spatial effect, which is related to the spatial interaction existing between geographic locations that takes place in <u>a particular moment in time</u>".

Since space and time are not necessarily neutral dimensions, it is reasonable to expect that the computation of spatial dependence should be adjusted to account for the time dimension. This is of paramount relevance when the data has a considerable time dimension, and moreover Dubé and Legros (2013a) show that ignoring the temporal dimension could lead to misinterpretation of the 'real' measure of spatial dependence over time. To study the spatial dependence in a non-static manner, one could simply calculate the Moran's index for different discrete time periods and apply the spatial Markov techniques (Rey 2001; Rey 2014) to determine whether there is significant spatio-temporal interaction (Carracedo *et al.* 2018). Nonetheless, scant attention has been paid to the establishment of a generalised spatial dependence measure in a situation where values of spatial units are collected over time.

Having said that, there are some propositions that deal with the presence of both dimensions in the data. Some previous authors consider separate spatial and temporal weight matrices to control for spatial effects, temporal effects and indirect spatio-temporal effects (Pace *et al.* 1998; Pace *et al.* 2000; Sun *et al.* 2005). On the other hand, a large number of researchers attempt to develop the so-called *spatio-temporal weight matrix* by blending spatial and temporal matrices together, and then plug this integrated matrix into the Global Moran's I index calculation formula (see, for example, Huang *et al.* 2010; Dubé and Legros 2013a, 2013b; Yu 2014; Lee and Li 2017). Despite being straightforward and less computationally intensive, the development of such a matrix is subject to a critical issue of whether the joint effect of spatial and temporal autocorrelation is additive, multiplicative or of a different nature (Bertazzon 2003; Lee and Li 2017). Moreover, these methods consider spatial observations and temporal values separately by incorporating temporal lags into the spatial weight matrix.

Here, taking an alternative approach this analysis follows a recent literature strand in which the measurement of spatio-temporal autocorrelation between two spatial units depends on the similarity of the two time series (Porat *et al.* 2012; Gao *et al.* 2019). In order to understand the rationale behind this approach, it is useful to look at the Global Moran's I index (Moran 1950a) for cross-sectional data:

$$I = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \times \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{7.16}$$

where x_i and x_j are values of observations at locations i and j respectively, $\bar{x}$ is the mean value of all observations, N is the number of observations, and W is the spatial weight matrix. When spatial observations are time series, the measurement of the deviation in the Global Moran's I index in equation (7.9) is no longer applicable. If x_i is a time series, the element $(x_i - \bar{x})$ becomes the deviation between time series at location i and the global mean series. This situation requires a measurement of similarity between two time series. Porat *et al.* (2012) replace the attribute deviation in the original Global Moran's I calculation with the Pearson correlation coefficient so that the element $(x_i - \bar{x})$ in equation (7.9) is substituted by the Pearson correlation statistic between the temporal sequences of x_i at location i and the average series. Nonetheless, this measurement is criticised for the assumptions of normal distribution and linear relationship between variables. Analysing the spatio-temporal autocorrelation for human mobility data, Gao *et al.* (2019) employ the adaptive temporal dissimilarity which is able to capture proximity both on values and on behaviours of the series.

In this research, the deviation of two time series is measured from the perspective of temporal trend proximity. Specifically, the element $(x_i - \bar{x})$ is replaced with the Kendall rank correlation coefficient (or Kendall's tau coefficient) (Kendall 1955) between time series. The proposed spatio-temporal autocorrelation measure is calculated by:

$$I^* = \frac{N}{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}} \times \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}\psi_i\psi_j}{\sum_{i=1}^{N}{\psi_i}^2} \tag{7.17}$$

where ψ_i denotes Kendall correlation coefficient between time series at location i and the average series.

Calculation of the Kendall's tau is referred to Section 3.2.2 (Chapter 3). To recap, Kendall correlation is based on the ranks of observations and measures the strength and direction of the monotonic relationship between two variables. It makes use of the idea of *concordance*. Two random variables are concordant if large (small) values of one are related to large (small) values of the other. When large (small) values of one are related to small (large) values of the other, the random variables are discordant. Kendall's tau has a critical advantage over the traditional Pearson correlation. The former is a non-parametric test and does not rely on the assumptions of the underlying distribution. In addition, Pearson's coefficient measures linear association whereas higher (absolute) value of Kendall's correlation indicates that there is a monotonic (but not necessarily linear) relationship between two variables (Puth *et al.* 2015). Kendall's tau is considered a better alternative to the other non-parametric rank correlation coefficient – Spearman's rho, when the sample size is small and there are many tied ranks (Legendre and Legendre 1998).

In the fuzzy clustering context, an extension of the Global Moran's I statistic – the Fuzzy Moran index (FM) is introduced in Coppi *et al.* (2010) and D'Urso *et al.* (2019a). The purpose is to assess the spatial autocorrelation of clustering partitions in which units belong to multiple clusters with varying degrees of membership. In the case of univariate time series, the (univariate) spatial autocorrelation for the generic k-th cluster is given as:

$$\rho_k = \frac{(X_{comp} - \bar{X}_{comp})' {U_k}^{0.5} S {U_k}^{0.5} (X_{comp} - \bar{X}_{comp})}{(X_{comp} - \bar{X}_{comp})' {U_k}^{0.5} diag(S'S) {U_k}^{0.5} (X_{comp} - \bar{X}_{comp})} \tag{7.18}$$

where X_{comp} is the 'compromise' vector (weighted mean of the time series over the T time occasions), $\bar{X}_{comp}$ is the N-vector with elements equal to the average of the X_{comp} values over N units, U_k is the square diagonal matrix of order N of the membership degrees of the k-th cluster, and S is the proximity matrix. $diag(.)$ is the operator that creates a diagonal matrix whose elements in the main diagonal are the same as those of the square matrix in the argument. Here, every diagonal element of $diag(S'S)$ contains the number of neighbouring areas of the associated spatial unit. In general, the rationale behind the Fuzzy Moran index is that in order to account for the membership structure of the k-th cluster, the elements of S are scaled by the membership degrees of the spatial units in the cluster involved. Conveniently, equation (7.18) can be rewritten as:

$$\rho_k = \frac{tr[(QX_{comp})' {U_k}^{0.5} S {U_k}^{0.5} (QX_{comp})]}{tr[(QX_{comp})' {U_k}^{0.5} diag(S'S) {U_k}^{0.5} (QX_{comp})]} \tag{7.19}$$

where $Q = I_N - \frac{1_N 1_N'}{N}$ is the centring operator in which I_N is an identity matrix of order N and 1_N is a column-vector of order N with unit elements.

However, information related to the temporal variations among time series is not considered in equation (7.19) since the vector X_{comp} represents average data over T time occasions. Furthermore, the fact that the clustering objects are space-time units requires a measurement of spatio-temporal autocorrelation so that both temporal and spatial dimensions of the data are adequately dealt with. In this regard, the Fuzzy Moran's index is generalised by modifying the deviation ($X_{comp} - \bar{X}_{comp}$). The Generalised Fuzzy Moran index is defined as:

$$FM^* = \frac{tr[\Psi' U_k^{0.5} S U_k^{0.5} \Psi]}{tr[\Psi' U_k^{0.5} diag(S'S) U_k^{0.5} \Psi]} \tag{7.20}$$

where Ψ is the N-vector with element $\Psi[i]$ being the Kendall's correlation coefficient between time series at location i and the mean series. Similar to the Fuzzy Moran's index, the Generalised Fuzzy Moran index ranges between -1 and 1. The value of 1 indicates perfect spatio-temporal autocorrelation, i.e. units are close in 'space' and exhibit similar temporal behaviour; the autocorrelation measure decreases as units become farther in 'space' and/or 'time', while the value of 0 indicates no autocorrelation. Broadly speaking, the higher the FM* index, the higher the spatio-temporal autocorrelation, the better the assignment of the (space-time) units to the clusters.

Simulations

In this section, two examples are used to illustrate the performance and main features of the Generalised Fuzzy Moran index. For the sake of comparison, in each simulation the performance of the Generalised Fuzzy Moran index is compared with some existing measures in the literature including the Global Moran's I index and the Fuzzy Moran index.

Simulation 1: space-time data without clustering

Consider 12 time series of length $T \in \{10,20,50\}$ and assume that the time series are generated via the following copula model:

$$C(x_1, \dots, x_{12}) = C_1(x_1, x_2, x_3). C_2(x_4, x_5, x_6). C_3(x_7, x_8, x_9). C_4(x_{10}, x_{11}, x_{12}) \tag{7.21}$$

where C_i $(i = 1, \dots, 4)$ are copulas belonging to the Clayton family with a pairwise Kendall's τ in {0.01, 0.75}. In this setup, the time series are divided into four groups that are independent from each other, and the dependence among time series within each group is controlled by the copula parameter. When $\tau = 0.75$, the time series from each group are relatively close in terms of temporal dependence, whereas setting $\tau = 0.01$ gives a sample consisting of 12 time series that are relatively far in temporal proximity.

Assume that there is additional information on the spatial relationship among time series units. To control for the various degree of proximity in 'space', two proximity matrices S1 and S2 are considered. The first proximity matrix (S1) is defined in such a way that $s_{ij} = 1$ if the *i*-th and *j*-th time series belong to the same copula and $s_{ij} = 0$ otherwise. In other words, two time series linked by the same copula are spatially close to each other. The proximity matrix S2 is defined in such a way that $s_{ij} = 0$ when both $i, j \in \{2, \dots, 12\}$ and $s_{1j} = 1$ for every $j \neq 1$. This configuration allows the first time series to be spatially close to the other time series which are however far from each other. Overall, the generated time series units are spatially closer following the proximity matrix S1 than the proximity matrix S2.

For each replication $R = 1, \dots, 200$, the time series from model (7.21) are generated and the Generalised Fuzzy Moran index is computed to determine the spatio-temporal autocorrelation of the sample. Results are reported in Tables 7.1 and 7.2. From the inspection of the two tables, some comments can be made as follows.

- Given all other parameters fixed, the Generalised Fuzzy Moran index returns a higher value than the Global Moran's I statistic.
- For all other parameters fixed, the spatio-temporal autocorrelation increases as the length T of the time series increases; however, the Global Moran's I index is almost unchanged. For example, using the proximity matrix S1 for various values of T, the Global Moran's I statistic is in the vicinity of 0.88 when $\tau = 0.75$ and -0.08 when $\tau = 0.01$. Hence, the spatial

autocorrelation is unaffected by changing the length of the time series whereas the spatio-temporal autocorrelation is affected.

- Fixing the spatial relationship among time series (by either S1 or S2), reducing the dependence parameter τ from 0.75 to 0.01 results in a lower Generalised Fuzzy Moran value. Thus, with the same proximity information, the computed spatio-temporal autocorrelation decreases as the time series are further in terms of temporal dependence.
- For time series generated from the same dependence parameter τ (either 0.75 or 0.01) and of the same length, values of the Generalised Fuzzy Moran index calculated using S1 is higher than those calculated using S2. Therefore, with the same level of temporal dependence, the computed spatio-temporal autocorrelation is higher when time series units are spatially closer and declines as units become farther in the 'space'.

Table 7.1 Results of Global Moran's I and Generalised Fuzzy Moran indices for simulated data from model (7.21) with proximity matrix S1.

Time series length	Global Moran's I index		Generalised Fuzzy Moran index	
	$\tau = 0.75$	$\tau = 0.01$	$\tau = 0.75$	$\tau = 0.01$
$T = 10$	0.8769	- 0.0838	0.9315	0.3724
$T = 20$	0.8837	- 0.0731	0.9733	0.6133
$T = 50$	0.8802	- 0.0755	0.9918	0.8156

Note: Mean values over R = 200 replications.

Table 7.2 Results of Global Moran's I and Generalised Fuzzy Moran indices for simulated data from model (7.21) with proximity matrix S2.

Time series length	Global Moran's I index		Generalised Fuzzy Moran index	
	$\tau = 0.75$	$\tau = 0.01$	$\tau = 0.75$	$\tau = 0.01$
$T = 10$	-0.0888	-0.0896	0.6354	0.2732
$T = 20$	-0.0894	-0.0909	0.8432	0.5299
$T = 50$	-0.0915	-0.0908	0.9402	0.7326

Note: Mean values over R = 200 replications.

Simulation 2: clustering of time series with and without spatial information

In the second experiment, the Generalised Fuzzy Moran index is computed for post-cluster units and its performance is compared with the Fuzzy Moran index. To this end, 12 time series of length $T = 50$ generated through the copula models in (7.21) are considered and the spatial information among time series is summarised by proximity matrices S1 and S2. For each proximity matrix, the COFUST algorithm is applied for the data sample setting the tuning spatial coefficient in the range of [0.7, 1]. For each replication, the Fuzzy Moran and Generalised Fuzzy Moran statistics are calculated for the four-cluster solution. Results are reported in Tables 7.3 and 7.4.

When clusters are obtained merely by the behaviour of time series, i.e. no spatial information is considered ($\beta = 1$), the Generalised Fuzzy Moran index differs significantly from the Fuzzy Moran index. Particularly when $\tau = 0.01$, the temporal dependence among the generated time series is low and the time series in each of the four clusters are randomly distributed in time while being close in 'space' due to the setup of the proximity matrix S1. As a result, the spatio-temporal autocorrelation measure is approximately 0. On the other hand, the spatial autocorrelation detected by the Fuzzy Moran index still gives a value of nearly 0.5. The influence of the temporal dimension on the Generalised Fuzzy Moran index is therefore apparent.

The results of clustering time series with spatial information are obtained by setting $\beta \neq 1$. For either proximity matrix, one could arrive at different conclusion if he/she chooses the optimal value of the spatial tuning coefficient β based on the values of Generalised Fuzzy Moran and Fuzzy Moran indices. For example, in the case of S1 and $\tau = 0.75$, the Fuzzy Moran index reaches its maximal value when $\beta = 1$, indicating that $\beta = 1$ being the optimal value at which the spatial autocorrelation among post-cluster units is maximised. However, the spatio-temporal autocorrelation measure reaches its maximum (0.6263) at $\beta = 0.9$. Thus, failing to account for both spatial and temporal dimensions in the data varying across space and over time could lead to incorrect results.

Table 7.3 Results of Fuzzy Moran and Generalised Fuzzy Moran indices for simulated data from model (7.29) with length T=50 and proximity matrix S1.

Clustering tuning parameter	Fuzzy Moran index		Generalised Fuzzy Moran index	
	$\tau = 0.75$	$\tau = 0.01$	$\tau = 0.75$	$\tau = 0.01$
$\beta = 1$	0.9868	0.4870	0.5790	-0.0431
$\beta = 0.9$	0.9548	0.4374	0.6236	-0.0690
$\beta = 0.8$	0.7526	0.4278	0.5722	-0.0289
$\beta = 0.7$	0.5460	0.4323	0.4976	-0.0482

Note: Mean values over R = 200 replications.

Table 7.4 Results of Fuzzy Moran and Generalised Fuzzy Moran indices for simulated data from model (7.29) with length T=50 and proximity matrix S2.

Clustering tuning parameter	Fuzzy Moran index		Generalised Fuzzy Moran index	
	$\tau = 0.75$	$\tau = 0.01$	$\tau = 0.75$	$\tau = 0.01$
$\beta = 1$	0.1234	0.3902	0.0405	-0.0350
$\beta = 0.9$	0.2243	0.4172	0.0220	-0.0339
$\beta = 0.8$	0.3409	0.0267	0.0094	-0.0033
$\beta = 0.7$	0.0118	0.0144	0.0084	-0.0009

Note: Mean values over R = 200 replications.

To sum up, the above two simulations demonstrate the performance and main features of the Generalised Fuzzy Moran index. On the one hand, the proposed index extends the classical Global Moran's I statistics to compute the spatial autocorrelation of time series data. On the other hand, it can be considered as an extension of the Fuzzy Moran index to measure the spatio-temporal autocorrelation of post-cluster units in a space-time clustering procedure. In this analysis, the Generalised Fuzzy Moran index will be adopted to assist the selection of final clustering partition in the COFUST clustering algorithm when several spatial coefficients (β) are considered.

7.3 Data

In order to explore how the global diets have evolved over time, data collated from the Food Balance Sheet (FBS) are utilised. The FBS contains annual time series on energy supply (measured as kcal/person/day) of total calories as well as primary food commodities for over 200 countries and territories (FAO 2019c). An overview of the FBS data is given in Section 4.3.4 (Chapter 4). Despite being repeatedly mentioned throughout the narrative, it is worth emphasising again that the consumption-level waste (i.e. food that is wasted at retail, restaurants, and household) is not incorporated in the supply figures, therefore the FBS data represents *food available for consumption* rather than actual food consumption. In the subsequent empirical analysis, the terms *caloric consumption*, *food consumption* and *diet* should be interpreted as *food available for consumption*. Another key limitation of the FBS is the reliability of data coverage and data quality for less developed countries (FAO 2018). In spite of these caveats, the FBS data has been widely utilised in empirical studies as it is the only source of standardised food consumption information that enables longitudinal comparisons between a large set of countries (Vilarnau *et al.* 2019). This chapter utilises the same data set as in Chapter 6, which consists of 118 annual univariate time series on daily per capital total calories available for consumption (hereafter referred to as *per capital daily calories*) from 1961 to 2013.

For the cluster profiling step, information on apparent consumption of main food aggregates is gathered from the FBS (FAO 2019c). Data on GDP per capita are collected from the FAOSTAT (FAO 2019b) and other data are retrieved from the World Bank database (World Bank 2020e).

To examine how the global diets have evolved with respect to some common guidelines on healthy diets, a suitable index to measure the healthiness of a country's diet is needed. Among a number of dietary quality indices in the existing literature (see Section 4.2, Chapter 4), the *Mediterranean Adequacy Index* (MAI) is selected in this analysis. This index was originally developed by Fidanza *et al.* (2004) to measure the adherence to the Mediterranean diet in two Italian cohorts of the Seven Countries Study. The benefits as well as drawbacks of the MAI are discussed in Section 4.3 (Chapter 4). To recap, two major advantages include the simplicity of calculation with readily available data and the usefulness in comparing the trends in food availability over time.

In terms of calculation, MAI is created as a quotient between the sum of calories from typical Mediterranean diet (healthy foods) and the sum of calories from non-typical Mediterranean diet (less healthy foods). Overall, the higher the MAI value, the greater the adherence to Mediterranean dietary patterns, the healthier the diet. Here, the MAI is computed as:

$$MAI = \frac{Calories\ (good)}{Calories\ (bad)} = \frac{cereals + starchy\ roots + vegetables + fruits + fish + vegetable\ oils}{milk + meat + eggs + animal\ fats + sugar \& sweetener} \tag{7.22}$$

7.4 Empirical results

7.4.1 Results of the copula-based fuzzy time series clustering algorithm

To assess the similarities among national diets over time, cluster analysis is utilised to identify groups of countries that share common characteristics in the past trends of calorie availability.

To help visualise the data, 118 univariate time series of per capita daily calories are illustrated in Figure 7.2. Each row of the heatmap represents a univariate time series, and the line graph below the heatmap plots the aggregate series. The heatmap format presents the data (calorie availability in kcal/capita/day) by a palette of green: the darker the shading the greater the calories, with time on the horizonal axis running from left to right. Overall, a switch from white shades to green over the period 1961-2013 is observed for most countries. Since the majority of time series start off with white shades (low calories), evolve to light green (medium calories) in the middle and end with dark green shades (high calories), a predominant upward trend is witnessed. This trend is confirmed by the gradual increase of the aggregate series (except for a little stagnation in the mid-1980s and early-1990s) in the line graph. However, some countries seem to buck this rising trend as shown by the transition from green shades (high calories) in 1961 to white (low calories) in 2013. This decrease in calorie availability over the past half a century is not surprising given a handful of negative growth rates observed in Figure

6.8. The historical calorie changes are therefore uneven across countries. Additionally, the period 1980-1990 is when the calorie changes (the switch from white to green or the reverse) became most apparent.

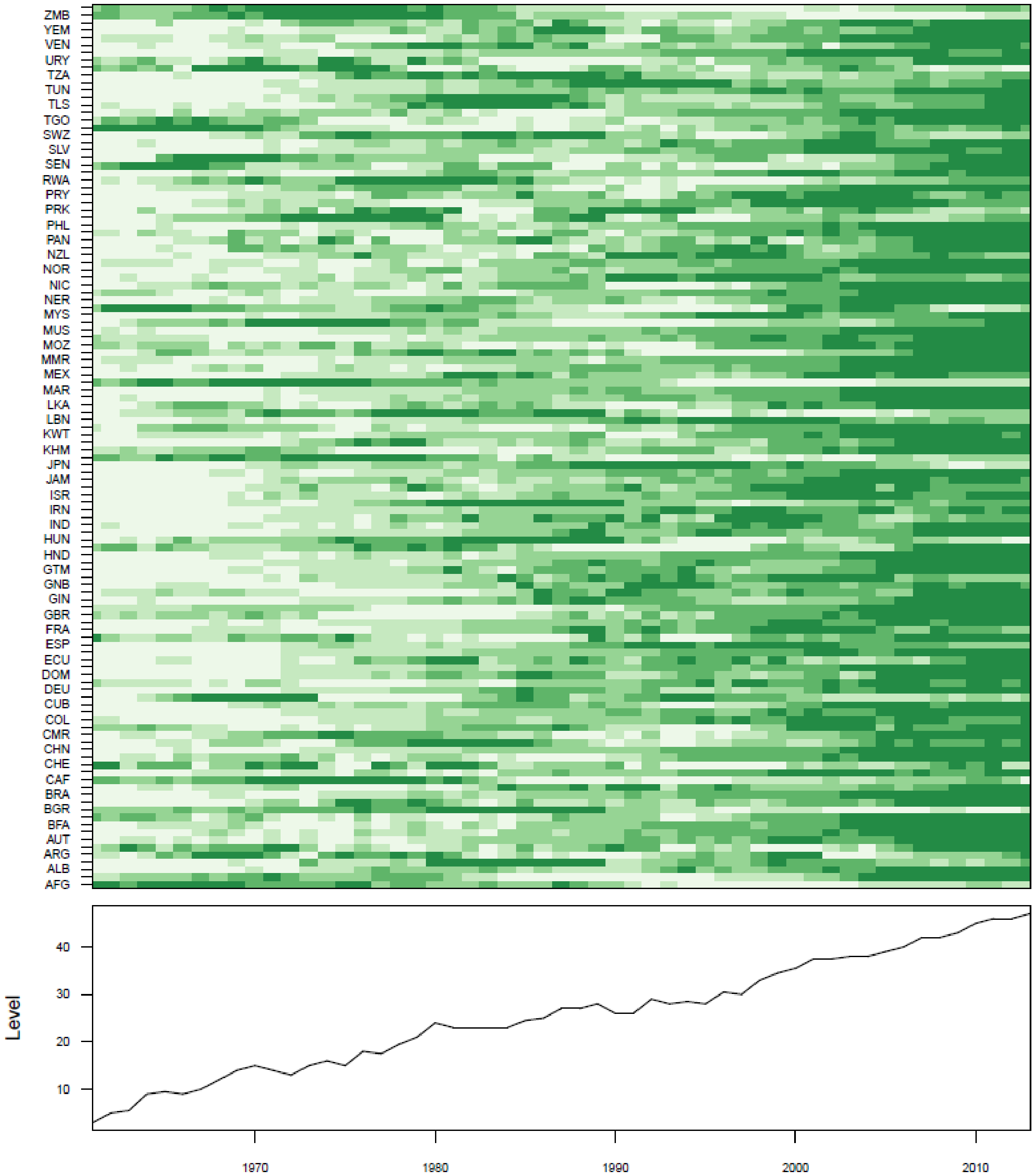


Figure 7.2 Time series plot of daily per capita calories for 118 countries, 1961-2013.

On the other hand, Figure 7.2 demonstrates the complexity of the data which varies both across countries and over time: 118 annual time series spanning over 53 years. The complexity of this data set requires an exploratory tool to describe and summarise the data meaningfully. A simple naïve approach is to derive the average figures by imposing a rule; for example, to aggregate the countries by geography (say continent) or economic criteria (say income level). While this approach has the virtue of simplicity, it takes no account of proximate countries (in terms of location or development) with different diets. For instance, the calorie figure of the USA and Mexico are as dissimilar as the figure of Japan and Germany despite their geographical and economic proximity respectively. In such circumstances, techniques that categorise countries more flexibly offer some appeals. One such tool is *cluster analysis* – a data description method that classifies countries into different clusters so that countries within the same cluster are similar to each other but dissimilar to countries in other clusters.

In this analysis, the novel copula-based time series fuzzy clustering algorithm (Disegna *et al.* 2017) is adopted to detect different clusters characterised by similarities in the historical trends of caloric consumption. The copula function helps to detect the dependence among time series whilst the fuzzy approach allows a country to belong to multiple clusters with varying degrees of membership. Indeed, fuzzy clustering is an attractive method as it allows the possibility that individuals within a country do not consume the same diet and therefore a number of diets coexist within a single country. While for some countries the notion of a national diet is a reasonable rule of thumb, for many other countries the coexistence of different diets will be more appropriate. The overlapping nature of fuzzy clustering gives us an edge in exploring different diets within a country – an impossible task by naively averaging the national diets by geography (say continent) or economic criteria (say income level).

To derive the clustering partition, the algorithm relies on the information related to the copula-based dependence among the clustering variables (time series in this case). As discussed in Remark 1 (Section 7.2.1), the clustering algorithm is conducted in two scenarios, when the clustering variables are original data ('Trend analysis') and pre-filtered data ('Fluctuation analysis'). To recap, 'Trend analysis' classifies countries that share common evolutions in caloric consumption while 'Fluctuation analysis' clusters countries that experience common deviations from the trend in caloric consumption.

Before performing the cluster analysis, it is useful to have an idea about what kind of relationship exists in the data. Figure 7.3 visualises the pairwise Kendall's correlation among the clustering time series in heatmap format: blue (red) indicates positive (negative) correlation and the darker the shade the stronger the correlation. The heatmap related to 'Fluctuation analysis' on the right-hand side is mainly made up of white or very light red/blue shades that denote correlation measures close to zero, implying low pairwise dependence among pre-filtered data (detrended series). In contrast, the majority of observed series employed in 'Trend analysis' are positively correlated as shown by the dominance of red shades whilst a minority are negatively correlated. The inspection of Figure 7.3 thus reveals a clear and significant dependence structure among the clustering variables used for 'Trend analysis' but a more random pattern (i.e. no association) among those used for 'Fluctuation analysis'. Clustering results from two analyses are discussed in turn below.

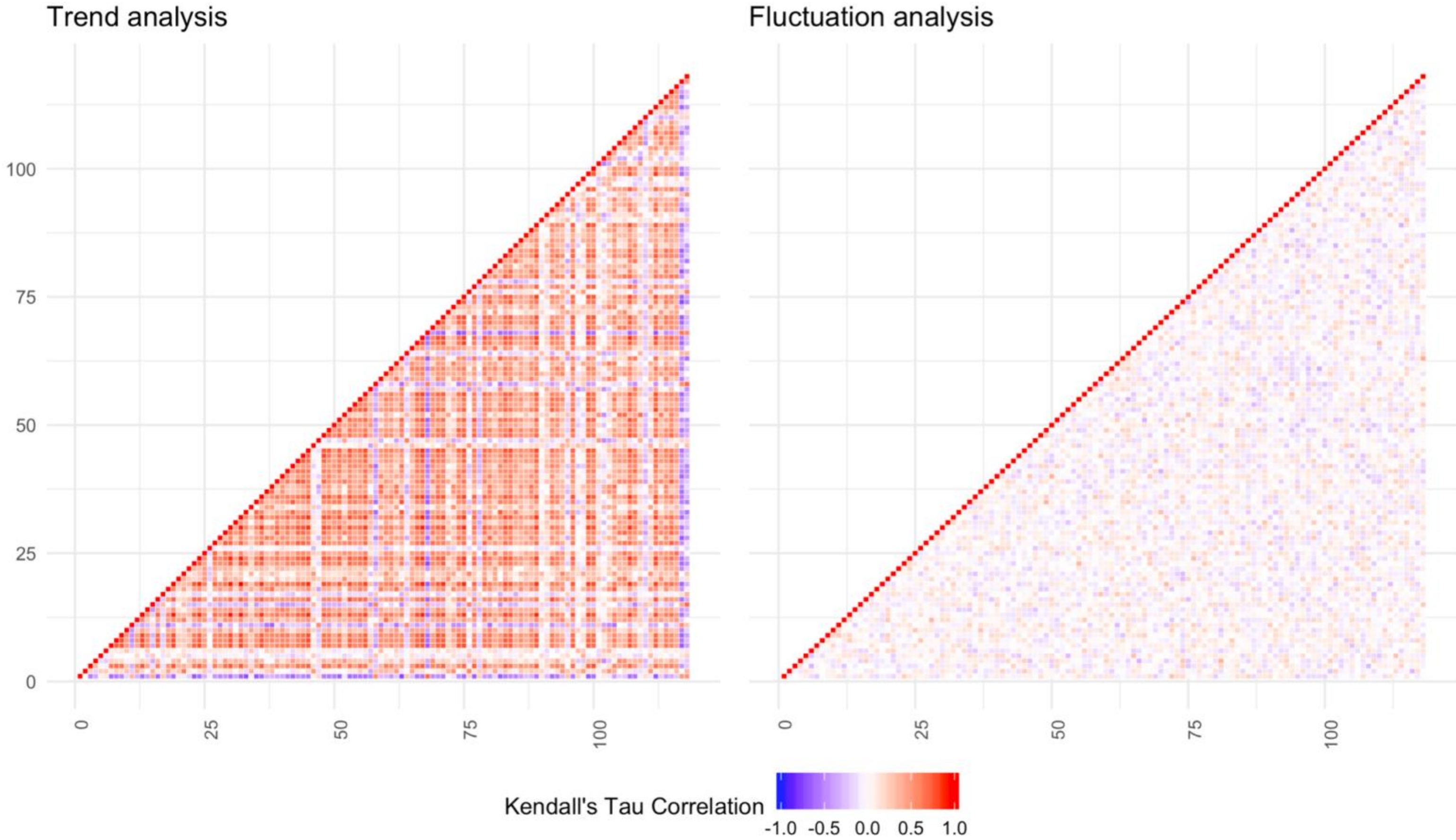


Figure 7.3 Heatmap of the pairwise Kendall's correlation between clustering variables.

Trend analysis: clustering variables are original data (observed series)

Figure 7.4 shows values of the FS and XB cluster validity index for varying number of clusters K from 2 to 10. The optimal number of clusters is suggested not by the level of the indices but by the changes of the series (the largest value among the peaks in FS and the smallest value among the troughs in XB). According to Figure 7.4, the two indices exhibit clear mirroring behaviours: the FS peaks while the XB troughs at $K = 2$, $K = 4$, and $K = 6$. It is evident that two-cluster solution is the best partition, followed by four-cluster and six-cluster solution. Thus, the clustering results of two-cluster, four-cluster and six-cluster partitions are interpreted in the subsequent discussion.

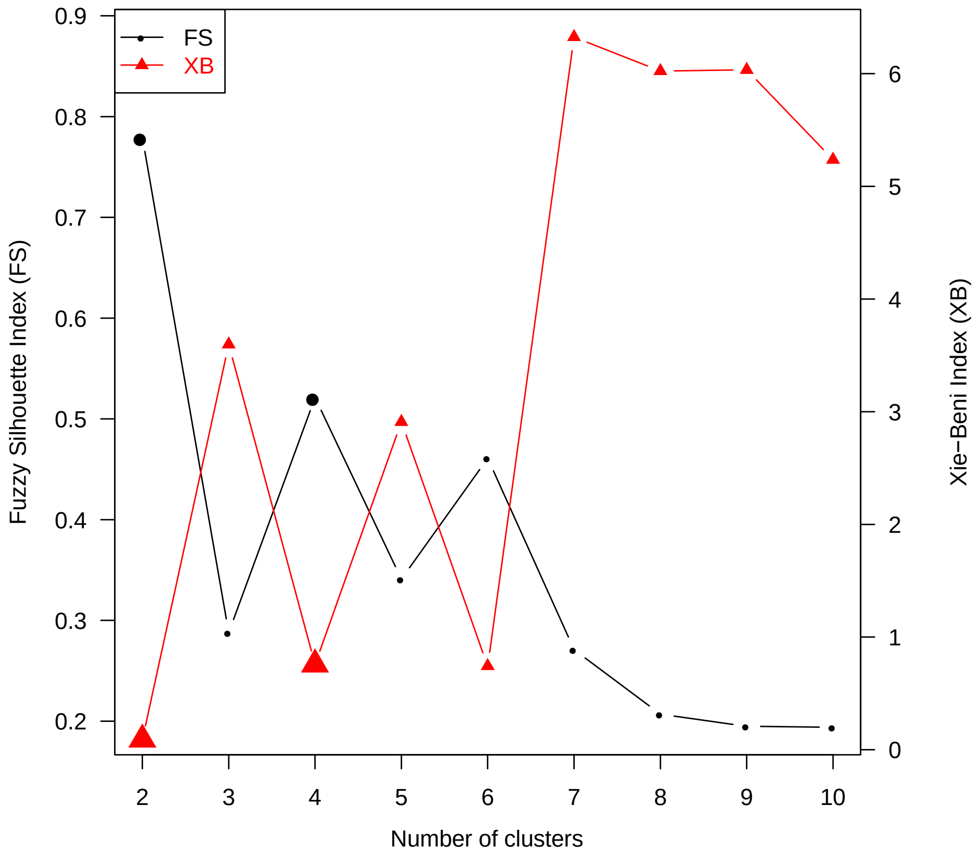


Figure 7.4 FS and XB validity index values for each cluster partition K from 2 to 10 (Trend analysis).

Fixing $K = 2$, the clustering algorithm identifies two clusters (CL) represented by China (CL1) and Zimbabwe (CL2), which account for 94% and 6% of the world population respectively. Figure 7.5 illustrates the cluster medoids and the membership degree of each country belonging to two clusters. The different colour shades for the membership degree perfectly illustrate the philosophy of fuzzy clustering: each country belongs to all clusters with varying membership proportions. In each cluster, only the cluster representative (denoted by the red colour) belongs 100% to that cluster, the medoids of other clusters (coloured in pink shades) have the membership degree of 0 and the remaining countries have a membership proportion in the range of (0, 1). The degree to which all other countries belong to each of these clusters is indicated by the shade of blue: the darker the blue shade, the higher the membership proportion. As expected from the dominant size of CL1, the majority of countries belong to CL1 with high membership degrees. The darkest shade of blue in CL1 is witnessed across Asia, Europe, Northern Africa, and Americas. On the other hand, only few countries are associated to CL2 with a membership proportion higher than 50%, and the most notable ones include Switzerland, Australia, and some countries scattered in Southern Africa.

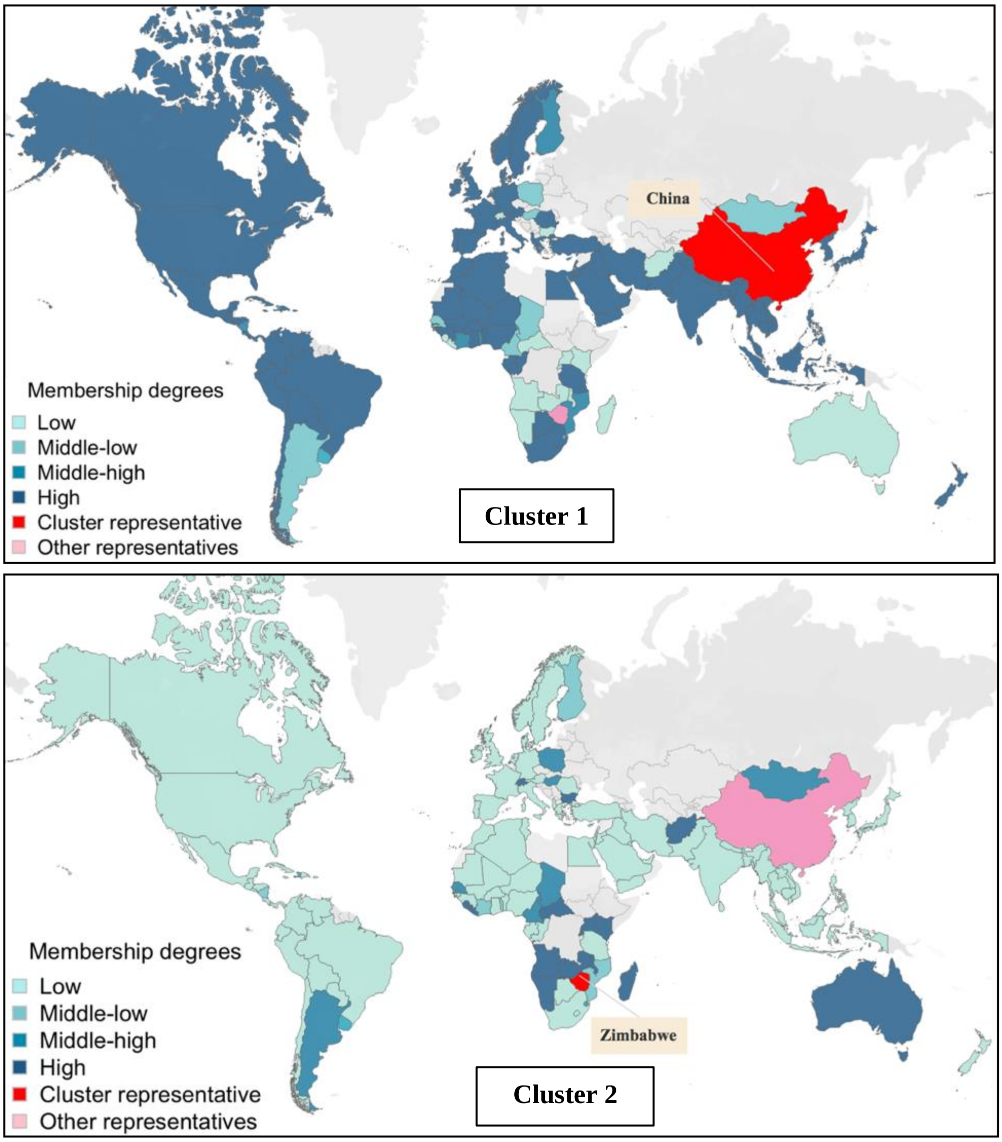


Figure 7.5 Cluster membership degrees ($K = 2$, Trend analysis).

Figure 7.6 plots the weighted average calories from 1961 to 2013. The solid line, denoting the calorie availability of CL1, shows a gradual increase over the past half a century – a shift that 94% of the global population has experienced. In contrast, the calorie availability of CL2, represented by the dashed line, was stagnant over the first two decades of the period, declined markedly to a trough in the early 1990s but has risen quickly and ultimately reached the level comparable to the initial calorie level in 1961. The inspection of Figure 7.6 thus helps make sense of the clustering results: CL1 includes countries (or precisely segments of the population) whose calorie availability has increased dramatically over the past 50 years (most of the world), and CL2 represents the minority whose calorie figure may have fallen. Even though rising calorie consumption is the dominant trend worldwide, 'Trend analysis' detects another trend which is minor but ineluctable and evident in Figure 6.8 – reducing calorie consumption. Will these two trends remain, or will new dietary patterns emerge when the number of clusters K increases?

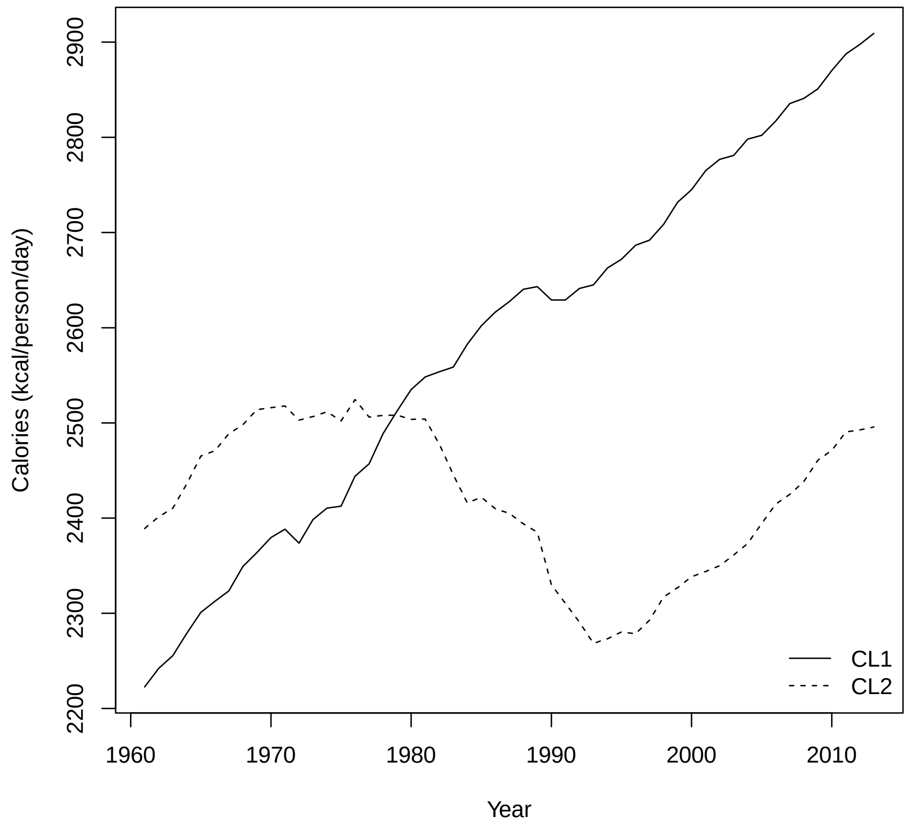
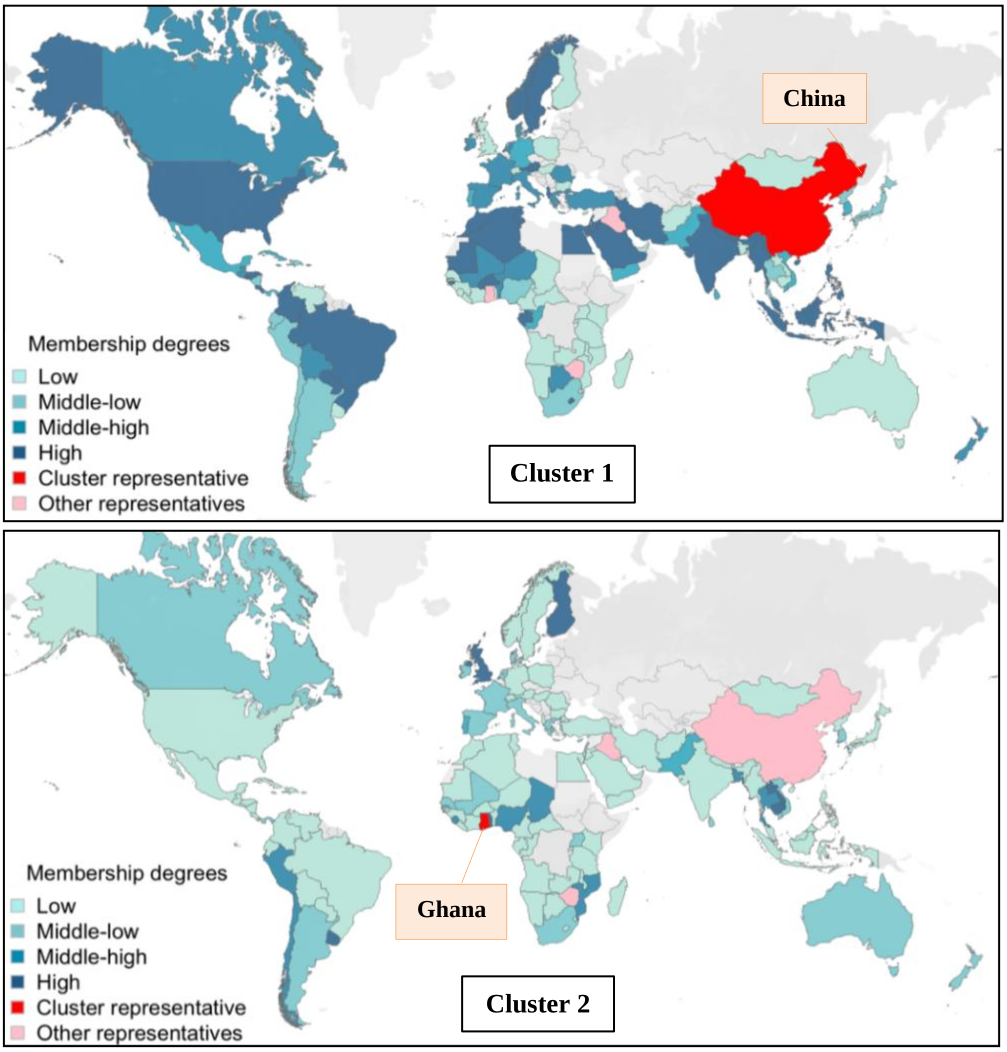


Figure 7.6 Weighted average daily per capita calories, 1961-2013 ($K = 2$, Trend analysis).

When $K = 4$, 'Trend analysis' reveals four clusters of varying sizes. Almost three quarters of the world population are classified into CL1, approximately 10% in CL2 and CL3 each, and CL4 is the niche cluster accounting for less than 5% of the global population. Figure 7.7 illustrates the geographical dispersion of each cluster. In each world map, the cluster is represented by a country representative ("medoid") coloured in red, and the calorie trajectory of which is most exemplar of the cluster. The four cluster medoids are China, Ghana, Iraq, and Zimbabwe respectively. The membership degree of other countries belonging to each of these clusters is shown by the shade of blue: the darker the colour the higher the membership proportion. Clearly, the dark shade of blue dominates in CL1, indicating the relatively high membership degree of most countries belonging to this cluster. The number of countries with high membership degree to other clusters is significantly lower, and the most noticeable are the UK and Finland in CL2, Venezuela and Tanzania in CL3, and Switzerland in CL4. For these countries, the notion of a single national dietary evolution is a fair approximation. Although the grouping of Ghana and the UK in the same cluster may sound questionable, each of these countries might have a group of people with similar taste and preference for calories. It may also reflect the exercise of free will in obesogenic environments leading to similar outcomes despite seemingly different settings. Obesity is after all a universal phenomenon. Yet, it is not hard to find other countries (for example Australia) whose membership proportion spreads almost equally in all clusters. In such circumstances, the idea of a national dietary trend is more of a 'fuzzy' nature and the presence of multiple dietary evolutions in a single country will be more appropriate.



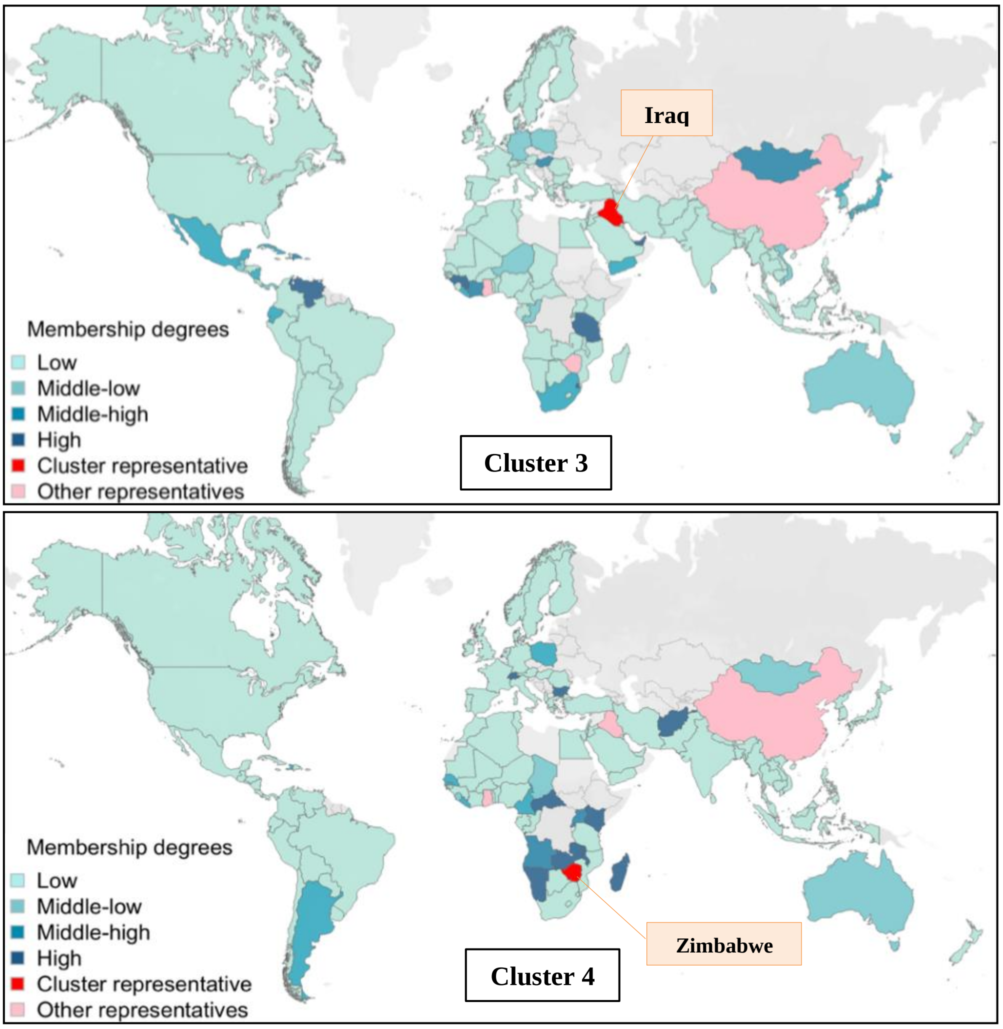


Figure 7.7 Cluster membership degrees ($K = 4$, Trend analysis).

Figure 7.8 shows the changes of the weighted average calories over the past 50 years. Some features seem noteworthy. First, the evolution of calories is not homogenous across clusters. Rising calorie content is the dominant but not the only trend. In general, the daily per capita calories have improved over the past half a century from a minimum of 2,200 to approximately 2,400 kcal/capita/day. This increase was mainly driven by CL1 which was once the least calorific but grew relentlessly and became the most calorific cluster in 2013. Despite rising quickly during the first decade of the period, the calorie availability of CL2 declined in the early 1970s before bouncing back in 1985 at a more or less similar speed as before the plummet. The calorie availability of CL3, which closely mimicked that of CL1 for the first three decades of the period, dropped suddenly in 1989 and recovered its growing trend in late 1990s. Unlike the first three clusters, the calorie figure associated with CL4 was oscillating up and down between 2,200 and 2,400 kcal/capita/day and ended up being the smallest among four clusters in 2013. Therefore, CL4 could represent those countries that experienced declining calorie consumption over the past half a century (as shown by the negative growth rates in Figure 6.8).

Overall, when $K = 4$, 'Trend analysis' identifies four dietary patterns: a monotonic rising trend (CL1), two rising trends with a dip (CL2-3), and a reversal (CL4). Although the four dietary trends were heading on distinct paths to different levels of calories, they have evolved in the same upward direction and at a similar pace during the last couple of decades. Compared to the two-cluster partition, the four-cluster solution keeps the niche cluster whose calorie consumption declined over the past half a century but divides the dominant cluster characterised by increasing calories. However, there is still a large cluster represents 75% of the world population whose calorie consumption has risen constantly over the time. Perhaps this cluster will be further split up when the number of cluster increases.

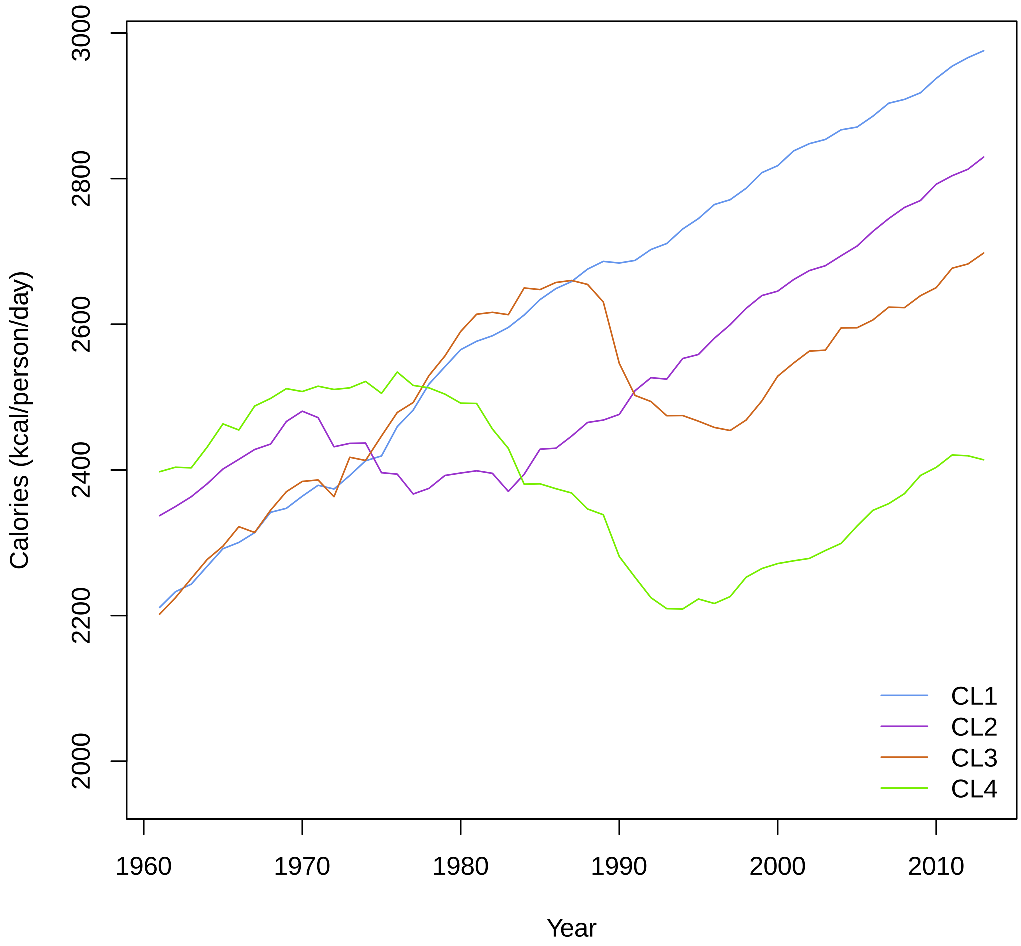


Figure 7.8 Weighted average daily per capita calories, 1961-2013 ($K = 4$, Trend analysis).

Fixing $K = 6$, cluster analysis identifies six clusters of varying sizes. Similar to the previous two scenarios, a large cluster (CL2) accounting for over half of the global population does exist. Other considerable clusters include CL3 and CL4 which respectively make up 15% and 10% of the world population whereas the remaining clusters (CL1, CL5 and CL6) are niche clusters. Figure 7.9 shows the geographical dispersion of each cluster. The dark shade of blue dominates in CL2, indicating the relatively high membership degree of most countries to this cluster. Countries predominantly included in CL2 are the USA, Canada, and Brazil. The dark shades of blue associated with CL3 – the second largest cluster belong to Japan, the Netherlands and some Mediterranean countries such as Spain, Italy and Greece. The number of countries with high membership to other clusters is significantly lower, for example Cameroon (CL1), the UK and Finland (CL4), Tanzania (CL5), Kenya and Switzerland (CL6).

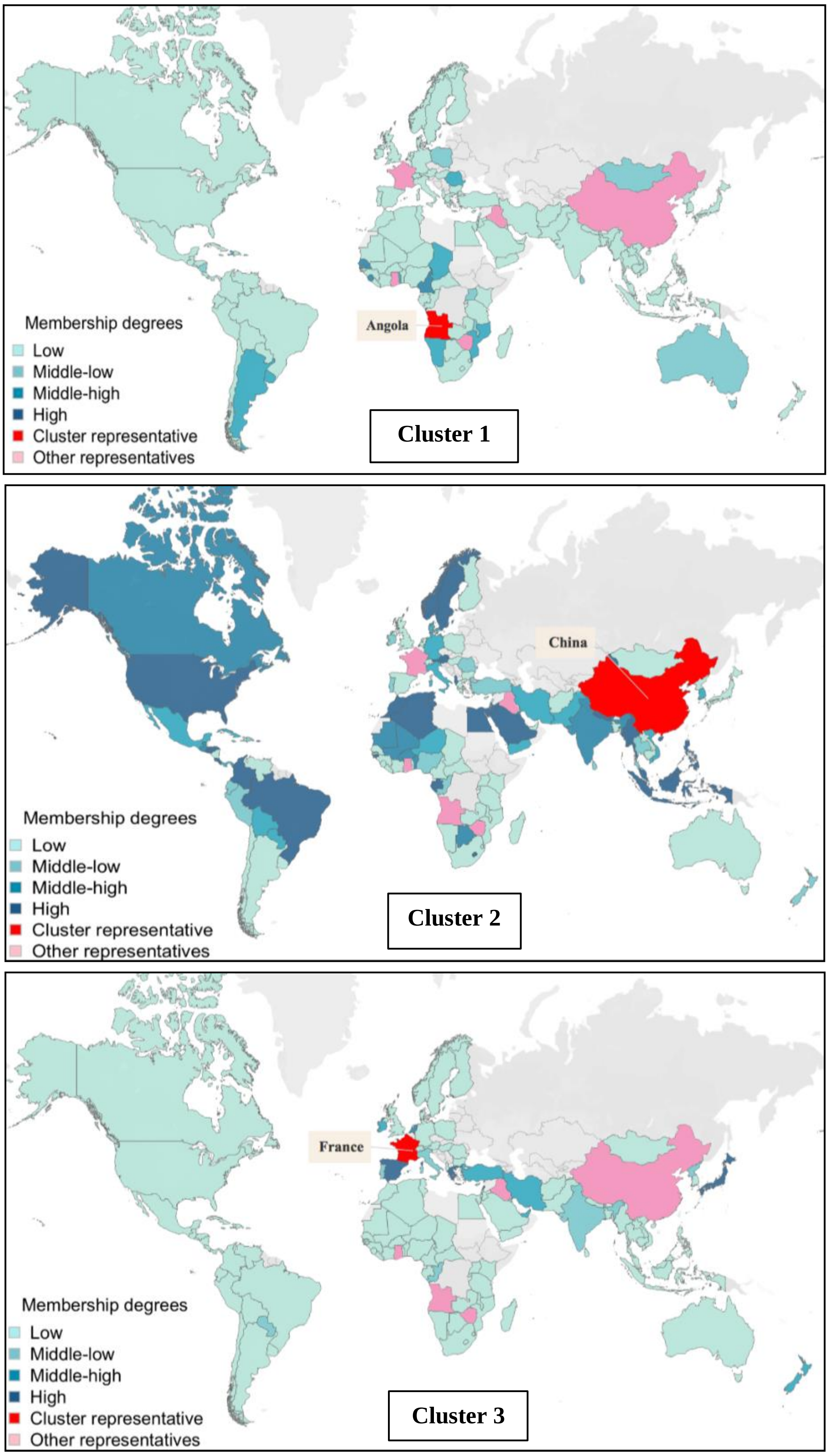


Figure 7.9 Cluster membership degrees ($K = 6$, Trend analysis).

To better understand the obtained clusters, Figure 7.10 plots the changes of the weighted average calories. In general, the clustering algorithm identifies a monotonic rising trend (CL2), a declining trend (CL6) and other rising trends with reversal. The diet of CL2 was once the least calorific in 1961 but grew steadily by almost a third and became one of the most calorific in 2013. Departing from the same lowest initial level of calories, CL5 mimicked the behaviour of CL2 closely before declining dramatically in the mid-1980s but has picked up since the late 1990s. With Iraq being the medoid and including countries such as Tanzania, Swaziland and Venezuela with high membership degrees, CL5 seems to represent those countries whose food consumption was severely affected by war or political instability. Showing some disruption in the 1970s, the caloric consumption series of CL4 which is largely associated with upper-middle-income countries (Peru, Chile, Thailand and Bangladesh to name a few) recovered in the early 1980s and is approaching the level similar to the majority of the global population. In contrast to these clusters, CL6 includes those countries whose calorie consumption has declined over the past half a century.

In addition to the previously identified four trends, allowing for a larger number of clusters K reveals two new dietary patterns corresponding to CL1 and CL3. The former is characterised by stagnating calorie consumption until the 1980s followed by a trough in the early 1990s before rebounding strongly since then. Comprising of only a small number of sub-Saharan African countries (those with the membership proportions of over 60% include Cameroon, Senegal, Sierra Leone), this cluster might represent diets of poor countries that were badly hit by the food crisis in the 1970s but have taken advantage of the globalisation process accelerated in the 1990s. Interestingly, CL3, despite its historical highest level of calories, has shown signs of stabilising calories since the early 2000s. Remaining at the mot calorific level, the calorie consumption associated with CL3 has levelled off for the past 15 years. Overall, what is clear from Figure 7.10 is that for the past two decades, all clusters except CL3 have experienced more calorific diets. The fact that the most calorific cluster has ceased to gain its calorie content gives evidence for the shift from Pattern 4 to Pattern 5 of the nutrition transition model which features behavioural changes for a better diet. Nonetheless, diet is not only about the quantity of calories consumed but also about what type of calories. It is therefore necessary to examine this stabilising pattern with regard to dietary components.

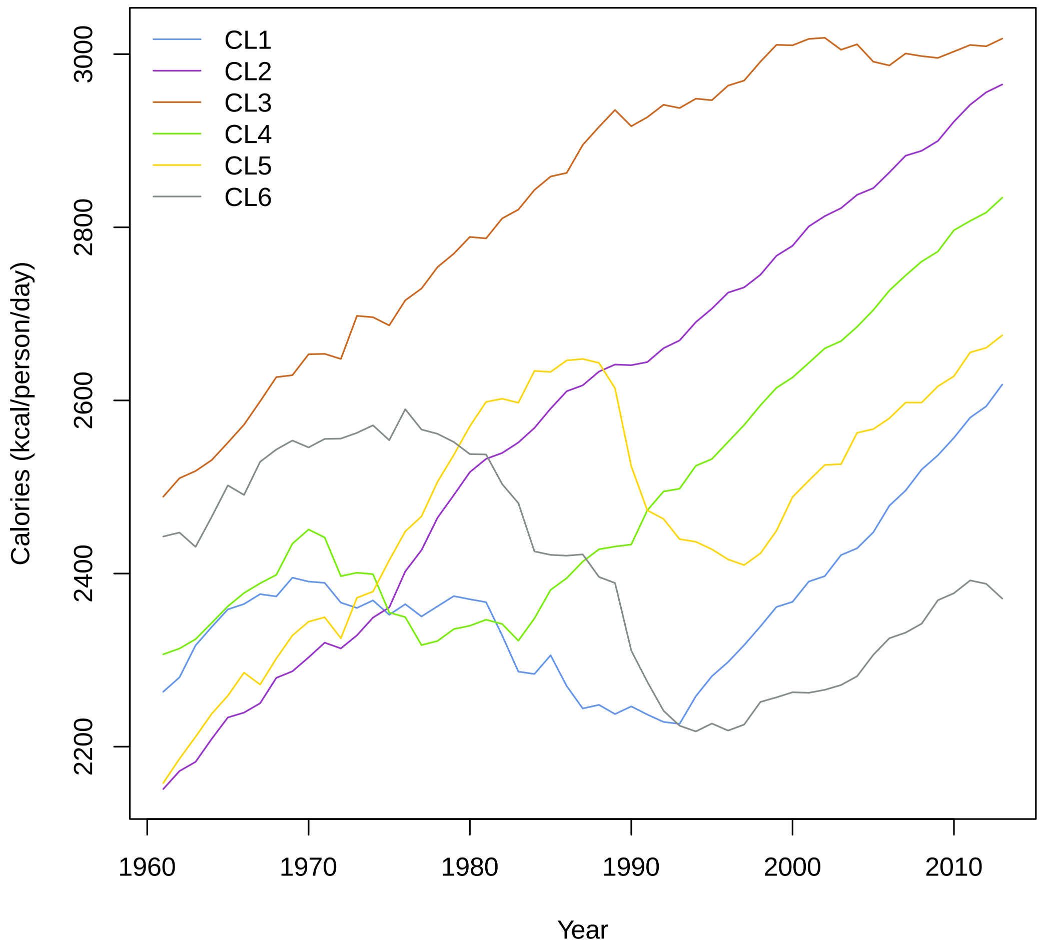


Figure 7.10 Weighted average daily per capita calories, 1961-2013 ($K = 6$, Trend analysis).

Next, the diet of CL3 is investigated for eleven main food groups, namely meat (including eggs), animal fat, milk, sugar, fish, vegetable oils, pulses, cereals, starchy roots, fruits, and vegetables. These food groups represent the composition of any diet and make up the total calorie figure. Figure 7.11 shows the energy contribution of these food groups in terms of deviation from the global average. The horizontal axis runs from left to right, with zero indicating that the consumption is equivalent to the global average, positive numbers indicating above average and negative numbers below average. Because the switching behaviour in the diet of CL3 seems to occur between the late 1990s and the early 2000s, the dietary characteristics in Figure 7.11 are examined for two sub-periods: before 2000 (denoted by the turquoise shade) and after 2000 (denoted by the red shade). In either period, the negative values of the bars suggest that the consumption of pulses, cereals and starchy roots is always below the global average whereas the positive values of the bars for the remaining food groups suggest the consumption above the average. Being rich in animal-source foods whilst low in cereals, roots and pulses, the diet of CL3 carries main elements of the 'Western diet' except the high consumption of fish and vegetables. On the other hand, it can be seen from the size of the bars between two sub-periods that after 2000 CL3 has increased the consumption of vegetable oils, fish, starchy roots, and meat but reduced the consumption of animal fat, milk, sugars, vegetables, pulses and cereals. Thus, the stabilisation of the per capita daily calories during the past 15 years is mainly driven by both lower consumptions of less healthy foods (animal fat, milk, and sugars) and higher consumptions of healthier ones (fish and roots).

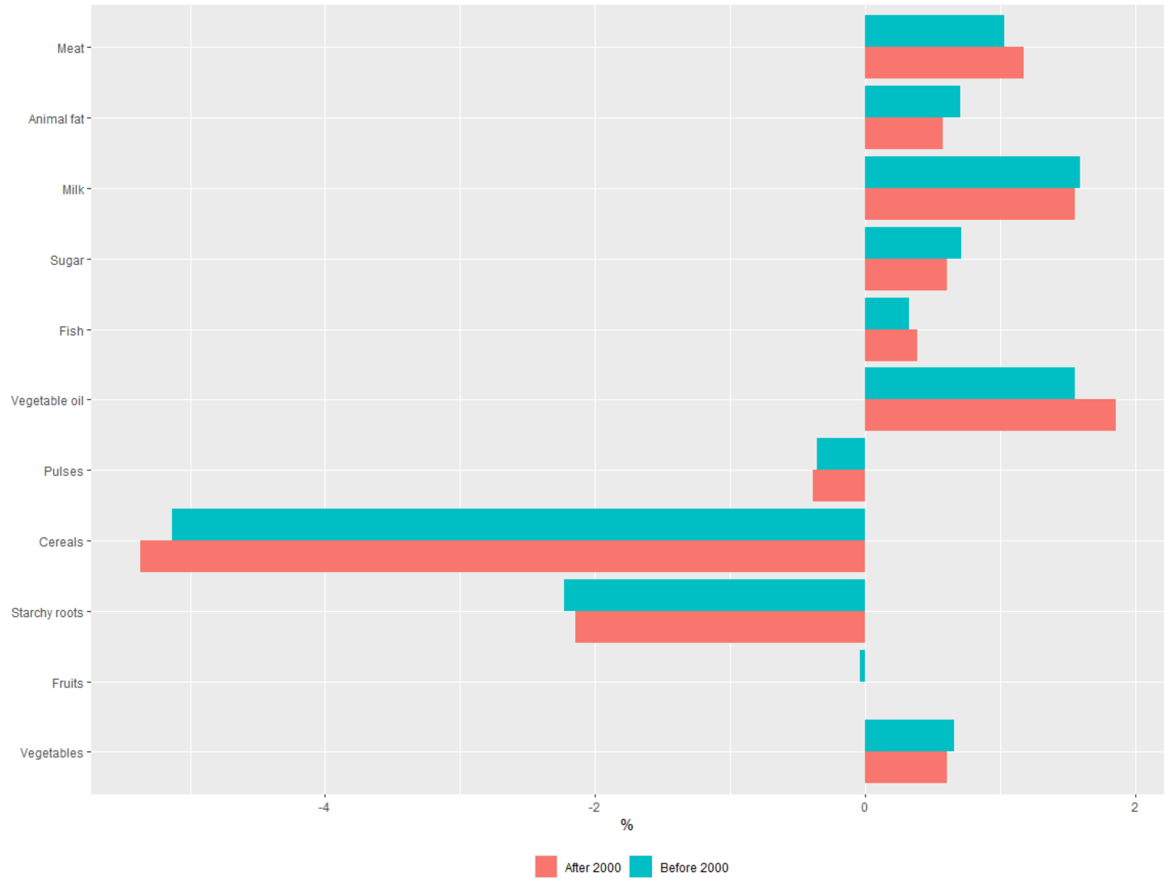


Figure 7.11 Changes in dietary composition of CL3 (Trend analysis).

To sum up, the ‘Trend analysis’ identifies different dietary evolutions representing common dietary trends among 118 countries. Regardless of the number of clusters (K), cluster analysis detects a predominant trend – monotonic rising calorie consumption over time which is experienced by the majority of the global population and another trend which is minor but ineluctable – reducing calorie consumption. Increasing the number of clusters enables the largest cluster to be split up into various dietary evolutions, all of which can be described as rising trends with a reversal, albeit to varying extents. Inspection of the geographical distribution of cluster membership and the changes of the weighted average calories helps to make sense of the clustering results. Setting $K = 6$ reveals a unique cluster (CL3) representing about 10% of the world population whose diet, despite its highest level of calorie consumption, is the only to not become more calorific over the past 15 years. For this cluster, the stabilisation of calorie consumption shows evidence of progressing beyond Pattern 4 of the nutrition transition. Importantly, such a behaviour is largely attributed to the lower consumption of less healthy foods (animal fats, sugars and milk).

Of perhaps more concern is that the other 90% of global population appears to be on a trajectory path of consuming an ever-increasing more calorific diet with no evidence of a slowdown. The largest cluster, CL2, is only 100 kilocalories less than CL3 in 2013 yet exhibits no signs of stabilisation observed for CL3. Starting at the lowest level of calories half a century ago, this cluster is now at the consumption level of CL3 in the mid-1990s. Extrapolating from this historical growth rate, CL2 would likely overtake CL3 to become the most calorific cluster in about ten years' time. The calorie content of this diet is accelerating, so is the rate of overweight and obesity. As CL2 is predominantly represented by populated countries such as China, the United States, Brazil, India, and Indonesia, the dire prediction that "a third of the global population will be overweight or obese by 2030" is not completely without foundation (Global Panel on Agriculture and Food Systems for Nutrition 2016). Given the dietary origin of obesity, the unceasing rise in calorie consumption acts as an important 'canary in the coal mine' for understanding the rise of obesity in populations with hitherto lower rates, for example Indonesia and India. Furthermore, there is no reason why the consumption level of CL3 represents a ceiling. Considering the current situation, CL2 and potentially others could overtake CL1 in the future.

Fluctuation analysis: clustering variables are pre-filtered data (detrended series) [3]

Figure 7.12 shows the values of FS and XB cluster validity index calculated for the number of clusters K ranging from 2 to 10. The FS exhibits a strong upward tendency and its value consistently increases when $K \geq 4$. This is a sign that the FS value is likely to go up and reaches the peak of a very large K ($K > 10$); therefore, the FS index cannot help to determine an optimal number of clusters. Turning to the XB series, three troughs are observed at $K = 2$, $K = 5$, and $K = 8$. Comparing the smaller XB values at these points reveals that the two-cluster solution is the best partition, followed by the five-cluster partition. Therefore, clustering results related to two-cluster and five-cluster solutions will be discussed subsequently.

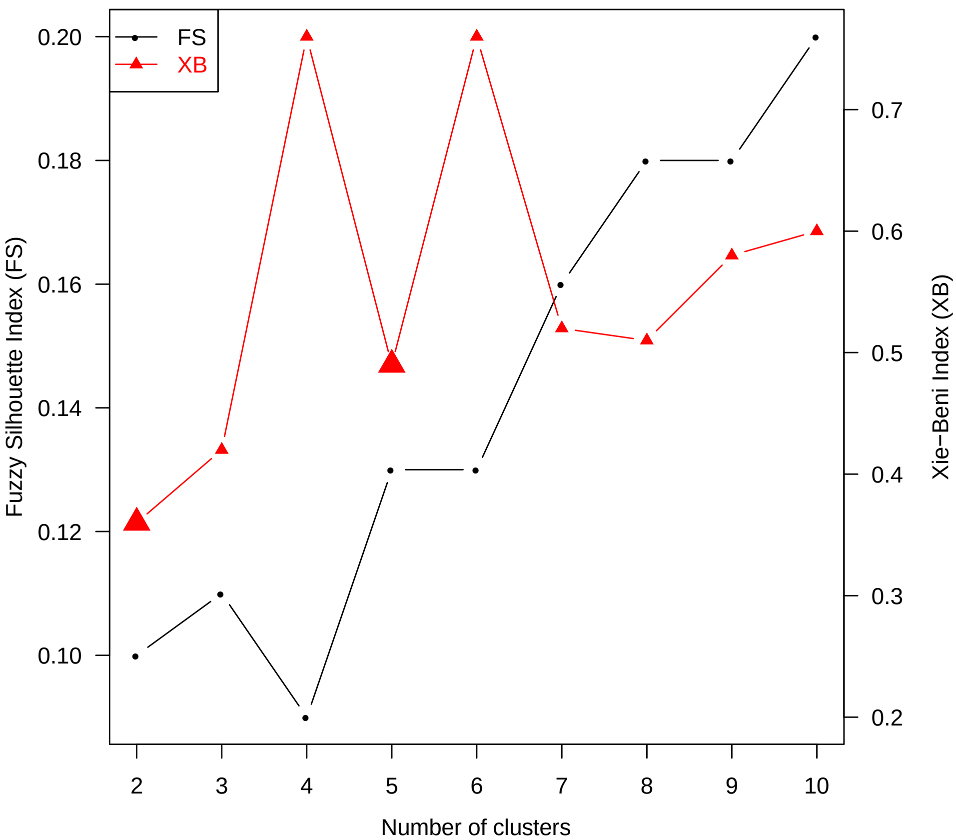


Figure 7.12 FS and XB validity index values for each cluster partition K from 2 to 10 (Fluctuation analysis).

Fixing $K = 2$, the two clusters obtained by the clustering algorithm are represented by Germany (CL1) and Namibia (CL2), accounting for 58% and 42% respectively of the world population. It is worth mentioning that this division differs significantly from the 90:10 split observed in the 'Trend

[3] The material contained in this section has been published in Le et al. (2020).

analysis'. Figure 7.13 illustrates the cluster representatives and the membership degrees of each country belonging to two clusters. The membership degree to which countries belong to each of these clusters is indicated by the shade of blue: the darker the shade the higher the membership degree. While most countries in Asia Pacific, Africa and Northern America tend to belong to two clusters with equal membership proportions, some exceptions are witnessed, for example India and Mexico are predominantly included in CL1 or New Zealand in CL2. A clear contrast is observed in Europe where Central European countries with Norway and Sweden are largely associated with CL1 with high membership proportions whereas those in Western Europe and Southern Europe mainly belong to CL2.

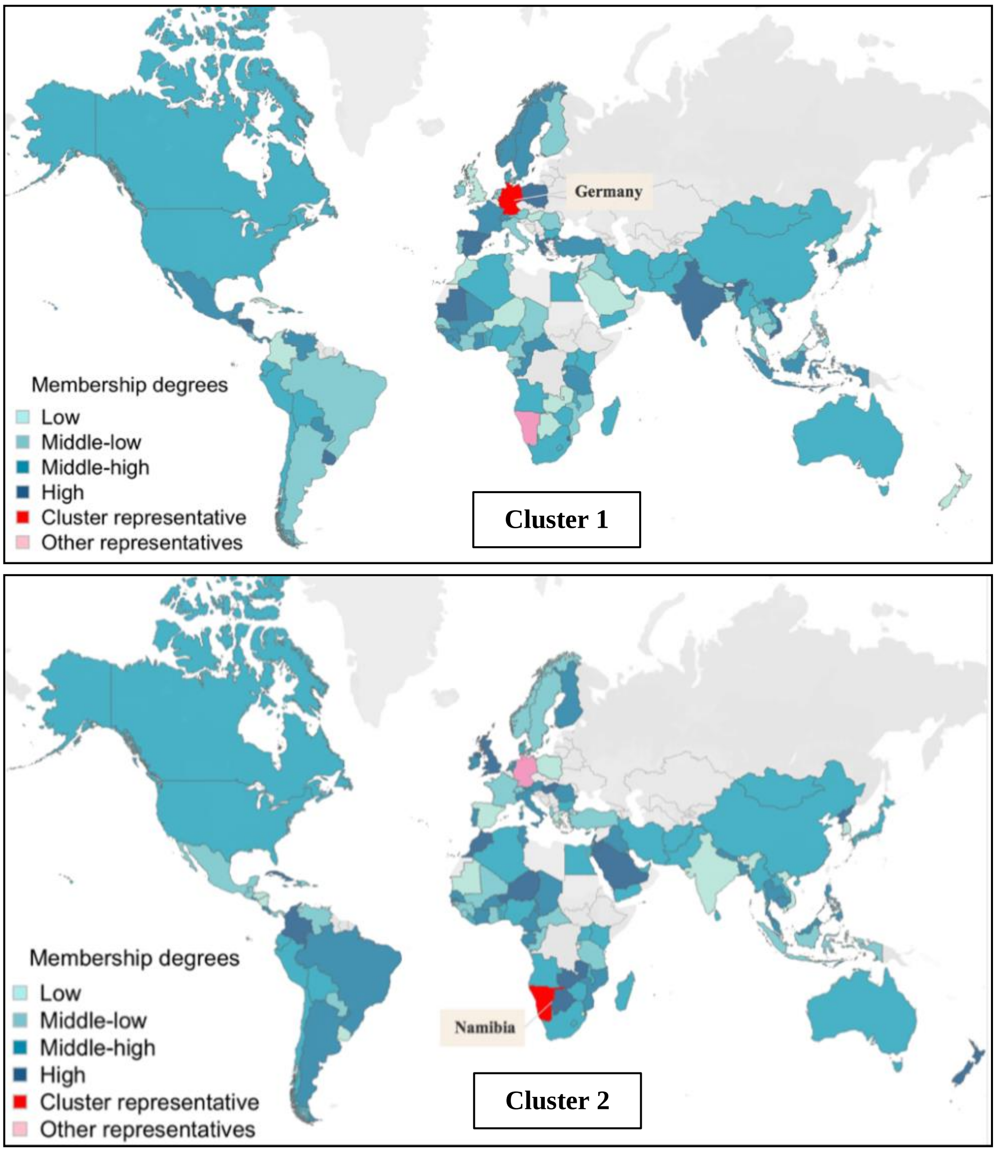


Figure 7.13 Cluster membership degrees ($K = 2$, Fluctuation analysis).

Figure 7.14 plots the weighted average calories from 1961 to 2013, with a strong increase in total calories over the past half a century, from about 2,250 kcal/person/day in 1961 to above 2,800 kcal/person/day in 2013. In particular, the two series are closely resembled in shape and in level over the period under examination – an unsurprising feature providing the somewhat 50/50 membership proportion of most countries belonging to two clusters. From the practical perspective, this alike behaviour of the two weighted average series can cause difficulties in policymaking recommendations. When the clustering results from the best partition are insufficiently informative, the analysis of the second-best partition is often recommended (D'Urso *et al.* 2019a).

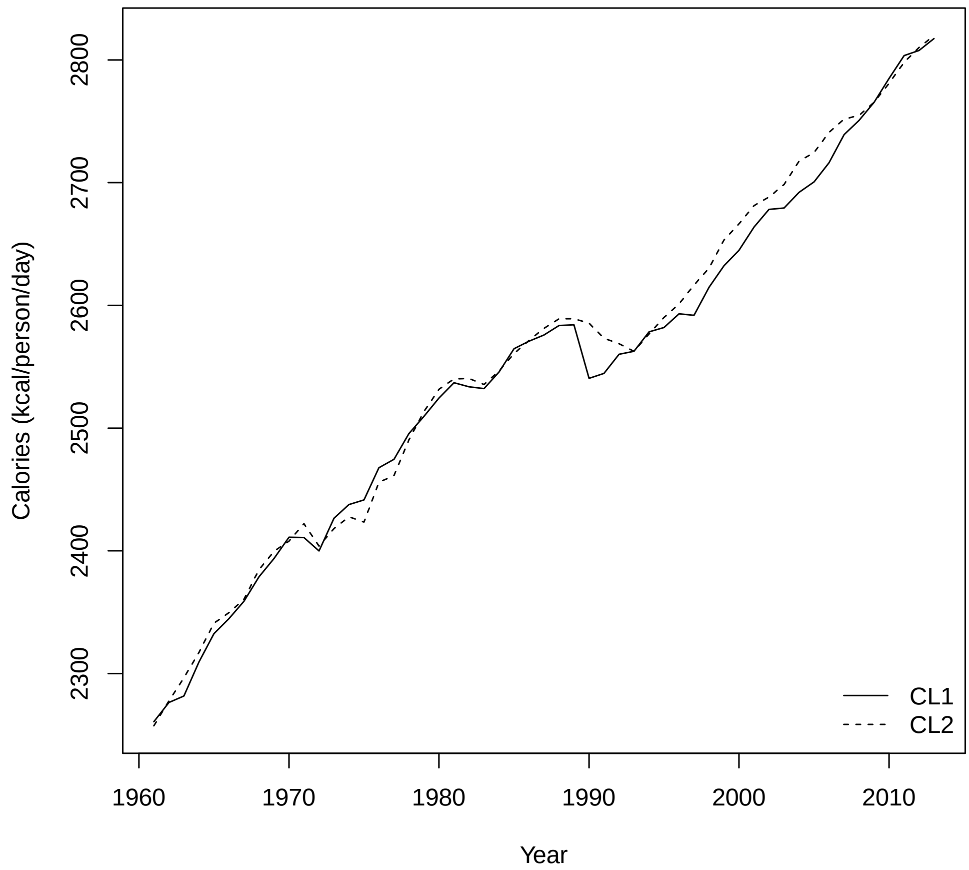
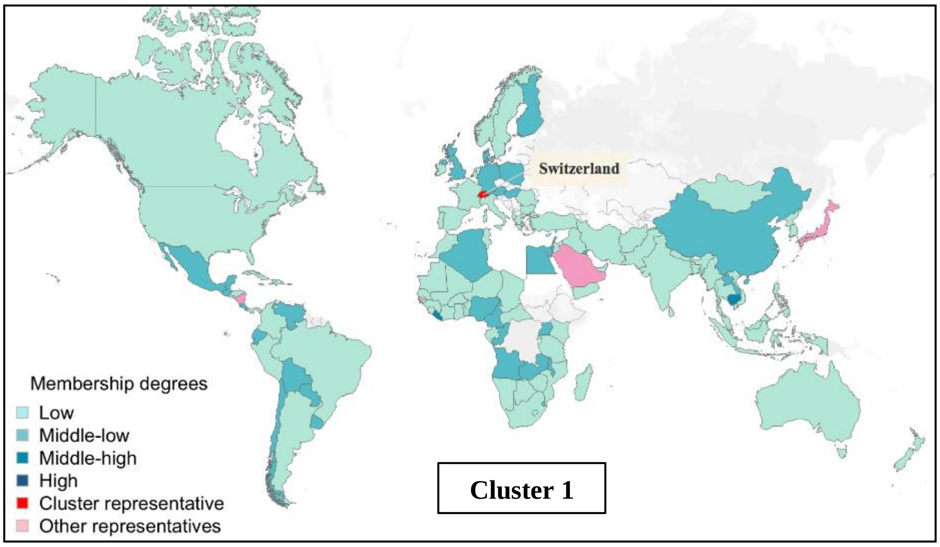
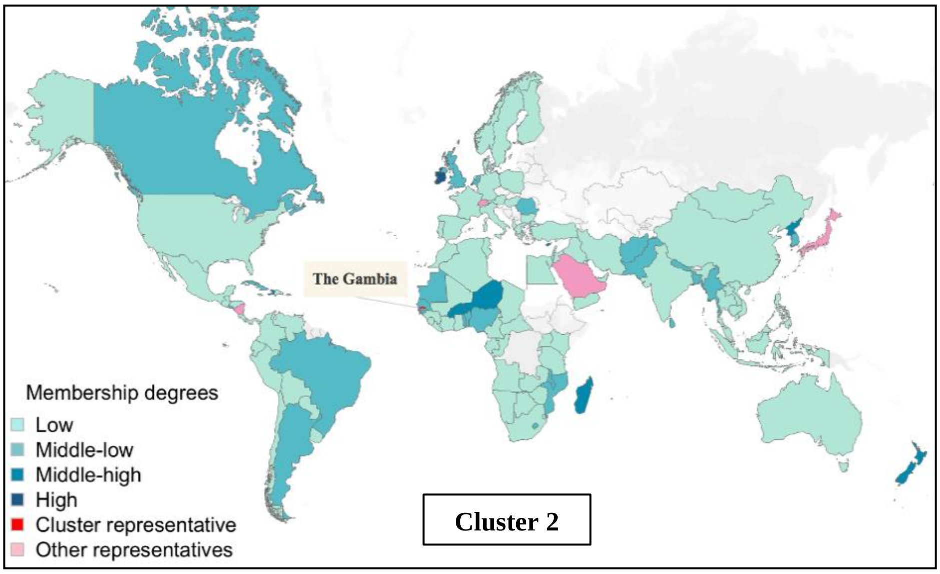
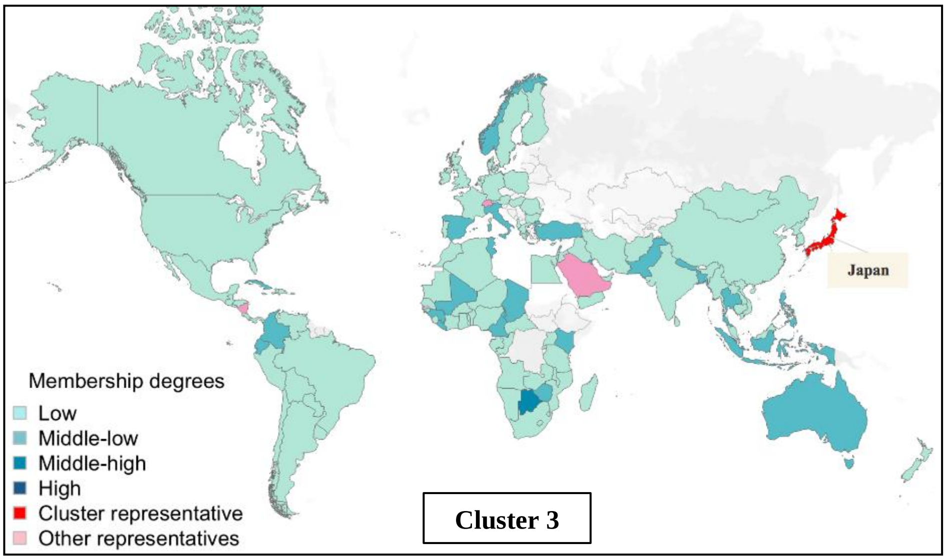


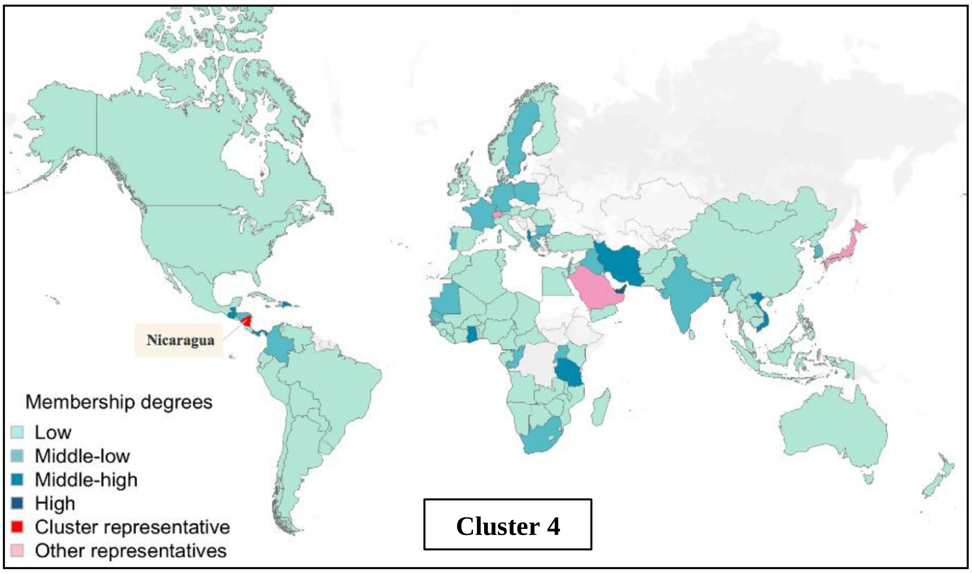
Figure 7.14 Weighted average daily per capita calories, 1961-2013 ($K = 2$, Fluctuation analysis).

When $K = 5$, the clustering algorithm returns five clusters represented by the following countries: Switzerland, The Gambia, Japan, Nicaragua, and Saudi Arabia. These clusters are of comparable sizes in terms of the world population. Figure 7.15 illustrates the cluster representatives and the membership degrees of each country belonging to five clusters. Some countries (depicted in the darkest blue shade) are predominantly associated with a single cluster (for instance, Ireland in CL2, Italy in CL3 and Vietnam in CL4), implying that for these countries the notion of a national diet may be a reasonable approximation. For many other countries which equally belong to multiple clusters, the coexistence of different dietary types is closer to reality. For example, Australia appears in CL3 and CL5, the UK in CL1 and CL2.


Switzerland
Membership degrees
Low
Middle-low
Middle-high
High
Cluster representative
Other representatives
Cluster 1


The Gambia
Membership degrees
Low
Middle-low
Middle-high
High
Cluster representative
Other representatives
Cluster 2


Japan
Membership degrees
Low
Middle-low
Middle-high
High
Cluster representative
Other representatives
Cluster 3



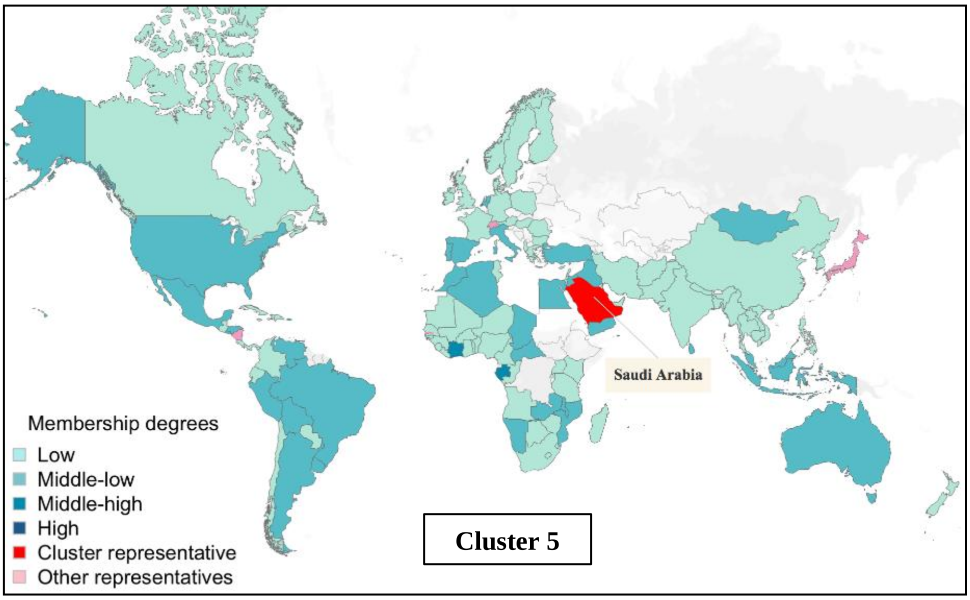


Figure 7.15 Cluster membership degrees ($K = 5$, Fluctuation analysis).

Figure 7.16 shows the changes of the weighted average calories from 1961 to 2013. Overall, two comments can be made. First, calorie availability has increased in all clusters during the past half a century on average by over 20% from around 2,250 to approximately 2,800 kcal/person/day, and the rising speed is more robust after 1992. However, there was a noticeable dip in 1990. In a FAO/WHO joint publication (WHO 2003, pp.14-15), what happens to the calorie availability in early 1990s is attributed to a decline in transition countries: "The increase in the world average consumption would have been higher but for the declines in the transition economies that occurred in the 1990s". As countries contribute to five clusters with roughly equal membership proportions, the declining pattern is observed for all five weighted average series. Since all clusters have become more calorific over time, it seems that the 'Fluctuation analysis' is unable to identify the niche cluster characterising reduced calorie consumption previously seen in the 'Trend analysis'. This finding can be explained by the choice of the clustering variables which are pre-filtered data (detrended series) for the 'Fluctuation analysis', and as a result, some systematic information related to the continued variability of time series such as direction (upward/downward) and speed of change is removed by pre-filtering.

Another important message from Figure 7.16 is that the gaps between calorie trajectories are becoming narrower, indicating sigma convergence. Moreover, this converging pattern is more evident after 1990, suggesting a rise in the convergence speed in the last couple of decades. Figure 7.16 also implies that countries with lower initial calories in 1961 tend to exhibit higher growth rates and countries are approaching towards a steady-state level (the 'catching-up' effect in beta convergence literature). To illustrate, the calorie content of CL4 was the second smallest among five clusters in 1961 yet grew quickly and joined the most calorific group (with CL1 and CL3) in 2013. In particular, the smallest average annual growth rate of CL1 (0.43%) gives evidence that clusters are converging towards CL1 which can be described as the most calorific cluster.

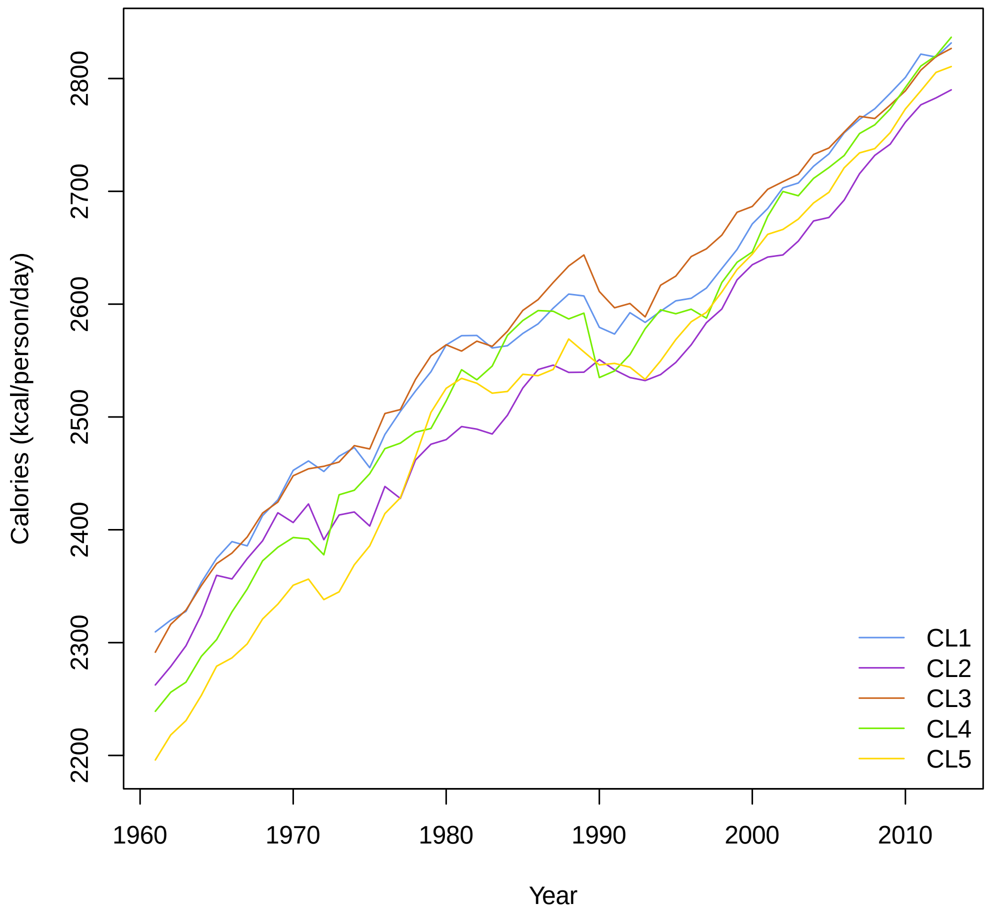
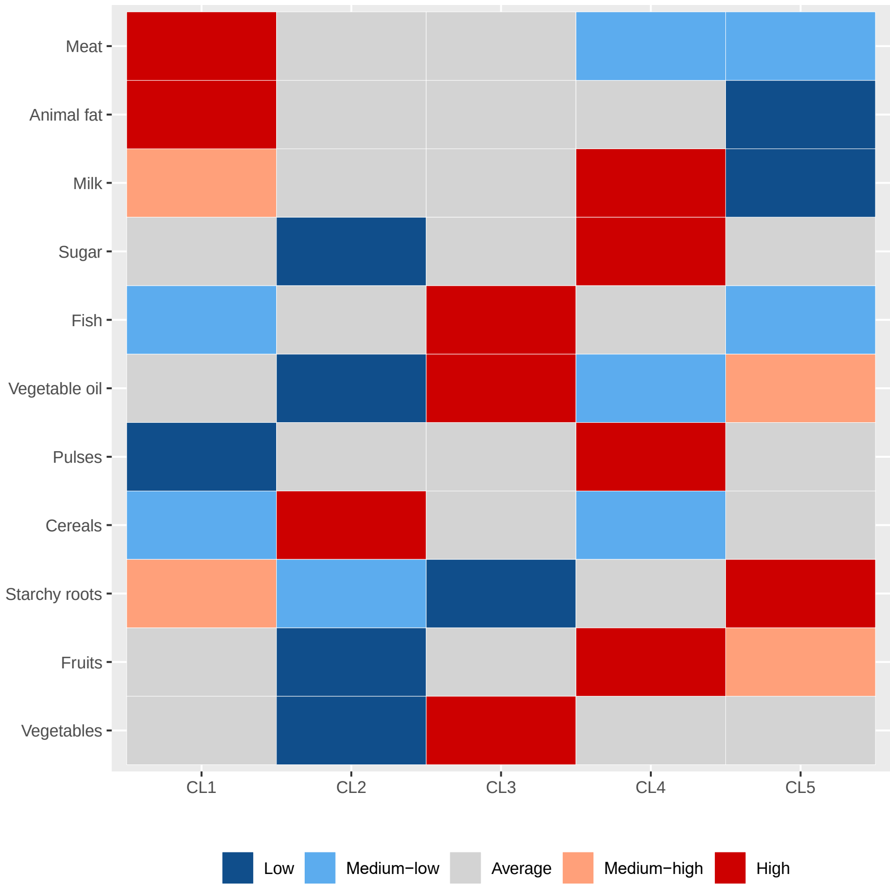


Figure 7.16 Weighted average daily per capita calories, 1961-2013 ($K = 5$, Fluctuation analysis).

So far, clusters are defined solely based on the historical calorie availability. The question is whether it is possible to attach a meaningful dietary type (based on diet composition) to each cluster. Figure 7.17 compares the average dietary characteristics of five clusters in terms of eleven food aggregates (meat (including eggs), animal fat, milk, sugars, fish, vegetable oil, pulses, cereals, starchy roots, fruits, and vegetables). These food groups approximately represent any diet and make up the calories available for consumption. With red (blue) indicating food groups consumed more (less) than the global average, this provides an indication of the dietary composition of each cluster that has been identified. CL1 is best described as having a typical 'Western diet' with high intakes of meat and animal fat, fairly high intakes of dairy products but low intakes of cereals and pulses. Diet of CL2, comprising mostly of basic staples such as cereals but little of sugars, vegetable oils, fruits or vegetables, could be deemed as the 'Traditional diet'. CL3 enjoys a diet rich in fish, vegetable oils, and vegetables. These dietary characteristics resemble the 'Mediterranean diet'. Regarding CL4, most of the calories come from milk, sugars, pulses, and fruits whereas the energy contribution of cereals and vegetable oils remains limited. This diet is coined 'Tropical diet'. CL5's diet emphasises starchy roots, but is loosely based on animal fats and milk, thus is named 'Vegetarian diet'. While the labels used here are merely convenient descriptors, the fact that clusters map quite neatly into broad dietary types provides a convenient way to categorise the diets that underlie national food availability figures. It also provides a basis to assess the health implications of these major dietary types subsequently.



Note: High/low indicates the value outside the range (global average ± 1 standard deviation);
Medium-high/medium-low indicates the value inside the range (global average ± 0.5 standard deviation).

Figure 7.17 Heat map of dietary composition of the five clusters.

In order to examine the healthiness of these dietary types, the MAI values are calculated. Figure 7.18 illustrates the changes of MAIs over the period 1961-2013. A downward trend is evident in all clusters but at varying speeds. Over the past 20 years, all diets except the 'Traditional diet' have been worsening sharply. Indeed, the declining rate is most rapid for the seemingly least healthy diets. On the other hand, the deterioration in MAIs seems to slow down after 1995 as compared to the historical pattern in the 1960s. Another striking feature from Figure 7.18 is that the MAIs peaked up in 1975 for all clusters. This could be due to two effects: price effect and income effect. The former refers to a rise in the price of some food commodities, hence lowering the consumption of unhealthy food (for instance sugars). The latter refers to a fall in income, creating a shift towards cheaper foods (for example cereals or vegetables).

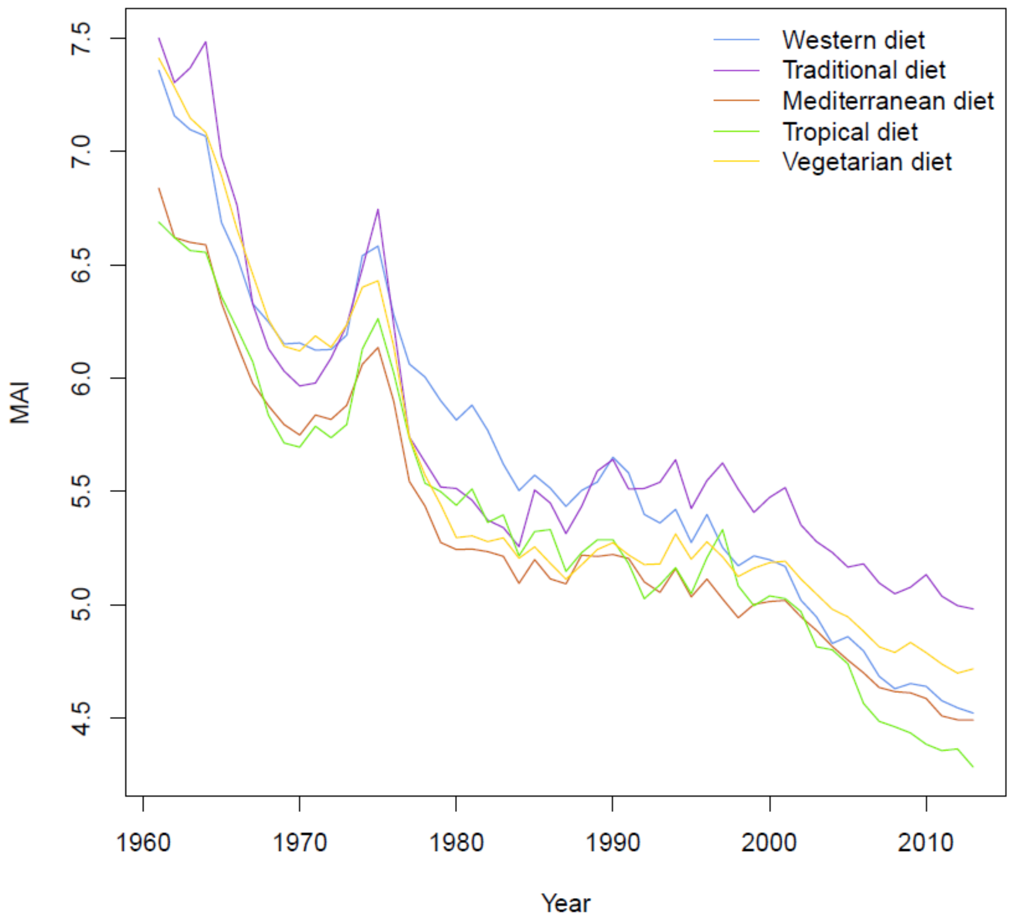


Figure 7.18 Mediterranean Adequacy Index for the five clusters, 1961-2013.

An explanation for the declining healthiness of five dietary types is put forward in Figure 7.19. All diets are replacing carbohydrates with fat, equating with a move from plant-based to animal-based foods and providing evidence for the *nutrition transition*. Globally, diets have become increasingly unhealthy owing to the confluence of two related facts: not only are we eating more calories, typically in processed forms but that these calories are increasingly from animal-source foods. Further, the decline in energy expenditure over time due to sedentary lifestyles allows the effect of these dietary changes to become apparent. The obesity problem has both quantity and quality dimensions.

It is clearly observed from Figure 7.19 that all diets have behaved and changed in the same way. While this uniform behavior is not unexpected from the nutritional point of view, some technicalities are worth mentioning. The clustering inputs for this Fluctuation analysis are detrended time series based on not the level of calories, but shocks to calorie consumption (for example, due to natural disaster, conflicts, war, economic crisis, or political instability). Because the drivers for dietary changes are often common across countries, the derived clusters behave similarly in many aspects. While the analysis solely utilises one variable – total calorie availability, fuzzy clustering is the best attempt to reflect the different dynamics within each country. However, unless the clustering algorithm is explicitly spatial, the role of different spatial dynamics across geographies would not be sufficiently accounted for.

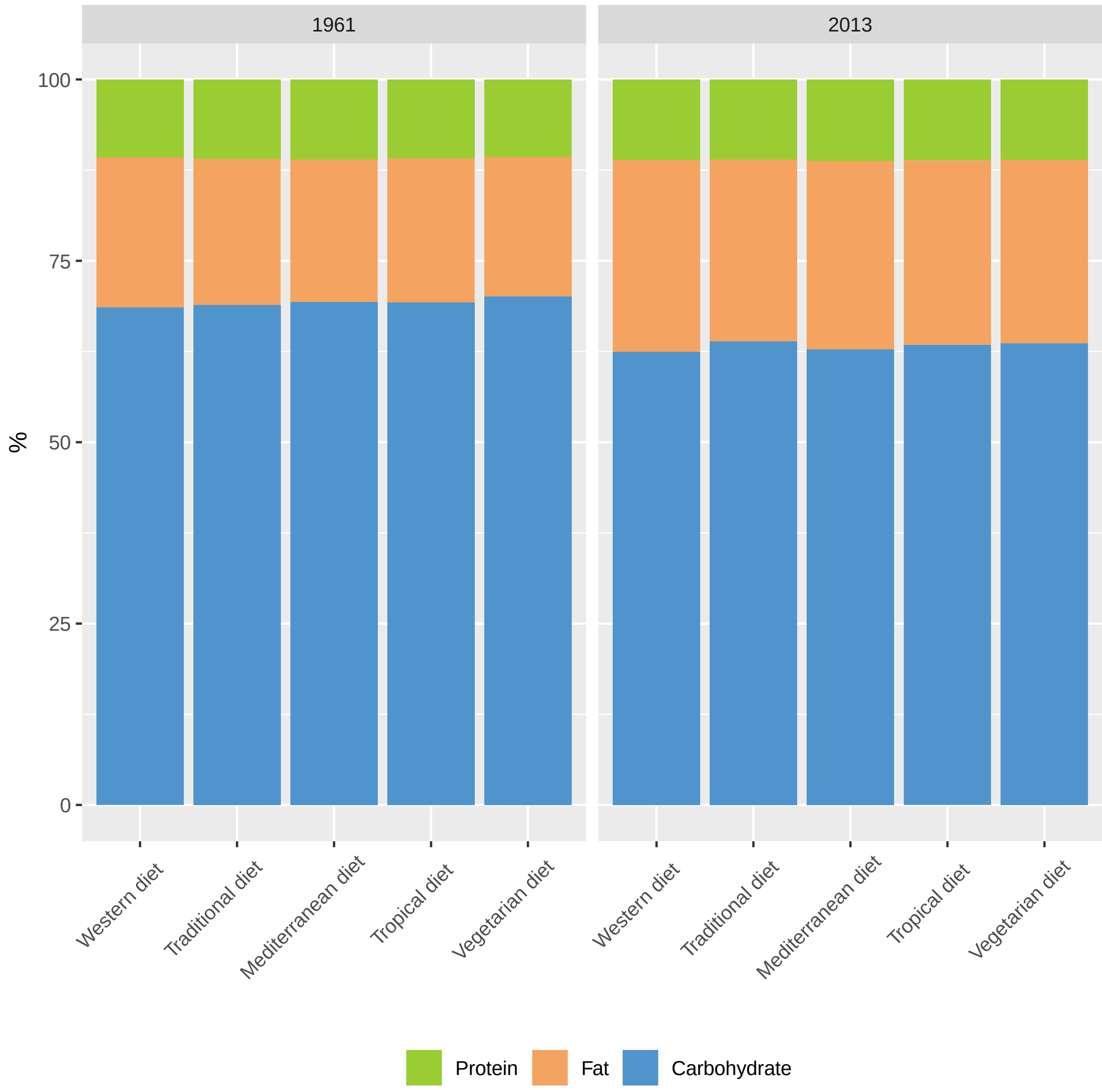


Figure 7.19 Macronutrient composition of the five clusters, 1961 and 2013.

As a pessimistic picture of deteriorating dietary healthiness emerges worldwide, a crucial question to ask is: "Do those dietary changes couple with any pay-off in health status?". Table 7.5 reveals rather optimistic news: rising life expectancy in all clusters. Nonetheless, the figures for HALE (healthy-adjusted life expectancy) tell a different story: the world is enjoying shorter years living at "full health". A new-born nowadays expects to gain somewhat five years as compared with babies born at the start of the millennium; nevertheless, these longer lives are predicted to be spent in poor health. Even though poorer diet is certainly not everything to blame for, wide-ranging evidence points out that globally over one-fifth of all adult mortality being linked to dietary factors (Afshin *et al.* 2019).

Table 7.5 Changes in Life Expectancy (LE) and Health Adjusted Life Expectancy (HALE) in years since 2000 (weighted averages by country).

	Western diet	Traditional diet	Mediterranean diet	Tropical diet	Vegetarian diet	*p*-value
LE[a] (years at birth)	5.2 [66.0-71.2]	5 [65.0-70.0]	5 [66.1-71.1]	4.7 [67.1-71.8]	4.8 [65.3-70.1]	0.508
HALE[b] (years at birth)	-5 [63.0-58.0]	-4.9 [62.1-57.2]	-4.9 [62.9-58.0]	-4.5 [63.6-59.1]	-4.8 [62.1-57.3]	0.501

Note: Repeated measures ANOVA test is used to determine whether the weighted mean value differs among the clusters identified. All test results are not significant unless indicated otherwise. Figures inside square brackets represent values for the beginning and ending period.

*, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.

[a] change between 2000-2013; [b] change between 2000-2015.

Data are retrieved from the World Bank (World Bank 2020e).

In summary, the 'Fluctuation analysis' identifies five clusters based on common shocks in the calorie availability. Although the algorithm solely utilises the information on the patterns of changing calorie consumption, the identified clusters match neatly with distinguished nutritional composition of diets. Nonetheless, these clusters do not vary much in either the behaviour of the calorie trajectories or the implications for health. All clusters strongly exhibit an increasing trend in the calorie availability while the inspection of Figure 6.8 reveals a decrease in calorie availability over the period for approximately 10% of the study sample. On the other hand, the large *p*-values of the repeated measures of ANOVA test in Table 7.5 imply that the weighted average changes in both life expectancy and healthy-adjusted life expectancy among five clusters are not statistically different. Overall, the effect of detrending is apparent.

On a final note, as cluster analysis is an exploratory tool to summarise data meaningfully, there is no right or wrong answer (when one decides to interpret the clustering results corresponding to a larger or smaller number of clusters). That said, some cluster validity statistics (for example, the fuzzy silhouette index and the Xie-Beni index) can offer some guidance on the optimal number of clusters from a technical point of view. In this research, these indices tend to suggest two-cluster solution being the best clustering solution in terms of both internal homogeneity and external heterogeneity. Yet, an analysis with K=2 is not sufficiently informative as one cluster is often more predominant accounting for a substantial proportion of the global population. Therefore, the number of clusters was increased gradually to say K=4 and K=6 in the order suggested by the cluster validity statistics and a more nuanced picture is revealed. The primary purpose is to see what happens when more clusters are allowed in the model: do the clusters further break up? and if they do, which ones? Taking the 'Trend analysis' as an example, the clustering solution with K=2 corresponds to two trends: rising calorie consumption over time versus reducing calorie consumption. Increasing K to 4 and 6, the former cluster is further split up into smaller clusters representing monotonomic rising trend and rising trends with a dip. In this case, showing the different clustering solutions for varying K can be thought of as a kind of sensitivity check.

7.4.2 Results of the COFUST clustering algorithm

The empirical study in Chapter 6 discovers a spatial autocorrelation process among national food consumption patterns. It has shown that the rate of convergence is underestimated when spatial effects are ignored, and it is therefore of crucial importance to appropriately handle such a relationship if it exists in the data. Thus, motivated, the second part of this empirical analysis brings in the spatial context and aims to identify agglomerations of countries characterised by similarity of patterns in food consumption considering the particular spatial relationship.

To this aim, the univariate time series of 118 countries and their spatial relationship (denoted by the proximity matrix) are analysed by the COFUST clustering algorithm. The proximity matrix S is specified in the same fashion as the matrix W^c which is the only one to yield significant spatial linkages among the proposed proximity measures. As discussed in Section 7.2.2, the clustering algorithm is performed for actual data ('Spatial trend analysis') and pre-filtered data ('Spatial fluctuation analysis'). Although countries within the same cluster are required to be close in the degree of economic development, the 'Spatial trend analysis' identifies those with similar evolution in the calorie consumption whereas the 'Spatial fluctuation analysis' detects those with similar deviation from the trend.

Spatial trend analysis

Figure 7.20 summarises values of the FS index calculated for the number of clusters K ranging from 2 to 10 and for β from 1 (no spatial information) to 0.5. The reason for choosing such a range of β values is that the influence of spatial information is too high for any value of β lower than 0.5 (Disegna *et al.* 2017). The FS trajectories show that the FS index is highest when $K = 2$ for varying β values, suggesting two-cluster solution be the best partition. From the managerial and practical point of view, a two-cluster solution sometimes cannot be fully informative, and the analysis of the second-best partition should be taken (D'Urso *et al.* 2019). As can be seen from Figure 7.20, the second peak in FS trajectories seems to vary for different values of β, making it impossible to determine the optimal values for K and β simultaneously.

In order to choose K and β, the heuristic procedure mentioned in Section 7.2.2 is adopted. The optimal number of clusters K is selected when no spatial information is considered (i.e. $\beta = 1$). Fixing K, the value of β that maximises the spatial autocorrelation among post-cluster units is chosen. This approach allows the cluster formation to be largely based on the dependence among time series, and the spatial information only fine-tunes the final cluster partition, perhaps by some adjustments in the cluster membership degrees.

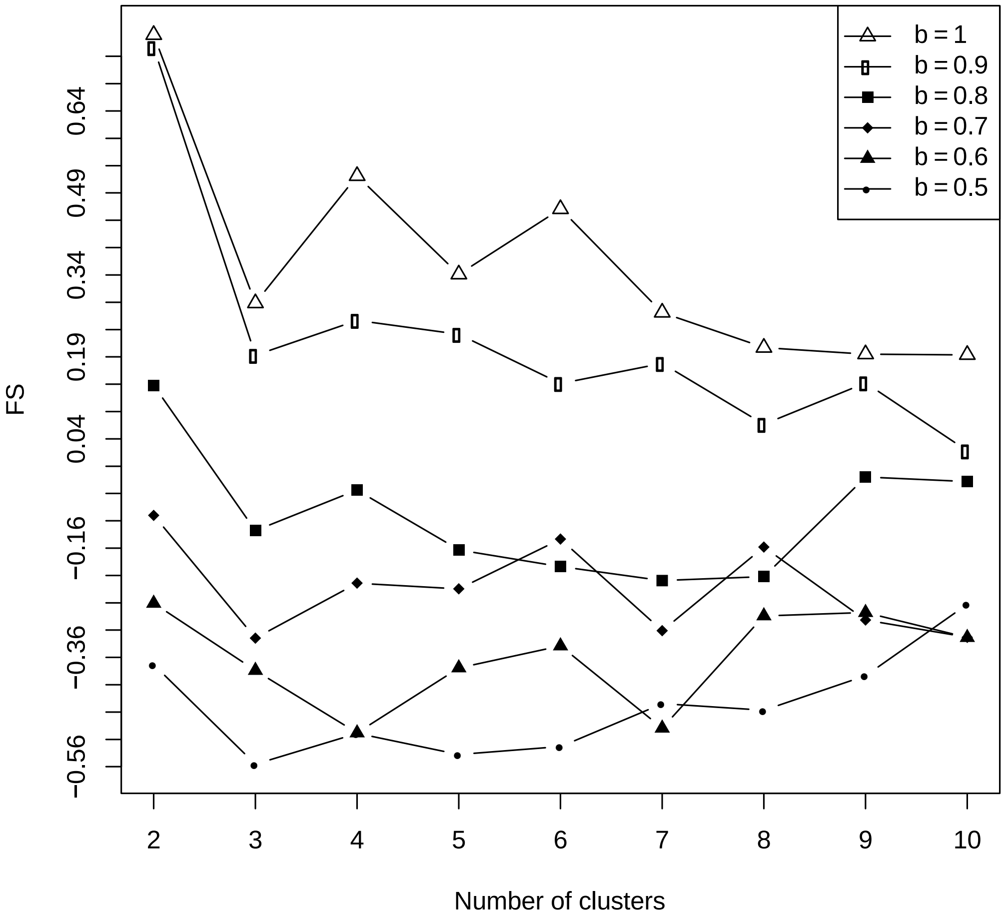


Figure 7.20 Values of the Fuzzy Silhouette index for each cluster partition for K varying from 2 to 10 and β from 1 to 0.5 (Spatial trend analysis).

As can be seen in Figure 7.20, the FS trajectory for $\beta = 1$ reaches the peaks at $K = 2,\ K = 4$ and $K = 6$, suggesting that two-cluster partition is the best solution followed by four-cluster and six-cluster partitions. Let K take the value of 2, 4 and 6, the Generalised Fuzzy Moran index is computed for a range of β from 1 to 0.5 to measure the spatio-temporal autocorrelation among post-cluster units in each scenario. The results are plotted in Figure 7.21. The figure's inspection reveals that the spatio-temporal autocorrelation is maximised when $\beta = 0.7$ for two-cluster, $\beta = 0.5$ for four-cluster solution, and $\beta = 0.9$ for six-cluster solution. Thus, the subsequent discussion will focus on the clustering results obtained from the following combinations: ($K = 2$ and $\beta = 0.7$), ($K = 4$ and $\beta = 0.5$), ($K = 6$ and $\beta = 0.9$).

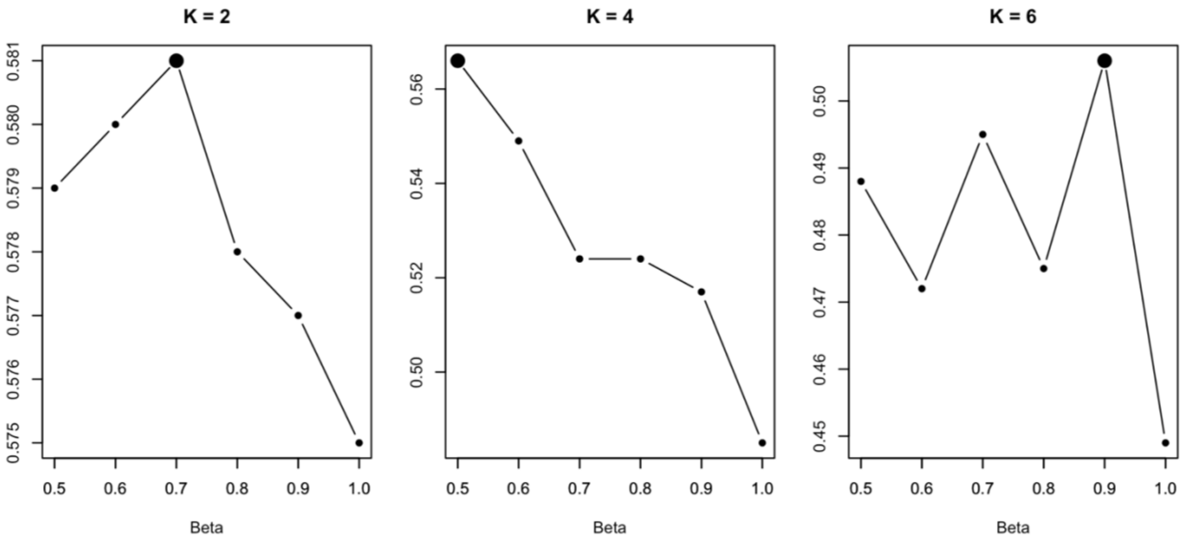


Figure 7.21 Values of the Generalised Fuzzy Moran index for β ranging from 1 to 0.5.

Fixing $K = 2$, the COFUST algorithm returns two clusters (CL) of comparable sizes, each represents approximately half of the world population. Figure 7.22 illustrates the membership of countries belonging to two clusters by the shade of blue: the darker the colour, the higher the membership proportion. In each world map, the country representative ("medoid"), whose calorie trajectory is the most exemplar of the cluster, is coloured in red while the medoids of other clusters are denoted by a shade of pink. CL1 is mainly made up of developed countries across Europe, Northern America, Australia and some developing countries scattered across other regions. On the other hand, many developing countries in Africa, Central America and China are associated with CL2 with very high membership degrees.

It is shown in Table 7.6 that cluster medoids change and the niche cluster which account for only 6% of the global population disappears when the spatial information is considered. As discussed in the previous chapter, this cluster corresponds to countries that experience a fall in calorie availability as opposed to the predominant rising trend experienced by the majority. Without spatial information the time series clustering algorithm in Chapter 6 reveals two major dietary trends: increasing versus decreasing calories over the past 50 years (CL1 and CL2 respectively in Figure 7.23b). Considering spatial information, the COFUST algorithm (in Figure 7.23a) detects two clusters: one corresponds to a monotonic increase of calorie consumption (CL1) similar to the non-spatial model and one to an increase with a dip (CL2). Specifically, the difference between CL2 in the spatial and non-spatial models lies in the starting value. The former is associated with the least calorific diet in 1961 whereas the latter with the initially most calorific diet. It was shown in Chapter 6 that the group of countries witnessing falling calories includes both poor countries (say Afghanistan) and the rich ones (say Switzerland). In this chapter, since the objective of cluster formulation is to not only maximise the similarity in the historical trends but also consider those with comparable incomes, such requirements reduce the possibility of grouping together countries with wide income disparities in one cluster.

Table 7.6 Cluster solutions with and without spatial information ($K = 2$).

Cluster	Without spatial information (β = 1)		With spatial information (β = 0.7)	
	Medoid	World population (%)	Medoid	World population (%)
CL1	China	93.7	Honduras	55.1
CL2	Zimbabwe	6.3	Trinidad and Tobago	44.9

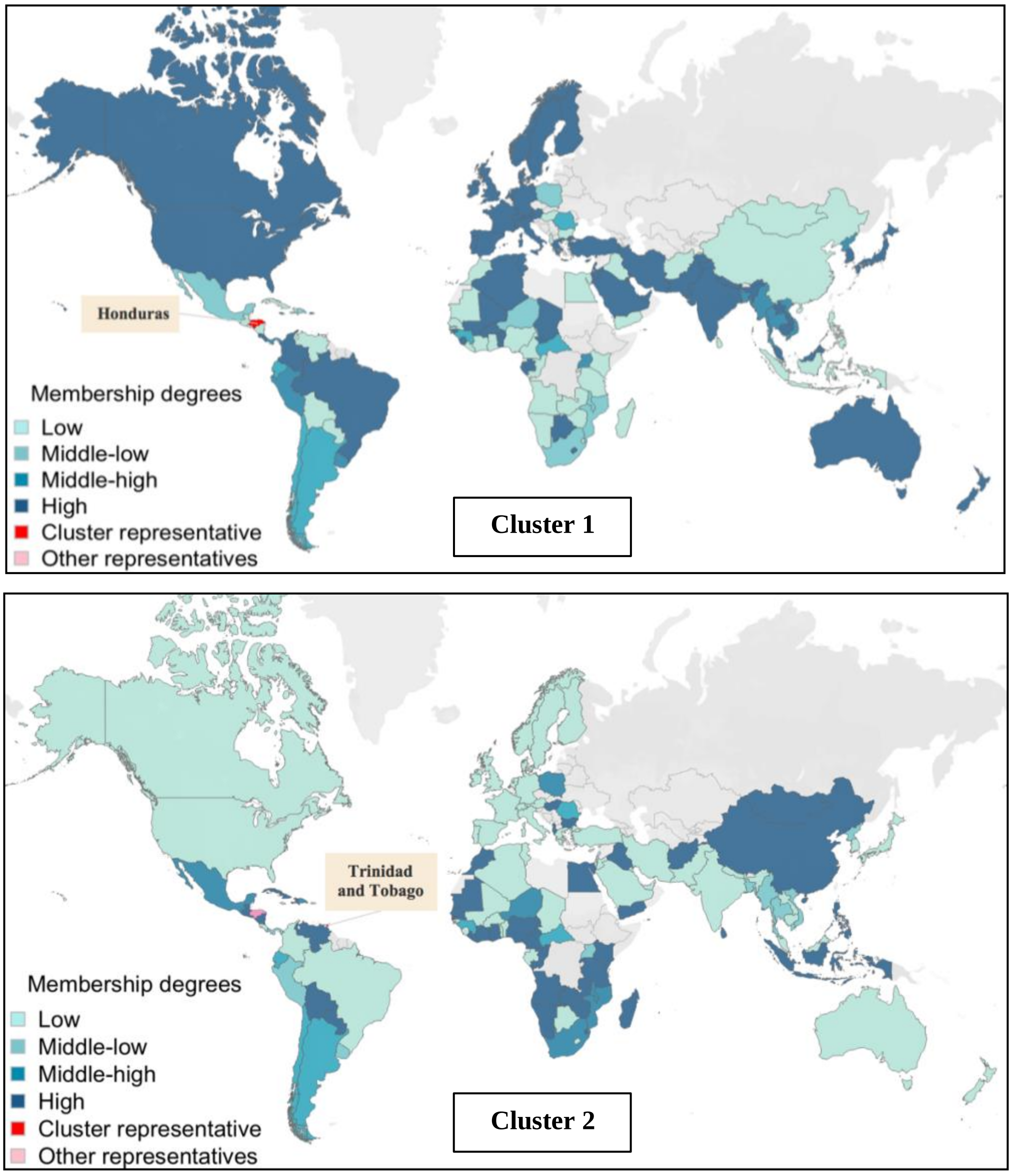


Figure 7.22 Cluster membership degrees when $K = 2$ and $\beta = 0.7$.

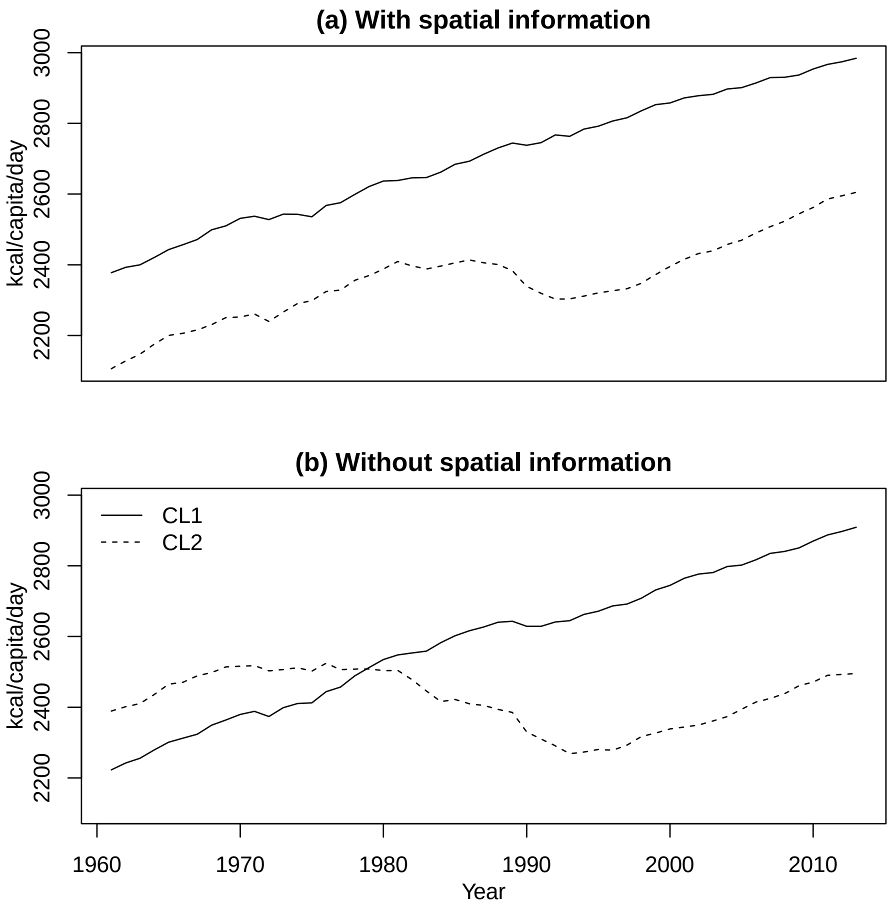
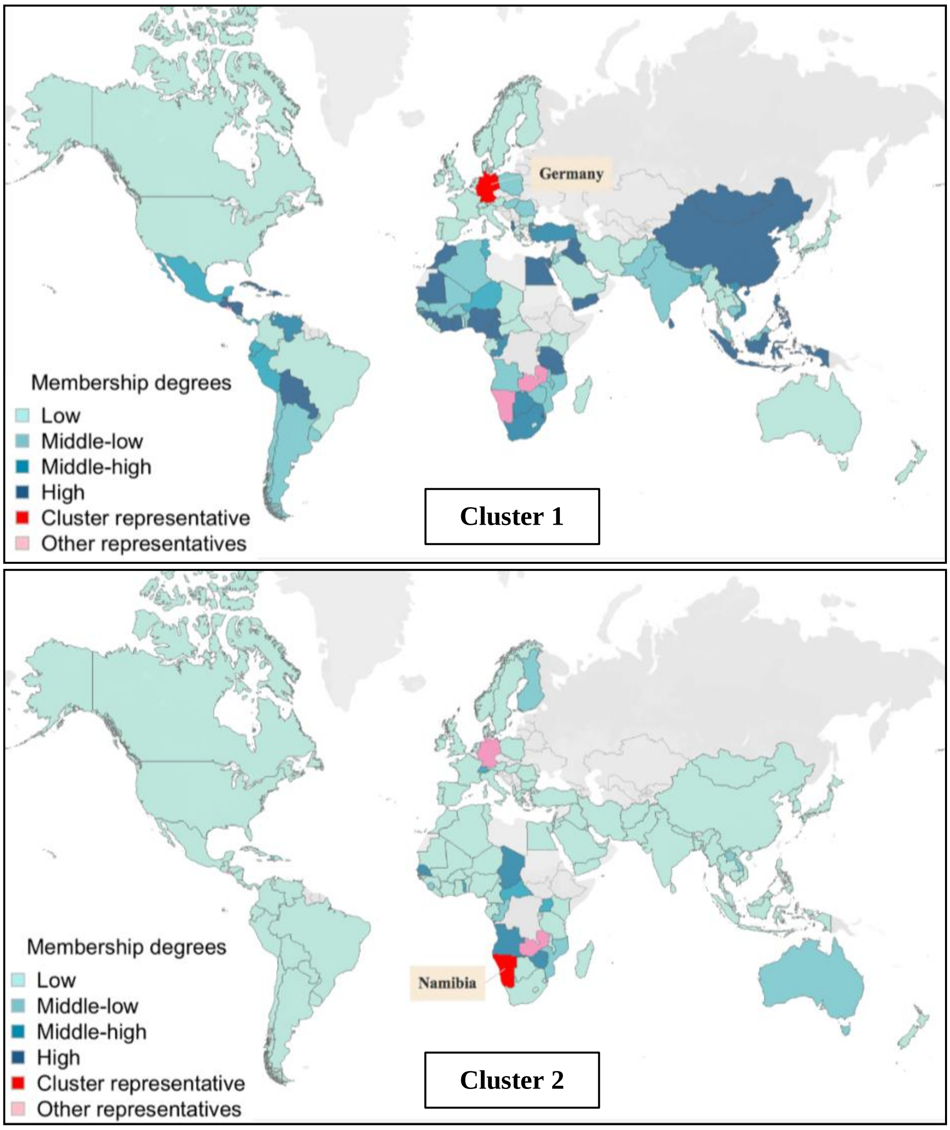


Figure 7.23 Weighted average daily per capita calories, 1961-2013 ($K = 2$ and $\beta = 0.7$).

Regarding the four-cluster solution, Table 7.7 summarises the clustering results. All cluster medoids change when the spatial information is considered. CL1 accounts for the majority of the world population, CL3 is the second largest cluster whereas CL2 and CL4 are niche clusters – each represents about 3% of the global population. The geographical distribution of membership degrees of each country along with the cluster medoids is illustrated in Figure 7.24. There is a concentration of high membership proportions in CL1 and CL3 as expected from their biggest cluster sizes. For many countries, the notion of a national diet is a reasonably valid assumption due to the predominance of membership degrees in one cluster, for example China and Indonesia (CL1), or Canada and the USA (CL3). However, other countries seem to subscribe to the idea of different diets coexisting as reflected by the fuzzy clustering approach. This is best demonstrated by for instance Mexico which equally belongs to CL1 and CL3 with roughly 50-50 membership proportion. When the space-time clustering algorithm is performed, one would expect to see the influence of the 'space' in the final cluster solution, i.e. countries with similar incomes tend to be grouped together. This feature is witnessed in some clusters. To illustrate, CL3 is made up of many Northern and Western European countries (say the UK, Norway) and their high-income peers (Japan, New Zealand, Canada to name a few). Yet, exceptions are spotted for example Bulgaria – a middle-income country is mainly associated with CL4 alongside some low-income countries (say Afghanistan or Madagascar). Thus, the inclusion of the spatial information does not compel final clusters to be made solely by countries with alike income levels.

Table 7.7 Cluster solutions with and without spatial information ($K = 4$).

Cluster	Without spatial information (β = 1)		With spatial information (β = 0.5)	
	Medoid	World population (%)	Medoid	World population (%)
CL1	China	73.8	Germany	50.5
CL2	Ghana	13.4	Namibia	2.5
CL3	Iraq	8.6	El Salvador	43.8
CL4	Zimbabwe	4.2	Zambia	3.2



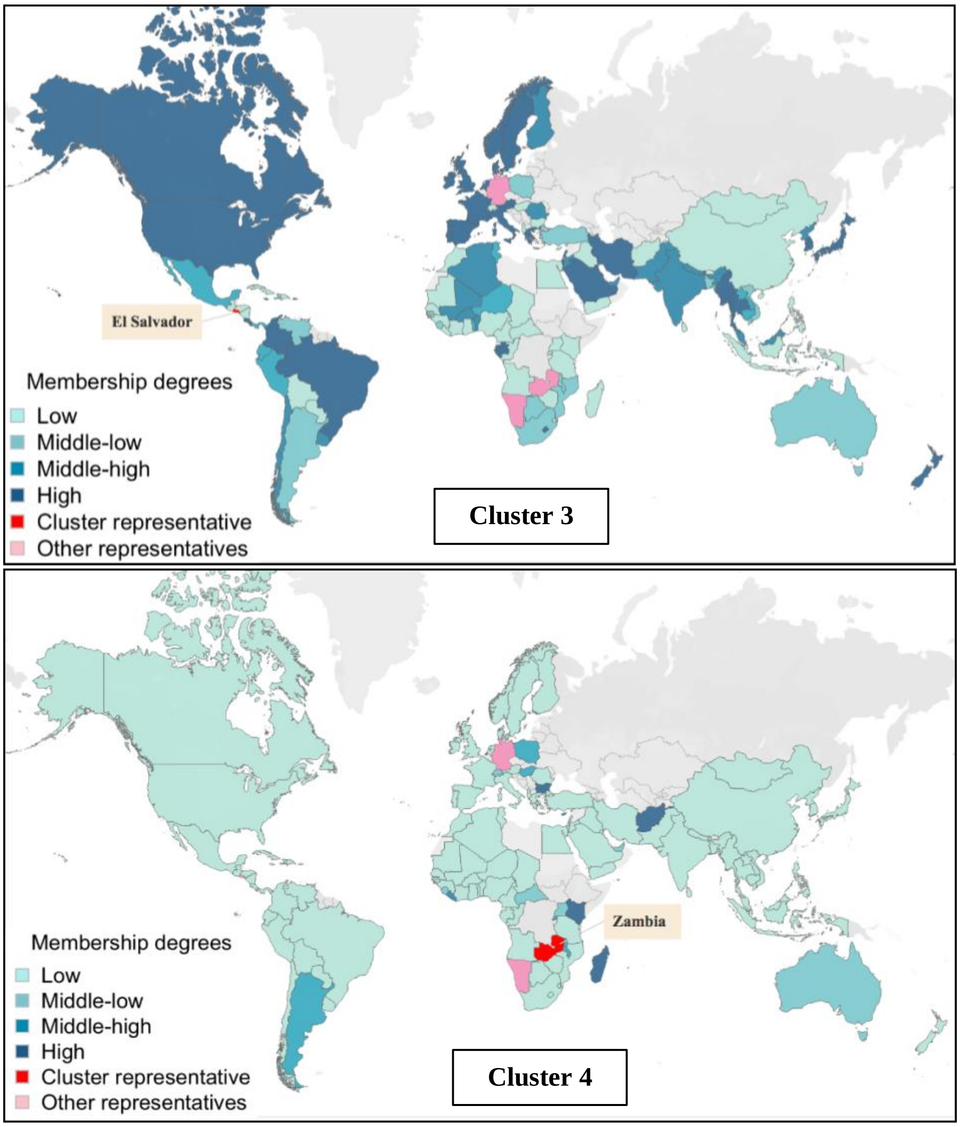


Figure 7.24 Cluster membership degrees when $K = 4$ and $\beta = 0.5$.

To better understand the four clusters obtained by the space-time algorithm, the trajectories of calorie consumption are plotted together in Figure 7.25a. Compared with the two-cluster solution, raising the number of clusters K to four does not impact the largest cluster which is characterised by a monotonic rising calorie consumption (CL3). Instead, the other cluster is split up into two bullish trends with a dip in the 1970s and 1990s (CL2 and CL1 respectively) and a reversal (CL4). Half of the world population experiences a high-calorie diet (CL3) whose calorie content has increased dramatically while four in ten individuals experience a diet (CL1) that evolves from a very low initial calorie level to the second most calorific in 2013. An upward trend is also witnessed in CL2 but with a long delay. In fact, the calorie consumption of CL2 was stagnating for the first three decades and began to shoot up strongly in the early 1990s at a more or less similar speed with the majority. Furthermore, since CL2 comprises of a small number of sub-Saharan African countries such as Angola, Chad, Zimbabwe with the membership proportion of over 60%, this cluster might resemble the diet of poor countries that were hit by the food crisis in the 1970s but have taken advantage of the globalisation process accelerated in the 1990s. CL4, on the other hand, exhibits an ‘odd’ trend approximating a diet that was once the most calorific but has reversed the increasing tendency since the 1970s. It should be noted that over the past 15 years the calorie consumption of CL1 and CL3 is characterised by a rising trend that seems unceasing, whilst that of CL4 has improved modestly and is about to converge with the calorie level of CL2. Interestingly, this pattern is not observed in Figure 7.25b in which clusters share the common theme of surging calorie consumption (to distinct levels) since the mid-1990s.

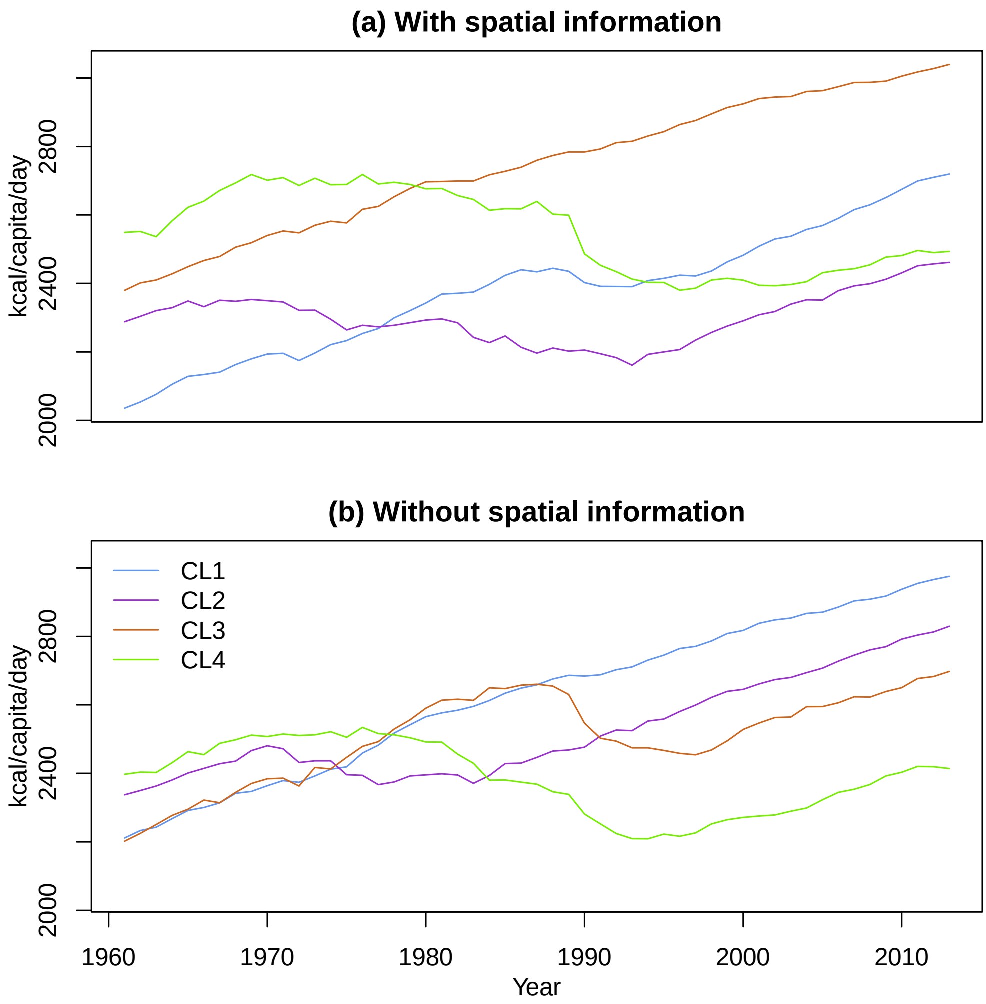
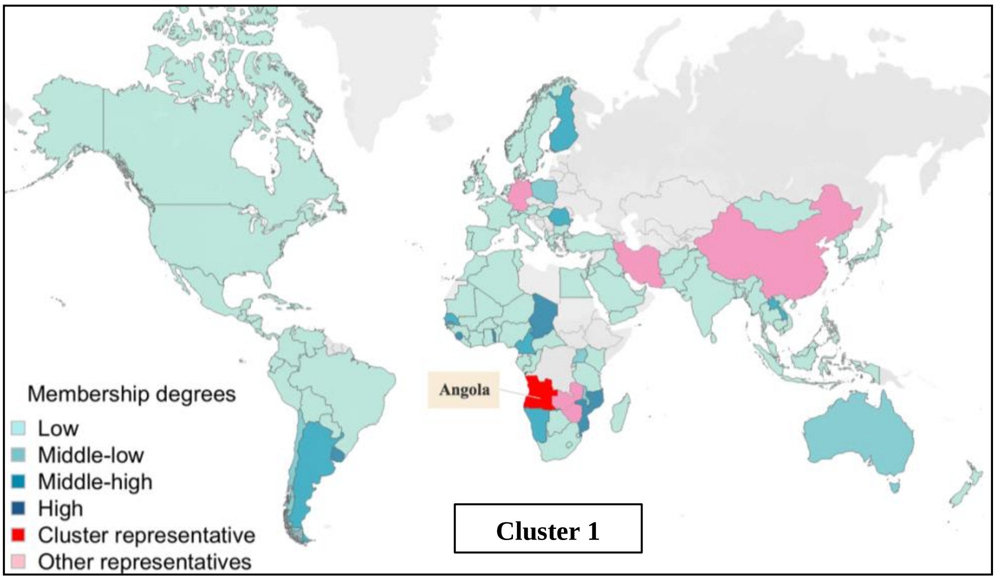


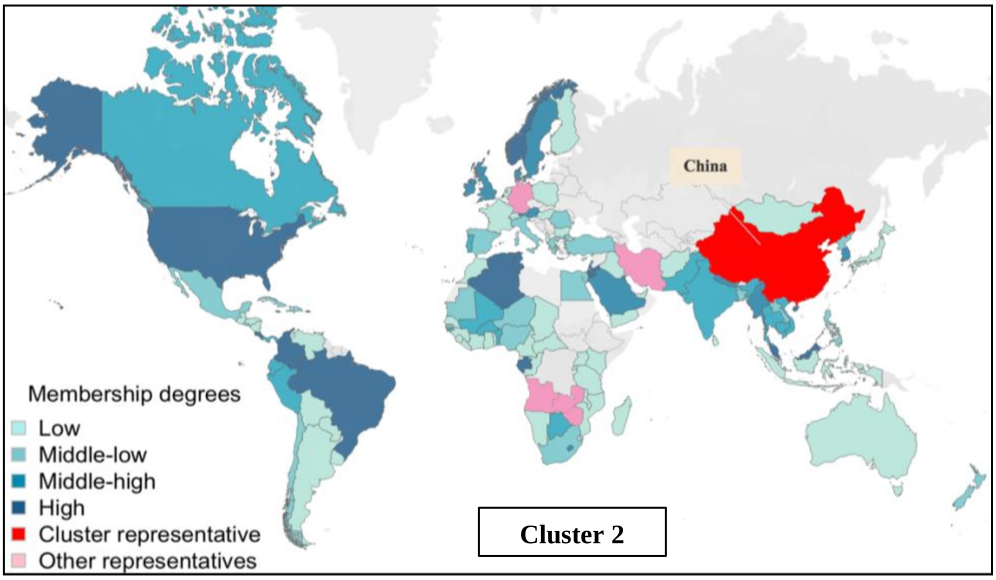
Figure 7.25 Weighted average daily per capita calories, 1961-2013 ($K = 4$ and $\beta = 0.5$).

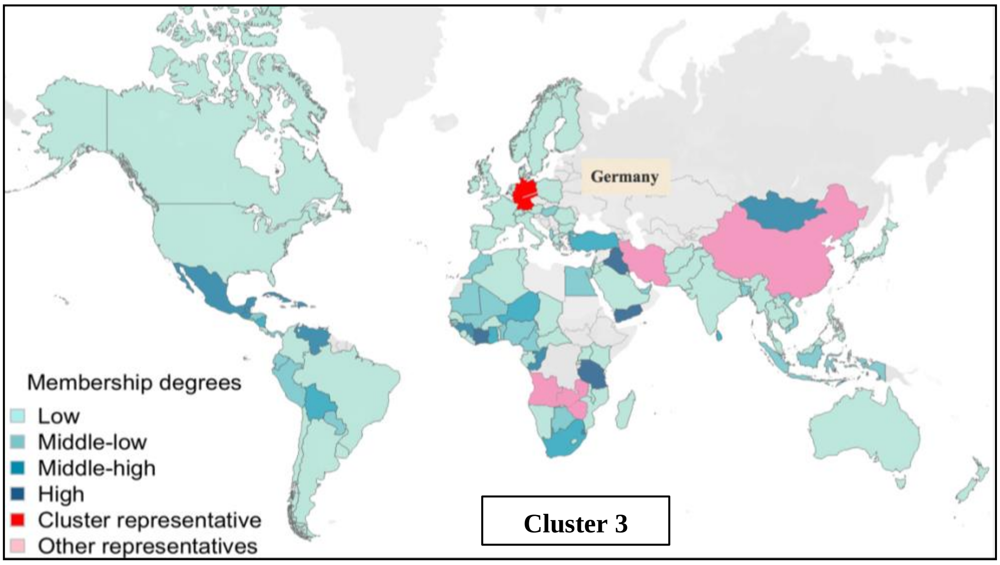
Fixing $K = 6$, the COFUST algorithm identifies six clusters of varying sizes. The cluster characteristics are shown in Table 7.8. Except for CL3-5, the medoids related to other clusters remain unchanged when the spatial information is incorporated; however, some adjustments regarding cluster sizes are observed. Similar to the previous two scenarios, a large cluster (CL2) accounting for about half of the global population does exist. Other considerable clusters include CL4 and CL3 which respectively make up 21% and 16% of the world population whereas the rest are niche clusters. Figure 7.26 shows the geographical dispersion of each cluster. The dark shade of blue dominates in CL2, indicating the relatively high membership degree of most countries to this cluster. Countries which predominantly belong to CL2 include the United States, the United Kingdom, Norway and Brazil. The number of countries with high membership to other clusters is significantly lower. The significant ones include Uruguay (CL1), Mexico (CL3), Bulgaria (CL5) and Kenya (CL6). Specifically, CL4 is a healthy cluster being largely associated with New Zealand, the Netherlands, Japan and Mediterranean countries. Here, the influence of the spatial information is found in the fuzzy cluster membership degree. To illustrate, the non-spatial algorithm in Chapter 6 assigns Switzerland and Afghanistan to CL6 both with high membership proportions (64% and 95% respectively); however, the spatial algorithm dramatically reduces the membership degree of Afghanistan to 54% while maintaining the high membership of Switzerland. In this analysis, β is chosen at 0.9, which indicates to some extent that clusters are defined based 90% on the behaviour of historical calorie trajectories and 10% on the spatial relationship among countries. This value returns a cluster solution quite similar to that obtained by the non-spatial algorithm, and the spatial information fine-tunes the cluster membership degree instead of overpowering the information conveyed by the calorie time series.

Table 7.8 Cluster solutions with and without spatial information ($K = 6$).

Cluster	Without spatial information (β = 1)		With spatial information (β = 0.9)	
	Medoid	World population (%)	Medoid	World population (%)
CL1	Angola	3.2	Angola	3.7
CL2	China	63.4	China	55.6
CL3	France	14.6	Germany	15.6
CL4	Ghana	9.7	Iran	21.1
CL5	Iraq	6.1	Zambia	1.4
CL6	Zimbabwe	3.0	Zimbabwe	2.6


Angola
Membership degrees
Low
Middle-low
Middle-high
High
Cluster representative
Other representatives
Cluster 1


China
Membership degrees
Low
Middle-low
Middle-high
High
Cluster representative
Other representatives
Cluster 2


Germany
Membership degrees
Low
Middle-low
Middle-high
High
Cluster representative
Other representatives
Cluster 3

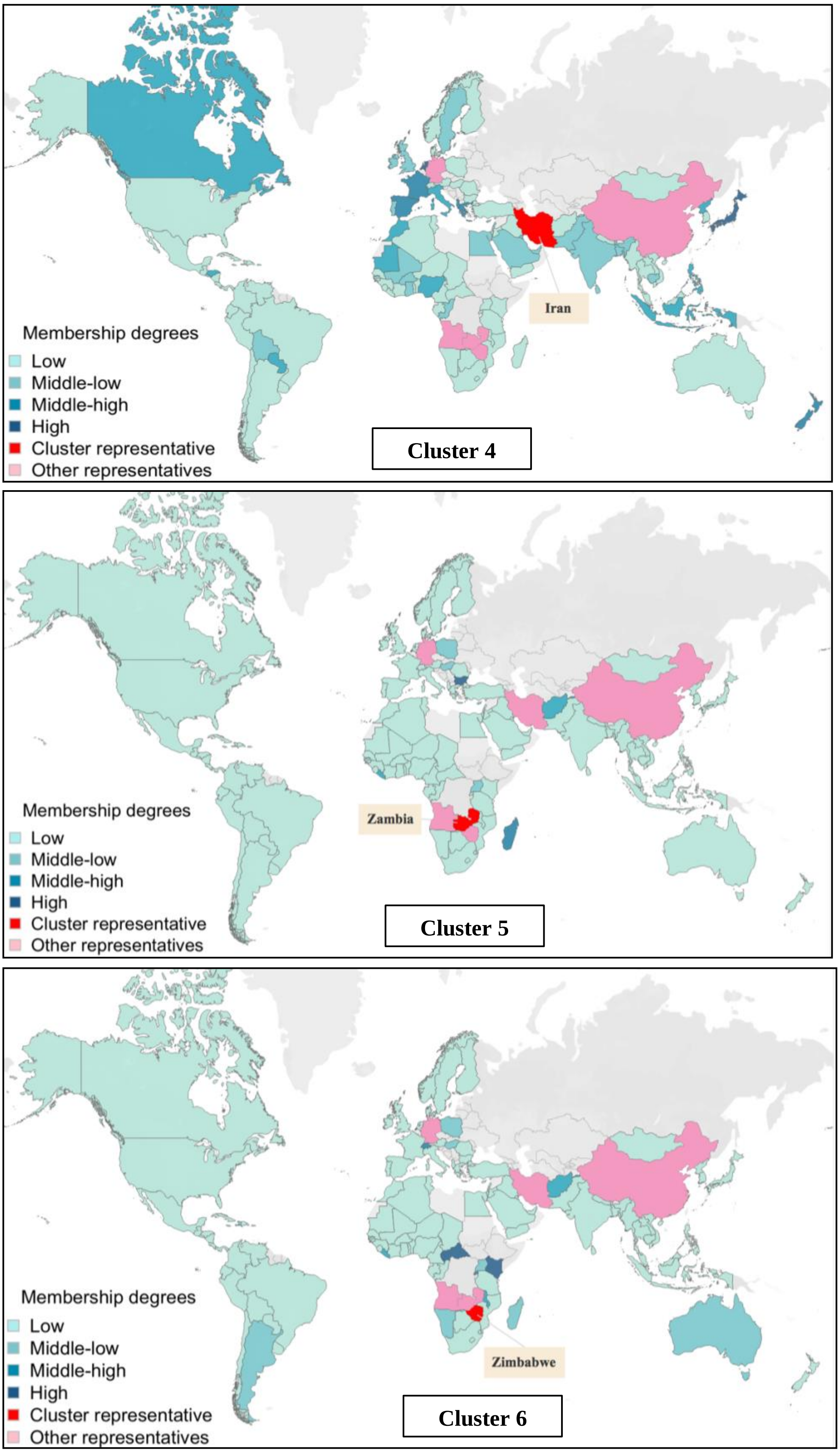


Figure 7.26 Cluster membership degrees when $K = 6$ and $\beta = 0.9$.

For a deeper understanding of the obtained clusters, Figure 7.27a shows the six trajectories of calorie consumption. Some features seem noteworthy. In addition to a monotonic increasing trend (CL2), a decreasing trend (CL5) and two rising trends with a reversal in the 1970s and the 1990s (CL1 and CL3 respectively), the COFUST algorithm identifies two additional trends corresponding to the calorie consumption of CL4 and CL6. The former, despite the substantial growth in the first four decades, started to slow down and lag behind CL2 in the early 2000s, and the latter is characterised by a cyclical fluctuation. Thus, CL4 represents a unique group of countries (more precisely a population segment) that can be labelled as once being the most calorific but has shown signs of stabilising calorie consumption since 2000 that is indicative of the transition to Pattern 5 of the NTM. Another feature standing out from Figure 7.27a is the convergence among dietary trends. Specifically, the calorie consumption of CL2 and CL4 has become increasingly similar over the past 50 years and both approach 3,000 kilocalories in 2013. CL5-6, departing from two distinct levels of calories in 1961, have bridged the gap and end up being the least calorific clusters in 2013 at roughly 2,400 kilocalories.

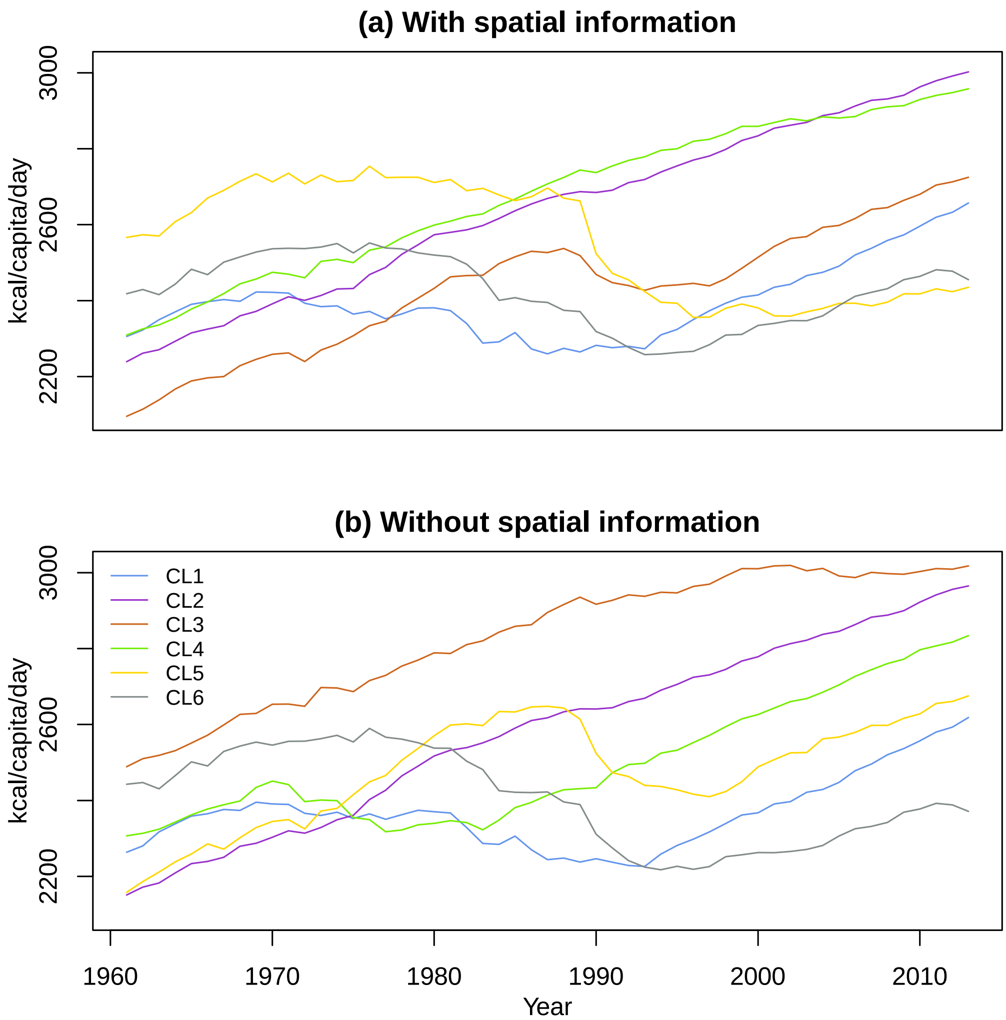


Figure 7.27 Weighted average daily per capita calories, 1961-2013 ($K = 6$ and $\beta = 0.9$).

For the purpose of comparison, Figure 7.27b shows the trajectories of calorie consumption by six clusters obtained by the non-spatial time series clustering algorithm in Chapter 6. Overall, considering the spatial proximity in the clustering procedure exerts a more profound influence on the level of the trajectories than the direction. In other words, the shape of the trajectories is comparable in both panels (a) and (b) of Figure 7.27, however, the most markedly recognised feature from panel (a) is not merely the existence of varying dietary trends but that these trends are converging on distinct levels of calories. Why is the picture of convergence less noticeable in panel (b)? The 'Trend analysis' in Chapter 6 seeks to identify clusters based on the dependence among the original time series and therefore the clusters in Figure 7.27b are dissimilar in the evolution of calorie consumption (say upward/downward direction, growth rate, turning point) due to the strong influence of the trend component. For the 'Spatial trend analysis', the spatial information adds some emphasis on the level ("richer countries tend to exhibit higher levels of calories") and this is further translated into converging patterns. Hence, the ability to embed the information on the level of calorie series is a virtue.

Nevertheless, the main advantage of the spatial clustering algorithm over the non-spatial procedure is the ability to account for the environment conditions of food consumption. The identified clusters not only show similarity in the calorie trajectories but also share homogeneous contextual factors for dietary behaviour. Understanding these factors could assist the design and implementation of policies for better diets. This is an important task as the conditional beta convergence predicts that countries with similar structural parameters would eventually converge on the same level of calories. Among the six clusters identified in Figure 7.27a, CL4 (the green line) is of particular interest since it is the only cluster that has shown signs of stabilising calories over the past 15 years. Next, this analysis examines how the countries characterised by CL4 are like in terms of development indicators.

The six clusters are profiled against a range of structural variables. Table 7.9 presents the weighted average values over the period 2000-2013 obtained using the fuzzy membership degrees of countries belonging to the clusters as weights. Countries associated with CL4 tend to have higher potential for agricultural production (shown by the largest percentage of arable land) and higher agricultural labour productivity (measured by agricultural value per worker). Regarding urban development, CL4, being the most urbanised cluster, at the same time has the lowest density urban areas. Despite the general consensus that the more urbanised a country the more 'Westernised' its dietary patterns, the finding presented in Table 7.9 seems to suggest that urban density matters. As argued by Cockx *et al.* (2019), living in a higher density area is associated with a profound shift from traditional diets towards increased consumption of readily prepared and processed foods. Since CL4 is the unique cluster showing the behavioural changes for a better diet, one would expect this cluster to have the better health status. As can be seen from Table 7.9, the population segment represented by CL4 indeed lives longer and healthier than any other cluster, denoted by the highest life expectancy and health adjusted life expectancy (HALE). Furthermore, they tend to be better educated and associated with higher rates of Internet usage. Overall, these findings confirm the importance of improved education and access to knowledge since the behavioural changes characterised by the shift to Pattern 5 of the nutrition transition model only apply for the group of more educated and health-aware individuals.

Of perhaps the more worrisome concern is CL2 (represented by the purple line in Figure 7.27a) which has overtaken CL4 to be the most calorific cluster since 2003. While it might seem intuitive to relate this rise in calorie consumption of CL2 with accelerating obesity rates, one should not forget the other side of the equation. That is, the increase in calorie consumption of CL2 since 2003 beyond that of CL4 could be explained by the rising demand of calorie expenditure via physical activities. CL2 includes China that has a large portion of the population participating in agriculture, fishery and manufacturing activities and these jobs require more calories than the Westernised economies in Europe and Northern America (CL4).

As shown in Table 7.9, CL2 shares several similar characteristics with CL4, including high agricultural labour productivity, high degree of urbanisation, high life expectancy, high proportion of Internet users, and high educational attainment rate. The distinguished feature of this cluster is the smallest percentage of arable land. The low capacity for agricultural production has made CL2 become more reliant on international food trade, and therefore more susceptible to the inflow of calorie-dense foods that are usually high in fat, salt, sugars and in processed forms. As the option of farming more land is limited in most countries, the emphasis should be placed on increasing agricultural productivity and raising crop yield.

Table 7.9 Profiling of clusters by development indicators over the period 2000-2013.

Variables	CL1	CL2	CL3	CL4	CL5	CL6
Arable land (%) ***	16.4	14.8	15.1	16.9	16.4	16.3
Agricultural value per worker (2010 US$) ***	16,393.3	17,045.3	7,179.2	20,738.3	11,683.9	12,235.0
Urban population (%) ***	51.3	61.0	52.8	61.0	48.0	42.1
Population in the largest cities (% of urban population) ***	35.3	31.7	31.9	29.8	33.5	35.7
Life expectancy (years at birth) ***	62.9	71.1	66.3	71.7	63.9	60.4
HALE (years at birth) ***	56.0	62.7	59.2	63.4	56.9	53.9
Years of schooling ***	6.2	7.8	6.7	7.9	7.1	6.3
Internet users (%) ***	17.1	28.1	14.9	29.5	17.6	15.9

Note: Weighted percentage and weighted mean are reported. Repeated measures ANOVA test is used to determine whether the weighted mean value significantly differs among the clusters.
*, ** and *** denote statistical significance at the 10%, 5% and 1% level, respectively.
Data are retrieved from the World Bank (World Bank 2020e).

Spatial fluctuation analysis

Figure 7.28 summarises values of the FS index calculated for the number of clusters K ranging from 2 to 10 and for β from 1 (no spatial information) to 0.5. The range of values for β is motivated by the fact the influence of the spatial information in the clustering procedure would be overestimated for a lower β value (from 0.4 to 0) (Disegna *et al.* 2017). The trajectories of the FS index give indications of the optimal number of clusters K – where the FS reaches its peaks. It is clear from Figure 7.28 that the FS value increases quickly as the number of clusters rises and all FS trajectories peak up at $K = 10$ regardless of the β value. This behaviour seemingly suggests the best partition is $K = 10$, however implies that the 'Spatial fluctuation analysis' could not arrive at any final cluster solution. The monotonic rising behaviour of the FS index indicates that it would likely to go up if the maximal number of clusters is not restricted to 10. This result is surprising given the meaningful clustering results from the 'Fluctuation analysis' in Section 7.4.1, and there are reasons to believe this reflects more of a technical problem related to the performance of the COFUST algorithm than implications for food consumption.

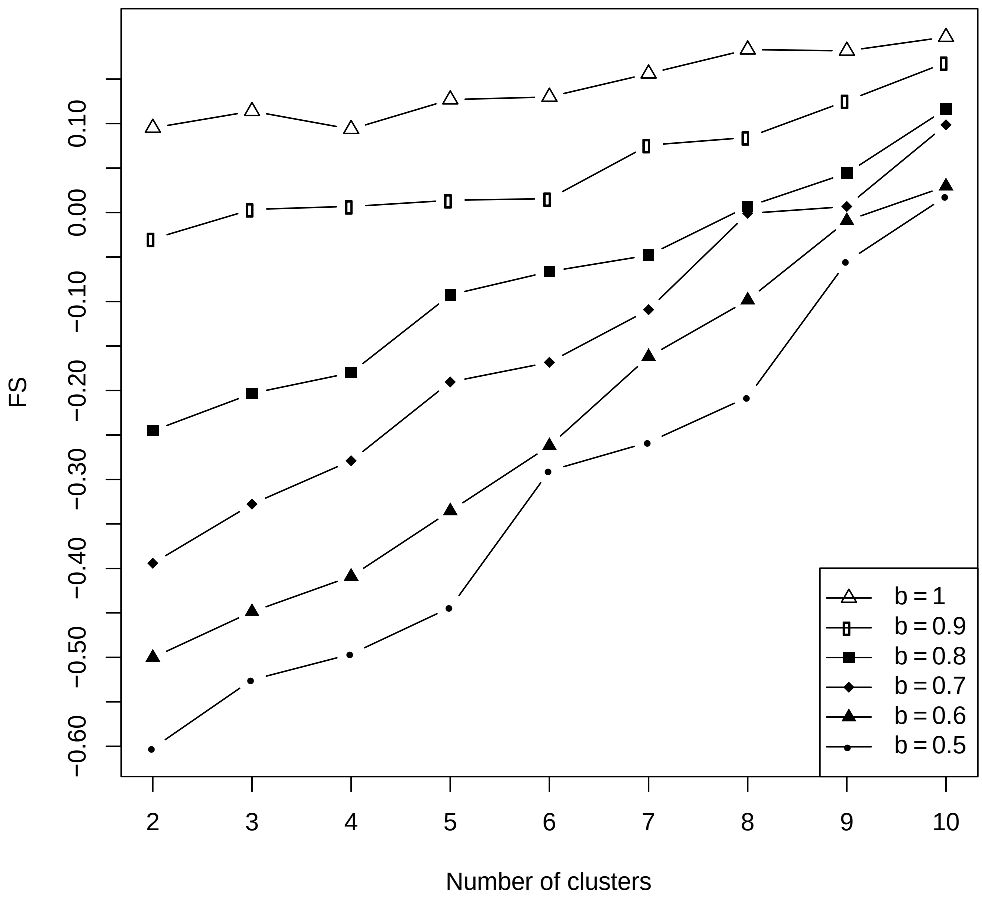


Figure 7.28 Values of the Fuzzy Silhouette index for each cluster partition for *K* varying from 2 to 10 and β from 1 to 0.5 (Spatial fluctuation analysis).

To recap, the 'Spatial fluctuation analysis' aims to detect groups of countries that exhibit common deviations from the trend whilst being spatially close. The input information is a combination of the dependence among detrended time series (pre-filtered data) and the spatial information. As the spatial information is the same for both 'Spatial trend analysis' and 'Spatial fluctuation analysis', the deficiency of the latter should be attributed to the dependence among detrended time series. As a matter of fact, this dependence measure was shown to be very weak since the pairwise Kendall's correlation figures are close to zero (Figure 7.3). Apparently, the pre-filtering step transformed the original data into stationary time series which likely exhibit random patterns regarding the temporal variation. It is therefore more difficult to identify a common structure in the pre-filtered data than the original data, and introducing the spatial information appears to add further complication. In the original paper, the authors of the COFUST algorithm (Disegna *et al.* 2017) recommend decomposing the original data and utilising only the residuals in the clustering task. The purpose is to avoid any overwhelming influence of seasonality inherent in tourism data. As tourism demand can exhibit substantial variations from one month to the next due to seasonal effects, it is possible to observe some common patterns of deviations from the trend (as represented by the residuals of the decomposed time series). In this study, annual data are employed and as a matter of fact the total caloric consumption figure does not alter significantly from one year to the next, resulting in minor deviations from the trend and decreasing the likelihood of common patterns existing.

These findings demonstrate that the trend is the most important component of food consumption data and removing it would result in a loss of critical information including the common dietary evolutions (i.e. how calorie consumption series behaved throughout the period). Without such information, it is not possible to identify a common structure in the data even after considering the spatial relationship among countries.

7.5 Chapter conclusion

This chapter aims to ascertain the similarities among national diets by the means of cluster analysis. First, an innovative copula-based time series fuzzy clustering algorithm is employed. Fuzzy clustering is an attractive technique since it allows the possibility that individuals within a country do not consume the same diet and therefore several diets/dietary trends coexist within a single country.

The clustering algorithm is performed on the observed series ('Trend analysis') and the detrended series ('Fluctuation analysis'). The former returns clusters of countries with common evolutions in calorie consumption whereas the latter groups countries with similar deviations from the trend. 'Trend analysis' successfully identifies a predominant trend – rising calorie relentlessly over time and a minor but ineluctable trend experienced by the minority of the global population – decreasing calorie consumption. In particular, increasing the number of clusters reveals a unique cluster that has shown indications for the shift from Pattern 4 to Pattern 5 of the nutrition transition. This is the only cluster whose calorie consumption has stabilised over the past 15 years. On the other hand, clustering results from ‘Fluctuation analysis’ are more nuanced. Five clusters are identified corresponding to five distinct dietary types, namely ‘Western diet’, ‘Traditional diet’, ‘Mediterranean diet’, ‘Tropical diet’ and ‘Vegetarian diet’. These clusters however do not differ significantly in the behaviour of the weighted calorie series and all show a rapid rising tendency over the past half a century. Further evidence indicates a decline in the healthiness of these dietary types and that all clusters demonstrate the replacement of carbohydrates for fat that is indicative of the nutrition transition.

Acknowledging that the dietary changes described by the *nutrition transition* include both quantitative and qualitative aspects, the utilisation of original data and pre-filtered data in cluster analysis helps to uncover patterns related to quantities and composition of calories. Not all calories are created equal. While the four dietary types other than the ‘Western diet’ has continuously increased their calorie content mostly from higher energy-dense diets with greater roles of fat as well as foods of animal origin and processed forms, the most calorific ‘Western diet’ has slightly reduced the calorie consumption itself. Perhaps such behavioural changes might arise as a unique segment of the population has switched to adopt a healthier dietary style from one of the remaining clusters.

Given the pessimistic picture of worsening dietary quality, what are the implications for health? The results show that even though we are living longer we are not necessarily living healthier. The evidence provided in this study sends a strong message to policymakers: the existing efforts are not adequate in reverting the slumping diets that are observed worldwide.

The second part of the empirical study adds in the spatial context and aims to identify agglomerations of countries characterised by similarities in patterns of food consumption while countries in the same cluster are close in the sense of economic development. The innovative Copula-based Fuzzy K-Medoids Space-Time clustering algorithm (COFUST) is applied. In order to assist the selection of the final partition, the Generalised Fuzzy Moran index is utilised to select the spatial coefficient that maximises the spatio-temporal autocorrelation among post-cluster units. The space-time clustering algorithm is performed for actual data (‘Spatial trend analysis’) and pre-filtered data (‘Spatial fluctuation analysis’). In the latter case, the inspection of the Fuzzy Silhouette plot reveals no natural common patterns in the data. This finding confirms that the trend is the most important component of food consumption data and de-trending removes some critical information regarding the common dietary evolutions.

The ‘Spatial trend analysis’ identifies two, four, and six clusters. The interpretation of cluster membership degree shows that including the spatial information, even though does not force the final clusters to be made solely by countries of similar income levels, influences the clustering results dramatically. Several cluster medoids change; therefore, policy targeting strategies will differ depending on whether to include the spatial information or not. The evolution of calorie trajectories reveals a clearer picture of convergence than the non-spatial algorithm. The COFUST algorithm still identifies groups of countries with similar dietary evolutions; yet, the ability to embed the information on the level of calorie series is a virtue. Another key advantage of the ‘Spatial trend analysis’ over the ‘Trend analysis’ is the ability to identify clusters which are homogeneous in the structural conditions of food consumption. The obtained clusters are profiled against a range of development indicators and evidence suggests that only the segment of more educated and health-aware populations exhibits the behavioural changes towards better diets. This finding emphasises the importance of improved education and access to knowledge that are crucial in raising nutrition awareness of populations.

However, this study is not without limitations. First, the FAO’s Food Balance Sheet data should be interpreted as *food available for human consumption* rather than *food consumption* as food waste is not accounted for. This kind of ‘apparent consumption’ data tends to mask other issues such as undernutrition and hunger that often coexist with overnutrition. Second, the COFUST algorithm returns clusters to which the membership proportions of a country are fixed, and it is not feasible to identify countries that switch the membership degrees between clusters. If future study could accommodate cluster-switching behaviour, it would be able possible to uncover countries whose diets have become better over time. Third, the logic of pulling some countries into the same cluster (say Ghana, the UK and Finland in CL2 regarding the Trend analysis) is not obvious. Even though this odd cluster might reflect the existence of a group of consumers with similar tastes for calories across countries, it could just be the algorithm identifying similar patterns spuriously. Next, traditional spatial measure (say geographical distance) is static but economic proximity measure (say income) is dynamic. Future research could deal with the situation in which the spatial information varies over time. Finally, income is considered as the only proximity in this analysis whilst other factors (for instance, education level, weather condition, common spoken language, etc) may contribute to the similarities in diets around the world as well. In future research, these measures could be simultaneously embedded into the clustering procedure as different levels of proximity.

Chapter 8

Discussion and conclusion

8.1 Chapter introduction

According to the World Health Organisation (WHO), the prevalence of obesity has nearly tripled worldwide since 1975 (WHO 2017a). Such is the extent of the problem that even in countries where undernutrition and infectious diseases are prevalent, obesity and overweight have become issues of pressing public policy concern (Popkin *et al.* 2020). Notwithstanding the interplay of multiple factors ranging from genetic predisposition to environmental influences that contribute to weight gain (Ralston *et al.* 2018), dietary changes represent a major driver of these disturbing weight trends (Hawkes *et al.* 2017). Against this backdrop, the aim of this research is to analyse the evolution of global patterns of food consumption over the past half a century. Specifically, the focus lays on the identification of different dietary trends as well as dietary types and the consequences that they imply for obesity and global health.

After setting out research background, research motivations and research questions in Chapter 1, this thesis starts with a literature review on the nutrition transition and the associated dietary convergence (Chapter 2). Critical developments of the nutrition transition model (NTM) are pointed out and a clear emphasis is placed on the *nutrition transition* which refers to the shift from traditional diets towards the ‘Western’ diet rich in fat, sugars, meat and processed foods but low in fibre, and accompanied by increasingly sedentary lifestyles (Pattern 4 of the NTM). Of paramount interest is the shift to Pattern 5 of the NTM where individuals, realising the high burdens of the highly calorific diet in Pattern 4, change back to healthier diets. However, so far Pattern 5 remains hypothetical and any evidence for the existence of such behavioural changes is needed to assist policy implementation. While drivers for the nutrition transition are manifold and often involve a wide range of economic, social, and cultural factors, the impacts of underlying global forces such as globalisation, rising income, urbanisation, and female participation in the labour force are well documented in the literature. As a manifestation of the nutrition transition, global diets are predicted to be more similar as countries develop and become further globalised. The second half of Chapter 2 reviews empirical evidence for the increasing similarity in food consumption patterns across national borders. Despite a number of previous research attempts exploring convergence in the consumption of calories and certain food groups, studies on this topic are dated, mainly focus on developed countries, and only few studies make connection to the nutrition transition literature. Some authors examine food consumption patterns at the global level; yet, formal convergence tests are generally missing.

Chapter 3 presents a review of the literature on cluster analysis – a data exploratory technique that organises units into homogeneous groups (clusters) so that the within-cluster dissimilarity is minimised whilst the between-cluster dissimilarity is maximised. The review covers various clustering methods for static data and data that vary across time or/and space. In terms of defining cluster boundary, clustering technique can be either crisp or fuzzy. Crisp clustering methods assign units into non-overlapping clusters, whereas fuzzy clustering allows overlapping clusters so that units can belong to multiple clusters with varying degree of membership between 0 (absolutely does not belong) and 1 (absolutely belongs). There is an overview on spatial clustering which requires units included in a cluster to be not only similar to each other but also spatially close. A later section of Chapter 3 demonstrates the usefulness of cluster analysis in food economics by summarising relevant previous studies. From this arise a number of research gaps to be explored. Even though food consumption data often vary temporally, the time dependent nature of the data has not been appropriately addressed in the clustering procedure. Next, the relationships between environment, food consumption and health are embedded in a spatial context; nonetheless, previous studies have largely neglected the role of the space, which can assist the selection of the final cluster solution.

Chapter 4 discusses methods of quantifying diet quality focusing on pre-defined diet quality indices. Aiming to quantify the degree of adherence to a particular dietary guideline, diet quality indices can be classified into: (i) those based on national/international nutritional recommendations and (ii) those based on the Mediterranean diet. Revisiting the methodology for index development, it is shown that the arbitrariness entailing the selection of components to be included in the index, the cut-off values for each component, and the weights assigned to each component, can explain many similarities and differences among existing indices. Whilst these indices (and the adapted versions) have been increasingly used to assess the association between the quality of diets and risk of various health outcomes, their predictive capacity is generally comparable. In this research, the world's diets are first defined by cluster analysis, and a pre-defined diet quality index is then exploited to measure the healthiness. Particularly, the Mediterranean Adequacy Index (MAI) is selected based on data availability, its construction characteristics and numerous advantages shown in previous studies.

Chapter 5 provides a description of the Food Balance Sheet (FBS) data and depicts various trends regarding the caloric consumption and the dietary composition over the past 50 years. These include a steady rise in food availability worldwide, a robust increase in fat and energy contribution of vegetable oils and meat, but a plummet in the contribution of cereals and starchy roots – all of which can be characterised by the *nutrition transition*. Chapter 5 also describes some past trends regarding the rising prevalence of obesity which no country, region, income group or gender is immune to, and offers interesting insights into the positive correlation between caloric consumption and economic status.

Chapter 6 examines the dietary convergence in the light of beta and sigma convergence methods using global data, estimates the speed of convergence and probes into the role played by income. The spatial component is introduced, and empirical analysis is carried out to assess whether adding a spatial dimension improves the results of convergence analysis or not. In order to ascertain the similarities of diets across countries, Chapter 7 employs innovative time series and space-time clustering algorithms. The clustering results uncover several dietary trends and dietary types around the world. The relevance of the empirical findings discussed in Chapter 6 and Chapter 7 will be elaborated in the next section.

Having summarised the purpose and content of the preceding chapters, the rest of this chapter proceeds as follows. Section 8.2 presents a summary of the empirical results. Section 8.3 and Section 8.4 respectively consider the value this research contributes to policy implementations and the existing literature. Section 8.5 points out limitations of this study and Section 8.6 suggests potential venue for future research. Section 8.7 concludes with final remarks regarding the significance of this research.

8.2 Summary of key findings

Despite being thoroughly discussed in Chapter 6 and Chapter 7, the empirical findings are reviewed subsequently in line with the proposed research objectives.

Research objective 1: Dietary convergence

The first research objective is to test for the convergence in global patterns of food consumption and to measure how the convergence process is influenced by structural factors. This research examines per capita daily calories available for consumption (food availability) data for 118 countries over the period 1961-2013 using sigma and beta convergence methodologies.

Sigma convergence refers to a narrowing dispersion of calories across countries over time, which is assessed via the coefficient of variation. Beta convergence refers to the phenomenon that countries with higher initial levels of calorie consumption experience slower rates of calorie growth than those with lower initial levels (the “catching-up” effect). This type of convergence is assessed by the sign and statistical significance of the beta coefficient in the econometric model regressing the calorie growth rates over a period of time on the initial level of calories. Beta convergence can be unconditional or conditional. Unconditional beta convergence assumes that countries eventually converge to the same global steady state equilibrium, and it is shown by a negative association between average growth rates and initial calorie levels even if no other explanatory variables are included in the model. Conditional beta convergence argues that convergence occurs only if the structural conditions are identical and that each country approaches its own unique equilibrium. Countries with the initially low level of calories do not necessarily exhibit higher growth rates of calories, but countries that are further from their own steady-state level exhibit faster growth.

Results of sigma convergence reveal a reduction in the coefficient of variation of calorie availability over the past half a century that is indicative of sigma convergence. However, the decline in the coefficient variation is uneven with a break in around 1998. The period 1999-2013 is characterised by strong sigma convergence whereas the pre-1998 period by both convergence and divergence.

Regression results of unconditional beta convergence model report a significantly negative beta coefficient, implying that countries with lower levels of initial calories tend to exhibit higher growth rates of calorie consumption. Countries bridge the gap between the current levels of calories and the global steady-state levels by on average 1.6% annually.

But is this convergence process homogeneous across countries? To demonstrate how economic development could influence countries to converge on a path to different steady-state levels (“club convergence), the study sample is divided into low-, lower-middle-, upper-middle- and high-income countries. Dummy variables with high-income countries being the reference category were included in the beta convergence model. All dummy coefficients are significantly negative, suggesting that countries in the long run converge to different growth rates of calories which tend to be higher for high-income countries than for lower-income countries. What about the rate of convergence? Adding interaction terms (dummy variable multiplied by the initial level of calories) in the model shows that low-income countries are converging at the fastest pace and the rate of convergence reduces as income rises. Specifically, low-income countries converge at an annual speed of 4.1% whereas high-income countries converge at 3.1% per annum.

A conditional beta convergence model was specified by incorporating structural parameters that are considered driving forces behind dietary changes. These variables include arable land (% of land area), GDP per capita growth, urban population (% of total population) and female labour force participation rate. After controlling for these variables, the beta coefficient is still significantly negative. Results of conditional beta convergence confirm positive effects of arable land, income growth and urban population on the calorie growth rate but negative effects of female employment. The latter could be explained by the improved health awareness and better nutritional knowledge thanks to higher levels of education and training for women. In addition, the magnitudes of the coefficients point to rising income being the most influential determinant of dietary changes.

Due to the break in the coefficient of variation, the beta convergence was examined by splitting the data into two sub-periods 1961-1998 and 1999-2013. The unconditional beta convergence specification reveals a faster convergence process for the past 15 years, and when income dummies are included in the beta convergence equation neither the dummy variables nor their interaction terms with the initial calorie level are significant in the latter period. Thus, income plays an important role in the early stage of dietary convergence. When national incomes reach a certain threshold, income alone cannot explain the path countries are converging to different steady-state levels. When the conditional beta convergence specification is applied, a less robust convergence process is reported in the latter period. The annual rate of convergence (1.82%) is much lower than in the former period (3.18%). Perhaps the structural parameters are closer to their steady states during the last decade and as a result, countries are converging to their own equilibrium levels of calories at a slower speed. Regarding the structural parameters, agroecological, social and demographic indicators exert a stronger impact in the initial period, however, economic factors have become a more important determinant of the dietary convergence since the Millennium.

Research objective 2: Dietary trends

The second research objective is to identify common dietary trends and dietary types around the world. As the convergence analysis signals convergence in the calorie availability across national borders, cluster analysis was employed to ascertain the similarities among national diets. An innovative copula-based time series fuzzy clustering algorithm was applied. Fuzzy clustering allows the possibility that individuals within a country do not consume the same diet and therefore several diets/dietary trends coexist within a single country. To reveal the twin aspects of food consumption, the clustering algorithm was performed on the observed time series ('Trend analysis') and the detrended series ('Fluctuation analysis').

The 'Trend analysis' detects clusters of countries with common evolutions in calorie consumption representing common dietary trends among 118 countries. Regardless of the number of clusters, 'Trend analysis' detects a predominant trend – monotonic rising calorie consumption which is experienced by the majority of the global population and another trend which is minor but ineluctable – reducing calorie consumption. Increasing the number of clusters enables the largest cluster to be split up into smaller clusters which can be described as rising trends with a reversal, albeit to varying extents. Particularly, the six-cluster solution reveals a unique cluster representing about 10% of the world population whose diet, despite its historically highest level of calorie consumption, is the only one which did not become more calorific over the past 15 years. For this cluster, the stabilisation of calorie consumption shows evidence of progressing beyond Pattern 4 of the nutrition transition. Importantly, such a behaviour is largely attributed to the lower consumption of less healthy foods (such as animal fats, sugar and milk).

On the other hand, the 'Fluctuation analysis' detects groups of countries with similar deviations from the trend and the results are more nuanced. Although the algorithm solely utilises the information on the variation of calorie consumption, the obtained clusters match neatly with distinguished nutritional composition of diets. Five identified clusters are labelled corresponding to five distinct dietary types, namely 'Western diet', 'Traditional diet', 'Mediterranean diet', 'Tropical diet' and 'Vegetarian diet'. Nonetheless, these clusters do not vary much in either the behaviour of the weighted calorie series, and all of them show a rapid rising tendency over the past 50 years. Further evidence indicates a decline in the healthiness of these dietary types (as measured by the MAI) and that all clusters demonstrate the replacement of carbohydrates for fats that is indicative of the nutrition transition.

Research objective 3: Incorporating a spatial dimension

The last research objective is to probe into the role of the space on dietary convergence and dietary changes. To account for both conventional and non-conventional measures of the space, three proxies for spatial relationship were exploited: (i) geographical distance, (ii) geographical contiguity, (iii) economic contiguity. The proposal of an economic proximity measure (average GDP per capita during 1970-2013) points to income level (rather than geographical closeness) that is driving the similarities in global diets. The Global Moran's I test of spatial dependence shows that economic proximity is the only to yield significant and positive spatial autocorrelation. In other words, countries with similar income levels tend to also have similar diets. However, as indicated by the Local Moran's I statistics, the calorie content of a country is not always in line with its neighbours (in this context neighbours should be interpreted as countries of the same development level). Thus, income is an important predictor for the (dis)similarities among diets globally, but other factors such as migration, heterogeneous consumer tastes and socio-cultural influences might explain local deviations from the global patterns.

Since theoretically it has been shown that ignoring the spatial dependence can cause biased OLS estimators, it is important to account for this phenomenon in the beta convergence model. The spatial diagnosis LM tests confirm positive spatial autocorrelation among the error terms in the unconditional beta convergence model and posit that the spatial dependence stems from unmodeled country-specific factors (national effects) rather than spill-overs. As shown in Table 8.1, the alternative spatial model reports the annual convergence speed of 4% and that it takes about 17 years for countries to eliminate half of the current disparity with the steady state. Compared to the non-spatial unconditional beta convergence model, the consideration of spatial interactions leads to a convergence process that is about three times as fast. These findings confirm that countries are open to a range of demographic and exhibit spatial proximity in dietary convergence.

Results of the spatial conditional beta convergence model point to spatial autocorrelation in terms of spill-overs. The inclusion of spatial effects reveals a slightly faster convergence process at an annual speed of 4.4% as compared to 3.99% obtained from the non-spatial model. Particularly, the negative sign of the spatial autoregressive parameter demonstrates negative spill-over feedback effects which might imply behavioural changes due to the adverse health consequences of the highly calorific diet. Hence, being surrounded by countries with substantial growth rates of calories negatively affects the country's own growth rate.

The existence of spatial dependence in calorie consumption was also investigated for two sub-periods 1961-1998 and 1999-2013. The Global Moran's I statistic becomes insignificant for the latter period in either unconditional or conditional beta convergence model, rejecting the possibility of spatial interactions. Here, the influence of other countries might still be strong over the past 15 years but not confined to countries within the proximate "neighbourhood". Regarding the former period 1961-1998, the inclusion of spatial interactions influences the results of unconditional beta convergence greatly. This is evident in the more robust convergence process with the annual convergence speed approximately three times as fast (1.15% versus 3.18%) and a half-life measure about three times smaller (22 years versus 60 years). The effects of spatial interactions on the conditional beta convergence however are less significant. Both non-spatial and spatial models indicate an annual convergence speed of slightly above 3% and it takes roughly 21 years to eliminate half of the current disparity with the steady state.

Table 8.1 Comparing convergence speed: with and without spatial effects.

	Period	Without spatial effects	With spatial effects
Unconditional beta convergence	1961-2013	1.63 (42.64)	4.00 (17.3)
	1961-1998	1.15 (60.05)	3.18 (21.81)
	1999-2013	1.84 (37.59)	1.85 (37.52)
Conditional beta convergence	1961-2013	3.99 (17.37)	4.40 (15.74)
	1961-1998	3.18 (21.82)	3.41 (20.31)
	1999-2013	1.82 (38.09)	-

Note: Convergence speed per year (in %) is reported. Half-life measures are inside parentheses.

Does the spatial relationship influence the clustering results? The innovative Copula-based Fuzzy K-Medoids Space-Time clustering algorithm (COFUST) was applied to identify agglomerations of countries characterised by similarity of patterns in food consumption whilst countries in the same cluster are close in terms of economic development. Similar to the non-spatial model, the space-time clustering algorithm was performed for observed time series ('Spatial trend analysis') and detrended series ('Spatial fluctuation analysis').

The 'Spatial trend analysis' identifies two-, four-, and six-cluster solutions. In any case, the interpretation of cluster membership discloses that the inclusion of spatial relationship, though does not force the obtained clusters to be made solely by countries of similar income levels, affects the clustering results dramatically. Table 8.2 demonstrates the alteration of several cluster medoids and further analysis shows a clearer picture of convergence among the weighted average calorie trajectories. The obtained clusters were profiled against a range of development indicators and results suggest that only the segment of more educated and health-aware populations exhibits the behavioural changes towards better diets.

Regarding the 'Spatial fluctuation analysis', the inspection of the Fuzzy Silhouette plot indicates no common patterns in the data, i.e. the algorithm does not identify any meaningful cluster solution. This finding confirms that trend is the most important component of food consumption data and de-trending removes some critical information related to common dietary evolutions.

Table 8.2 Comparing clustering results: with and without spatial effects.

Cluster solution	Without spatial effects	With spatial effects
$K = 2$	China (93.7) Zimbabwe (6.3)	Honduras (55.1) Trinidad and Tobago (44.9)
$K = 4$	China (73.8) Ghana (13.4) Iraq (8.6) Zimbabwe (4.2)	Germany (50.5) Namibia (2.5) El Salvador (43.8) Zambia (3.2)
$K = 6$	Angola (3.2) China (63.4) France (14.6) Ghana (9.7) Iraq (6.1) Zimbabwe (3.0)	Angola (3.7) China (55.6) Germany (15.6) Iran (21.1) Zambia (1.4) Zimbabwe (2.6)

Note: Only clustering solutions for 'Trend analysis' and 'Spatial trend analysis' are reported.
The figures inside parentheses denote cluster size in terms of weighted global population.

8.3 Policy implications and recommendations

The abovementioned findings have both academic and societal impacts. The key conclusions that are relevant for policymaking purposes are highlighted as below.

The world is eating badly

More advanced data analysis techniques enable us to improve the understanding of what people eat and why it matters for global health. Of great concern is the finding that 90% of global population appears to be on a trajectory path of consuming an ever-increasing more calorific diet with little evidence of a slowdown. Extrapolating from the past trends, the largest cluster would likely to become also the most calorific cluster in ten years' time. The calorie content of such a diet is accelerating, as are overweight and obesity. As this cluster is predominantly represented by populated countries such as China, the United States, Brazil, India, and Indonesia, the dire prediction that "a third of the global population will be overweight or obese by 2030" is consistent with the findings presented in Chapter 7 (Global Panel on Agriculture and Food Systems for Nutrition 2016). Given the dietary origin of obesity, the unceasing rise in calorie consumption acts as a 'canary in the coal mine' for predicting the looming obesity crisis in populations with hitherto lower rates, for example Indonesia and India.

An important finding is that not only we are consuming more calories than 50 years ago, but we are deriving a lower proportion of these calories from cereals or carbohydrates and more from fat. The obesity problem has both quantity and quality dimensions. As the move from plant-based to animal-based foods is evident for all dietary types, dietary healthiness has deteriorated with no exceptions. Even though the declining rate is most rapid for the seemingly least healthy diets, the historical experience of developing countries suggests that the unhealthy "Westernisation" of diets can, if unchecked, "proceed rapidly and irreversibly" (Raghunathan *et al.* 2021).

The message is clear: the world is eating badly, and this can pose a significant threat to achieving nutrition targets and "ending malnutrition in all forms". Thus, current attempts to improve diets are obviously inadequate and existing efforts need to be redoubled. Ambitious and transformative nutrition commitments, especially those that are specific, measurable, achievable, relevant, and time-bound (SMART), need to be renewed and expanded, and their accountability needs to be regularly rechecked by an international system of governance.

Good health is not possible without good nutrition

Given the pessimistic picture of worsening dietary quality worldwide, this study provides a clear-cut evidence that diets consumed by the majority of global population are inadequate for good nutrition. In line with previous research (Afshin *et al.* 2019; Willett *et al.* 2019), the results presented in Chapter 7 confirm that the nutritional content of the foods consumed reverberates into the health status of the population. Even though we are living longer, we are not necessarily living healthier. This sobering fact resonates with the crucial role of diet quality in determining human nutritional status and demonstrates how damaging the highly calorific 'Western' diet is to our body. The findings of this research lend supports to the policy recommendation that nutrition must be fully integrated in health systems and investments in nutrition should be prioritised, especially those targeting the communities most affected by malnutrition.

While improving diets alone is not sufficient to stop malnutrition, it is a necessary condition for reducing disability and death from malnutrition. In fact, healthy diets have been recommended as "double-duty" actions to tackle both under- and over-nutrition (Hawkes *et al.* 2020). In the era of COVID-19, poor diets pose almost an equal threat to life as cancer or old age, and the high rates of obesity and diet-related diseases contributed to the UK's appalling death rate (National Food Strategy 2020). The pandemic reminds us that strong immunity and resistance to communicable diseases originates from healthy diets (Dupouy and Gurinovic 2020). It is more critically important now than ever to promote good nutrition and sound well-being to populations, starting with the promotion of healthy and nutritious diets.

How to define a universal healthy diet is an important stepping stone towards this goal, yet it remains the utmost challenge owing to the different nutritional needs of people determined by age, gender, disease profiles, and physical activity levels. To address this problem, the EAT-Lancet Commission developed a global reference diet based on the best evidence available on foods, dietary patterns and health outcomes (Willett *et al.* 2019). Nonetheless, this diet is shown to be unaffordable to at least 1.6 billion people in the world (Hirvonen *et al.* 2020). Even more disheartening is the conclusion that nutrient-adequate diets are out of reach for the poorest (Bai *et al.* 2021). Economic growth is thus needed for a larger proportion of the populations to afford more nutritious foods. All stakeholders, including governments and businesses, need to take more concerted action to ensure food systems and food environments are delivering healthy diets that are affordable, accessible, and desirable for all.

Diets and dietary changes: one size does not fit all

Although there is increasing scientific consensus over the convergence of national diets onto the Western-style diet, this research demonstrated that diets and their trends are very heterogenous. Statistical evidence supports the presumption of converging caloric consumption, however, results of convergence analysis showed that the convergence process is not uniform across countries. Instead, countries at different stages of economic development experience the paths towards distinct levels of calories and at varying speeds. While diet monitoring is equally vital in every nation, the findings presented in this research call for scrupulous attention to low-income countries at the early stage of development as they exhibit the most worrying trends. Overall, dietary convergence is best described as the tendency of diets becoming more similar to rather than converging onto a single international norm.

Hence, an important policy implication is to discourage overly prescriptive global-level policies which may not be effective in some countries and populations. The design, development and implementation of healthy diet promoting programmes should be context specific. For this purpose, it is useful to scrutinise and understand the structural development parameters that facilitate the dietary shift. Some influential indicators include resource availability, rising income, globalisation, food trade, urbanisation, and female employment. Understanding these factors better can offer clear and specific guidance to policymakers on effective evidence-based policies (Hawkes *et al.* 2013).

Take the case of results from the space-time cluster analysis in this research. To elucidate the differences in environment conditions of food consumption between clusters, the identified clusters were profiled against a range of agroecological, demographic and socio-economic characteristics. Despite several similar characteristics between the largest cluster and the so-called 'Western diet' cluster, the former has overtaken the latter to be the most calorific since 2003 and is thus worth scrutinising. It is noted that this largest cluster is associated with the lowest percentage of arable land (representing resource availability). The low capacity for agricultural production makes this cluster become more dependent on international food trade, and as a result more susceptible to the inflow of calorie-dense foods that are high in fat, salt, sugars and in processed forms. Policies aimed at improving the diet consumed by this cluster would not make sense unless the issue of limited availability of agricultural resources is addressed. As the option of farming more land is not viable in most countries, policy interventions should be placed on increasing agricultural productivity and raising crop yield.

Healthier diets are at least possible

A key advantage of using longitudinal data is the ability to explore various dietary trends and to assess how diets have changed over time. Results of the innovative time series cluster analysis reveal a unique cluster representing about 10% of the world population whose diet, despite its highest historical level of calorie consumption, is the only to not become more calorific over the past 15 years. For this cluster, the stabilisation of calorie consumption shows evidence of progressing beyond Pattern 4 of the nutrition transition model. Furthermore, it was found that the behavioural change has been largely attributed to the reduced consumption of less healthy foods such as animal fat, sugars, and milk since the Millennium. This empirical evidence echoes the findings of previous research (Imamura *et al.* 2015; Bentham *et al.* 2020) regarding the shift towards decreased consumption of animal-sourced foods and sugar in many high-income nations. The findings presented in this research therefore help to clear up the lingering doubt about the existence of Pattern 5 by showing that the behavioural change towards healthier diets is feasible.

That said, for a "Great Food Transformation" to healthy diets by 2050, it is estimated that the global consumption of healthy foods such as fruits, vegetables, nuts and legumes must be more than doubled whereas the consumption of unhealthy foods such as red meat and sugars must be reduced by more than half (Willett *et al.* 2019). These radical dietary shifts lend real urgency for the strategic implementation and realisation of widespread, multi-sector, multi-level actions to change what people are eating. Lessons from successes must be learnt and scaled up everywhere.

Since food choices are complex and are driven by numerous determinants at local, regional, national and international levels (Benthem de Grave *et al.* 2020), policy interventions should focus on factors that are shown to affect food consumption and dietary changes positively. For example, the findings offered in Chapter 7 suggest that only the segment of more educated and health-aware populations exhibits the behavioural changes towards better diets. These results bring to the fore the importance of improved education and access to knowledge that are vital in raising nutrition awareness of populations. Increasing educational level of population will lead to better nourished populations, as established by Rippin *et al.* (2020).

Spatial interactions matter

It is evident that economically close countries tend to consume similar diets and tend to converge more rapidly to similar levels of calories. Therefore, policymakers can obtain biased estimates of the speed of convergence and arrive at incorrect conclusion if spatial interactions are neglected. Results of cluster analysis vary substantially as observed in the changes of cluster medoids; therefore, policy targeting strategies will differ depending on whether to consider the spatial relationship or not.

Undoubtedly, this research contributes to the ongoing debate on the need to take into account spatial interactions among national food policies. The findings demonstrate that a random shock to a specific country not only affects the growth rate of calories in the respective country but also diffuses throughout the entire 'neighbourhood' because of the spatial dependence. Movements away from a steady state equilibrium induced by a shock are not restricted to the corresponding country but apply to a set of economically adjacent countries. The globalisation of food consumption is thus apparent.

On the same policy implication, the empirical evidence that dietary evolutions are *spatially* dependent provides a basis for the development of group-specific interventions targeting populations at risk of worsening diets. While these policy measures are often place-based, this study lays foundations for the implementation of coherent food policies beyond geographical boundaries.

The relevance of spatial dimension cannot be overlooked especially when a measure of economic proximity is adopted. As low-income countries become further developed, they may direct their attention and efforts towards improving their infrastructure characteristics to a degree similar to those of high-income countries. Since the convergence speed is augmented with spatial effects, two countries with similar characteristics are expected to display greater interaction and would eventually converge to similar levels of calories. Although the economic disparities between poor and rich countries might seem too wide to be bridged in the near future, the half-life obtained in Chapter 6 signals that it will not be long until the world will be drowning in highly calorific unhealthy diets.

Convergence (or lack of) and spatial interactions may be useful to draw implications for policies, but they could be the outcome of past policies (an example is given in Section 6.4.3). It is worth emphasising that the current thesis is about highlighting the association between income growth and food consumption, but it cannot offer much evidence in terms of causation owing to the limitations of the data used. As countries develop and become further globalised and industrialised, their dietary patterns become more similar across countries and tend to resemble the 'Western' diet. Taking the level of economic development into account, the analysis reveals strong convergence of food consumption during the first three decades under review but less so over the past two decades. The lack of convergence (and spatial interactions) has important policy implications especially when the influence of other countries might still be strong but not confined to countries within the proximate 'neighbourhood'. As dietary patterns that are once characterised by wealthier countries are no longer limited to the West, the detrimental effects of such diets for human health (including rising rates of obesity) are spreading over the world. However, the relationship between income, food consumption and obesity is difficult to disentangle from the range of other factors that might impact food choices including but not limiting to food price, education background, and cultural influences. For example, in some cultures, food is merely considered a source of energy for the body while in other cultures food plays an important part in social bonding and religious experiences. The wider socioecological determinants of change in nutrition-related behaviours need to be acknowledged and understood if one wants to make any conclusion about the causal relationship between income, food consumption and obesity. That is not to mention the other side of the coin – reducing physical activity.

8.4 Contributions to knowledge

Overall, this research contributes to the literatures across food economics, global health, and computational statistics in several ways. Mentioned the publication.

To begin with, this study broadens the current knowledge of the nutrition transition literature by proffering statistical evidence to support the notion of global dietary convergence. While there was much anecdotal evidence for the converging patterns of food consumption across countries as a result of the nutrition transition, this research is among a few studies that have formally tested this convergence hypothesis by the means of sigma and beta convergence methodologies. In particular, the examination of convergence clubs based on nonparametric methods and conditional beta convergence in Chapter 6 addresses a key limitation of the majority of cross-sectional studies in the assumption of a single global steady-state equilibrium. Results advocate the existence of different convergence dynamics among countries of distinct levels of development and illustrate how structural variables such as resource availability, rising income, urbanisation, and female employment can determine the heterogeneous paths that countries are converging to.

Linked to the above contribution, this study adds to the scholarly debate on the rising obesity by offering evidence to support the theory of modernisation. Findings show that during the process of economic development, countries are experiencing the domestic nutrition transition and as a consequence national diets are becoming more alike. Results of the spatial beta convergence models demonstrate that growing income plays an important role in the dietary convergence across countries. As indicated by the modernisation theory, economic development is closely related to structural factors such as urbanisation and female participation in the labour force. Results from the conditional beta convergence model corroborate such an argument since these development variables are shown to exert significant influences on the dietary changes. Over the past half a century, as countries further develop, their diets have resembled the so-called ‘Western’ diet high in fat, salt and sugar. As physical activity level tends to drop due to the modern lifestyles, the manifestation of high-calorie diets in expanding waistline is obvious.

From a methodological standpoint, this research contains some novelties in that it applied innovative clustering techniques to a large database of dietary intake. More advanced data analysis techniques enable us to improve the understanding of diets and why they matter for health status. Although cluster analysis is a common method to investigate food consumption behaviours, previous scholars largely neglect the time dependent nature of food consumption data that vary temporally. This research fills in this gap by using a copula-based fuzzy time series clustering algorithm. The copula function captures the dependence among time series and the clustering results correspond to various dietary trends. The utilisation of fuzzy clustering with the possibility to have multiple diets coexist within a single country represents another important innovation. Previous scholars mainly employ crisp clustering approach which reduces numerous diets within a country into one average national diet thus limits the understanding of individual differences in food consumption. While the national dietary patterns were clustered solely on the basis of total caloric consumption, the ability to identify distinct dietary types/trends inherent in a country is hence a great merit.

Another significant contribution of this research to the food economics literature lies in the examination of spatial approach in beta convergence model and cluster analysis. Particularly, the utilisation of an economic proximity measure (average GDP per capita) instead of traditional metrics for the space such as geographical distance or geographical contiguity has not previously been undertaken. Notwithstanding the nascent literature on spatial clustering of obesity prevalence and other health outcomes, this study is among the first attempts to assess the usefulness of taking into account the spatial dimension in food consumption studies. The findings presented in Chapters 6 and 7 show that ignoring spatial interactions can lead to biased estimates and incorrect interpretation of both convergence and clustering results. This analysis also demonstrated that the adoption of spatial econometric techniques for beta convergence modelling explicitly considers and helps explain the spatial dynamics in dietary convergence.

Finally, this research contributes to the growing strand in the spatial analysis literature by proposing a generalised measure of spatio-temporal autocorrelation – the Generalised Fuzzy Moran's index (GFM). This index was developed in response to the need of a generalised spatial dependence measure in a situation where values of spatial units are collected over time. Differing from the earlier scholars who consider spatial observations and temporal values separately, this study integrates the deviation of time series (measured by the Kendall rank correlation) to reveal the spatial pattern. On the one hand, the GFM index extends the classical Global Moran's I statistics to compute the spatial autocorrelation of time series data. On the other hand, the GFM index can be considered as an extension of the Fuzzy Moran index introduced in D'Urso *et al.* (2019a). In this respect, it can be of practical usefulness in space-time clustering procedure as to assist the selection of final clustering partition when several spatial coefficients are considered. In Chapter 7, two experiments were conducted to exhibit the performance and confirm the validity of the GFM index in discovering the spatio-temporal patterns. By comparing with some existing measures of spatial dependence, this study showed that failing to account for both spatial and temporal dimensions in space-time data leads to incorrect results.

8.5 Limitations of the research

Despite being explained in detail in previous chapters, some limitations merit discussion.

Validity of the data series: Food availability, not food consumption

This research focuses on the total caloric consumption and rely on the Food Balance Sheet (FBS) data provided by the FAO. These data are collected from country reports of production, imports and exports of food that are often subject to collection errors (FAO 2018). The data are converted according to global estimates of food loss due to livestock feed, seed use, industrial use, and losses during transportation. Lacking precision in production or consumption measure, the data give a general picture of consumption trends that can imply food available for consumption for most nations (FAO 2017b). In fact, they have many shortcomings that might hinder the interpretation of the results.

The foremost caveat of the FBS is that the data do not imply actual food intakes/ consumption as the food waste at household, retail and restaurants is not incorporated into the supply figures (Vilarnau *et al.* 2019). According to Anand *et al.* (2015), almost a third of all food produced for human consumption is wasted before being consumed, therefore the FBS likely overestimates the actual food consumption quantity. Prior studies report that the FBS data are at least 20% higher than the true dietary intake (Grünberger 2014; Popkin and Reardon 2018).

The inaccuracy of the underlying data sources (such as data on crops, production, storage, and losses) limit the data reliability, particularly in less developed countries (Choudhury and Headey 2017; Desiere *et al.* 2018). Even though the FAO recently updated its methodology by improving the estimates of specific sources (such as data on stocks, food, feed, loss), the information related to food waste at the consumer level is still being revised (FAO 2019a). Also, the figures on home production and harvest of wild plants are not taken into account (Da Silva *et al.* 2009). This is more of a problem in developing countries and rural areas where a large number of people depend on local or small-scale agricultural production and wild foods (Beal *et al.* 2017).

Another caveat worth mentioning refers to the lack of disaggregated information on food supplies by age, gender, education, or socio-economic levels since the estimates are provided per capita (Del Gobbo *et al.* 2015). Representing an overall average amount of food availability within a country, the FBS can mask the coexistence of over-consumption and under-consumption among disparate groups of people living in the same country (Beal *et al.* 2017; FAO 2017b). It is not rare to find (developing) countries in which one portion of the population consumes a certain food at an average level whereas another portion of the population consumes the same food at a level higher than average. In this case, the use of countrywide average data would not truly indicate the consumption level of either group. This caveat likely increases the uncertainty in analysing countries with very unequal income distributions. Therefore, it is crucial to acknowledge that the results obtained by average diets in this research only represent a partial picture of the story. While fuzzy clustering is an attempt to reveal the diets underlying country averages, it is something of a second to best to detailed dietary data on a global scale.

The absence of food prices in the conditional beta convergence model

Despite being mentioned in the theoretical framework in Figure 2.3, food prices are not incorporated in the conditional beta convergence model in Chapter 6. While income is a significant driver of the nutrition transition, without the relative cost of food, it is impossible to determine the extent to which individuals substitute one food for another. However, in the same spirit with Azzam (2020), food prices are not considered in the conditional beta convergence model in this study due to the absence of historical data of food prices corresponding to the food categories listed in the Food Balance Sheet. In principle, one could use Consumer Prices Food Index to proxy the cost of diets. Yet, there are some shortcomings in employing such a proxy. First, even though the FAO provides a standardised set of Food CPI data (FAO 2021a) for as many as 198 countries over the world, the sampling methodology can differ greatly across countries. For example, some countries report indices for urban areas only, others for specific household groups instead of country-wide averages. Second, the FAO's Food CPI data only date back to year 2000, thus leaving an insufficiently long time series for the convergence analysis which examines food availability starting from 1961. Third, the Food CPI considers the prices of a basket of goods and services that are typically purchased by specific groups of households, but not the FBS food aggregates that make up the total calorie availability figure under examination.

The MAI's ability to capture dietary healthiness

While the MAI has the virtue of simplicity and the ability to depict major shifts in the trends of food availability, its validity at household level is debatable. A major drawback is the inconclusive categorisation of food groups as Mediterranean and non-Mediterranean (Chang *et al.* 2017). For example, Balanza *et al.* (2007) question whether potatoes and seed oils are considered typical components of the Mediterranean diet or not.

In addition, the MAI reduces the Mediterranean diet as a whole to a list of food products regardless of the frequency of consumption as well as the proportion of food items (Finardi *et al.* 2018). To illustrate, both red meat and white meat are not separated but aggregated in the broader category of meat regardless of the nutritional differences between the two. Following the MAI's calculation, the same importance is given to all food groups, independently of how they align with the Mediterranean diet recommendations or the evidence-based influence of these foods on the diet-disease relationship.

Lastly, the inclusion of certain food items, for instance eggs and milk in the MAI's calculation may undermine the accuracy of the obtained values. This is partly because the FBS data represents raw food ingredient rather than the food groups in which it is found such as cakes and pastries, and partly because of the controversial effects of dairy products consumption on health (Vilarnau *et al.* 2019).

The COFUST's ability to handle time series data exhibiting weak dependence

The 'Fluctuation analysis' identified five meaningful clusters whose dietary characteristics match nicely with common dietary types around the world. Since some of these dietary types seem to carry the spatial relevance for example the tropical diet associated with Latin America or the cereal-based diet usually found among African countries, one would expect such spatial patterns to be more clearly defined in a spatial model. Surprisingly, the 'Spatial fluctuation analysis' could not arrive at any final cluster solution as indicated by the monotonic rising behaviour of the Fuzzy Silhouette index. There are good reasons to believe this reflects more of a technical problem related to the performance of the COFUST algorithm than implications for food consumption.

To recap, the 'Spatial fluctuation analysis' aims to detect groups of countries that exhibit common deviations from the trend whilst being spatially close. The input information of the cluster analysis is a combination of the dependence among detrended time series (pre-filtered data) and the spatial information. As the spatial information remains the same for both 'Spatial trend analysis' and

‘Spatial fluctuation analysis’, the deficiency of the latter should therefore be attributed to the dependence among detrended time series. As a matter of fact, this dependence measure was shown to be very weak since the pairwise Kendall’s correlation figures are close to zero (Figure 7.3). Apparently, the pre-filtering step transformed the original data into stationary time series which likely exhibit random patterns regarding the temporal variation. It is therefore more difficult to identify a common structure in the pre-filtered data than the original data, and introducing the spatial information appears to add further complication. In the original paper, the authors of the COFUST algorithm (Disegna *et al.* 2017) recommend decomposing the original data and utilising only the residuals in the clustering task. The purpose is to avoid any overwhelming influence of seasonality inherent in tourism data. As tourism demand can exhibit substantial variations from one month to the next due to seasonal effects, it is possible to observe some common patterns of deviations from the trend (as represented by the residuals of the decomposed time series). This study however resorts to annual data and broadly speaking the total caloric consumption figure does not alter significantly from one year to the next, resulting in minor deviations from the trend and decreasing the likelihood of common patterns existing.

Time-varying nature of the space metric

Spatial analysis deals with interactions of units (say countries, regions, municipalities or cities) in space, where the space can be geographical, social, economic or cultural in nature. An important element in spatial analysis is the spatial weight matrix, which summarises the spatial relationship among units. Conventionally, geographical measures of spatial dependence (say distance in kilometres) are preferred since not only geography is a good proxy for transportation costs and technological transfer, but also geographical measures are static and exogeneous. Therefore, the spatial weight matrix is usually time invariant. Nevertheless, the application of spatial econometric models spans over a range of disciplines beyond geographical and regional studies so that the notion of space would be less meaningful if it is limited to geography. A growing number of scholars have adopted non-geographical measures of spatial dependence that can be a proximity measure in terms of economic, socio-economic, trade, demographic, or climatic variables (Ahmad and Hall 2017; Hao and Wu 2020). This development of the spatial literature invalidates the time invariance property of conventional spatial weight matrices. As the non-geographical measures of spatial dependence change over time, the resulting spatial weight matrix does vary over time. This applies to the economic proximity measure adopted in this study.

Even though this issue was attacked by the use of average GDP per capita during 1970-2013, the time-varying nature of the economic proximity measure and the associated spatial weight matrix requires greater attention. While most countries in the study sample have always belonged to a certain income category for example Canada in the richest group and Cambodia in the poorest group, for others the variations in GDP per capita could be considerable. To illustrate, China was associated with the 1st quantile income in 1970 but the 3rd income quantile in 2013. Using average incomes would mask this switching behaviour and might impair the results of relevant spatial analyses (for example the Global Moran's I test for spatial dependence). Of importance is the conclusion of Lee and Yu (2012) that ignoring time varying feature of spatial weight matrices can cause estimation bias.

8.6 Further research

The limitations listed in the previous section present many opportunities for extending the scope of this research. Some of the directions for future studies are outlined subsequently.

First, acknowledging the lack of information related to food waste, future research could incorporate this information in the FBS data by adjusting the calories from food items for waste using the waste factors as it has been done in Behrens *et al.* (2017) or Azzam (2020). Even after modifying the data for food waste, it remains crucial to account for foods consumed away from home since these tend to be processed, high in sugars and unhealthy fat, and they represent an important dimension of the nutrition transition. Although sales data can inform about food away from home expenditures, consumption data are hard to find and unlikely to be available for a wide range of countries. Thus, one cannot overemphasise the need to collect and collate more accurate consumption data without which it would be difficult to get a comprehensive picture of what people are eating around the world.

Second, future studies could examine the robustness in measuring the healthiness of diets using MAI versus other diet quality indices, especially those constructed from both nutrients and food groups.

Third, the beta convergence analysis in Chapter 6 could be replicated to demonstrate how other structural conditions can affect the convergence process. Whilst the dummies representing income levels were included and results showed that the convergence path varies between countries of different levels of economic development, new groupings based on for example population size or language spoken could be considered.

Next, this thesis opens up a relatively new area of research for analysing the spatial relationship in the context of food consumption. Future studies could consider a distance in terms of geopolitical framework (for example, countries belonging to the European Union, Latin America including Mexico and the Caribbean, South-East Asia, Mercosur economies) or sub-continental measures of geographical proximity. Income was considered as the only proximity in this study, however, other factors such as education attainment, weather condition, demographic structure, language spoken may as well contribute to the (dis)similarities in diets around the world. Future research could examine these factors as separate proxies for spatial information or embed them simultaneously into the beta convergence model or cluster analysis as different levels of proximity.

In addition, since the COFUST is unable to identify meaningful clusters when the dependence among time series is weak (as in the case of clustering pre-filtered data), future studies could try an alternative space-time clustering algorithm. A suggestion would be the algorithm introduced in D'Urso *et al.* (2019a) that computes the dynamic time warping distance between time series instead of using the copula function.

Finally, further studies could deal with the situation in which spatial information and hence spatial dependence varies over time. One possibility is to conduct panel data models and let spatial weight matrices be time-varying. The panel data setting also allows the estimation of a spatial econometric model with time-varying coefficients in a spatial panel model. It is worth mentioning that when spatial proximity is constructed as the function of an economic or social distance, spatial weights could be endogenous as well as changing over time. In such scenario, endogenous weights might hinder the usual estimation procedure and cause estimation bias. If time-varying endogenous spatial weight matrices are employed, future research could adopt the estimation method suggested by Qu *et al.* (2017).

Despite the growing research interest in the estimation of spatial econometric model with time-varying spatial dependence, the clustering literature counterpart is still in its infancy. The development of a space-time clustering algorithm that can handle time-varying spatial information is thus a promising research venue for future studies. In addition, an extension of the Generalised Fuzzy Moran's index with time-varying spatial proximity would be needed.

8.7 Final remarks

To conclude, the author would like once again to re-emphasise the significance of this study. Motivated by the fact that diets represent a common cause of and contribute to a wide range of common diseases, this research aims to explore the global patterns of food consumption over time and the indications they imply for obesity and global health. To this aim, the daily per capita calories available for human consumption in 118 countries during 1961-2013 are empirically scrutinised by a range of quantitative methods including econometric convergence tests, clustering techniques and spatial analysis. The empirical evidence generates several important policy insights which can be used as guidance towards the adoption of more robust measures and more concerted actions to promote healthy diets and good nutrition. Further, this research updates and substantially extends the current knowledge on nutrition transition and dietary convergence using both updated data and new methods.

Overall, this study highlights the importance of diets as both a cause and a solution of the global burden of malnutrition, particularly the obesity epidemic. Empirical results unambiguously reveal that the world is eating badly, and this has adverse impacts on nutrition and health outcomes. On an optimistic note, the move towards healthier diets is at least possible, but we need to act now! This is a real challenge for the 21st century and is what the thesis intends to highlight through impactful research.

Appendix A Nutritional, demographic and epidemiological profiles of five patterns of the nutrition transition.

		Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 5
Nutritional profile	*Diet*	Plants, wild animals, varied diet	Cereals predominant, less varied diet	More starchy staples, fruits, vegetables, animal proteins, low variety diet continues	More animal fat, sugars, processed foods, less fibre	Higher-quality fat, reduced refined carbohydrates, more whole grains, fruits, vegetables
Demographic profile	*Fertility and mortality*	Low fertility; high mortality; short life expectancy	High natural fertility; high infant and maternal mortality; short life expectancy	Fertility remained constant, then declined; mortality declined slowly, then rapidly; population grew steadily and then exploded	Life expectancy hits unique levels (ages 60-70); huge decline and fluctuations in fertility	Life expectancy increases to ages 70-80; disability-free period increases
	Age structure	Young population	Young, very few elderly	Young, the shift to older population begins	Rapid increase in the proportion of elderly people	Increase in the proportion of elderly people (over 75 ages)
	Residency patterns	Rural, low density	Rural, a few small crowded cities	Mostly rural, move to cities increases, international migration begins, megacities develop	Dispersal of urban population decrease in rural green space	Lower-density cities rejuvenate, increase in urbanisation of rural areas encircling cities
Epidemiological profile	*Morbidity*	Infectious diseases, no epidemics	Epidemics, endemic disease, deficiency disease, starving	Tuberculosis, smallpox infection, parasitic diseases, polio, weaning disease expand and later decline	Chronic disease related to diet and pollution, but infectious disease declines	Increases in health promotion, rapid decline in cardiovascular disease, slower change in age-specific cancer profile

Source: Author's summarisation based on Popkin (2006).

Appendix B Summary of main diet quality indices

B.1 Based on dietary recommendations

Study	Index (Abbreviation)	Country /region	Dietary method	Index components		Index range	Weight	Original results
				Food groups /Nutrients /Both	Number of components			
Patterson *et al.* (1994)	Diet Quality Index (DQI-1994)	the US	24-hour recall and 2-day food records	Both	8	[0, 16]	Unequal	Reflects diet quality
Kennedy *et al.* (1995)	Healthy Eating Index (HEI-1995)	the US	24-hour recall and 2-day food records	Both	10	[0, 100]	Equal	Reflects nutrient adequacy
Huijbregts *et al.* (1997)	Healthy Diet Indicator (HDI-1997)	Europe	DHM	Both	9	[0, 9]	Equal	Negatively associated with mortality
Haines *et al.* (1999)	Diet Quality Index-Revised (DQI-R-1999)	the US	24-hour recall	Both	10	[0, 100]	Equal	Reflects diet quality
Löwik *et al.* (1999)	Dietary Quality Index Nutrient Based (DQINB-1999)	the Netherlands	2-day food records	Nutrients	5	[0, 5]	Equal	Not specified
McCullough *et al.* (2000a); McCullough *et al.* (2000b)	Healthy Eating Index-Frequency Questionnaire (HEI-f-2000)	the US	FFQ	Both	10	[0, 100]	Equal	Weakly associated with risk of major CHDs
Stookey *et al.* (2000)	Chinese Diet Quality Index (CH-DQI-2000)	China	24-hour recall	Both	10	[-74 ,56]	Unequal	Associated with BMI

Osler *et al.* (2001)	Healthy Food Index (HFI-2001)	Denmark	FFQ	Food groups	4	[0, 4]	Equal	Inversely associated with all-cause and cardiovascular mortality
Fitzgerald *et al.* (2002)	Diet Quality Score (DQS-2002)	Canada	24-hour recall	Nutrients	17	[0, 17]	Equal	Inversely related to cancer
Harnack *et al.* (2002)	Dietary Guidelines Index (DGI-2002)	the US	FFQ	Both	9	[0, 18]	Equal	Negatively associated with the risk of cancer among post-menopausal women
McCullough *et al.* (2002)	Alternate Healthy Eating Index (AHEI-2002)	the US	FFQ	Both	9	[2.5, 87.5]	Equal	Low correlation with the risk of CVDs and other CHDs. Not associated with cancer
Seymour *et al.* (2003)	Diet Quality Index (DQI-2003)	the US	FFQ	Both	8	[0, 16]	Unequal	Unrelated to cancer mortality Limited predictive capacity for mortality
Newby *et al.* (2003)	Diet Quality Index-Revised (DQI-R-2003)	the US	FFQ and 1-week food records	Both	10	[0, 100]	Equal	Reflects diet quality
Kim *et al.* (2003)	Diet Quality Index-International (DQI-I-2003)	Worldwide	24-hour recall	Both	17	[0, 100]	Unequal	Can compare diet quality across countries
Dynesen *et al.* (2003)	Danish Healthy Diet Index (D-HDI-2003)	Denmark	FFQ	Food groups	5	[0, 15]	Equal	Not specified
Maynard *et al.* (2005)	Healthy Diet Score (HDS-2005)	the UK	FFQ	Both	12	[0, 12]	Equal	Not specified
Shatenstein *et al.* (2005)	Canadian Healthy Eating Index (C-HEI-2005)	Canada	FFQ	Both	9	[0, 100]	Equal	Reflects diet quality of Canadian population
Fogli-Cawley *et al.* (2006)	Dietary Guidelines for Americans Adherence Index (DGAI-2005)	the US	FFQ	Both	20	[0, 20]	Equal	Not specified
Bazelmans *et al.* (2006)	Healthy Food and Nutrient Index (HFNI-2006)	Belgium	1-day food record	Both	8	[0, 8]	Equal	Associated with mortality for men
Toft *et al.* (2007)	Diet Quality Score (DQS-2007)	Denmark	FFQ	Food groups	4	[0, 12]	Equal	Negatively associated with the risk of ischaemic heart disease

Mazzocchi *et al.* (2008)	Recommendation Compliance Index (RCI-2008)	Worldwide	Food Balance Sheet	Both	7	[0, 1]	Unequal	Not specified
Lee *et al.* (2008)	Overall Dietary Index (ODI)	Taiwan	FFQ	Both	9	[0, 100]	Equal	Cannot predict blood pressure or hypertension
Lee *et al.* (2008)	Overall Dietary Index-revised (ODI-R-2008)	Taiwan	FFQ	Both	9	[0, 100]	Equal	Associated with risk of obesity
Fung *et al.* (2008)	Dietary Approaches to Stop Hypertension (DASH-2008)	the US	FFQ	Both	8	[8, 40]	Equal	Significantly associated with lower risk of CHDs and stroke in women
Guenther *et al.* (2008)	Healthy Eating Index-2005 (HEI-2005)	the US	24-hour recall	Both	12	[0, 100]	Unequal	Not specified
Kant *et al.* (2009)	Dietary Behaviour Score (DBS-2009)	the US	FFQ	Food groups	6	[0, 36]	Equal	Inversely associated with mortality
Von Ruesten *et al.* (2010)	German Food Pyramid Index (GFPI-2010)	Germany	FFQ	Food groups	8	[0, 110]	Equal	Not associated with the risk of CHDs
Cade *et al.* (2011)	Healthy Diet Indicator (HDI-2011)	the UK	FFQ	Both	10	[0, 10]	Equal	Not strongly associated with the risk of breast cancer
Drake *et al.* (2011)	Diet Quality Index-Swedish Nutrition Recommendation (DQI-SNR-2011)	Sweden	DHM	Both	6	[0, 6]	Equal	Not specified
Chiuve *et al.* (2012)	Alternate Healthy Eating Index (AHEI-2010)	the US	FFQ	Both	11	[0, 110]	Equal	Negatively associated with the risk of CHDs
Knudsen *et al.* (2012)	Danish Diet Quality Index (D-DQI-2012)	Denmark	Food diary	Both	6	[0, 6]	Equal	Reflects diet quality of Danish population
Berentzen *et al.* (2013)	Healthy Diet Indicator (HDI-2013)	the Netherlands	FFQ	Both	7	[0, 7]	Equal	Not associated with overall cancer risk
Guenther *et al.* (2013)	Healthy Eating Index-2010 (HEI-2010)	the US	Not specified	Both	12	[0, 100]	Unequal	Not specified
Zarrin *et al.* (2013)	Aussie-DQI	Australia	24-hour recall, FFQ	Both	11	[0, 120]	Equal	Inversely associated with cancer mortality among men

Kanerva *et al.* (2014)	Baltic Sea Diet Score (BSDS-M-2014)	Finland	FFQ	Both	9	[0, 9]	Equal	Reflects diet quality, nutrient adequacy
Kanerva *et al.* (2014)	Baltic Sea Diet Score (BSDS-Q-2014)	Finland	FFQ	Both	9	[0, 25]	Equal	Reflects diet quality
Radwan *et al.* (2015)	Obesity-specific Healthy Eating Index (OS-HEI)	the US	2-day food records	Food groups	7	[0, 100]	Unequal	Can predict obesity prevalence
National Cancer Institute (2018)	Healthy Eating Index-2015 (HEI-2015)	the US	Not specified	Both	13	[0, 100]	Unequal	Not specified
Kuczmarski *et al.* (2019)	Mean Adequacy Ratio (MAR)	the US	24-hour recall	Nutrients	17	[0, 100]	Equal	Not associated with risk for malnutrition
Gémez *et al.* (2019)	Diet Quality Score (DQS-2019)	Latin America	24-hour recall	Both	17	[0, 199]	Equal	Not specified
Moraeus *et al.* (2020)	Swedish Healthy Eating Index for Adolescents 2015 (SHEIA15)	Sweden	24-hour recall	Both	7	[0, 9]	Equal	Not associated with overweight and obesity

B.2 Based on the Mediterranean diet

Study	Index (Abbreviation)	Country/ region	Dietary method	Index components		Index range	Weight	Original results
				Food groups /Nutrients /Both	Number of components			
Trichopoulou *et al*. (1995)	Mediterranean Diet Score (MDS-1995)	Greece	FFQ	Both	8	[0, 8]	Equal	Inversely associated with overall mortality
Alberti-Fidanza *et al*. (1999)	Mediterranean Adequacy Index (MAI-1999)	Italy	DHM	Food groups	12	[0, 99.99]	Equal	Not specified
Gerber *et al*. (2000)	Mediterranean Diet Quality Index (Med-DQI-2000)	South France	FFQ	Both	7	[0, 14]	Equal	Not associated with cholesterol, Inversely associated with vitamin E, omega-3 fatty acids and beta-carotene
Haveman-Nies *et al*. (2001)	Mediterranean Diet Score (MDS-2001)	the US, Europe	FFQ and modified DHM	Both	8	[0, 8]	Equal	Not specified
Haveman-Nies *et al*. (2002)	Mediterranean Diet Score (MDS-2002)	Europe	modified DHM	Both	7	[0, 7]	Equal	Associated with higher survival among old people
Trichopoulou *et al*. (2003)	Mediterranean Diet Score (MDS-2003)	Greece	FFQ	Both	9	[0, 9]	Equal	Significantly associated with all types of mortality
Goulet *et al*. (2003)	Mediterranean Score (MS-2003)	Canada	FFQ	Food groups	11	[0, 44]	Equal	Associated with beneficial modifications in metabolic and anthropometric variables in women
Knoops *et al*. (2004)	Mediterranean Diet Score (MDS-2004)	Europe	DHM	Both	8	[0, 8]	Equal	Associated with lower rate of mortality
Trichopoulou *et al*. (2005)	Modified Mediterranean Diet Score (mMDS-2005)	Europe	FFQ, food record	Both	14	[0, 9]	Equal	Associated with higher survival among old people

Fung *et al.* (2005)	Alternate Mediterranean Diet Score (aMED-2005)	the US	FFQ	Both	9	[0, 9]	Equal	Might be associated with risk of diabetes and CVDs
Gerber (2006)	Mediterranean Diet Quality Index (Med-DQI-2006)	France	FFQ	Both	7	[0, 14]	Equal	Inversely associated with alpha, beta-carotene, vitamin E, EPA, DHA
Panagiotakos *et al.* (2007)	MedDietScore-2007	Greece	FFQ	Food groups	11	[0, 55]	Equal	Associated with reduced risks of hypertension, obesity, hypercholesterolemia diabetes mellitus
Rumawas *et al.* (2009)	Mediterranean-Style Dietary Pattern Score (MSDPS-2009)	the US	FFQ	Food groups	13	[0, 100]	Equal	Reflects diet quality
Buckland *et al.* (2009)	Relative Mediterranean Diet (rMED-2009)	Spain	Dietary history questionnaire	Food groups	9	[0, 18]	Equal	Negatively associated with the risk of CHDs
Buckland *et al.* (2010)	Relative Mediterranean Diet (rMED-2010)	Europe	Dietary history questionnaire, FFQ, food record	Food groups	9	[0, 18]	Equal	Negatively associated with the risk of incident gastric adenocarcinoma
Cade *et al.* (2011)	Mediterranean Diet Score (MDS-2011)	the UK	FFQ	Both	10	[0, 10]	Equal	Not associated with the risk of breast cancer
Dominguez *et al.* (2013)	Mediterranean Adherence Diet Screener (MEDAS-2013)	Spain	FFQ	Food groups	13	[0, 13]	Unequal	Inversely associated with mortality
Buckland *et al.* (2013)	Adapted Relative Mediterranean Diet (arMED-2013)	Europe	Dietary history questionnaire, FFQ, food record	Food groups	8	[0, 16]	Equal	Negatively associated with the risk of breast cancer
Yang *et al.* (2014)	Modified Mediterranean Diet Score (mMDS-2014)	the US	Lifestyle questionnaire	Food groups	10	[0, 42]	Equal	Inversely associated with metabolic syndrome, cholesterol, and weight gain
El Kinany *et al.* (2020)	Modified Mediterranean Diet Score (MMD-2020)	Morocco	FFQ	Both	12	[0, 12]	Equal	Associated with reduced overweight/obesity risk

Abbreviations: FFQ = Food frequency questionnaire, DHM = Dietary history method, CHDs = Chronic diseases, CVDs = cardiovascular diseases, EPA = Eicosapentaenoic acid, DHA = Docosahexaenoic acid, BMI = Body mass index.

Glossary of Terms

a posteriori	A data-driven method that applies factor or cluster analysis to derive common underlying food consumption patterns within a population.
a priori	A method that evaluates the dietary healthiness (through pre-defined diet quality indices/scores) and categorises individuals according to the extent to which their dietary patterns comply with dietary guidelines.
Beta convergence	The process in which poor countries grow faster than rich countries at an earlier stage before converging to grow at similar rates in the long term.
Body Mass Index	A ratio of weight-for-height commonly used to classify underweight, normal weight, overweight and obesity in adults.
Cluster analysis	A technique to discover groups (or clusters) of objects/units so that objects/units in the same cluster are more similar to one another whilst dissimilar from objects/units in other clusters.
Crisp clustering	Allows each unit to only belong to one cluster and the cluster membership is either 0 or 1.
Demographic transition	The shift from a pattern of high fertility and mortality to one of low fertility and mortality.
Dietary assessment	Methods that estimate the consumption of food and nutrients at national, household and individual level.
Double burden of malnutrition	The coexistence of undernutrition along with overweight and obesity or diet-related non-communicable diseases, within individuals, households and populations, and across the life course (WHO 2017b).
Epidemiological transition	The shift from a high prevalence of infectious diseases to one of a high prevalence of chronic and degenerative diseases associated with urban-industrial lifestyles.
Food environment	The interface where people interact with the wider food system to acquire and consume foods (Turner *et al.* 2020).
Food system	Any activity that influences the production, processing, distribution, retailing, and consumption of food.
Foreign direct investment	A type of cross-border investment in which an investor residing in a country buys or establishes a lasting interest in controlling over assets in another country (OECD iLibrary 2020).

Fuzzy clustering	Allows a unit to belong to multiple clusters with varying degrees of membership between 0 and 1.
Globalisation	A process of greater integration within the world economy, through movements of goods and services, capital, technology and (to a lesser extent) labour, which leads increasingly to economic decisions being influenced by global conditions (Jenkins 2004).
Half-life	The number of years required for progress halfway towards the steady-state level when convergence is assumed to have been achieved.
Hunger	An uncomfortable or painful sensation caused by insufficient food consumption (FAO 2013).
Macronutrients	Fats, proteins and carbohydrates (starch, fibre, sugar) that are needed for a wide range of bodily functions and processes.
Malnutrition	Deficiencies, excesses or imbalances in a person's intake of energy and/or nutrients (WHO 2020).
Micronutrient	Vitamins, minerals and certain other substances required by the body in smaller amounts for normal physiological function.
Micronutrient deficiency	Not getting enough of one or more micronutrients.
Non-communicable diseases	Diseases that are not passed from person to person. They are often long lasting and generally progress slowly, for example cardiovascular diseases, cancer, chronic respiratory diseases and diabetes (WHO 2018).
Nutrient-dense food	Food with a high content of nutrients with respect to its mass or volume (FAO 2013).
Nutrition transition	A model describing the shifts in diets, physical activities and causes of disease that accompany changes in economic development, lifestyle, urbanisation, and demography.
Obesity	A state of excessive accumulation of fats, measured by a Body Mass Index greater than or equal to 30 kg/m^2.
Overnutrition	A result of excessive food intake relative to dietary nutrient requirements (FAO 2013).
Overweight	A body mass index between 25 and 30 kg/m^2.
Processed foods	Foods that have been altered from its raw state (OECD 2021).
Sigma convergence	The reduction in cross-sectional dispersion over time.
Spatial dependence	The existence of a functional relationship between a phenomenon happening in a location and what happens in other locations (Anselin 1988).

Stunting	Low height-for-age.
Temporal dependence	The phenomenon that values of a variable depend on past values of the same variable.
Trade liberalisation	The process of reducing or removing restrictions on international trade.
Ultra-processed foods	Formulations of ingredients that result from a series of industrial processes (Monteiro *et al.* 2019a).
Undernutrition	Includes wasting (low weight-for-height), stunting (low height-for-age) and underweight (low weight-for-age) (WHO 2020).
Underweight	Low weight-for-age.
Urbanisation	The shift from a population that is dispersed across small rural areas in which agriculture is the dominant economic activity towards one where the population is concentrated in larger, dense urban settlements (National Research Council 2003).
Wasting	Low weight-for-height.

List of References

Abdulai, A. and Aubert, D., 2004. Nonparametric and parametric analysis of calorie consumption in Tanzania. *Food Policy*, 29 (2), 113-129.

Abramovitz, M., 1986. Catching up, forging ahead, and falling behind. *The Journal of Economic History*, 46 (2), 385-406.

Adhikari, R. and Putnam, K. J., 2020. Comovement in the commodity futures markets: An analysis of the energy, grains, and livestock sectors. *Journal of Commodity Markets*, 18, 100090.

Afonso, H., LaFleur, M. and Alarcon, D., 2015. *Inequality measurement*. Department of Economic and Social Affairs.

Afshin, A., Sur, P. J., Fay, K. A., Cornaby, L., Ferrara, G., Salama, J. S., Mullany, E. C., Abate, K. H., Abbafati, C., Abebe, Z., Afarideh, M., Aggarwal, A., Agrawal, S., Akinyemiju, T., Alahdab, F., Bacha, U., Bachman, V. F., Badali, H., Badawi, A., Bensenor, I. M., Bernabe, E., Biryukov, S. H., Biadgilign, S. K. K., Cahill, L. E., Carrero, J. J., Cercy, K. M., Dandona, L., Dandona, R., Dang, A. K., Degefa, M. G., Zaki, M. E. S., Esteghamati, A., Esteghamati, S., Fanzo, J., Farinha, C. S. E. S., Farvid, M. S., Farzadfar, F., Feigin, V. L., Fernandes, J. C., Flor, L. S., Foigt, N. A., Forouzanfar, M. H., Ganji, M., Geleijnse, J. M., Gillum, R. F., Goulart, A. C., Grosso, G., Guessous, I., Hamidi, S., Hankey, G. J., Harikrishnan, S., Hassen, H. Y., Hay, S. I., Hoang, C. L., Horino, M., Islami, F., Jackson, M. D., James, S. L., Johansson, L., Jonas, J. B., Kasaeian, A., Khader, Y. S., Khalil, I. A., Khang, Y.-H., Kimokoti, R. W., Kokubo, Y., Kumar, G. A., Lallukka, T., Lopez, A. D., Lorkowski, S., Lotufo, P. A., Lozano, R., Malekzadeh, R., Marz, W., Meier, T., Melaku, Y. A., Mendoza, W., Mensink, G. B. M., Micha, R., Miller, T. R., Mirarefin, M., Mohan, V., Mokdad, A. H., Mozaffarian, D., Nagel, G., Naghavi, M., Nguyen, C. T., Nixon, M. R., Ong, K. L., Pereira, D. M., Poustchi, H., Qorbani, M., Rai, R. K., Razo-Garcia, C., Rehm, C. D., Rivera, J. A., Rodriguez-Ramirez, S., Roshandel, G., Roth, G. A., Sanabria, J., Sanchez-Pimienta, T. G., Sartorius, B., Schmidhuber, J., Schutte, A. E., Sepanlou, S. G., Shin, M.-J., Sorensen, R. J. D., Springmann, M., Szponar, L., Thorne-Lyman, A. L., Thrift, A. G., Touvier, M., Tran, B. X., Tyrovolas, S., Ukwaja, K. N., Ullah, I., Uthman, O. A., Vaezghasemi, M., Vasankari, T. J., Vollset, S. E., Vos, T., Vu, G. T., Vu, L. G., Weiderpass, E., Werdecker, A., Wijeratne, T., Willett, W. C., Wu, J. H., Xu, G., Yonemoto, N., Yu, C. and Murray, C. J. L., 2019. Health effects of dietary risks in 195 countries, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 393 (10184), 1958-1972.

Aghabozorgi, S., Shirkhorshidi, A. S. and Wah, T. Y., 2015. Time-series clustering – A decade review. *Information Systems*, 53, 16-38.

Ahmad, M. and Hall, S. G., 2017. Economic growth and convergence: Do institutional proximity and spillovers matter? *Journal of Policy Modeling*, 39 (6), 1065-1085.

Aizenman, J. and Brooks, E., 2008. Globalisation and taste convergence: The cases of wine and beer. *Review of International Economics*, 16 (2), 217-233.

Alberti, A., Fruttini, D. and Fidanza, F., 2009. The Mediterranean Adequacy Index: Further confirming results of validity. *Nutrition, Metabolism and Cardiovascular Diseases*, 19 (1), 61-66.

Alberti-Fidanza, A. and Fidanza, F., 2004. Mediterranean Adequacy Index of Italian diets. *Public Health Nutrition*, 7 (7), 937-941.

Alberti-Fidanza, A., Fidanza, F., Chiuchiu, M. P., Verducci, G. and Fruttini, D., 1999. Dietary studies on two rural Italian population groups of the Seven Countries Study. 3. Trend of food and nutrient intake from 1960 to 1991. *European Journal of Clinical Nutrition*, 53 (11), 854.

Aldenderfer, M. S. and Blashfield, R. K., 1984. *Cluster analysis: Quantitative applications in the social sciences*. Beverly Hills: Sage Publication.

Aleksandrowicz, L., Green, R., Joy, E. J. M., Smith, P. and Haines, A., 2016. The impacts of dietary change on greenhouse gas emissions, land use, water use, and health: A systematic review. *PloS One*, 11 (11), e0165797.

Alexandratos, N. and Bruinsma, J., 2012. *World agriculture towards 2030/2050: The 2012 Revision*. Rome: Agricultural Development Economics Division, Food and Agricultural Organisation of the United Nations (FAO).

Aljuraiban, G. S., Gibson, R., Oude Griep, L. M., Okuda, N., Steffen, L. M., Van Horn, L. and Chan, Q., 2019. Perspective: The application of a priori diet quality scores to cardiovascular disease risk - A critical evaluation of current scoring systems. *Advances in Nutrition*, 11 (1), 10-24.

Alkerwi, A., 2014. Diet quality concept. *Nutrition*, 30 (6), 613-618.

Alonso, A. M., Berrendero, J. R., Hernández, A. and Justel, A., 2006. Time series clustering based on forecast densities. *Computational Statistics & Data Analysis*, 51 (2), 762-776.

Alonso, A. M. and Maharaj, E. A., 2006. Comparison of time series using subsampling. *Computational Statistics & Data Analysis*, 50 (10), 2589-2599.

Alsunni, A. A., 2015. Energy drink consumption: Beneficial and adverse health effects. *International Journal of Health Sciences*, 9 (4), 468.

Ambagna, J. J., Dury, S. and Dop, M. C., 2019. Estimating trends in prevalence of undernourishment: advantages of using HCES over the FAO approach in a case study from Cameroon. *Food Security*, 11 (1), 93-107.

Ambroise, C., Dang, M. and Govaert, G., 1997. Clustering of spatial data by the EM algorithm. *geoENV I - Geostatistics for environmental applications.* Springer, 493-504.

Ameye, H. and Swinnen, J., 2019. Obesity, income and gender: The changing global relationship. *Global Food Security*, 23, 267-281.

Anand, R., 2011. A study of determinants impacting consumers food choice with reference to the fast food consumption in India. *Society and Business Review*, 6 (2), 176-187.

Anand, S. S., Hawkes, C., De Souza, R. J., Mente, A., Dehghan, M., Nugent, R., Zulyniak, M. A., Weis, T., Bernstein, A. M. and Krauss, R. M., 2015. Food consumption and its impact on cardiovascular disease: Importance of solutions focused on the globalised food system: A report from the workshop convened by the World Heart Federation. *Journal of the American College of Cardiology*, 66 (14), 1590-1614.

Anderson, K., 2010. Globalisation's effects on world agricultural trade, 1960-2050. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365 (1554), 3007-3021.

Angulo, A. M., Gil, J. M. and Gratia, A., 2001. Calorie intake and income elasticities in EU countries: A convergence analysis using cointegration. *Papers in Regional Science*, 80 (2), 165-187.

Annoni, P., de Dominicis, L. and Khabirpour, N., 2019. Location matters: A spatial econometric analysis of regional resilience in the European Union. *Growth and Change*, 50 (3), 824-855.

Anselin, L., 1988. *Spatial econometrics: methods and models*. Dordrecht: Kluwer.

Anselin, L., 1992. Space and applied econometrics: Introduction. *Regional Science and Urban Economics*, 22 (3), 307-316.

Anselin, L., 1994. *Testing for spatial dependence in linear regression models: A review*. Morgantown: West Virginia University.

Anselin, L., 1995. Local indicators of spatial association - LISA. *Geographical Analysis*, 27 (2), 93-115.

Anselin, L., 2002. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27 (3), 247-267.

Anselin, L., 2003. Spatial Econometrics. *A Companion to Theoretical Econometrics*, 310-330.

Anselin, L., 2006. Spatial econometrics. *In:* Mills, T. C. and Patterson, K., eds. *Palgrave handbook of econometrics.* Basingstoke: Palgrave, 901-969.

Anselin, L., 2010. Thirty years of spatial econometrics. *Papers in Regional Science*, 89 (1), 3-25.

Anselin, L., 2013. *Spatial econometrics: methods and models*. Vol. 4. Springer Science & Business Media.

Anselin, L. and Bera, A., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. *In:* Ullah, A. and Giles, D., eds. *Handbook of Applied Economics Statistics.* New York: Marcel Dekker, 237-289.

Anselin, L., Bera, A. K., Florax, R. and Yoon, M. J., 1996. Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26 (1), 77-104.

Anselin, L. and Florax, R., 1995. Small sample properties of tests for spatial dependence in regression models: Some further results. *In:* Anselin, L. and Florax, R., eds. *New directions in spatial econometrics.* Berlin: Springer-Verlag, 21-74.
Anselin, L. and Rey, S. J., 1997. Introduction to the special issue on spatial econometrics. *International Regional Science Review*, 20 (1 & 2), 1-7.
Arbia, G., 2006. *Spatial econometrics: statistical foundations and applications to regional convergence*. Berlin: Springer.
Arimond, M. and Ruel, M. T., 2004. Dietary diversity is associated with child nutritional status: Evidence from 11 demographic and health surveys. *The Journal of Nutrition*, 134 (10), 2579-2585.
Arsenault, J. E., Hijmans, R. J. and Brown, K. H., 2015. Improving nutrition security through agriculture: an analytical framework based on national food balance sheets to estimate nutritional adequacy of food supplies. *Food Security*, 7 (3), 693-707.
Arvaniti, F. and Panagiotakos, D. B., 2008. Healthy indexes in public health practice and research: A review. *Critical Reviews in Food Science and Nutrition*, 48 (4), 317-327.
Athanasopoulos, G., Ahmed, R. A. and Hyndman, R. J., 2009. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, 25 (1), 146-166.
Atkin, D., 2016. The caloric costs of culture: Evidence from Indian migrants. *American Economic Review*, 106 (4), 1144-1181.
Atkinson, A. and Bourguignon, F., 2015. *Handbook of income distribution*. 1st edition edition. Vol. Volume 2A. North Holland: Elsevier.
Atkinson, A. B., 1970. On the measurement of inequality. *Journal of Economic Theory*, 2 (3), 244-263.
Atkinson, A. B., 1983. *The Economics of Inequality*. 2nd edition. Oxford: Clarendon Press.
Austin, S. B., Melly, S. J., Sanchez, B. N., Patel, A., Buka, S. and Gortmaker, S. L., 2005. Clustering of fast-food restaurants around schools: A novel application of spatial statistics to the study of food environments. *American Journal of Public Health*, 95 (9), 1575-1581.
Azzam, A., 2020. Is the world converging to a 'Western diet'? *Public Health Nutrition*, 24 (2), 309-317.
Bach, A., Serra-Majem, L., Carrasco, J. L., Roman, B., Ngo, J., Bertomeu, I. and Obrador, B., 2006. The use of indexes evaluating the adherence to the Mediterranean diet in epidemiological studies: A review. *Public Health Nutrition*, 9 (1a), 132-146.
Bach-Faig, A., Fuentes-Bol, C., Ramos, D., Carrasco, J. L., Roman, B., Bertomeu, I. F., Cristia, E., Geleva, D. and Serra-Majem, L., 2011. The Mediterranean diet in Spain: Adherence trends during the past two decades using the Mediterranean Adequacy Index. *Public Health Nutrition*, 14 (4), 622-628.
Badiane, O., Odjo, S. and Collins, J., 2018. *Africa Agriculture Trade Monitor Report*. Washington, DC: International Food Policy Research Institute (IFPRI).
Bagnall, A., Lines, J., Bostrom, A., Large, J. and Keogh, E., 2017. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31 (3), 606-660.
Bahadur, K. C., Dias, G. M., Veeramani, A., Swanton, C. J., Fraser, D., Steinke, D., Lee, E., Wittman, H., Farber, J. M. and Dunfield, K., 2018. When too much isn't enough: Does current food production meet global nutritional needs? *PLoS One*, 13 (10), e0205683.
Bai, J., Seale Jr, J. L. and Wahl, T. I., 2020. Meat demand in China: To include or not to include meat away from home? *The Australian Journal of Agriculture and Resource Economics*, 64 (1), 150-170.
Bai, Y., Alemu, R., Block, S. A., Headey, D. and Masters, W. A., 2021. Cost and affordability of nutritious diets at retail prices: Evidence from 177 countries. *Food Policy*, 99, 101983.
Bailey, R. and Harper, D. R., 2015. *Reviewing interventions for healthy and sustainable diets*. London: Royal Institute of International Affairs.
Baker, P. and Friel, S., 2014. Processed foods and the nutrition transition: Evidence from Asia. *Obesity Reviews*, 15 (7), 564-577.
Baker, P., Friel, S., Schram, A. and Labonte, R., 2016. Trade and investment liberalisation, food systems change and highly processed food consumption: A natural experiment contrasting the soft-drink markets of Peru and Bolivia. *Globalisation and Health*, 12, 24.

Baker, P., Kay, A. and Walls, H., 2014. Trade and investment liberalisation and Asia's noncommunicable disease epidemic: A synthesis of data and existing literature. *Globalisation and Health*, 10 (1), 66.

Baker, P., Machado, P., Santos, T., Sievert, K., Backholer, K., Hadjikakou, M., Russell, C., Huse, O., Bell, C., Scrinis, G., Worsley, A., Friel, S. and Lawrence, M., 2020. Ultra-processed foods and the nutrition transition: Global, regional and national trends, food systems transformations and political economy drivers. *Obesity Reviews*, 21 (12), e13126.

Balanza, R., García-Lorda, P., Pérez-Rodrigo, C., Aranceta, J., Bonet, M. B. and Salas-Salvadó, J., 2007. Trends in food availability determined by the Food and Agriculture Organisation's food balance sheets in Mediterranean Europe in comparison with other European areas. *Public Health Nutrition*, 10 (2), 168-176.

Baldos, U. L. C. and Hertel, T. W., 2014. Global food security in 2050: the role of agricultural productivity and climate change. *Australian Journal of Agricultural and Resource Economics*, 58 (4), 554-570.

Bansal, S. and Zilberman, D., 2020. Macrorelationship between average life expectancy and prevalence of obesity: Theory and evidence from global data. *Agricultural Economics*, 51 (3), 403-427.

Baraldi, A. and Blonda, P., 1999. A survey of fuzzy clustering algorithms for pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29, 786-801.

Baraldi, L. G., Steele, E. M., Canella, D. S. and Monteiro, C. A., 2018. Consumption of ultra-processed foods and associated sociodemographic factors in the USA between 2007 and 2012: Evidence from a nationally representative cross-sectional study. *BMJ Open*, 8 (3), e020574.

Barlow, P., Loopstra, R., Tarasuk, V. and Reeves, A., 2020. Liberal trade policy and food insecurity across the income distribution: An observational analysis in 132 countries, 2014-2017. *The Lancet Global Health*, 8 (8), e1090-e1097.

Barlow, P., McKee, M., Basu, S. and Stuckler, D., 2017. The health impact of trade and investment agreements: A quantitative systematic review and network co-citation analysis. *Globalisation and Health*, 13, 13.

Barro, R. J. and Sala-i-Martin, X., 1992. Convergence. *Journal of political Economy*, 100 (2), 223-251.

Basford, K. E. and McLachlan, G. J., 1985. The mixture method of clustering applied to three-way data. *Journal of Classification*, 2 (1), 109-125.

Batista, G. E., Wang, X. and Keogh, E. J., 2011. A complexity-invariant distance measure for time series, *The 2011 SIAM international conference on data mining* (pp. 699-710): SIAM, Mesa.

Baumol, W. J., 1986. Productivity growth, convergence, and welfare: what the long-run data show. *The American Economic Review*, 1072-1085.

Baumont, C., Ertur, C. and Le Gallo, J., 2003. Spatial convergence clubs and the European regional growth process, 1980–1995. *In:* Fingleton, B., ed. *European regional growth.* Berlin: Springer, 131-158.

Bazelmans, C., De Henauw, S., Matthys, C., Dramaix, M., Kornitzer, M., De Backer, G. and Levêque, A., 2006. Healthy food and nutrient index and all cause mortality. *European Journal of Epidemiology*, 21 (2), 145-152.

BCFN, 2012. *Obesity - The impacts on public health and society*. Barilla Center for Food & Nutrition.

Beal, T., Massiot, E., Arsenault, J. E., Smith, M. R. and Hijmans, R. J., 2017. Global trends in dietary micronutrient supplies and estimated prevalence of inadequate intakes. *PloS One*, 12 (4), e0175554.

Beck, N., Gleditsch, K. S. and Beardsley, K., 2006. Space is more than geography: Using spatial econometrics in the study of political economy. *International Studies Quarterly*, 50 (1), 27-44.

Bedoui, R., Braiek, S., Guesmi, K. and Chevallier, J., 2019. On the conditional dependence structure between oil, gold and USD exchange rates: Nested copula based GJR-GARCH model. *Energy Economics*, 80, 876-889.

Behrens, P., Kiefte-de Jong, J. C., Bosker, T., Rodrigues, J. F. D., De Koning, A. and Tukker, A., 2017. Evaluating the environmental impacts of dietary recommendations. *Proceedings of the National Academy of Sciences*, 114 (51), 13412-13417.

Belasco, W., 2008. *Food: The key concepts*. Oxford: Berg Publishers.

Bell, K. P. and Bockstael, N. E., 2000. Applying the Generalized-Moments estimation approach to spatial problems involving microlevel data. *The Review of Economics and Statistics*, 82 (1), 72-82.

Bell, W., Lividini, K. and Masters, W. A., 2021. Global dietary convergence from 1970 to 2010 altered inequality in agriculture, nutrition and health. *Nature Food*, 2 (3), 156-165.

Bellemare, M. F., Çakir, M., Peterson, H. H., Novak, L. and Rudi, J., 2017. On the measurement of food waste. *American Journal of Agricultural Economics*, 99 (5), 1148-1158.

Bellu, L. G. and Liberati, P., 2006a. *Describing income inequality: Theil index and Entropy class indexes*. Rome: Food and Agriculture Organisation of the United Nations (FAO).

Bellu, L. G. and Liberati, P., 2006b. *Policy impacts on inequality: Welfare based measures of inequality - The Atkinson index*. Rome: The Food and Agriculture Organisation of the United Nations (FAO).

Bennett, M. K., 1941. Wheat in national diets. *Wheat Studies*, 18 (02), 37-76.

Bentham, J., Singh, G. M., Danaei, G., Green, R., Lin, J. K., Stevens, G. A., Farzadfar, F., Bennett, J. E., Di Cesare, M. and Dangour, A. D., 2020. Multidimensional characterisation of global food supply from 1961 to 2013. *Nature Food*, 1 (1), 70-75.

Benthem de Grave, R., Rust, N. A., Reynolds, C. J., Watson, A. W., Smeddinck, J. D. and Souza Monteiro, D. M., 2020. A catalogue of UK household datasets to monitor transitions to sustainable diets. *Global Food Security*, 24, 100344.

Benton, T., Fairweather, D., Graves, A., Harris, J., Jones, A., Lenton, T., Norman, R., O'Riordan, T., Pope, E. and Tiffin, R., 2017. *Environmental tipping points and food system dynamics: Main report*. UK: The Global Food Security programme.

Berdanier, C. D., Dwyer, J. T. and Heber, D., 2014. *Handbook of Nutrition and Food*. 3rd edition.: CRC Press.

Berentzen, N. E., Beulens, J. W., Hoevenaar-Blom, M. P., Kampman, E., Bueno-de-Mesquita, H. B., Romaguera-Bosch, D., Peeters, P. H. M. and May, A. M., 2013. Adherence to the WHO's healthy diet indicator and overall cancer risk in the EPIC-NL cohort. *PLoS One*, 8 (8), e70535.

Berg, A. and Krueger, A. O., 2003. Trade, growth, and poverty: A selective survey, *Annual World Bank Conference on Development Economics* (pp. 47-91). Washington DC: The World Bank.

Berndt, D. J. and Clifford, J., 1994. Using dynamic time warping to find patterns in time series, *AAAI-94 Workshop on Knowledge Discovery in Databases*.

Bernstein, M. A., Tucker, K. L., Ryan, N. D., O'Neill, E. F., Clements, K. M., Nelson, M. E., Evans, W. J. and Singh, M. A. F., 2002. Higher dietary variety is associated with better nutritional status in frail elderly people. *Journal of the American Dietetic Association*, 102 (8), 1096-1104.

Bertail, P. and Caillavet, F., 2008. Fruit and vegetable consumption patterns: a segmentation approach. *American Journal of Agricultural Economics*, 90 (3), 827-842.

Bertazzon, S., 2003. Spatial and temporal autocorrelation in innovation diffusion analysis. *In:* Kumar, V., Gavrilova, M. L., Tan, C. J. K. and L'Ecuyer, P., eds. *Computational Science and Its Applications, ICCSA 2003.* Berlin, Heidelberg: Springer, 23-32.

Bezdek, J., 1974. Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology*, 1, 57-71.

Bezdek, J. C., 1981. *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press.

Bigna, J. J. and Noubiap, J. J., 2019. The rising burden of non-communicable diseases in sub-Saharan Africa. *The Lancet Global Health*, 7 (10), e1295-e1296.

Bindiya, M. V., Unnikrishman, A. and Poulouse, J. K., 2013. Spatial clustering algorithms - An overview. *Asian Journal of Computer Science and Information Technology*, 3 (1), 1-8.

Bivand, R. S., 1984. Regression modeling with spatial dependence: An application of some class selection and estimation methods. *Geographical Analysis*, 16 (1), 25-37.

Blandford, D., 1984. Changes in food consumption patterns in the OECD area. *European Review of Agricultural Economics*, 11 (1), 43-64.

Blouin, C., Chopra, M. and van der Hoeven, R., 2009. Trade and social determinants of health. *The Lancet*, 373 (9662), 502-507.

Bodirsky, B. L., Dietrich, J. P., Martinelli, E., Stenstad, A., Pradhan, P., Gabrysch, S., Mishra, A., Weindl, I., Le Mouël, C., Rolinski, S., Baumstark, L., Wang, X., Waid, J. L., Lotze-Campen,

H. and Popp, A., 2020. The ongoing nutrition transition thwarts long-term targets for food security, public health and environmental protection. *Scientific Reports*, 10 (1), 19778.
Borkowski, B., Dudek, H. and Szczesny, W., 2008. Food consumption convergence within Europe: A panel data analysis. *Polish Journal of Enviromental Studies*, 18 (5B), 41-47.
Borlaug, N. E., 1971. *The green revolution, peace and humanity* [online]. Available from: https://www.nobelprize.org/prizes/peace/1970/borlaug/lecture/ [Accessed 25th August 2021].
Bosu, W. K., 2015. An overview of the nutrition transition in West Africa: Implications for non-communicable diseases. *Proceedings of The Nutrition Society*, 74 (4), 466-477.
Bouis, H. E. and Haddad, L. J., 1992. Are estimates of calorie income elasticities too high - a recalibration of the plausible range. *Journal of Development Economics*, 39 (2), 333-364.
Bourne, L. T., Lambert, E. V. and Steyn, K., 2002. Where does the black population of South Africa stand on the nutrition transition? *Public Health Nutrition*, 5 (1a), 157-162.
Bradatan, C., 2003. Cuisine and cultural identity in Balkans. *Anthropology of East Europe Review*, 21 (1), 43-47.
Bradshaw, J. M., Ensor, S., Lorenz, H., Spence, C. and Hucker, G., 2019. Underestimating the true impact of obesity. *The Lancet Public Health*, 4 (1), e16.
Braha, K., Cupák, A., Pokrivčák, J., Qineti, A. and Rizov, M., 2017. Economic analysis of the link between diet quality and health: Evidence from Kosovo. *Economics & Human Biology*, 27, 261-274.
Breda, J., Castro, L. S. N., Whiting, S., Williams, J., Jewell, J., Engesveen, K. and Wickramasinghe, K., 2020. Towards better nutrition in Europe: Evaluating progress and defining future directions. *Food Policy*, 96, 101887.
Brož, V. and Kočenda, E., 2018. Dynamics and factors of inflation convergence in the European Union. *Journal of International Money and Finance*, 86, 93-111.
Brunelle, T., Dumas, P. and Souty, F., 2014. The impact of globalisation on food and agriculture: The case of the diet convergence. *The Journal of Environment & Development*, 23 (1), 41-65.
Brusco, M. J. and Cradit, J. D., 2001. A variable-selection heuristic for K-means clustering. *Psychometrika*, 66 (2), 249-270.
Bryngelsson, D., Wirsenius, S., Hedenus, F. and Sonesson, U., 2016. How can the EU climate targets be met? A combined analysis of technological and demand-side changes in food and agriculture. *Food Policy*, 59, 152-164.
Buckland, G., Agudo, A., Luján, L., Jakszyn, P., Bueno-de-Mesquita, H. B., Palli, D., Boeing, H., Carneiro, F., Krogh, V. and Sacerdote, C., 2010. Adherence to a Mediterranean diet and risk of gastric adenocarcinoma within the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort study. *The American Journal of Clinical Nutrition*, 91 (2), 381-390.
Buckland, G., González, C. A., Agudo, A., Vilardell, M., Berenguer, A., Amiano, P., Ardanaz, E., Arriola, L., Barricarte, A. and Basterretxea, M., 2009. Adherence to the Mediterranean diet and risk of coronary heart disease in the Spanish EPIC Cohort Study. *American Journal of Epidemiology*, 170 (12), 1518-1529.
Buckland, G., Travier, N., Cottet, V., Gonzalez, C. A., Luján-Barroso, L., Agudo, A., Trichopoulou, A., Lagiou, P., Trichopoulos, D. and Peeters, P. H., 2013. Adherence to the Mediterranean diet and risk of breast cancer in the European prospective investigation into cancer and nutrition cohort study. *International Journal of Cancer*, 132 (12), 2918-2927.
Budayan, C., Dikmen, I. and Birgonul, M. T., 2009. Comparing the performance of traditional cluster analysis, self-organizing maps and fuzzy C-means method for strategic grouping. *Expert Systems with Applications*, 36 (9), 11772-11781.
Burggraf, C., Kuhn, L., Zhao, Q. R., Teuber, R. and Glauben, T., 2015. Economic growth and nutrition transition: An empirical analysis comparing demand elasticities for foods in China and Russia. *Journal of Integrative Agriculture*, 14 (6), 1008-1022.
Burggraf, C., Teuber, R., Brosig, S. and Meier, T., 2018. Review of a priori dietary quality indices in relation to their construction criteria. *Nutrition reviews*, 76 (10), 747-764.
Burridge, P., 1980. On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42 (1), 107-108.

Buttrey, S. E. and Karo, C., 2002. Using k-nearest-neighbor classification in the leaves of a tree. *Computational Statistics & Data Analysis*, 40 (1), 27-37.
Béné, C., Oosterveer, P., Lamotte, L., Brouwer, I. D., de Haan, S., Prager, S. D., Talsma, E. F. and Khoury, C. K., 2019. When food systems meet sustainability - Current narratives and implications for actions. *World Development*, 113, 116-130.
Cabral, R. and Castellanos-Sosa, F. A., 2019. Europe's income convergence and the latest global financial crisis. *Research in Economics*, 73 (1), 23-34.
Cade, J. E., Taylor, E. F., Burley, V. J. and Greenwood, D. C., 2011. Does the Mediterranean dietary pattern or the Healthy Diet Index influence the risk of breast cancer in a large British cohort of women? *European journal of clinical nutrition*, 65 (8), 920.
Caiado, J. and Crato, N., 2010. Identifying common dynamic features in stock returns. *Quantitative Finance*, 10 (7), 797-807.
Caiado, J., Crato, N. and Pena, D., 2006. A periodogram-based metric for time series classification. *Computational Statistics and Data Analysis*, 50 (10), 2668-2684.
Caiado, J., Crato, N. and Pena, D., 2009. Comparison of times series with unequal length in the frequency domain. *Communications in Statistics: Simulation & Computation*, 38 (3), 527-540.
Caiado, J., Maharaj, E. A. and D'Urso, P., 2015. Time series clustering. *In:* Hennig, C., Meila, M., Murtagh, F. and Rocci, R., eds. *Handbook of Cluster Analysis.* Chapman and Hall/CRC, 241-264.
Cameron, A. C. and Trivedi, P. K., 2009. *Microeconometrics Using Stata*. Texas: Stata Press.
Campello, R. J. G. B. and Hruschka, E. R., 2006. A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, 157 (21), 2858-2875.
Carmone, F. J., Kara, A. and Maxwell, S., 1999. HINoV: a new model to improve market segment de fi nition by identifying noisy variables. *Journal of Marketing Research*, 36 (4), 501-509.
Carracedo, P., Debón, A., Iftimi, A. and Montes, F., 2018. Detecting spatio-temporal mortality clusters of European countries by sex and age. *International Journal for Equity in Health*, 17 (1), 38.
Casini, L., Contini, C., Marone, E. and Romano, C., 2013. Food habits. Changes among young Italians in the last 10 years. *Appetite*, 68, 21-29.
Cassisi, C., Montalto, P., Aliotta, M. A. and Pulvirenti, A., 2012. Similarity measures and dimensionality reduction techniques for time series data mining. *Advances in Data Mining Knowledge Discovery and Applications*, 71-96.
Cebeci, Z., 2019. Comparison of internal validity indices for fuzzy clustering. *Journal of Agricultural Informatics*, 10 (2), 1-14.
CEPII, 2019. *GeoDist* [online]. Research and Expertise on the World Economy. Available from: http://www.cepii.fr/cepii/en/bdd_modele/presentation.asp?id=6 [Accessed 12th June 2019].
Chandon, P. and Wansink, B., 2012. Does food marketing need to make us fat? A review and solutions. *Nutrition Reviews*, 70 (10), 571-593.
Chang, T. F. M., Lepellere, M. A., Iseppi, L. and De Lorenzo, A., 2017. Food styles and the dynamics of the Mediterranean Adequacy Index. *New Medit*, 16 (3), 28-38.
Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A., 2014. NbClust: An R package for determining the relevant number of clusters in a dataset. *Journal of Statistical Software*, 61, 1-36.
Chasco, C. Y. and López, F. A. H., 2008. Is spatial dependence an instantaneous effect? Some evidence in economic series of Spanish provinces. *Estadistica Espanola*, 50, 101-118.
Chauvin, N. D., Mulangu, F. and Porto, G., 2012. *Food production and consumption trends in sub-Saharan Africa: Prospects for the transformation of the agricultural sector*. New York, NY: United Nations Development Programme, Regional Bureau for Africa.
Chen, D., Jaenicke, E. C. and Volpe, R. J., 2016. Food environments and obesity: household diet expenditure versus food deserts. *American Journal of Public Health*, 106 (5), 881-888.
Chen, J. S. and Hsu, C. H. C., 1999. The use of logit analysis to enhance market segmentation methodology. *Journal of Hospitality and Tourism Research*, 23 (3), 268-283.
Chen, Q. and Marques-Vidal, P., 2007. Trends in food availability in Portugal in 1966–2003. *European Journal of Nutrition*, 46 (7), 418-427.
Chen, Y., 2012. On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, 5 (2), 65-72.

Chen, Y., 2013. New approaches for calculating Moran’s index of spatial autocorrelation. *PloS One*, 8 (7), e68336.

Cheng, Z., Jiang, L., Liu, D. and Zheng, Z., 2018. Density based spatio-temporal trajectory clustering algorithm, *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3358-3361).

Chiuve, S. E., Fung, T. T., Rimm, E. B., Hu, F. B., McCullough, M. L., Wang, M., Stampfer, M. J. and Willett, W. C., 2012. Alternative dietary indices both strongly predict risk of chronic disease. *The Journal of Nutrition*, 142 (6), 1009-1018.

Chocholatá, M. and Furková, A., 2017. Does the location and the institutional background matter in convergence modelling of the EU regions? *Central European Journal of Operations Research*, 25 (3), 679-697.

Choudhury, S. and Headey, D., 2017. What drives diversification of national food supplies? A cross-country analysis. *Global Food Security*, 15, 85-93.

Cirera, X. and Masset, E., 2010. Income distribution trends and future food demand. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365 (1554), 2821-2834.

Clements, K. W. and Si, J., 2018. Engel’s law, diet diversity, and the quality of food consumption. *American Journal of Agricultural Economics*, 100 (1), 1-22.

Clements, K. W., Wu, Y. and Zhang, J., 2006. Comparing international consumption patterns. *Empirical Economics*, 31 (1), 1-30.

Cliff, A. D. and Ord, J. K., 1969. The problem of spatial autocorrelation. *In:* Scott, A. J., ed. *London Papers in Regional Science 1, Studies in Regional Science.* London: Pion, 25-55.

Cliff, A. D. and Ord, J. K., 1973. *Spatial Autocorrelation*. London: Pion.

Cliff, A. D. and Ord, J. K., 1981. *Spatial processes: Models and applications*. London: Pion.

Clément, M., Levasseur, P., Seetahul, S. and Piaser, L., 2021. Does inequality have a silver lining? Municipal income inequality and obesity in Mexico. *Social Science & Medicine*, 272, 113710.

Coates, J., Rogers, B. L., Blau, A., Lauer, J. and Roba, A., 2017. Filling a dietary data gap? Validation of the adult male equivalent method of estimating individual nutrient intakes from household-level data in Ethiopia and Bangladesh. *Food Policy*, 72, 27-42.

Cockx, L., Colen, L. and De Weerdt, J., 2018. From corn to popcorn? Urbanisation and dietary change: Evidence from rural-urban migrants in Tanzania. *World Development*, 110, 140-159.

Cockx, L., Colen, L., De Weerdt, J. and Paloma, G. Y., 2019. *Urbanisation as a driver of changing food demand in Africa: Evidence from rural-urban migration in Tanzania*. Sevilla, Spain: European Commission, Joint Research Centre (JRC).

Colozza, D. and Avendano, M., 2019. Urbanisation, dietary change and traditional food practices in Indonesia: A longitudinal analysis. *Social Science & Medicine*, 233, 103-112.

Cook, S. and Winfield, T., 2013. Crime across the States: Are US crime rates converging? *Urban Studies*, 50 (9), 1724-1741.

Cook, S. J., Hays, J. C. and Franzese, R. J., 2015. Model specification and spatial interdependence. *The 2015 Texas A&M Conference on Innovations in Comparative Policital Methodology*, Texas.

Coppi, R. and D'Urso, P., 2003. Three-way fuzzy clustering models for LR fuzzy time trajectories. *Computational Statistics & Data Analysis*, 43, 149-177.

Coppi, R. and D'Urso, P., 2006. Fuzzy unsupervised classification of multivariate time trajectories with the Shannon entropy regularization. *Computational Statistics & Data Analysis*, 50, 1452-1477.

Coppi, R., D’Urso, P. and Giordani, P., 2010. A fuzzy clustering model for multivariate spatial time series. *Journal of Classification*, 27, 54-88.

Corduas, M., 2010. Mining time series data: a selective survey. *Data Analysis and Classification.* Springer, 355-362.

Cormack, R. M., 1971. A review of classification. *Journal of the Royal Statistical Society. Series A (General)*, 321-367.

Cornelsen, L., Alarcon, P., Häsler, B., Amendah, D. D., Ferguson, E., Fèvre, E. M., Grace, D., Dominguez-Salas, P. and Rushton, J., 2016. Cross-sectional study of drivers of animal-source food consumption in low-income urban areas of Nairobi, Kenya. *BMC Nutrition*, 2 (1), 70.

Cornil, Y. and Chandon, P., 2016. Pleasure as an ally of healthy eating? Contrasting visceral and Epicurean eating pleasure and their association with portion size preferences and wellbeing. *Appetite*, 104, 52-59.

Corrado, L. and Fingleton, B., 2012. Where is the economics in spatial econometrics? *Journal of Regional Science*, 52 (2), 210-239.

Costa-Font, J., Fabbri, D. and Gil, J., 2010. Decomposing cross-country differences in levels of obesity and overweight: does the social environment matter? *Social Science & Medicine*, 70 (8), 1185-1193.

Costa-Font, J. and Gil, J., 2008. What lies behind socio-economic inequalities in obesity in Spain? A decomposition approach. *Food Policy*, 33 (1), 61-73.

Costa-Font, J. and Mas, N., 2016. 'Globesity'? The effects of globalization on obesity and caloric intake. *Food Policy*, 64, 121-132.

Coulston, A. M., Boushey, C. J., Ferruzzi, M. G. and Delahanty, L. M., 2017. *Nutrition in the Prevention and Treatment of Disease*. 4th edition. London: Academic Press.

Cowling, K., Stuart, E. A., Neff, R. A., Vernick, J., Magraw, D. and Pollack Porter, K., 2020. The relationship between joining a US free trade agreement and processed food sales, 2002–2016: A comparative interrupted time-series analysis. *Public Health Nutrition*, 23 (9), 1609-1617.

Cressie, N. A. C., 1993. *Statistics for spatial data*. New York: Wiley.

Crimmins, E. M., Saito, Y. and Ingegneri, D., 1989. Changes in life expectancy and disability-free life expectancy in the United States. *Population and Development Review*, 235-267.

Cuadrado, C., Dunstan, J., Silva-Illanes, N., Mirelman, A. J., Nakamura, R. and Suhrcke, M., 2020. Effects of a sugar-sweetened beverage tax on prices and affordability of soft drinks in Chile: A time series analysis. *Social Science & Medicine*, 245, 112708.

D'Urso, P., 2004. Fuzzy C-Means Clustering Models for Multivariate Time-Varying Data: Different Approaches. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12 (3), 287-326.

D'Urso, P., 2005. Fuzzy clustering for data time arrays with inlier and outlier time trajectories, *IEEE Transactions on Fuzzy Systems* (pp. 583).

D'Urso, P., Cappelli, C., Di Lallo, D. and Massari, R., 2013. Clustering of financial time series. *Physica A*, 392 (9), 2114-2129.

D'Urso, P., De Giovanni, L. and Massari, R., 2016. GARCH-based robust clustering of time series. *Fuzzy Sets and Systems*, 305, 1-28.

D'Urso, P., De Giovanni, L. and Massari, R., 2018. Robust fuzzy clustering of multivariate time trajectories. *International Journal of Approximate Reasoning*, 99, 12-38.

D'Urso, P. and Maharaj, E. A., 2009. Autocorrelation-based fuzzy clustering of time series. *Fuzzy Sets and Systems*, 160 (24), 3565-3589.

D'Urso, P., Maharaj, E. A. and Alonso, A. M., 2017. Fuzzy clustering of time series using extremes. *Fuzzy Sets and Systems*, 318, 56-79.

Da Rocha, B. R. S., Rico-Campà, A., Romanos-Nanclares, A., Ciriza, E., Barbosa, K. B. F., Martínez-González, M. A. and Martín-Calvo, N., 2020. Adherence to Mediterranean diet is inversely associated with the consumption of ultra-processed foods among Spanish children: the SENDO project. *Public Health Nutrition*, 1-10.

Da Silva, R., Bach-Faig, A., Quintana, B. R., Buckland, G., de Almeida, M. D. V. and Serra-Majem, L., 2009. Worldwide variation of adherence to the Mediterranean diet, in 1961–1965 and 2000–2003. *Public Health Nutrition*, 12 (9A), 1676-1684.

Dapi, L. N., Omoloko, C., Janlert, U., Dahlgren, L. and Håglin, L., 2007. "I eat to be happy, to be strong, and to live." Perceptions of rural and urban adolescents in Cameroon, Africa. *Journal of Nutrition Education and Behaviour*, 39 (6), 320-326.

Datar, A., Nicosia, N. and Shier, V., 2014. Maternal work and children's diet, activity, and obesity. *Social Science & Medicine*, 107, 196-204.

Dave, D., Doytch, N. and Kelly, I. R., 2016. Nutrient intake: A cross-national analysis of trends and economic correlates. *Social Science & Medicine*, 158, 158-167.

De Luca, G. and Zuccolotto, P., 2011. A tail dependence-based dissimilarity measure for financial time series clustering. *Advances in Data Analysis and Classification*, 5 (4), 323-340.

De Luca, G. and Zuccolotto, P., 2017. Dynamic tail dependence clustering of financial time series. *Statistical Papers*, 58 (3), 641-657.

De Miguel-Etayo, P., Moreno, L. A., Santabárbara, J., Martín-Matillas, M., Azcona-San Julian, M. C., Marti del Moral, A., Campoy, C., Marcos, A., Garagorri, J. M., López-Belmonte, G., Delgado, M., Aparicio, V., Carbonell, A., Agil, A., Silva, D. R., Pérez-Ballesteros, C., Piqueras, M. J., Chillón, P., Tercedor, P., Martín-Lagos, J. A., Martín-Bautista, E., Pérez-Expósito, M., Garófano, M., Aguilar, M. J., Fernández-Mayorga, A., Sánchez, P., Wärnberg, J., Puertollano, M. A., Gómez-Martínez, S., Zapatera, B., Nova, E., Romeo, J., Díaz, E. L., Pozo, T., Morandé, G., Villaseñor, A., Madruga, D., Muñoz, R., Veiga, O., Villagra, A., Martínez-Gómez, D., Vaquero, M. P., Pérez-Granados, A. M., Navas-Carretero, S., Martí, A., Azcona-SanJulian, C., Moleres, A., Rendo, T., Marqués, M., Miranda, M. G., Martínez, J. A., Redondo-Figuero, C., García-Fuentes, M., DeRufino, P., González-Lamuño, D., Amigo, T., Sanz, R., Romero, P., Rodríguez, G., Bueno, G., Mesana, M. I., Vicente-Rodríguez, G., Fernández, J., Rey, P., Muro, C., Tomás, C., Calle, M. E. and Barrios, L., 2019. Diet quality index as a predictor of treatment efficacy in overweight and obese adolescents: The EVASYON study. *Clinical Nutrition*, 38 (2), 782-790.

de Ridder, D., Kroese, F., Evers, C., Adriaanse, M. and Gillebaart, M., 2017. Healthy diet: Health impact, prevalence, correlates, and interventions. *Psychology & Health*, 32 (8), 907-941.

De Sousa, J., Mayer, T. and Sihra, E., 2018. *Market integration and convergence in consumption patterns*. FREIT (Forum for Research in International Trade).

de Soysa, I. and de Soysa, A. K., 2017. Do globalisation and free markets drive obesity among children and youth? An empirical analysis, 1990–2013. *International Interactions*, 44 (1), 88-106.

Debela, B. L., Demmler, K. M., Klasen, S. and Qaim, M., 2020. Supermarket food purchases and child nutrition in Kenya. *Global Food Security*, 25, 100341.

DeFries, R., Fanzo, J., Remans, R., Palm, C., Wood, S. and Anderman, T. L., 2015. Metrics for land-scarce agriculture. *Science*, 349 (6245), 238-240.

Dekker, L. H., Rijnks, R. H., Strijker, D. and Navis, G. J., 2017. A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands – the Lifelines cohort study. *International Journal of Behavioral Nutrition and Physical Activity*, 14 (1), 166.

Del Gobbo, L. C., Khatibzadeh, S., Imamura, F., Micha, R., Shi, P., Smith, M., Myers, S. S. and Mozaffarian, D., 2015. Assessing global dietary habits: A comparison of national estimates from the FAO and the Global Dietary Database. *The American Journal of Clinical Nutrition*, 101 (5), 1038-1046.

Delgado, C. L., 2003. Rising consumption of meat and milk in developing countries has created a new food revolution. *The Journal of Nutrition*, 133 (11), 3907S-3910S.

Desiere, S., Hung, Y., Verbeke, W. and D'Haese, M., 2018. Assessing current and future meat and fish consumption in Sub-Sahara Africa: Learnings from FAO Food Balance Sheets and LSMS household survey data. *Global Food Security*, 16, 116-126.

Development Initiatives, 2018. *2018 Global Nutrition Report*.

Development Initiatives, 2020. *2020 Global Nutrition Report: Action on equity to end malnutrition*. Bristol, UK: Development Initiatives.

Di Lascio, F. and Giannerini, S., 2012. A Copula-Based Algorithm for Discovering Patterns of Dependent Observations. *Journal of Classification*, 29 (1), 50-75.

Di Lascio, F. M. L. and Disegna, M., 2017. A copula-based clustering algorithm to analyse EU country diets. *Knowledge-Based Systems*, 132, 72-84.

Dimitrov, A. and Atanasova, G., 1964. Food Consumption in Bulgaria. *Eastern European Economics*, 3 (1), 60-67.

Disegna, M., D'Urso, P. and Durante, F., 2017. Copula-based fuzzy clustering of spatial time series. *Spatial Statistics*, 21 (Part A), 209-225.

Dithmer, J. and Abdulai, A., 2017. Does trade openness contribute to food security? A dynamic panel analysis. *Food Policy*, 69, 218-230.

Djokoto, J. G., 2012. Effects of foreign direct investment inflows into agriculture on food security in Ghana. *International Journal of Innovation and Sustainable Development*, 3 (2), 81-92.

Doak, C. M., Adair, L. S., Bentley, M., Monteiro, C. and Popkin, B. M., 2005. The dual burden household and the nutrition transition paradox. *International Journal of Obesity*, 29 (1), 129-136.

Dogbe, W., 2021. Can poverty status explain obesity in developing countries? Evidence from Ghana. *Agribusiness*, 37 (2), 409-421.

Dolnicar, S., 2003. Using cluster analysis for market segmentation - Typical misconceptions, established methodological weaknesses and some recommendations for improvement. *Australasian Journal of Market Research*, 11 (2), 5-12.

Dolnicar, S., Grabler, K., Grun, B. and Kulnig, A., 2011. Key drivers of airline loyalty. *Tourism Management*, 32 (5), 1020-1026.

Dolnicar, S., Grün, B. and Leisch, F., 2016. Increasing sample size compensates for data problems in segmentation studies. *Journal of Business Research*, 69 (2), 992-999.

Dolnicar, S., Grün, B., Leisch, F. and Schmidt, K., 2014. Required sample sizes for data-driven market segmentation analyses in tourism. *Journal of Travel Research*, 53 (3), 296-306.

Dolnicar, S., Kaiser, S., Lazarevski, K. and Leisch, F., 2012. Biclustering overcoming data dimenstionality problems in market segmentation. *Journal of Travel Research*, 51 (1), 41-49.

Dominguez, L. J., Bes-Rastrollo, M., De la Fuente-Arrillaga, C., Toledo, E., Beunza, J. J., Barbagallo, M. and Martinez-Gonzalez, M. A., 2013. Similar prediction of total mortality, diabetes incidence and cardiovascular events using relative-and absolute-component Mediterranean diet score: the SUN cohort. *Nutrition, Metabolism and Cardiovascular Diseases*, 23 (5), 451-458.

Doorslaer, E. v. and Koolman, X., 2004. Explaining the differences in income-related health inequalities across European countries. *Health Economics*, 13 (7), 609-628.

Drake, I., Gullberg, B., Ericson, U., Sonestedt, E., Nilsson, J., Wallström, P., Hedblad, B. and Wirfält, E., 2011. Development of a diet quality index assessing adherence to the Swedish nutrition recommendations and dietary guidelines in the Malmö Diet and Cancer cohort. *Public Health Nutrition*, 14 (5), 835-845.

Dreher, A., 2006. Does globalization affect growth? Evidence from a new index of globalization. *Applied Economics*, 38 (10), 1091-1110.

Drewnowski, A., 2009. Obesity, diets, and social inequalities. *Nutrition Reviews*, 67 (suppl_1), S36-S39.

Drewnowski, A. and Darmon, N., 2005. Food choices and diet costs: an economic analysis. *The Journal of Nutrition*, 135 (4), 900-904.

Drewnowski, A., Hanks, A. S. and Smith, T. G., 2010. International trade, food and diet costs, and the global obesity epidemic. *In:* Hawkes, C., Blouin, C., Henson, S., Drager, N. and Dubé, L., eds. *Trade, food, diet and health: Perspectives and policy options.* Wiley-Blackwell, 77-90.

Drewnowski, A. and Popkin, B. M., 1997. The nutrition transition: New trends in the global diet. *Nutrition Reviews*, 55 (2), 31-43.

Drukker, D. M., Egger, P. and Prucha, I. R., 2013. On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econometric Reviews*, 32 (5-6), 686-733.

Du, S., Mroz, T. A., Zhai, F. and Popkin, B. M., 2004. Rapid income growth adversely affects diet quality in China - particularly for the poor! *Social Science & Medicine*, 59 (7), 1505-1515.

Dubois, P., Griffith, R. and Nevo, A., 2014. Do prices and attributes explain international differences in food purchases? *American Economic Review*, 104 (3), 832-867.

Dubé, J. and Legros, D., 2013a. A spatio-temporal measure of spatial dependence: An example using real estate data. *Papers in Regional Science*, 92 (1), 19-30.

Dubé, J. and Legros, D., 2013b. Dealing with spatial data pooled over time in statistical models. *Letters in Spatial and Resource Sciences*, 6 (1), 1-18.

Dudek, H., 2014. Do shares of food expenditure in the European Union converge? A country-level panel data analysis. *Economic Computation and Economic Cybernetics Studies and Research*, 48 (4), 245-260.

Dufour, D. L., Bender, R. L. and Reina, J. C., 2015. Local trends in diet in urban Colombia, 1990–1995 to 2008: Little evidence of a nutrition transition among low-income women. *American Journal of Human Biology*, 27 (1), 106-115.

Dunn, J. C., 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3, 32-57.

Dunn, J. C., 1974. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, 4 (1), 95-104.

Dupouy, E. and Gurinovic, M., 2020. Sustainable food systems for healthy diets in Europe and Central Asia: Introduction to the special issue. *Food Policy*, 96, 101952.
Durante, F., Pappadà, R. and Torelli, N., 2014. Clustering of financial time series in risky scenarios. *Advances in Data Analysis and Classification*, 8 (4), 359-376.
Durante, F., Pappadà, R. and Torelli, N., 2015. Clustering of time series via non-parametric tail dependence estimation. *Statistical Papers*, 56 (3), 701-721.
Durante, F. and Sempi, C., 2016. *Principles of copula theory*. CRC Press.
Durbin, J., 1954. Errors in variables. *Review of the International Statistical Institute*, 22, 23-32.
Dynesen, A. W., Haraldsdóttir, J., Holm, L. and Astrup, A., 2003. Sociodemographic differences in dietary habits described by food frequency questions—results from Denmark. *European Journal of Clinical Nutrition*, 57 (12), 1586.
Dyson, T., 2011. The role of the demographic transition in the process of urbanisation. *Population and Development Review*, 37, 34-54.
Díaz Dapena, A., Rubiera-Morollon, F. and Paredes, D., 2019. New approach to economic convergence in the EU: A multilevel analysis from the spatial effects perspective. *International Regional Science Review*, 42 (3-4), 335-367.
Díaz-Bonilla , E. and Thomas, M., 2015. *Why some are more equal than others: Country typologies of food security*. Rome: Food and Agriculture Organisation of the United Nations.
Díaz-Bonilla , E., Thomas, M., Robinson, S. and Cattaneo, A., 2000. *Food security and trade negotiations in the World Trade Organization*. International Food Policy Research Institute (IFPRI).
d'Amour, C. B., Pandey, B., Reba, M., Ahmad, S., Creutzig, F. and Seto, K. C., 2020. Urbanisation, processed foods, and eating out in India. *Global Food Security*, 25, 100361.
D'Urso, P., De Giovanni, L., Disegna, M. and Massari, R., 2019a. Fuzzy clustering with spatial–temporal information. *Spatial Statistics*, 30, 71-102.
D'Urso, P., De Giovanni, L. and Massari, R., 2019b. Trimmed fuzzy clustering of financial time series based on dynamic time warping. *Annals of Operations Research*, 1-17.
D'Urso, P., Di Lallo, D. and Maharaj, E. A., 2013. Autoregressive model-based fuzzy clustering and its application for detecting information redundancy in air pollution monitoring networks. *Soft Computing*, 17 (1), 83-131.
D'Urso, P., Manca, G., Waters, N. and Girone, S., 2019c. Visualising regional clusters of Sardinia's EU supported agriculture: A Spatial Fuzzy Partitioning Around Medoids. *Land Use Policy*, 83, 571-580.
D'Urso, P. and Vichi, M., 1998. Dissimilarities between trajectories of a three-way longitudinal data set. *Advances in data science and classification*. Springer, 585-592.
Eaton, S. B. and Konner, M., 1985. Paleolithic nutrition: A consideration of its nature and current implications. *New England Journal of Medicine*, 312 (5), 283-289.
Eaton, S. B., Konner, M. and Shostak, M., 1988. Stone Agers in the fast lane: Chronic degenerative diseases in evolutionary perspective. *The American Journal of Medicine*, 84 (4), 739-749.
Echouffo-Tcheugui, J. B. and Ahima, R. S., 2019. Does diet quality or nutrient quantity contribute more to health? *The Journal of Clinical Investigation*, 129 (10), 3969-3970.
El Kinany, K., Deoula, M. M. S., Hatime, Z., Boudouaya, H. A., Atassi, M., El Asri, A., Benslimane, A., Nejjari, C., Ibrahimi, S. A. and Lagiou, P., 2020. Modified Mediterranean diet score adapted to a southern Mediterranean population and its relation to overweight and obesity risk. *Public Health Nutrition*, 1-7.
Elhorst, J. P., 2001. Dynamic models in space and time. *Geographical Analysis*, 33 (2), 119-140.
Elhorst, J. P., 2010. Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, 5 (1), 9-28.
Elhorst, J. P., 2014. *Spatial Econometrics: From cross-sectional data to spatial panels*. Heidelberg: Springer-Verlag.
Elsner, K. and Hartmann, M., 1998. *Convergence of food consumption patterns between Eastern and Western Europe*. Germany: Institute of Agricultural Development in Central and Eastern Europe.
Engle-Stone, R. and Brown, K. H., 2015. Comparison of a household consumption and expenditures survey with nationally representative food frequency questionnaire and 24-hour dietary recall

data for assessing consumption of fortifiable foods by women and young children in Cameroon. *Food and Nutrition Bulletin*, 36 (2), 211-230.
Erbe Healy, A., 2014. Convergence or difference? Western European household food expenditure. *British Food Journal*, 116 (5), 792-804.
Ernst, D. and Dolnicar, S., 2018. How to avoid random market segmentation solutions. *Journal of Travel Research*, 57 (1), 69-82.
Eshete, H., Abebe, Y., Loha, E., Gebru, T. and Tesheme, T., 2017. Nutritional status and effect of maternal employment among children aged 6–59 months in Wolayta Sodo town, southern Ethiopia: A cross-sectional study. *Ethiopian Journal of Health Sciences*, 27 (2), 155-162.
Esling, P. and Agon, C., 2013. Time-Series Data Mining. *ACM Computing Surveys*, 45 (1), 1-34.
Ester, M., Kriegel, H. P., Sander, J. and Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise (34. Vol. 96, pp. 226-231).
Estivill-Castro, V. and Yang, J., 2004. Fast and robust general purpose clustering algorithms. *Data Mining Knowledge Discovery*, 8 (2), 127-150.
Euromonitor International, 2013. *Packaged Food in Malaysia*. London: Euromonitor International.
Euromonitor International, 2020. *Modern grocery retailers in world* [online]. Euromonitor International. Available from:
Everett, M., 2006. *Foreign Direct Investment: An analysis of its significance*. Central Bank of Ireland.
Everitt, B. S., Landau, S., Leese, M. and Stahl, D., 2011. *Cluster Analysis* [Bibliographies
Non-fiction]. 5th edition. Chichester, West Sussex, U.K: John Wiley & Sons.
Ezzati, M. and Riboli, E., 2013. Behavioral and dietary risk factors for noncommunicable diseases. *New England Journal of Medicine*, 369 (10), 954-964.
Fan, S., 2020. Reflections of food policy evolution over the last three decades. *Applied Economic Perspectives and Policy*, 42 (3), 380-394.
FAO, 2001. *Food Balance Sheets: A handbook*. Rome: Food and Agriculture Organisation of the United Nations (FAO).
FAO, 2003. Chapter 3: Calculation of the energy content of foods - Energy conversion factors. *Food energy - methods of analysis and conversion factors.* Rome, Italy: Food and Agriculture Organisation of the United Nations.
FAO, 2013. *Glossary of commonly used nutrition terms* [online]. Rome: FAO. Available from: http://www.fao.org/faoterm/collection/nutrition/en/ . [Accessed 24 February 2021].
FAO, 2015. *Global Initiative on Food Loss and Waste Reduction*. Rome: Food and Agricultural Organisation of the United Nations.
FAO, 2017a. *Food consumption expenditure* [online]. Rome: Food and Agriculture Organisation of the United Nations (FAO). Available from: www.fao.org/fileadmin/templates/ess/documents/food.../ShareOfFood_en.xls . [Accessed 8th August 2021].
FAO, 2017b. *Guidelines for the compliation of Food Balance Sheets*. Rome: Food and Agriculture Organisation.
FAO, 2018. *Dietary Assessment: A resource guide to method selection and application in low resource settings*. Rome: Food and Agriculture Organisation of the United Nations.
FAO, 2019a. *Key differences between New and Old Food Balance Sheet (FBS) Methodology*. Rome, Italy: FAO.
FAO, 2019b. *Macro Indicators* [online]. Food and Agricultural Organisation of the United Nations. Available from: http://www.fao.org/faostat/en/#data/MK [Accessed 14th June 2019].
FAO, 2019c. *The Food Balance Sheet* [online]. Food and Agricultural Organisation. Available from: http://www.fao.org/faostat/en/#data/FBSH [Accessed 12th June 2019].
FAO, 2020a. *Crop Perspects and Food Situation, March 2020: Quarterly Global Report*. Rome: The Food and Agriculture Organisation of the United Nations (FAO).
FAO, 2020b. *Food Balance Sheets and the Food Consumption Survey: A Comparison of Methodologies and Results* [online]. Rome, Italy: The Food and Agriculture Organisation of the United Nations (FAO). Available from: http://www.fao.org/economic/the-statistics-division-ess/methodology/methodology-systems/food-balance-sheets-and-the-food-consumption-survey-a-comparison-of-methodologies-and-results/en/ [Accessed 12th November 2020].

FAO, 2021a. *Consumer Price Indices* [online]. Rome: Food and Agriculture Organisation of the United Nations (FAO). Available from: http://www.fao.org/faostat/en/#data/CP/metadata [Accessed 22nd August 2021].

FAO, 2021b. *The State of Food Security and Nutrition in the World* [online]. Rome: Food and Agricultural Organisation of the United Nations (FAO). Available from: http://www.fao.org/publications/sofi/2020/en/ [Accessed 20th February 2021].

FAO, IFAD, UNICEF, WFP and WHO, 2019. *The State of Food Security and Nutrition in the World 2019. Safeguarding against economic slowdowns and downturns*. Rome: The Food and Agricultural Organisation of the United Nations, FAO.

FAO, IFAD, UNICEF, WFP and WHO, 2020. *The State of Food Security and Nutrition in the World 2020. Transforming food systems for affordable healthy diets*. Rome: Food and Agriculture Organisation of the United Nations (FAO).

FAO and WHO, 2001. *Human vitamin and mineral requirement*. Rome: FAO.

FAO/WHO, 2014. *Conference Outcome Document: Rome Declaration on Nutrition*. Rome, Italy: Second International Conference on Nutrition.

Fenech, J. P. and Vosgha, H., 2019. Oil price and Gulf Corporation Council stock indices: New evidence from time-varying copula models. *Economic Modelling*, 77, 81-91.

Fidanza, F., Alberti, A., Lanti, M. and Menotti, A., 2004. Mediterranean Adequacy Index: Correlation with 25-year mortality from coronary heart disease in the Seven Countries Study. *Nutrition, Metabolism and Cardiovascular Diseases*, 14 (5), 254-258.

Fiedler, J. L., Lividini, K., Bermudez, O. I. and Smitz, M. F., 2012. Household Consumption and Expenditures Surveys (HCES): A primer for food and nutrition analysts in low-and middle-income countries. *Food and Nutrition Bulletin*, 33, S170-S184.

Fiedler, J. L., Martin-Prével, Y. and Moursi, M., 2013. Relative costs of 24-hour recall and Household Consumption and Expenditures Surveys for nutrition analysis. *Food and Nutrition Bulletin*, 34 (3), 318-330.

Filomeno, M., Bosetti, C., Garavello, W., Levi, F., Galeone, C., Negri, E. and La Vecchia, C., 2014. The role of a Mediterranean diet on the risk of oral and pharyngeal cancer. *British Journal of Cancer*, 111 (5), 981.

Finardi, C., Bucchini, L. and Turrini, A., 2018. "Mediterranean Diet 'reflections'". Estimating adherence to the Mediterranean diet through secondary data. *Progress in Nutrition*, 20 (3), 344-360.

Fingleton, B., 2003. Externalities, economic geography, and spatial econometrics: Conceptual and modeling developments. *International Regional Science Review*, 26 (2), 197-207.

Fischer, M. M. and Nijkamp, P., 2013. *Handbook of Regional Science*. Berlin: Springer.

Fitzgerald, A. L., Dewar, R. A. and Veugelers, P. J., 2002. Diet quality and cancer incidence in Nova Scotia, Canada. *Nutrition and Cancer*, 43 (2), 127-132.

Florax, R. J. G. M., Folmer, H. and Rey, S. J., 2003. Specification searches in spatial econometrics: The relevance of Hendry's methodology. *Regional Science and Urban Economics*, 33 (5), 557-579.

Fogli-Cawley, J. J., Dwyer, J. T., Saltzman, E., McCullough, M. L., Troy, L. M. and Jacques, P. F., 2006. The 2005 dietary guidelines for Americans adherence index: development and application. *The Journal of Nutrition*, 136 (11), 2908-2915.

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K. and West, P. C., 2011. Solutions for a cultivated planet. *Nature*, 478 (7369), 337-342.

Fongar, A., Gödecke, T. and Qaim, M., 2019. Various forms of double burden of malnutrition problems exist in rural Kenya. *BMC Public Health*, 19 (1), 1543.

Font, J. C., Fabbri, D. and Gil, J., 2010. Decomposing cross-country differences in levels of obesity and overweight: does the social environment matter? *Social Science & Medicine*, 70 (8), 1185-1193.

Ford, P. B. and Dzewaltowski, D. A., 2008. Disparities in obesity prevalence due to variation in the retail food environment: Three testable hypotheses. *Nutrition Reviews*, 66 (4), 216-228.

Fouedjio, F., 2016. A hierarchical clustering method for multivariate geostatistical data. *Spatial Statistics*, 18, 333-351.

Fox, A., Feng, W. and Asal, V., 2019. What is driving global obesity trends? Globalisation or "modernisation"? *Globalisation and Health*, 15 (1), 32.

Fransen, H. P. and Ocké, M. C., 2008. Indices of diet quality. *Current Opinion in Clinical Nutrition & Metabolic Care*, 11 (5), 559-565.
Frazão, E., Meade, B. G. S. and Regmi, A., 2008. *Converging patterns in global food consumption and food delivery systems*. Economic Research Service/USDA.
Fu, T. C., 2011. A review on time series data mining. *Engineering Applications of Artificial Intelligence*, 24 (1), 164-181.
Fukase, E. and Martin, W., 2017. *Economic growth, convergence, and world food demand and supply*. The World Bank. 1813-9450.
Fung, T. T., Chiuve, S. E., McCullough, M. L., Rexrode, K. M., Logroscino, G. and Hu, F. B., 2008. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. *Archives of Internal Medicine*, 168 (7), 713-720.
Fung, T. T., Isanaka, S., Hu, F. B. and Willett, W. C., 2018. International food group–based diet quality and risk of coronary heart disease in men and women. *The American Journal of Clinical Nutrition*, 107 (1), 120-129.
Fung, T. T., McCullough, M. L., Newby, P., Manson, J. E., Meigs, J. B., Rifai, N., Willett, W. C. and Hu, F. B., 2005. Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. *The American Journal of Clinical Nutrition*, 82 (1), 163-173.
Galeano, P. and Peña, D. P., 2000. Multivariate analysis in vector time series. *Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo*, 4 (4), 383-403.
Galor, O., 1996. Convergence? Inferences from theoretical models. *The Economic Journal*, 106 (437), 1056-1069.
Gao, Y., Cheng, J., Meng, H. and Liu, Y., 2019. Measuring spatio-temporal autocorrelation in time series data of collective human mobility. *Geo-spatial Information Science*, 22 (3), 166-173.
Garcia-Escudero, L. A. and Gordaliza, A., 1999. Robustness properties of k-means and trimmed k-means. *Journal of the American Statistical Association*, 94 (447), 956-969.
García-Dorado, S. C., Cornselsen, L., Smith, R. and Walls, H., 2019. Economic globalisation, nutrition and health: A review of quantitative evidence. *Globalisation and Health*, 15 (1), 1-19.
García-Escudero, L. A., Gordaliza, A., Matrán, C. and Mayo-Iscar, A., 2010. A review of robust clustering methods. *Advances in Data Analysis and Classification*, 4 (2-3), 89-109.
Garnett, T., Appleby, M. C., Balmford, A., Bateman, I. J., Benton, T. G., Bloomer, P., Burlingame, B., Dawkins, M., Dolan, L. and Fraser, D., 2014. *What is a sustainable healthy diet? A discussion paper*. Wageningen: CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).
Garnett, T. and Wilkes, A., 2014. Appetite for change: Social, economic and environmental transformations in China's food system. *Food Climate Research Network*.
Gartner, D. R., Taber, D. R., Hirsch, J. A. and Robinson, W. R., 2016. The spatial distribution of gender differences in obesity prevalence differs from overall obesity prevalence among US adults. *Annals of Epidemiology*, 26 (4), 293-298.
Gehlhar, M. and Regmi, A., 2005. *Factors shaping global food markets*. Economic Research Service/USDA.
Gerbens-Leenes, P. W., Nonhebel, S. and Krol, M. S., 2010. Food consumption patterns and economic growth. Increasing affluence and the use of natural resources. *Appetite*, 55 (3), 597-608.
Gerber, M., 2006. Qualitative methods to evaluate Mediterranean diet in adults. *Public Health Nutrition*, 9 (1a), 147-151.
Gerber, M. J., D Scali, J., Michaud, A., D Durand, M., Astre, C. M., Dallongeville, J. and Romon, M. M., 2000. Profiles of a healthful diet and its relationship to biomarkers in a population sample from Mediterranean southern France. *Journal of the American Dietetic Association*, 100 (10), 1164-1171.
Getis, A., 2007. Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37 (4), 491-496.
Ghodsee, K., 2004. Red Nostalgia? Communism, Women's emancipation, and economic transformation in Bulgaria. *L'Homme*, 15 (1), 33-46.
Gibbons, S. and Overman, H. G., 2012. Mostly pointless spatial econometrics? *Journal of Regional Science*, 52 (2), 172-191.
Gibson, R. S., 2005. *Principles of nutritional assessment*. Oxford university press, USA.

Gil, J. M., Gracia, A. and Pérez, L. P. Y., 1995. Food consumption and economic development in the European Union. *European Review of Agricultural Economics*, 22 (3), 385-399.

Gil-Alana, L. A., Ruiz-Alba, J. L. and Ayestarán, R., 2020. UK tourism arrivals and departures: seasonality, persistence and time trends. *Applied Economics*, 52 (46), 5077-5087.

Giles, E. L. and Brennan, M., 2014. Trading between healthy food, alcohol and physical activity behaviours. *BMC Public Health*, 14 (1), 1231.

Gilson, I. and Fouad, A., 2015. *Trade policy and food security: Improving access to food in developing countries in the wake of high world prices*. Washington, DC: The World Bank.

Giner, C. and Brooks, J., 2019. *Policies for encouraging healthier food choices*. Paris.

Gini, C., 1936. *On the measure of concentration with special reference to income and statistics*.

Giorgino, T., 2009. Computing and visualising dynamic time warping alignments in R: The dtw package. *Journal of statistical Software*, 31 (7), 1-24.

Gittleman, J. L. and Kot, M., 1990. Adaptation: Statistics and a null model for estimating phylogenetic effects. *Systematic Zoology*, 39 (3), 227-241.

Gleeson, D. and Labonté , R., 2020. Commodities harmful to health. *Trade Agreements and Public Health.* Singapore: Palgrave Pivot.

Global Panel on Agriculture and Food Systems for Nutrition, 2016. *Food systems and diets: Facing the challenges of the 21st century*. London, UK.

Godenau, D., Caceres-Hernandez, J. J., Martin-Rodriguez, G. and Gonzalez-Gomez, J. I., 2020. A consumption-oriented approach to measuring regional food self-sufficiency. *Food Security: The Science, Sociology and Economics of Food Production and Access to Food*, 12, 1049-1063.

Gordon, A. D., 1996. A survey of constrained classification. *Computational Statistics & Data Analysis*, 21 (1), 17-29.

Gordon, K. D., 1987. Evolutionary perspectives on human diet. *In:* Johnson, F. E., ed. *Nutritional Anthropology*. New York: Liss, 3-41.

Gorski, M. T. and Roberto, C. A., 2015. Public health policies to encourage healthy eating habits: Recent perspectives. *Journal of Healthcare Leadership*, 7, 81.

Goryakin, Y., Lobstein, T., James, W. P. T. and Suhrcke, M., 2015. The impact of economic, political and social globalization on overweight and obesity in the 56 low and middle income countries. *Social Science & Medicine*, 133, 67-76.

Gouel, C. and Guimbard, H., 2019. Nutrition transition and the structure of global food demand. *American Journal of Agricultural Economics*, 101 (2), 383-403.

Goulet, J., Lamarche, B., Nadeau, G. and Lemieux, S., 2003. Effect of a nutritional intervention promoting the Mediterranean food pattern on plasma lipids, lipoproteins and body weight in healthy French-Canadian women. *Atherosclerosis*, 170 (1), 115-124.

Gracia, A. and Albisu, L. M., 2001. Food consumption in the European Union: Main determinants and country differences. *Agribusiness: An International Journal*, 17 (4), 469-488.

Griffith, D. A., 1988. *Advanced spatial statistics*. Dordrecht: Kluwer.

Griffith, D. A., 2005. Effective geographic sample size in the presence of spatial autocorrelation. *Annals of the Association of American Geographers*, 95 (4), 740-760.

Griffith, R., Lluberas, R. and Lührmann, M., 2016. Gluttony and sloth? Calories, labor market activity and the rise of obesity. *Journal of the European Economic Association*, 14 (6), 1253-1286.

Griffith, R. and O'Connell, M., 2009. The use of scanner data for research into nutrition. *Fiscal Studies*, 30 (3/4), 339-365.

Grigg, D., 1995. The nutritional transition in Western Europe. *Journal of Historical Geography*, 21 (3), 247.

Grigg, D., 1998. Convergence in European diets: The case of alcoholic beverages. *GeoJournal*, 44 (1), 9-18.

Grigg, D., 1999. The changing geography of world food consumption in the second half of the twentieth century. *Geographical Journal*, 1-11.

Grubesic, T. H., Wei, R. and Murray, A. T., 2014. Spatial clustering overview and comparison: Accuracy, sensitivity, and computational expense. *Annals of the Association of American Geographers*, 104 (6), 1134-1156.

Grünberger, K., 2014. *Estimating food consumption patterns by reconciling Food Balance Sheets and Household Budget Surveys*. Rome: Food and Agriculture Organisation of the United Nations.

Guenther, P. M., Casavale, K. O., Reedy, J., Kirkpatrick, S. I., Hiza, H. A. B., Kuczynski, K. J., Kahle, L. L. and Krebs-Smith, S. M., 2013. Update of the healthy eating index: HEI-2010. *Journal of the Academy of Nutrition and Dietetics*, 113 (4), 569-580.

Guenther, P. M., Reedy, J. and Krebs-Smith, S. M., 2008. Development of the Healthy Eating Index-2005. *Journal of the American Dietetic Association*, 108 (11), 1896-1901.

Guthold, R., Stevens, G. A., Riley, L. M. and Bull, F. C., 2018. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1.9 million participants. *The Lancet Global Health*, 6 (10), e1077-e1086.

Gémez, G., Fisberg, R. M., Previdelli, J., Sales, C. H., Kovalskys, I., Fisberg, M., Herrera-Cuenca, M., Sanabria, L. Y. C., García, M. C. Y., Torres, R. G. P., Rigotti, A., Guajardo, V., Zimberg, I. Z., Chinnock, A., Murillo, A. G., Brenes, J. C. and Group, E. S., 2019. Diet quality and diet diversity in eight Latin American countries: Results from the latin american study of nutrition and health (ELANS). *Nutrients*, 11 (7), 1-17.

Gödecke, T., Stein, A. J. and Qaim, M., 2018. The global burden of chronic and hidden hunger: Trends and determinants. *Global Food Security*, 17, 21-29.

Haines, P. S., Siega-Riz, A. M. and Popkin, B. M., 1999. The Diet Quality Index revised: a measurement instrument for populations. *Journal of the American Dietetic Association*, 99 (6), 697-704.

Haining, R., 1990. *Spatial data analysis in the social and environmental sciences*. Cambridge: Cambridge University Press.

Haining, R., 2003. *Spatial data analysis: Theory and practice*. Cambridge: Cambridge University Press.

Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E., 2014. *Multivariate data analysis* [Non-fiction]. 7th edition edition. Harlow: Pearson Education Limited.

Hajibaba, H., Grün, B. and Dolnicar, S., 2019. Improving the stability of market segmentation analysis. *International Journal of Contemporary Hospitality Management*, 32 (4), 1393-1411.

Hajizadeh, M., Campbell, M. K. and Sarma, S., 2016. A spatial econometric analysis of adult obesity: Evidence from Canada. *Applied Spatial Analysis and Policy*, 9 (3), 329-363.

Hall, K. D., Guo, J., Dore, M. and Chow, C. C., 2009. The progressive increase of food waste in America and its environmental impact. *PloS One*, 4 (11), e7940.

Halleck Vega, S. and Elhorst, J. P., 2014. Modelling regional labour market dynamics in space and time. *Papers in Regional Science*, 93 (4), 819-841.

Hanandita, W. and Tampubolon, G., 2015. The double burden of malnutrition in Indonesia: Social determinants and geographical variations. *SSM - Population Health*, 1, 16-25.

Hao, Y. and Wu, H., 2020. The role of Internet development on energy intensity in China: Evidence from a spatial econometric analysis. *Asian Economics Letters*, 1 (1), 17194.

Harnack, L., Nicodemus, K., Jacobs Jr, D. R. and Folsom, A. R., 2002. An evaluation of the Dietary Guidelines for Americans in relation to cancer occurrence. *The American journal of clinical nutrition*, 76 (4), 889-896.

Harris, J., Chisanga, B., Drimie, S. and Kennedy, G., 2019. Nutrition transition in Zambia: Changing food supply, food prices, household consumption, diet and nutrition outcomes. *Food Security*, 11 (2), 371-387.

Harris, J., Nguyen, P. H., Tran, L. M. and Huynh, P. N., 2020. Nutrition transition in Vietnam: Changing food supply, food prices, household expenditure, diet and nutrition outcomes. *Food Security*, 12, 1141-1155.

Hathaway, R. and Bezdek, J., 1988. Recent convergence results for the fuzzy c-means clustering algorithms. *Journal of Classification*, 5, 237-247.

Hausman, J. A., 1978. Specification tests in econometrics. *Econometrica*, 46, 1251-1271.

Haveman-Nies, A., de Groot, L. P. G. M., Burema, J., Cruz, J. A. A., Osler, M. and van Staveren, W. A., 2002. Dietary quality and lifestyle factors in relation to 10-year mortality in older Europeans: the SENECA study. *American Journal of Epidemiology*, 156 (10), 962-968.

Haveman-Nies, A., Tucker, K. L., de Groot, L., Wilson, P. W. F. and Van Staveren, W. A., 2001. Evaluation of dietary quality in relationship to nutritional and lifestyle factors in elderly people of the US Framingham Heart Study and the European SENECA study. *European Journal of Clinical Nutrition*, 55 (10), 870.

Hawkes, C., 2005. The role of foreign direct investment in the nutrition transition. *Public Health Nutrition*, 8 (4), 357-365.

Hawkes, C., 2006. Uneven dietary development: Linking the policies and processes of globalisation with the nutrition transition, obesity and diet-related chronic diseases. *Globalisation and Health*, 2 (1), 4.
Hawkes, C., 2007. *Marketing food to children: Changes in the Global Regulatory Environment 2004-2006*. Geneva, Switzerland: World Health Organisation.
Hawkes, C., 2008. Dietary implications of supermarket development: A global perspective. *Development Policy Review*, 26 (6), 657-692.
Hawkes, C., 2010. The influence of trade liberalisation and global dietary change: The case of vegetable oils, meat and highly processed foods. *In:* Hawkes, C., Blouin, C., Henson, S., Drager, N. and Dubé, L., eds. *Trade, food, diet and health: perspectives and policy options.* Chichester: Wiley-Blackwell, 35-59.
Hawkes, C., Friel, S., Lobstein, T. and Lang, T., 2012. Linking agricultural policies with obesity and noncommunicable diseases: A new perspective for a globalising world. *Food Policy*, 37 (3), 343-353.
Hawkes, C., Harris, J. and Gillespie, S., 2017. Changing diets: Urbanisation and the nutrition transition. *Global Food Policy Report.* Washington, DC: International Food Policy Research Institute (IFPRI), 34-41.
Hawkes, C., Jewell, J. and Allen, K., 2013. A food policy package for healthy diets and the prevention of obesity and diet-related non-communicable diseases: The NOURISHING framework. *Obesity Reviews*, 14, 159-168.
Hawkes, C., Ruel, M. T., Salm, L., Sinclair, B. and Branca, F., 2020. Double-duty actions: Seizing programme and policy opportunities to address malnutrition in all its forms. *The Lancet*, 395 (10218), 142-155.
Hawkes, C., Smith, T. G., Jewell, J., Wardle, J., Hammond, R. A., Friel, S., Thow, A. M. and Kain, J., 2015. Smart food policies for obesity prevention. *The Lancet*, 385 (9985), 2410-2421.
Healthy Weight Commitment Foundation, 2020. *Creating Programs to Reduce Obesity that are Effective, Sustainable and Easy to Replicate* [online]. Available from: http://www.healthyweightcommit.org/programs/ [Accessed 19th June 2020].
Heiser, W. J. and Goenen, P. J. F., 1997. Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, 62 (1), 63-83.
Heller, M. C., Walchale, A., Heard, B. R., Hoey, L., Khoury, C. K., De Haan, S., Burra, D. D., Duong, T. T., Osiemo, J., Trinh, T. H. and Jones, A. D., 2020. Environmental analyses to inform transitions to sustainable diets in developing countries: case studies for Vietnam and Kenya. *The International Journal of Life Cycle Assessment*, 25 (7), 1183-1196.
Hendry, D. F., 1995. *Dynamic econometrics*. Oxford: Oxford University Press.
Heng, Y. and House, L. A., 2018. Cluster analysis for fruit consumption patterns: an international study. *British Food Journal*, 120 (9), 1942-1952.
Herforth, A., Johns, T., Creed-Kanashiro, H. M., Jones, A. D., Khoury, C. K., Lang, T., Maundu, P., Powell, B. and Reyes-Garcia, V., 2019. *Agrobiodiversity and feeding the world: More of the same will result in more of the same*. Cambridge, MA.
Herrmann, R. and Röder, C., 1995. Does food consumption converge internationally? Measurement, empirical tests and determinants. *European Review of Agricultural Economics*, 22 (3), 400-414.
Hertel, T. W., Eales, J. S. and Preckel, P. V., 1998. Changes in the structure of global food demand. *American Journal of Agricultural Economics*, 80 (5), 1042-1050.
Hirvonen, K., Bai, Y., Headey, D. and Masters, W. A., 2020. Affordability of the EAT–Lancet reference diet: A global analysis. *The Lancet Global Health*, 8 (1), e59-e66.
Hiza, H. A. B., Koegel, K. L. and Pannucci, T. E., 2018. Diet quality: The key to healthy eating. *Journal of the Academy of Nutrition and Dietetics*, 118 (9), 1583-1585.
Hoffman, J. R. and Falvo, M. J., 2004. Protein – which is best? *Journal of Sports Science & Medicine*, 3 (3), 118.
Holdsworth, M. and Landais, E., 2019. Urban food environments in Africa: Implications for policy and research. *Proceedings of the Nutrition Society*, 78 (4), 513-525.
Honkanen, P., 2010. Food preference based segments in Russia. *Food Quality and Preference*, 21 (1), 65-74.

Hu, F. B., 2002. Dietary pattern analysis: a new direction in nutritional epidemiology. *Current Opinion in Lipidology*, 13 (1), 3-9.
Hu, T. and Sung, S. Y., 2006. A hybrid EM approach to spatial clustering. *Computational statistics & data analysis*, 50 (5), 1188-1205.
Huang, B., Wu, B. and Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *International Journal of Geographical Information Science*, 24 (3), 383-401.
Huang, J. and Bouis, H. E., 2001. Structural changes in the demand for food in Asia: empirical evidence from Taiwan. *Agricultural Economics*, 26 (1), 57-69.
Huang, J. and David, C. C., 1993. Demand for cereal grains in Asia: the effect of urbanisation. *Agricultural Economics*, 8 (2), 107-124.
Huang, S., 2004. *Global trade patterns in fruits and vegetables*. United States Department of Agriculture.
Hubert, L. and Arabie, P., 1985. Comparing partitions. *Journal of Classification*, 2 (1), 193-218.
Hughey, S. M., Kaczynski, A. T., Porter, D. E., Hibbert, J., Turner-McGrievy, G. and Liu, J., 2018. Spatial clustering patterns of child weight status in a southeastern US county. *Applied Geography*, 99, 12-21.
Huijbregts, P., Feskens, E., Räsänen, L., Fidanza, F., Nissinen, A., Menotti, A. and Kromhout, D., 1997. Dietary pattern and 20 year mortality in elderly men in Finland, Italy, and The Netherlands: longitudinal cohort study. *Bmj*, 315 (7099), 13-17.
Huneault, L., Mathieu, M. È. and Tremblay, A., 2011. Globalisation and modernisation: An obesogenic combination. *Obesity Reviews*, 12 (5), e64-e72.
Hwang, H., Desarbo, W. S. and Takane, Y., 2007. Fuzzy clusterwise generalized structured component analysis. *Psychometrika*, 72, 181.
Hyseni, L., Atkinson, M., Bromley, H., Orton, L., Lloyd-Williams, F., McGill, R. and Capewell, S., 2017. The effects of policy actions to improve population dietary patterns and prevent diet-related non-communicable diseases: Scoping review. *European Journal of Clinical Nutrition*, 71 (6), 694-711.
IFPRI, 2015. *Global Nutrition Report*. Washington, DC: International Food Policy Research Institute.
IFPRI, 2016. *Global Nutrition Report 2016: From promise to impact: Ending malnutrition*. Washington, DC: International Food Policy Research Institute (IFPRI).
Imamura, F., Micha, R., Khatibzadeh, S., Fahimi, S., Shi, P., Powles, J., Mozaffarian, D. and Global Burden of Diseases Nutrition and Chronic Diseases Expert, G., 2015. Dietary quality among men and women in 187 countries in 1990 and 2010: A systematic assessment. *The Lancet Global Health*, 3 (3), e132-e142.
Ioannidou, S., Horváth, Z. and Arcella, D., 2020. Harmonised collection of national food consumption data in Europe. *Food Policy*, 101908.
Ivanova, L., Dimitrov, P., Ovcharova, D., Dellava, J. and Hoffman, D. J., 2006. Economic transition and household food consumption: A study of Bulgaria from 1985 to 2002. *Economics & Human Biology*, 4 (3), 383-397.
Izakian, H., Pedrycz, W. and Jamal, I., 2013. Clustering spatiotemporal data: An augmented fuzzy c-means. *IEEE Trans. Fuzzy Systems*, 21 (5), 855-868.
Izakian, H., Pedrycz, W. and Jamal, I., 2015. Fuzzy clustering of time series data using dynamic time warping distance. *Engineering Applications of Artificial Intelligence*, 39, 235-244.
Jacobs , D. R. J. and Steffen, L. M., 2003. Nutrients, foods, and dietary patterns as exposures in research: a framework for food synergy. *The American Journal of Clinical Nutrition*, 78 (3), 508S-513S.
Jacobs, K. and Sumner, D. A., 2002. *The Food Balance Sheets of the Food and Agriculture Organization: a review of potential ways to broaden the appropriate uses of the data*. Davis, California.
Jain, A. K. and Farrokhnia, F., 1991. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24 (12), 1167-1186.
Jain, A. K., Murty, M. N. and Flynn, P. J., 1999. Data Clustering: A Review. *ACM Computing Surveys*, 31 (3), 264-323.

Jajuga, K. and Walesiak, M., 2000. Standardisation of data set under different measurement scales. *In:* Decker, R. and Gaul, W., eds. *Classification and Information Processing at the Turn of the Millennium.* Heidelberg: Springer-Verlag, 105-112.
James, R. D. and Campbell, H. S., 2013. The effects of space and scale on unconditional beta convergence: Test results from the United States, 1970-2004. *GeoJournal*, 78 (5), 803-815.
James, R. D. and Campbell, H. S., 2014. The impact of space and scale on conditional convergence: Test results from the United States (1970-2004). *Annals of GIS*, 20 (1), 11-21.
Jehn, M. and Brewis, A., 2009. Paradoxical malnutrition in mother–child pairs: untangling the phenomenon of over-and under-nutrition in underdeveloped economies. *Economics & Human Biology*, 7 (1), 28-35.
Jenkins, R., 2004. Globalisation, production, employment and poverty: Debates and evidence. *Journal of International Development*, 16 (1), 1-12.
Ji, Q., Bouri, E., Roubaud, D. and Shahzad, S. J. H., 2018. Risk spillover between energy and agricultural commodity markets: A dependence-switching CoVaR-copula model. *Energy Economics*, 75, 14-27.
Jin, C., Chen, R., Cheng, D., Mo, S. and Yang, K., 2020. The dependency measures of commercial bank risks: Using an optimal copula selection method based on non-parametric kernel density. *Finance Research Letters*, 101706.
Joy, E. J. M., Young, S. D., Black, C. R., Ander, E. L., Watts, M. J. and Broadley, M. R., 2013. Risk of dietary magnesium deficiency is low in most African countries based on food supply data. *Plant and Soil*, 368 (1), 129-137.
Juul, F., Lin, Y., Deierlein, A. L., Vaidean, G. and Parekh, N., 2021. Trends in food consumption by degree of processing and diet quality over 17 years: Results from the Framingham Offspring Study. *British Journal of Nutrition*, 1-11.
Kalpakis, K., Gada, D. and Puttagunta, V., 2001. Distance measures for effective clustering of ARIMA time-series, *IEEE International Conference on Data Mining* (pp. 273-280). Los Alamitos, CA, USA, USA: IEEE.
Kamdar, T. and Joshi, A., 2000. *On creating adaptive web servers using weblog mining*. Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County.
Kanerva, N., Kaartinen, N. E., Schwab, U., Lahti-Koski, M. and Männistö, S., 2014. The Baltic Sea Diet Score: a tool for assessing healthy eating in Nordic countries. *Public Health Nutrition*, 17 (8), 1697-1705.
Kant, A. K., 1996. Indexes of overall diet quality: a review. *Journal of the American Dietetic Association*, 96 (8), 785-791.
Kant, A. K., 2004. Dietary patterns and health outcomes. *Journal of the American Dietetic Association*, 104 (4), 615-635.
Kant, A. K., Leitzmann, M. F., Park, Y., Hollenbeck, A. and Schatzkin, A., 2009. Patterns of recommended dietary behaviors predict subsequent risk of mortality in a large cohort of men and women in the United States. *The Journal of nutrition*, 139 (7), 1374-1380.
Karageorgou, D., Imamura, F., Zhang, J., Shi, P., Mozaffarian, D. and Micha, R., 2018. Assessing dietary intakes from household budget surveys: A national analysis in Bangladesh. *PloS One*, 13 (8), e0202831.
Kasman, S. and Kasman, A., 2020. Convergence in obesity and overweight rates across OECD countries: evidence from the stochastic and club convergence tests. *Empirical Economics*, 1-34.
Kassambara, A., 2017. *Practical Guide To Cluster Analysis in R*. STHDA.
Kaufman, L. and Rousseeuw, P., 2005. *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.
Kaufman, L. and Rousseeuw, P. J., 1990a. *Finding Groups in Data : An Introduction to Cluster Analysis* [Non-fiction]. New York: John Wiley & Sons, Inc.
Kaufman, L. and Rousseeuw, P. J., 1990b. *Finding groups in data: An introduction to cluster analysis*. New York: John Wiley & Sons, Inc.
Kaufman, L. and Rousseeuw, P. J., 2009. *Finding groups in data: An introduction to cluster analysis*. Vol. 344. New Jersey: John Wiley & Sons.

Kearney, J., 2010. Food consumption trends and drivers. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365 (1554), 2793-2807.

Kelejian, H. H. and Prucha, I. R., 1998. A generalised spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17 (1), 99-121.

Kelejian, H. H. and Prucha, I. R., 1999. A generalised moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40 (2), 509-533.

Kelejian, H. H. and Prucha, I. R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104 (2), 219-257.

Kelejian, H. H. and Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157 (1), 53-67.

Kelejian, H. H. and Robinson, D. P., 1992. Spatial autocorrelation: A new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics*, 22 (3), 317-331.

Kelly, M., 2016. The nutrition transition in developing Asia: Dietary change, drivers and health impacts. *In:* Jackson, P., Spiess, W. and Sultana, F., eds. *Eating, Drinking: Surviving.* Springer, Cham, 83-90.

Kendall, M. G., 1955. *Rank Correlation Methods*. London: Griffin.

Kennedy, E., Ohls, J., Carlson, S. and Fleming, K., 1995. The Healthy Eating Index: design and applications. *Journal of the American Dietetic Association*, 95 (10), 1103-1108.

Kennedy, G., Ballard, T. and Dop, M., 2011. *Guidelines for measuring household and individual dietary diversity*. Rome: Nutrition and Consumer Protection Division, Food and Agriculture Organisation of the United Nations.

Kennedy, G., Nantel, G. and Shetty, P., 2004. *Globalisation of food systems in developing countries: A synthesis of country case studies*. Rome, Italy: FAO Food and Nutrition.

Keogh, E. and Kasetty, S., 2003. On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Mining and knowledge discovery*, 7 (4), 349-371.

Keogh, E., Lonardi, S., Ratanamahatana, C. A., Wei, L., Lee, S.-H. and Handley, J., 2007. Compression-based data mining of sequential data. *Data Mining and Knowledge Discovery*, 14 (1), 99-129.

Keohane, R. O. and Nye, J. S., 2000. Globalisation: What's new? What's not?(And so what?). *Foreign Policy*, 104-119.

Khoo-Lattimore, C., Prayag, G. and Disegna, M., 2019. Me, my girls, and the ideal hotel: Segmenting motivations of the girlfriend getaway market using fuzzy C-medoids for fuzzy data. *Journal of Travel Research*, 58 (5), 774-792.

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., Rieseberg, L. H. and Struik, P. C., 2014. Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences*, 111 (11), 4001.

Kim, D., Lee, C. K. and Seo, D. Y., 2016. Food deserts in Korea? A GIS analysis of food consumption patterns at sub-district level in Seoul using the KNHANES 2008-2012 data. *Nutrition Research and Practice*, 10 (5), 530-536.

Kim, S., Haines, P. S., Siega-Riz, A. M. and Popkin, B. M., 2003. The Diet Quality Index-International (DQI-I) provides an effective tool for cross-national comparison of diet quality as illustrated by China and the United States. *The Journal of nutrition*, 133 (11), 3476-3484.

Kim, S., Moon, S. and Popkin, B. M., 2000. The nutrition transition in South Korea. *The American Journal of Clinical Nutrition*, 71 (1), 44-53.

Kimani-Murage, E. W., Muthuri, S. K., Oti, S. O., Mutua, M. K., Van De Vijver, S. and Kyobutungi, C., 2015. Evidence of a double burden of malnutrition in urban poor settings in Nairobi, Kenya. *PloS one*, 10 (6), e0129943.

King, M. L., 1981. A small sample property of the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43 (2), 263-264.

Kirk, D., 1996. Demographic transition theory. *Population Studies*, 50 (3), 361-387.

Kleiman, S., Ng, S. W. and Popkin, B., 2012. Drinking to our health: Can beverage companies cut calories while maintaining profits? *Obesity Reviews*, 13 (3), 258-274.

Knoops, K. T. B., de Groot, L. C., Kromhout, D., Perrin, A. E., Moreiras-Varela, O., Menotti, A. and Van Staveren, W. A., 2004. Mediterranean diet, lifestyle factors, and 10-year mortality in elderly European men and women: the HALE project. *JAMA*, 292 (12), 1433-1439.

Knoops, K. T. B., Fidanza, F., Alberti-Fidanza, A., Kromhout, D. and Van Staveren, W. A., 2006. Comparison of three different dietary scores in relation to 10-year mortality in elderly European subjects: the HALE project. *European Journal of Clinical Nutrition*, 60 (6), 746.

Knudsen, V., Fagt, S., Trolle, E., Matthiessen, J., Groth, M., Biltoft-Jensen, A., Sørensen, M. and Pedersen, A., 2012. Evaluation of dietary intake in Danish adults by means of an index based on food-based dietary guidelines. *Food & nutrition research*, 56 (1), 17129.

Knüppel, S., Norman, K. and Boeing, H., 2019. Is a single 24-hour dietary recall per person sufficient to estimate the population distribution of usual dietary intake? *The Journal of Nutrition*, 149 (9), 1491-1492.

Kontsevaya, A. V., Imaeva, A. E., Balanova, Y. A., Kapustina, A. V., Breda, J., Jewell, J. M., Salakhov, E. R., Drapkina, O. M. and Boyland, E., 2020. The extent and nature of television food advertising to children and adolescents in the Russian Federation. *Public Health Nutrition*, 23 (11), 1868-1876.

Kourlaba, G. and Panagiotakos, D. B., 2009. Dietary quality indices and human health: a review. *Maturitas*, 62 (1), 1-8.

Kourlaba, G., Polychronopoulos, E., Zampelas, A., Lionis, C. and Panagiotakos, D. B., 2009. Development of a diet index for older adults and its relation to cardiovascular disease risk factors: the Elderly Dietary Index. *Journal of the American Dietetic Association*, 109 (6), 1022-1030.

Krishnapuram, R. and Freg, C. P., 1992. Fitting an unknown number of lines and planes to image data through compatible cluster merging. *Pattern Recognition*, 25 (4), 385-400.

Krishnapuram, R., Joshi, A., Nasraoui, O. and Yi, L., 2001. Low-complexity fuzzy relational clustering algorithms for web mining. *IEEE Transactions on Fuzzy Systems*, 9 (4), 595-607.

Krishnapuram, R., Joshi, A. and Yi, L., 1999. A fuzzy relative of the k-medoids algorithm with application to web document and snippet clustering, *1999 IEEE International Fuzzy Systems* (Vol. 3, pp. 1281-1286).

Kromhout, D., Menotti, A., Alberti-Fidanza, A., Puddu, P. E., Hollman, P., Kafatos, A., Tolonen, H., Adachi, H. and Jacobs, D. R., Jr., 2018. Comparative ecologic relationships of saturated fat, sucrose, food groups, and a Mediterranean food pattern score to 50-year coronary heart disease mortality rates among 16 cohorts of the Seven Countries Study. *European Journal of Clinical Nutrition*, 72 (8), 1103-1110.

Kuczmarski, M. F., Shupe, E. S., Pohlig, R. T., Rawal, R., Zonderman, A. B. and Evans, M. K., 2019. A longitudinal assessment of diet quality and risks associated with malnutrition in socioeconomic and racially diverse adults. *Nutrients*, 11 (9), 1-17.

Kónya, I. and Ohashi, H., 2007. International consumption patterns among high-income countries: Evidence from the OECD data. *Review of International Economics*, 15 (4), 744-757.

La Vecchia, C. and Majem, L. S., 2015. Evaluating trends in global dietary patterns. *The Lancet Global Health*, 3 (3), e114-e115.

Lafuente-Rego, B. and Vilar, J. A., 2016. Clustering of time series using quantile autocovariances. *Advances in Data Analysis and Classification*, 10 (3), 391-415.

Lakdawalla, D. and Philipson, T., 2009. The growth of obesity and technological change. *Economics & Human Biology*, 7 (3), 283-293.

Landes, D., 1998. *The wealth and poverty of the nations. Why some are so rich and some so poor*. New York: W. W. Norton & Company.

Lang, T., 2009. Reshaping the food system for ecological public health. *Journal of Hunger & Environmental Nutrition*, 4 (3-4), 315-335.

Lang, T. and Heasman, M., 2015. *Food wars: the global battle for mouths, minds and markets*. 2nd edition. New York: Routledge.

Lang, T. and Rayner, G., 2007. Overcoming policy cacophony on obesity: An ecological public health framework for policymakers. *Obesity Reviews*, 8, 165-181.

Lanham-New, S. A., Hill, T. R., Gallagher, A. M. and Vorster, H. H., 2019. *Introduction to human nutrition*. 3rd edition.: John Wiley & Sons.

Lau, J. D., Elbaar, L., Chao, E., Zhong, O., Yu, C. R., Tse, R. and Au, L., 2020. Measuring overweight and obesity in Chinese American children using US, international and ethnic-specific growth charts. *Public Health Nutrition*, 23 (15), 2663-2670.

Law, C., Fraser, I. and Piracha, M., 2018. Nutrition transition and changing food preferences in India. *Journal of Agricultural Economics*.

Law, C., Green, R., Suneetha Kadiyala, B. S., Knai, C., Brown, K. A., Dangour, A. D. and Cornelsen, L., 2019. Purchase trends of processed foods and beverages in urban India. *Global Food Security*, 23, 191-204.

Le Gallo, J. and Sandy, D. E., 2006. Evaluating the temporal and the spatial heterogeneity of the European convergence process, 1980-1999. *Journal of Regional Science*, 46 (2), 269-288.

Le, T. H., Disegna, M. and Lloyd, T., 2020. National Food Consumption Patterns: Converging Trends and the Implications for Health. *EuroChoices*.

Leclercq, C., Allemand, P., Balcerzak, A., Branca, F., Sousa, R. F., Lartey, A., Lipp, M., Quadros, V. P. and Verger, P., 2019. FAO/WHO GIFT (Global Individual Food consumption data Tool): A global repository for harmonised individual quantitative food consumption studies. *Proceedings of the Nutrition Society*, 78 (4), 484-495.

Lee, J. and Li, S., 2017. Extending Moran's index for measuring spatiotemporal clustering of geographic events. *Geographical Analysis*, 49 (1), 36-57.

Lee, J., Yoo, S., Kim, H. and Chung, Y., 2018. The spatial and temporal variation in passenger service rate and its impact on train dwell time: A time-series clustering approach using dynamic time warping. *International Journal of Sustainable Transportation*, 12 (10), 725-736.

Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72 (6), 1899-1925.

Lee, L. F. and Yu, J., 2012. QML estimation of spatial dynamic panel data models with time varying spatial weights matrices. *Spatial Economic Analysis*, 7 (1), 31-74.

Lee, M. J., Popkin, B. M. and Kim, S., 2002. The unique aspects of the nutrition transition in South Korea: The retention of healthful elements in their traditional diet. *Public Health Nutrition*, 5 (1a), 197-203.

Lee, M. S., Lai, C. J., Yang, F. Y., Su, H. H., Yu, H. L. and Wahlqvist, M. L., 2008. A global overall dietary index: ODI-R revised to emphasize quality over quantity. *Asia Pacific Journal of Clinical Nutrition*, 17 (S1), 82-86.

Leenders, R. T. A. J., 2002. Modelling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks*, 24 (1), 21-47.

Legendre, P. and Legendre, L., 1998. *Numerical ecology*. 2nd edition. Amsterdam: Elsevier Science.

Legoherel, P. and Wong, K. K. L., 2012. Market segmentation in the tourism industry and consumer spending. *Journal of Travel & Tourism Marketing*, 20 (2), 15-30.

Leisch, F., 2006. A toolbox for K-centroids cluster analysis. *Computational Statistics and Data Analysis*, 51 (2), 526-544.

Leonard, T., Hughes, A. E., Donegan, C., Santillan, A. and Pruitt, S. L., 2018. Overlapping geographic clusters of food security and health: Where do social determinants and health outcomes converge in the U.S? *SSM - Population Health*, 5, 160-170.

LeSage, J. and Pace, R. K., 2014. Interpreting spatial econometric models. *In:* Fischer, M. M. and Nijkamp, P., eds. *Handbook of regional science.* Berlin: Springer.

LeSage, J. P., 1998. *Spatial econometrics*. Regional Research Institute, West Virginia University Morgantown, WV.

LeSage, J. P. and Pace, R. K., 2009. *Introduction to spatial econometrics*. Boca Raton: CRC Press Taylor & Francis Group.

Levin, D., Noriega, D., Dicken, C., Okrent, A. M., Harding, M. and Lovenheim, M., 2018. *Examining food store scanner data: A comparison of the IRI InfoScan data with other data sets, 2008-2012*. US Department of Agriculture, Economic Research Service.

Li, H., Chen, J. L., Li, G. and Goh, C., 2016. Tourism and regional income inequality: Evidence from China. *Annals of Tourism Research*, 58, 81-99.

Li, X. and Wang, R., 2016. Are US obesity rates converging? *Applied Economics Letters*, 23 (8), 539-543.

Liao, T. W., 2005. Clustering of time series data - A survey. *Pattern Recognition*, 38 (11), 1857-1874.

Liu, Y., Chen, J., Wu, S., Liu, Z. and Chao, H., 2018. Incremental fuzzy C medoids clustering of time series data using dynamic time warping distance. *PloS One*, 13 (5).

Lolayekar, A. P. and Mukhopadhyay, P., 2019. Spatial dependence and regional income convergence in India (1981–2010). *GeoJournal*, 84 (4), 851-864.

Lopez Barrera, E. and Hertel, T., 2021. Global food waste across the income spectrum: Implications for food prices, production and resource use. *Food Policy*, 98, 101874.

Lopez, R., 2004. Urban sprawl and risk for being overweight or obese. *American Journal of Public Health*, 94 (9), 1574-1579.

Lopez-Arana, S., Avendano, M., van Lenthe, F. J. and Burdorf, A., 2014. Trends in overweight among women differ by occupational class: Results from 33 low-and middle-income countries in the period 1992–2009. *International Journal of Obesity*, 38 (1), 97-105.

Lowe, C., Kelly, M., Sarma, H., Richardson, A., Kurscheid, J. M., Laksono, B., Amaral, S., Stewart, D. and Gray, D. J., 2021. The double burden of malnutrition and dietary patterns in rural Central Java, Indonesia. *The Lancet Regional Health – Western Pacific*, 14, 100205.

Lusk, J. L. and Brooks, K., 2011. Who participates in household scanning panels? *American Journal of Agricultural Economics*, 93 (1), 226-240.

Löwik, M. R. H., Hulshof, K. and Brussaard, J. H., 1999. Food-based dietary guidelines: some assumptions tested for The Netherlands. *British Journal of Nutrition*, 81 (S1), S143-S149.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, *The Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281-297). Berkeley, California: University of California Press.

Magkos, F., Tetens, I., Bügel, S. G., Felby, C., Schacht, S. R., Hill, J. O., Ravussin, E. and Astrup, A., 2020. The environmental foodprint of obesity. *Obesity*, 28 (1), 73-79.

Maharaj, E. A., 1996. A significance test for classifying arma models. *Journal of Statistical Computation & Simulation*, 54 (4), 305-331.

Maharaj, E. A. and D'Urso, P., 2010. A coherence-based approach for the pattern recognition of time series. *Physica A*, 389 (17), 3516-3537.

Maharaj, E. A. and D'Urso, P., 2011. Fuzzy clustering of time series in the frequency domain. *Information Sciences*, 181 (7), 1187-1211.

Maharaj, E. A., D'Urso, P. and Caiado, J., 2019. *Time series clustering and classification*. CRC Press.

Mangyo, E., 2008. Who benefits more from higher household consumption? The intra-household allocation of nutrients in China. *Journal of Development Economics*, 86 (2), 296-312.

Mankiw, N. G., Romer, D. and Weil, D. N., 1992. A contribution to the empirics of economic growth. *The Quarterly Journal of Economics*, 107 (2), 407-437.

Manski, C. F., 2000. Economic analysis of social interactions. *Journal of Economic Perspectives*, 14 (3), 115-136.

Manton, K. G. and Soldo, B. J., 1985. Dynamics of health changes in the oldest old: New perspectives and evidence. *The Milbank Memorial Fund Quarterly. Health and Society*, 206-285.

Marrón-Ponce, J. A., Sánchez-Pimienta, T. G., da Costa Louzada, M. L. and Batis, C., 2018. Energy contribution of NOVA food groups and sociodemographic determinants of ultra-processed food consumption in the Mexican population. *Public Health Nutrition*, 21 (1), 87-93.

Martin, W., 2018. *Food trade policy and the dietary transition*. OCP Plicy Center.

Martin, W., 2019. Economic growth, convergence, and agricultural economics. *Agricultural Economics*, 50, 7-27.

Mason, P. and Lang, T., 2017. *Sustainable diets: How ecological nutrition can transform consumption and the food system*. New York: Routledge.

Masters, W. A., Hall, A., Martinez, E. M., Shi, P., Singh, G., Webb, P. and Mozaffarian, D., 2016. The nutrition transition and agricultural transformation: A Preston curve approach. *Agricultural Economics*, 47 (S1), 97-114.

Mastorakou, D., Rabaeus, M., Salen, P., Pounis, G. and de Lorgeril, M., 2019. Chapter 9 - Mediterranean Diet: A Health-Protective Dietary Pattern for Modern Times. *In:* Pounis, G., ed. *Analysis in Nutrition Research.* Academic Press, 233-258.

Mathieu-Bolh, N. and Wendner, R., 2020. We are what we eat: Obesity, income, and social comparisons. *European Economic Review*, 128, 103495.

Maynard, M., Ness, A. R., Abraham, L., Blane, D., Bates, C. and Gunnell, D. J., 2005. Selecting a healthy diet score: lessons from a study of diet and health in early old age (the Boyd Orr cohort). *Public Health Nutrition*, 8 (3), 321-326.

Mazzocchi, M., Brasili, C. and Sandri, E., 2008. Trends in dietary patterns and compliance with World Health Organization recommendations: a cross-country analysis. *Public Health Nutrition*, 11 (5), 535-540.

McBratney, A. B. and Moore, A. W., 1985. Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, 35 (1-4), 165-185.

McChesney, R. W. and Schiller, D., 2003. *The political economy of international communications: Foundations for the emerging global debate about media ownership and regulation*. Geneva: United Nations Research Institute for Social Development.

McCracken, K. and Phillips, D. R., 2017. Demographic and epidemiological transition. *International Encyclopedia of Geography*, 1-8.

McCullough, M. L., Feskanich, D., Rimm, E. B., Giovannucci, E. L., Ascherio, A., Variyam, J. N., Spiegelman, D., Stampfer, M. J. and Willett, W. C., 2000a. Adherence to the Dietary Guidelines for Americans and risk of major chronic disease in men. *The American Journal of Clinical Nutrition*, 72 (5), 1223-1231.

McCullough, M. L., Feskanich, D., Stampfer, M. J., Giovannucci, E. L., Rimm, E. B., Hu, F. B., Spiegelman, D., Hunter, D. J., Colditz, G. A. and Willett, W. C., 2002. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. *The American Journal of Clinical Nutrition*, 76 (6), 1261-1271.

McCullough, M. L., Feskanich, D., Stampfer, M. J., Rosner, B. A., Hu, F. B., Hunter, D. J., Variyam, J. N., Colditz, G. A. and Willett, W. C., 2000b. Adherence to the Dietary Guidelines for Americans and risk of major chronic disease in women. *The American Journal of Clinical Nutrition*, 72 (5), 1214-1222.

Menard, M. and Eboueya, M., 2002. Extreme physical information and objective function in fuzzy clustering. *Fuzzy Sets and Systems*, 128, 285-303.

Menotti, A., Kromhout, D., Puddu, P. E., Alberti-Fidanza, A., Hollman, P., Kafatos, A., Tolonen, H., Adachi, H. and Jacobs Jr, D. R., 2017. Baseline fatty acids, food groups, a diet score and 50-year all-cause mortality rates. An ecological analysis of the Seven Countries Study. *Annals of Medicine*, 49 (8), 718-727.

Menyhert, B., 2020. *Peas in a pod: Households' food consumption patterns across the EU*. Joint Research Centre, European Commission.

Mertz, W., 1984. Foods and nutrients. *Journal of the American Dietetic Association (USA)*.

Metro, D., Tardugno, R., Papa, M., Bisignano, C., Manasseri, L., Calabrese, G., Gervasi, T., Dugo, G. and Cicero, N., 2018. Adherence to the Mediterranean diet in a Sicilian student population. *Natural Product Research*, 32 (15), 1775-1781.

Micha, R., Coates, J., Leclercq, C., Charrondiere, U. R. and Mozaffarian, D., 2018. Global Dietary Surveillance: Data Gaps and Challenges. *Food and nutrition bulletin*, 39 (2), 175-205.

Michail, N. A., 2020. Convergence of consumption patterns in the European Union. *Empirical Economics*, 58 (3), 979-994.

Mihalache-O'Keef, A. and Li, Q., 2011. Modernisation vs. dependency revisited: Effects of foreign direct investment on food security in less developed countries. *International Studies Quarterly*, 55 (1), 71-93.

Mikalsen, K. O., Bianchi, F. M., Soguero-Ruiz, C. and Jenssen, R., 2018. Time series cluster kernel for learning similarities between multivariate time series with missing data. *Pattern Recognition*, 76, 569-581.

Milford, A. B., Le Mouël, C., Bodirsky, B. L. and Rolinski, S., 2019. Drivers of meat consumption. *Appetite*, 141, 104313.

Miljkovic, D., Shaik, S., Miranda, S., Barabanov, N. and Liogier, A., 2015. Globalisation and obesity. *The World Economy*, 38 (8), 1278-1294.

Milligan, G. W. and Cooper, M. C., 1988. A study of standardisation of variables in cluster analysis. *Journal of classification*, 5, 181-204.

Mohammadi, H. and Ram, R., 2012. Cross-country convergence in energy and electricity consumption, 1971–2007. *Energy Economics*, 34 (6), 1882-1887.

Mohammadi, H. and Ram, R., 2017. Convergence in energy consumption per capita across the US states, 1970–2013: An exploration through selected parametric and non-parametric methods. *Energy Economics*, 62, 404-410.
Monfort, P., 2008. *Convergence of EU regions: Measures and evolution*. European Union Regional Policy.
Monteiro, C. A., Cannon, G., Lawrence, M., Costa Louzada, M. L. and Pereira Machado, P., 2019a. *Ultra-processed foods, diet quality, and health using the NOVA classification system*. Rome: Food and Agriculture Organisation of the United Nations (FAO).
Monteiro, C. A., Cannon, G., Levy, R. B., Moubarac, J.-C., Louzada, M. L. C., Rauber, F., Khandpur, N., Cediel, G., Neri, D. and Martinez-Steele, E., 2019b. Ultra-processed foods: What they are and how to identify them. *Public Health Nutrition*, 22 (5), 936-941.
Montero, M. D. P., Mora-Urda, A. I., Anzid, K., Cherkaoui, M. and Marrodan, M. D., 2017. Diet quality of Moroccan adolescents living in Morocco and in Spain. *Journal of Biosocial Science*, 49 (2), 173-186.
Montero, P. and Vilar, J. A., 2014. TSclust: An R Package for Time Series Clustering. *Journal of Statistical Software*, 62 (1), 1-42.
Moodie, R., Stuckler, D., Monteiro, C., Sheron, N., Neal, B., Thamarangsi, T., Lincoln, P. and Casswell, S., 2013. Profits and pandemics: prevention of harmful effects of tobacco, alcohol, and ultra-processed food and drink industries. *The Lancet*, 381 (9867), 670-679.
Moraeus, L., Lindroos, A. K., Warensjö Lemming, E. and Mattisson, I., 2020. Diet diversity score and healthy eating index in relation to diet quality and socio-demographic factors: Results from a cross-sectional national dietary survey of Swedish adolescents. *Public Health Nutrition*, 23 (10), 1754-1765.
Moran, P. A. P., 1950a. A test for the serial independence of residuals. *Biometrika*, 37 (1/2), 178-181.
Moran, P. A. P., 1950b. Notes on continuous stochastic phenomena. *Biometrika*, 37 (1/2), 17-23.
Moser, C., Barrett, C. and Minten, B., 2009. Spatial integration at multiple scales: Rice markets in Madagascar. *Agricultural Economics*, 40 (3), 281-294.
Moss, M., 2013. *Salt, sugar, fat: How the food giants hooked us*. New York: Penguin Random House.
Mozaffarian, D., Angell, S. Y., Lang, T. and Rivera, J. A., 2018a. Role of government policy in nutrition - barriers to and opportunities for healthier eating. *BMJ*, 361, k2426.
Mozaffarian, D., Angell, S. Y., Lang, T. and Rivera, J. A., 2018b. Role of government policy in nutrition—barriers to and opportunities for healthier eating. *BMJ*, 361, k2426.
Muhammad, A., D'Sourza, A., Meade, B., Renata, M. and Mozaffarian, D., 2017. *The influence of income and prices on global dietary patterns by country, age, and gender*. U.S. Department of Agriculture, Economic Research Service.
Mummert, A., Esche, E., Robinson, J. and Armelagos, G. J., 2011. Stature and robusticity during the agricultural transition: Evidence from the bioarchaeological record. *Economics & Human Biology*, 9 (3), 284-301.
Murakami, K., Livingstone, M. B. E., Fujiwara, A. and Sasaki, S., 2020. Application of the Healthy Eating Index-2015 and the Nutrient-Rich Food Index 9.3 for assessing overall diet quality in the Japanese context: Different nutritional concerns from the US. *PloS One*, 15 (1), e0228318.
Murray, S., Brock, S. and Seto, K. C., 2015. Urbanisation, food consumption and the environment. *In:* Seto, K. C., Solecki, W. D. and Griffith, C. A., eds. *The Routledge Handbook of Urbanisation and Global Environmental Change.* New York: Routledge.
Murtagh, F., 1985. A survey of algorithms for contiguity-constrained clustering and related problems. *The Computer Journal*, 28 (1), 82-88.
Muth, M. K., Okrent, A. M., Zhen, C. and Karns, S. A., 2019. *Using scanner data for food policy research*. London: Academic Press.
Muth, M. K., Sweitzer, M., Brown, D., Capogrossi, K., Karns, S., Levin, D., Okrent, A., Siegel, P. and Zhen, C., 2016. *Understanding IRI household-based and store-based scanner data*. US Department of Agriculture, Economic Research Service.
Mutlu, S. and Gracia, A., 2006. Spanish food expenditure away from home (FAFH): By type of meal. *Applied Economics*, 38 (9), 1037-1047.
Nandi, A., Sweet, E., Kawachi, I., Heymann, J. and Galea, S., 2014. Associations between macrolevel economic factors and weight distributions in low- and middle-income countries: A multilevel

analysis of 200 000 adults in 40 countries. *American Journal of Public Health*, 104 (2), e162-e171.

Nardocci, M., Leclerc, B. S., Louzada, M. L., Monteiro, C. A., Batal, M. and Moubarac, J. C., 2019. Consumption of ultra-processed foods and obesity in Canada. *Canadian Journal of Public Health*, 110 (1), 4-14.

National Food Strategy, 2020. *National Food Strategy: Part One* [online]. Available from: https://www.nationalfoodstrategy.org/partone/ [Accessed 8th August 2020].

National Research Council, 2003. *Cities transformed: Demographic change and its implications in the developing world*. Washington, DC: The National Academies Press.

National Research Council, 2005. *Improving data to analyse food and nutrition policies*. Washinton, DC: The National Academies Press.

Neelakantan, N., Koh, W. P., Yuan, J. M. and van Dam, R. M., 2018. Diet quality indexes are associated with a lower risk of cardiovascular, respiratory, and all-cause mortality among Chinese adults. *The Journal of Nutrition*, 148 (8), 1323-1332.

Neuman, M., Kawachi, I., Gortmaker, S. and Subramanian, S. V., 2014. National economic development and disparities in body mass index: A cross-sectional study of data from 38 countries. *PloS One*, 9 (6).

Neumayer, E. and Plümper, T., 2016. Spatial spill-overs from terrorism on tourism: Western victims in Islamic destination countries. *Public Choice*, 169 (3-4), 195-206.

Newby, P. K., Hu, F. B., Rimm, E. B., Smith-Warner, S. A., Feskanich, D., Sampson, L. and Willett, W. C., 2003. Reproducibility and validity of the Diet Quality Index Revised as assessed by use of a food-frequency questionnaire. *The American Journal of Clinical Nutrition*, 78 (5), 941-949.

Newby, P. K. and Tucker, K. L., 2004. Empirically derived eating patterns using factor or cluster analysis: a review. *Nutrition Reviews*, 62 (5), 177-203.

Newman, L., Kates, R. W., Matthews, R. and Millman, S., 1990. *Hunger in History*. Cambridge, MA: Basil Blackwell Ltd.

Ng, S. W. and Dunford, E., 2013. Complexities and opportunities in monitoring and evaluating US and global changes by the food industry. *Obesity Reviews*, 14, 29-41.

Ng, S. W., Slining, M. M. and Popkin, B. M., 2014. The Healthy Weight Commitment Foundation pledge: calories sold from US consumer packaged goods, 2007–2012. *American Journal of Preventive Medicine*, 47 (4), 508-519.

Nguyen, P. H., Scott, S., Headey, D., Singh, N., Tran, L. M., Menon, P. and Ruel, M. T., 2021. The double burden of malnutrition in India: Trends and inequalities (2006–2016). *Plos one*, 16 (2), e0247856.

Nicklas, T. A., O'Neil, C. E. and Fulgoni, V. L., III, 2012. Diet quality is inversely related to cardiovascular risk factors in adults. *The Journal of Nutrition*, 142 (12), 2112-2118.

Nie, P. and Sousa-Poza, A., 2014. Maternal employment and childhood obesity in China: Evidence from the China Health and Nutrition Survey. *Applied Economics*, 46 (20), 2418-2428.

Nowak, J. and Kochkova, O., 2011. Income, culture, and household consumption expenditure patterns in the European Union: Convergence or divergence? *Journal of International Consumer Marketing*, 23 (3-4), 260-275.

Nugent, R., Levin, C., Hale, J. and Hutchinson, B., 2020. Economic effects of the double burden of malnutrition. *The Lancet*, 395 (10218), 156-164.

Nuttall, F. Q., 2015. Body mass index: obesity, BMI, and health: a critical review. *Nutrition Today*, 50 (3), 117.

Oberlander, L., Disdier, A. C. and Etilé, F., 2017. Globalisation and national trends in nutrition and health: A grouped fixed-effects approach to intercountry heterogeneity. *Health Economics*, 26 (9), 1146-1161.

Oddo, V. M., Bleich, S. N., Pollack, K. M., Surkan, P. J., Mueller, N. T. and Jones-Smith, J. C., 2017. The weight of work: The association between maternal employment and overweight in low- and middle-income countries. *International Journal of Behavioral Nutrition and Physical Activity*, 14 (1), 66.

Odunitan-Wayas, F. A., Okop, K. J., Dover, R. V. H., Alaba, O. A., Micklesfield, L. K., Puoane, T., Levitt, N. S., Battersby, J., Meltzer, S. T. and Lambert, E. V., 2020. Food purchasing behaviour

of shoppers from different South African socio-economic communities: Results from grocery receipts, intercept surveys and in-supermarkets audits. *Public Health Nutrition*, 1-12.
OECD, 2019a. *Agricultural Policy Monitoring and Evaluation*. Paris.
OECD, 2019b. *The Heavy Burden of Obesity: The Economics of Prevention* [online]. Paris: OECD Publishing.
OECD, 2020. *Better agro-food policies are crucial to improve global food security* [online]. Paris: OECD. Available from: http://www.oecd.org/agriculture/topics/food-security/ [Accessed 18th December 2020].
OECD, 2021. *Making better policies for food systems*. Paris.
OECD iLibrary, 2020. *Foreign direct investment (FDI)* [online]. OECD. Available from: https://www.oecd-ilibrary.org/finance-and-investment/foreign-direct-investment-fdi/indicator-group/english_9a523b18-en [Accessed 5th March 2020].
OECD/FAO, 2020. *OECD-FAO Agricultural Outlook 2020-2029*. Rome/OECD Publishing, Paris: FAO.
Ogundari, K. and Abdulai, A., 2013. Examining the heterogeneity in calorie-income elasticities: A meta-analysis. *Food Policy*, 40, 119-128.
Ogundari, K. and Ito, S., 2015. Convergence and determinants of change in nutrient supply: Evidence from sub-Saharan African countries. *British Food Journal*, 117 (12), 2880-2898.
Oliver, M. A. and Webster, R., 1989. A geostatistical basis for spatial weighting in multivariate classification. *Mathematical Geology*, 21 (1), 15-35.
Olivier, J., Thoenig, M. and Verdier, T., 2008. Globalisation and the dynamics of cultural identity. *Journal of International Economics*, 76 (2), 356-370.
Omran, A. R., 1971. The epidemiologic transition: A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly*, 49 (4), 509-538.
Omran, A. R., 2005. The epidemiologic transition: a theory of the epidemiology of population change. *The Milbank Quarterly*, 83 (4), 731-757.
Ord, K., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70 (349), 120-126.
Ortiz-Bobea, A., Ault, T. R., Carrillo, C. M., Chambers, R. G. and Lobell, D. B., 2021. Anthropogenic climate change has slowed global agricultural productivity growth. *Nature Climate Change*, 11 (4), 306-312.
Osborne, T., 2005. Imperfect competition in agricultural markets: Evidence from Ethiopia. *Journal of Development Economics*, 76 (2), 405-428.
Osler, M., Heitmann, B. L., Gerdes, L. U., Jørgensen, L. M. and Schroll, M., 2001. Dietary patterns and mortality in Danish men and women: a prospective observational study. *British Journal of Nutrition*, 85 (2), 219-225.
Otoiu, A. and Titan, E., 2015. Socio-economic convergence in the EU at national and regional level. *Procedia Economics and Finance*, 23, 1090-1095.
Pace, R. K., Barry, R., Clapp, J. M. and Rodriquez, M., 1998. Spatiotemporal autoregressive models of neighborhood effects. *The Journal of Real Estate Finance and Economics*, 17 (1), 15-33.
Pace, R. K., Barry, R., Gilley, O. W. and Sirmans, C. F., 2000. A method for spatial–temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16 (2), 229-246.
Paci, L. and Finazzi, F., 2018. Dynamic model-based clustering for spatio-temporal data. *Statistics and Computing*, 28 (2), 359-374.
Paddock, J. R., 2017. Changing consumption, changing tastes? Exploring consumer narratives for food secure, sustainable and healthy diets. *Journal of Rural Studies*, 53, 102-110.
Panagiotakos, D. B., Pitsavos, C., Arvaniti, F. and Stefanadis, C., 2007. Adherence to the Mediterranean food pattern predicts the prevalence of hypertension, hypercholesterolemia, diabetes and obesity, among healthy adults; the accuracy of the MedDietScore. *Preventive Medicine*, 44 (4), 335-340.
Pandey, B., Reba, M., Joshi, P. K. and Seto, K. C., 2020. Urbanisation and food consumption in India. *Scientific Reports*, 10 (1), 17241.
Panico, S., Mattiello, A., Panico, C. and Chiodini, P., 2014. Mediterranean dietary pattern and chronic diseases. *Advances in nutrition and cancer*. Springer, 69-81.

Patterson, R. E., Haines, P. S. and Popkin, B. M., 1994. Diet quality index: capturing a multidimensional behavior. *Journal of the American Dietetic Association*, 94 (1), 57-64.
Pawitan, Y. and Huang, J., 2003. Constrained clustering of irregularly sampled spatial data. *Journal of Statistical Computation and Simulation*, 73 (12), 853-865.
Pawlak, K., 2016. Food security situation of selected highly developed countries against developing countries. *Journal of Agribusiness and Rural Development*, 2 (40).
Pereira, J. L., Mattei, J., Isasi, C. R., Van Horn, L., Carnethon, M. R., Daviglus, M. L., Perera, M. J., Sotres-Alvarez, D. and Fisberg, R. M., 2020. Diet quality, excess body weight and cardiometabolic risk factors in adolescents living in São Paulo, Brazil and in the USA: differences and similarities. *Public Health Nutrition*, 1-11.
Pertega, D. S. and Vilar, J. A., 2010. Comparing several parametric and nonparametric approaches to time series clustering: A simulation study. *Journal of Classification*, 27 (3), 333-362.
Petrova, S., Dimitrov, P., Willett, W. C. and Campos, H., 2011. The global availability of n-3 fatty acids. *Public Health Nutrition*, 14 (7), 1157-1164.
Petrovici, D. A., Ritson, C. and Ness, M., 2005. Exploring disparities and similarities in European food consumption patterns. *Cahiers d'Economie et de Sociologie Rurales*, 75, 23-49.
Pham, D. L., 2001. Spatial models for fuzzy clustering. *Computer Vision and Image Understanding*, 84 (2), 285-297.
Philipson, T. J. and Posner, R. A., 2008. Is the obesity epidemic a public health problem? A review of Zoltan J. Acs and Alan Lyles's obesity, business and public policy. *Journal of Economic Literature*, 46 (4), 974-982.
Phillips, P. C. B. and Sul, D., 2007. Transition modeling and econometric convergence tests. *Econometrica*, 75 (6), 1771-1855.
Phillips, P. C. B. and Sul, D., 2009. Economic transition and growth. *Journal of Applied Econometrics*, 24 (7), 1153-1185.
Piccolo, D., 1990. A DISTANCE MEASURE FOR CLASSIFYING ARIMA MODELS. *Journal of Time Series Analysis*, 11 (2), 153-164.
Pickett, K. E., Kelly, S., Brunner, E., Lobstein, T. and Wilkinson, R. G., 2005. Wider income gaps, wider waistbands? An ecological study of obesity and income inequality. *Journal of Epidemiology & Community Health*, 59 (8), 670-674.
Pietrzykowski, M., 2019. Convergence in GDP per capita across the EU regions - spatial effects. *Economics and Business Review*, 5 (2), 64-85.
Pingali, P., 2007. Westernisation of Asian diets and the transformation of food systems: Implications for research and policy. *Food Policy*, 32 (3), 281-298.
Placzek, O., 2021. *Socio-economic and demographic aspects of food security and nutrition*. Paris.
Popkin, B. and Ng, S. W., 2007. The nutrition transition in high- and low-income countries: What are the policy lessons? *Agricultural Economics*, 37, 199-211.
Popkin, B. M., 1993. Nutritional patterns and transitions. *Population and Development Review*, 138-157.
Popkin, B. M., 1999. Urbanisation, lifestyle changes and the nutrition transition. *World Development*, 27 (11), 1905-1916.
Popkin, B. M., 2001a. The nutrition transition and its relationship to demographic change. *Nutrition and health in developing countries.* Springer, 427-445.
Popkin, B. M., 2001b. The nutrition transition and obesity in the developing world. *The Journal of Nutrition*, 131 (3), 871S-873S.
Popkin, B. M., 2002a. An overview of the nutrition transition and its health implications: the Bellagio meeting. *Public Health Nutrition*, 5, 93-103.
Popkin, B. M., 2002b. The shift in stages of the nutrition transition in the developing world differs from past experiences! *Public Health Nutrition*, 5 (1A), 205-214.
Popkin, B. M., 2006a. Global nutrition dynamics: The world is shifting rapidly toward a diet linked with noncommunicable diseases. *The American Journal of Clinical Nutrition*, 84 (2), 289-298.
Popkin, B. M., 2006b. Technology, transport, globalisation and the nutrition transition food policy. *Food Policy*, 31 (6), 554-569.
Popkin, B. M., 2008. *The world is fat: the fads, trends, policies, and products that are fattening the human race*. New York: Avery-Penguin Group.

Popkin, B. M., 2015. Nutrition transition and the global diabetes epidemic. *Current Diabetes Reports*, 15 (9), 64.
Popkin, B. M., 2017. Relationship between shifts in food system dynamics and acceleration of the global nutrition transition. *Nutrition Reviews*, 75 (2), 73-82.
Popkin, B. M., 2021. Measuring the nutrition transition and its dynamics. *Public Health Nutrition*, 24 (2), 318-320.
Popkin, B. M., Adair, L. S. and Ng, S. W., 2012. Global nutrition transition and the pandemic of obesity in developing countries. *Nutrition Reviews*, 70 (1), 3-21.
Popkin, B. M., Corvalan, C. and Grummer-Strawn, L. M., 2020. Dynamics of the double burden of malnutrition and the changing nutrition reality. *The Lancet*, 395 (10217), 65-74.
Popkin, B. M. and Gordon-Larsen, P., 2004. The nutrition transition: Worldwide obesity dynamics and their determinants. *International Journal of Obesity*, 28 (S3), S2.
Popkin, B. M. and Gordon-Larsen, P., 2016. An international perspective on pediatric obesity. *In:* Goran, M. I. and Sother, M. S., eds. *Handbook of Pediatric Obesity: Etiology, Pathophysiology, and Prevention.* CRC Press, 53-66.
Popkin, B. M. and Reardon, T., 2018. Obesity and the food system transformation in Latin America. *Obesity Reviews*, 19 (8), 1028-1064.
Porat, I., Shoshany, M. and Frenkel, A., 2012. Two phase temporal-spatial autocorrelation of urban patterns: Revealing focal areas of re-urbanisation in Tel Aviv-Yafo. *Applied Spatial Analysis and Policy*, 5 (2), 137-155.
Porkka, M., Kummu, M., Siebert, S. and Varis, O., 2013. From food insufficiency towards trade dependency: A historical analysis of global food availability. *PloS One*, 8 (12), e82714.
Poti, J. M., Braga, B. and Qin, B., 2017. Ultra-processed food intake and obesity: What really matters for health - Processing or nutrient content? *Current Obesity Reports*, 6 (4), 420-431.
Poti, J. M., Mendez, M. A., Ng, S. W. and Popkin, B. M., 2015. Is the degree of food processing and convenience linked with the nutritional quality of foods purchased by US households? *The American Journal of Clinical Nutrition*, 101 (6), 1251-1262.
Prentice, A. M., 2018. The Double Burden of Malnutrition in Countries Passing through the Economic Transition. *Annals of Nutrition and Metabolism*, 72(suppl 3) (3), 47-54.
Public Health England, 2016. *The Eatwell Guide*. Public Health England in association with the Welsh Government, Food Standards Scotland and the Food Standards Agency in Northern Ireland.
Puth, M. T., Neuhäuser, M. and Ruxton, G. D., 2015. Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour*, 102, 77-84.
Putnam, J. J. and Allshouse, J. E., 1994. *Food Consumption, Prices, and Expenditures, 1970-93*. US Department of Agriculture, Economic Research Service.
Qin, K., Chen, Y., Zhan, Y. and Cheng, F., 2011. Spatial clustering considering spatio-temporal correlation, *2011 19th International Conference on Geoinformatics* (pp. 1-4).
Qiu, G., Liu, X., Amiranti, A. Y., Yasini, M., Wu, T., Amer, S. and Jia, P., 2020. Geographic clustering and region-specific determinants of obesity in the Netherlands. *Geospatial Health*, 15, 131-139.
Qu, X., Lee, L.-f. and Yu, J., 2017. QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Journal of Econometrics*, 197 (2), 173-201.
Quah, D., 1993. Galton's fallacy and tests of the convergence hypothesis. *The Scandinavian Journal of Economics*, 427-443.
Radwan, A., Gil, J. M., Variyam, J. N. and Creda–Upc–Irta, B., 2015a. A New, Obesity-specific Healthy Eating Index (OS-HEI) (202712.): European Association of Agricultural Economists.
Radwan, A., Gil, J. M., Variyam, J. N. and Creda–Upc–Irta, B., 2015b. A New, Obesity-specific Healthy Eating Index (OS-HEI), *EAAE-AAEA Joint Seminar 'Consumer Behaviour in a Changing World: Food, Culture, Society'*. Naples, Italy: European Association of Agricultural Economists.
Rae, A. N., 1998. The effects of expenditure growth and urbanisation on food consumption in East Asia: A note on animal products. *Agricultural Economics*, 18 (3), 291-299.
Raghunathan, K., Headey, D. and Herforth, A., 2021. Affordability of nutritious diets in rural India. *Food Policy*, 99, 101982.

Ralston, J., Brinsden, H., Buse, K., Candeias, V., Caterson, I., Hassell, T., Kumanyika, S., Nece, P., Nishtar, S. and Patton, I., 2018. Time for a new obesity narrative. *The Lancet*, 392 (10156), 1384-1386.

Rauber, F., da Costa Louzada, M. L., Steele, E. M., Millett, C., Monteiro, C. A. and Levy, R. B., 2018. Ultra-processed food consumption and chronic non-communicable diseases-related dietary nutrient profile in the UK (2008–2014). *Nutrients*, 10 (5), 587.

Rauber, F., Steele, E. M., Louzada, M. L. d. C., Millett, C., Monteiro, C. A. and Levy, R. B., 2020. Ultra-processed food consumption and indicators of obesity in the United Kingdom population (2008-2016). *PLoS One*, 15 (5), e0232676.

Ravallion, M., 2012. Why don't we see poverty convergence? *American Economic Review*, 102 (1), 504-523.

Ravuvu, A., Friel, S., Thow, A. M., Snowdon, W. and Wate, J., 2017. Monitoring the impact of trade agreements on national food environments: Trade imports and population nutrition risks in Fiji. *Globalisation and Health*, 13 (1), 33.

Rawal, V. and Navarro, D. K., 2019. *The global economy of pulses*. Rome: Food and Agriculture Organisation of the United Nations (FAO).

Rawashdeh, M. and Ralescu, A., 2012. Crisp and fuzzy cluster validity: Generalized intra-inter silhouette index, *2012 Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS)* (pp. 1-6).

Reardon, T. and Berdegué, J. A., 2002. The rapid rise of supermarkets in Latin America: Challenges and opportunities for development. *Development Policy Review*, 20 (4), 371-388.

Reardon, T., Timmer, C. P., Barrett, C. B. and Berdegue, J., 2003. The rise of supermarkets in Africa, Asia, and Latin America. *American Journal of Agricultural Economics*, 85 (5), 1140-1146.

Reardon, T., Timmer, C. P. and Minten, B., 2012. Supermarket revolution in Asia and emerging development strategies to include small farmers. *Proceedings of the National Academy of Sciences*, 109 (31), 12332-12337.

Reardon, T., Tschirley, D., Dolislager, M., Snyder, J., Hu, C. and White, S., 2014. *Urbanisation, diet change, and transformation of food supply chains in Asia*. Michigan State University: Global Center for Food Systems Innovation.

Reardon, T., Tschirley, D., Liverpool-Tasie, L. S. O., Awokuse, T., Fanzo, J., Minten, B., Vos, R., Dolislager, M., Sauer, C., Dhar, R., Vargas, C., Lartey, A., Raza, A. and Popkin, B. M., 2021. The processed food revolution in African food systems and the double burden of malnutrition. *Global Food Security*, 28, 100466.

Regmi, A. and Dyck, J., 2001. *Effects of urbanisation on global food demand*. Washington D.C: USDA-ERS.

Regmi, A. and Meade, B., 2013. Demand side drivers of global food security. *Global Food Security*, 2 (3), 166-171.

Regmi, A., Takeshima, H. and Unnevehr, L., 2008a. *Convergence in food demand and delivery: Do middle-income countries follow high-income trends*. Food Distribution Research Society. 0047-245X.

Regmi, A., Takeshima, H. and Unnevehr, L. J., 2008b. *Convergence in global food demand and delivery*. Economic Research Service, United States Department of Agriculture.

Regmi, A. and Unnevehr, L. J., 2006. Are diets converging globally? A comparison of trends across selected countries. *Journal of Food Distribution Research*, 37 (856-2016-57513), 14-21.

Remans, R., Wood, S. A., Saha, N., Anderman, T. L. and DeFries, R. S., 2014. Measuring nutritional diversity of national food supplies. *Global Food Security*, 3 (3-4), 174-182.

Rey, S., 2014. Spatial dynamics and space-time data analysis. *In:* Fischer, M. M. and Nijkamp, P., eds. *Handbook of Regional Science.* Berlin, Heidelberg: Springer, 1365-1383.

Rey, S. J., 2001. Spatial empirics for economic growth and convergence. *Geographical Analysis*, 33 (3), 195-214.

Rey, S. J. and Montouri, B. D., 1999. US regional income convergence: A spatial econometric perspective. *Regional Studies*, 33 (2), 143-156.

Reyes Matos, U., Mesenburg, M. A. and Victora, C. G., 2020. Socioeconomic inequalities in the prevalence of underweight, overweight, and obesity among women aged 20-49 in low- and middle-income countries. *International Journal of Obesity*, 44 (3), 609-616.

Rippin, H. L., Hutchinson, J., Greenwood, D. C., Jewell, J., Breda, J. J., Martin, A., Rippin, D. M., Schindler, K., Rust, P. and Fagt, S., 2020. Inequalities in education and national income are associated with poorer diet: Pooled analysis of individual participant data across 12 European countries. *PloS One*, 15 (5), e0232447.

Rodrigues, S. S. P., Caraher, M., Trichopoulou, A. and De Almeida, M. D. V., 2008. Portuguese households' diet quality (adherence to Mediterranean food pattern and compliance with WHO population dietary goals): trends, regional disparities and socioeconomic determinants. *European Journal of Clinical Nutrition*, 62 (11), 1263.

Román-Vinas, B., Barba, L. R., Ngo, J., Martínez-González, M. Á., Wijnhoven, T. M. A. and Serra-Majem, L., 2009. Validity of dietary patterns to assess nutrient intake adequacy. *British Journal of Nutrition*, 101 (S2), S12-S20.

Ronto, R., Wu, J. H. Y. and Singh, G. M., 2018. The global nutrition transition: Trends, disease burdens and policy interventions. *Public Health Nutrition*, 21 (12), 2267-2270.

Rosato, V., Temple, N. J., La Vecchia, C., Castellan, G., Tavani, A. and Guercio, V., 2019. Mediterranean diet and cardiovascular disease: a systematic review and meta-analysis of observational studies. *European Journal of Nutrition*, 58 (1), 173-191.

Rousham, E., Pradeilles, R., Akparibo, R., Aryeetey, R., Bash, K., Booth, A., Muthuri, S. K., Osei-Kwasi, H., Marr, C. M. and Norris, T., 2020. Dietary behaviours in the context of nutrition transition: A systematic review and meta-analyses in two African countries. *Public Health Nutrition*, 23 (11), 1948-1964.

Rousseeuw, P. J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Roy, T., 2020. *Reading the economic history of Afghanistan*. London: The London School of Economics and Political Science.

Ruel, M. T., 2003. Is dietary diversity an indicator of food security or dietary quality? A review of measurement issues and research needs. *Food and Nutrition Bulletin*, 24 (2), 231-232.

Rumawas, M. E., Dwyer, J. T., McKeown, N. M., Meigs, J. B., Rogers, G. and Jacques, P. F., 2009. The development of the Mediterranean-style dietary pattern score and its application to the American diet in the Framingham Offspring Cohort. *The Journal of nutrition*, 139 (6), 1150-1156.

Rupa, J. A., Umberger, W. J. and Zeng, D., 2019. Does food market modernisation lead to improved dietary diversity and diet quality for urban Vietnamese households? *Australian Journal of Agricultural and Resource Economics*.

Russell, J., Lechner, A., Hanich, Q., Delisle, A., Campbell, B. and Charlton, K., 2018. Assessing food security using household consumption expenditure surveys (HCES): A scoping literature review. *Public Health Nutrition*, 21 (12), 2200-2210.

Sadowski, A., 2019. Spatial and economic determinants of different food consumption patterns around the world. *Journal of Agribusiness and Rural Development*, 1 (51), 69-76.

Sahal Estimé, M., Lutz, B. and Strobel, F., 2014. Trade as a structural driver of dietary risk factors for noncommunicable diseases in the Pacific: An analysis of household income and expenditure survey data. *Globalisation and Health*, 10 (1), 48.

Sala-i-Martin, X. X., 1996a. Regional cohesion: Evidence and theories of regional growth and convergence. *European Economic Review*, 40, 1325-1352.

Sala-i-Martin, X. X., 1996b. The classical approach to convergence analysis. *The Economic Journal*, 106 (437), 1019-1036.

Sans, P. and Combris, P., 2015. World meat consumption patterns: An overview of the last fifty years (1961–2011). *Meat Science*, 109, 106-111.

Sares-Jäske, L., Knekt, P., Lundqvist, A., Heliövaara, M. and Männistö, S., 2017. Dieting attempts modify the association between quality of diet and obesity. *Nutrition Research*, 45, 63-72.

Scalvedi, M. L., Turrini, A. and Saba, A., 2018. Which dietary patterns are more likely to be associated with aspects of eco-sustainable food behaviours in Italy? *International Journal of Food Sciences and Nutrition*, 69 (6), 660-675.

Schmid, A., 2010. The role of meat fat in the human diet. *Critical Reviews in Food Science and Nutrition*, 51 (1), 50-66.

Schmidhuber, J. and Shetty, P., 2005. The nutrition transition to 2030. Why developing countries are likely to bear the major burden. *Acta Agriculturae Scandinavica, Section C - Food Economics*, 2 (3-4), 150-166.

Schmidhuber, J. and Traill, W. B., 2006. The changing structure of diets in the European Union in relation to healthy eating guidelines. *Public Health Nutrition*, 9 (5), 584-595.

Schneider, U. A., Havlík, P., Schmid, E., Valin, H., Mosnier, A., Obersteiner, M., Böttcher, H., Skalský, R., Balkovič, J. and Sauer, T., 2011. Impacts of population growth, economic development, and technical change on global food production and consumption. *Agricultural Systems*, 104 (2), 204-215.

Schoeller, D. A. and Westerterp, M., 2017. *Advances in the Assessment of Dietary Intake*. CRC Press.

Schram, A., Labonte, R., Baker, P., Friel, S., Reeves, A. and Stuckler, D., 2015. The role of trade and investment liberalisation in the sugar-sweetened carbonated beverages market: A natural experiment contrasting Vietnam and the Philippines. *Globalisation and Health*, 11 (1), 41.

Schwingshackl, L. and Hoffmann, G., 2015. Diet quality as assessed by the Healthy Eating Index, the Alternate Healthy Eating Index, the Dietary Approaches to Stop Hypertension score, and health outcomes: A systematic review and meta-analysis of cohort studies. *Journal of the Academy of Nutrition and Dietetics*, 115 (5), 780-800.

Sengul, H. and Sengul, S., 2006. Food consumption and economic development in Turkey and European Union countries. *Applied Economics*, 38 (20), 2421-2431.

Seymour, J. D., Calle, E. E., Flagg, E. W., Coates, R. J., Ford, E. S. and Thun, M. J., 2003. Diet quality index as a predictor of short-term mortality in the American Cancer Society Cancer Prevention Study II Nutrition Cohort. *American Journal of Epidemiology*, 157 (11), 980-988.

Shatenstein, B., Nadon, S., Godin, C. and Ferland, G., 2005. Diet quality of Montreal-area adults needs improvement: estimates from a self-administered food frequency questionnaire furnishing a dietary indicator score. *Journal of the American Dietetic Association*, 105 (8), 1251-1260.

Sheehy, T., Carey, E., Sharma, S. and Biadgilign, S., 2019. Trends in energy and nutrient supply in Ethiopia: a perspective from FAO food balance sheets. *Nutrition Journal*, 18 (1), 46.

Sheehy, T. and Sharma, S., 2010. The nutrition transition in Barbados: Trends in macronutrient supply from 1961 to 2003. *British Journal of Nutrition*, 104 (8), 1222-1229.

Sheehy, T. and Sharma, S., 2013. Trends in energy and nutrient supply in Trinidad and Tobago from 1961 to 2007 using FAO food balance sheets. *Public Health Nutrition*, 16 (9), 1693-1702.

Shekhar, S., Jiang, Z., Ali, R. Y., Eftelioglu, E., Tang, X., Gunturi, V. and Zhou, X., 2015. Spatiotemporal data mining: a computational perspective. *ISPRS International Journal of Geo-Information*, 4 (4), 2306-2338.

Shorrocks, A. F., 1984. Inequality decomposition by population subgroups. *Econometrica*, 52 (6), 1369-1385.

Shrimpton, R. and Rokx, C., 2012. *The double burden of malnutrition: A review of global evidence*. Washington: The World Bank.

Sievert, K., Lawrence, M., Naika, A. and Baker, P., 2019. Processed foods and nutrition transition in the Pacific: Regional trends, patterns and food system drivers. *Nutrients*, 11 (6), 1328.

Sklar, M., 1959. Fonctions de repartition an dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, 8, 229-231.

Skoufias, E., Di Maro, V., González-Cossío, T. and Ramirez, S. R., 2011. Food quality, calories and household income. *Applied Economics*, 43 (28), 4331-4342.

Smith, I. F., 2013. Sustained and integrated promotion of local, traditional food systems for nutrition security. *Diversifying Food and Diets.* Routledge, 154-171.

Smith, L. P., Ng, S. W. and Popkin, B. M., 2014. Resistant to the recession: US adults maintain cooking and away-from-home eating patterns during times of economic turbulence. *American Journal of Public Health*, 104, 840-846.

Smith, R., Kelly, B., Yeatman, H. and Boyland, E., 2019. Food marketing influences children's attitudes, preferences and consumption: A systematic critical review. *Nutrients*, 11 (4), 875.

Sobal, J., 1991. Obesity and socioeconomic status: a framework for examining relationships between physical and social variables. *Medical Anthropology*, 13 (3), 231-247.

Sojková , Z. and Matejková, E., 2001. Investigation of convergence theory evidence in European countries food consumption. *71rd Seminar of the European Association of Agricultural Economists (EAAE): The Food Consumer in the early 21st Century*, Zaragoza, Spain. EAAE.
Solow, R. M., 1956. A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70 (1), 65-94.
Sondermann, D., 2014. Productivity in the euro area: Any evidence of convergence? *Empirical Economics*, 47 (3), 999-1027.
Srinivasan, C. S., Irz, X. and Shankar, B., 2006. An assessment of the potential consumption impacts of WHO dietary norms in OECD countries. *Food Policy*, 31 (1), 53-77.
Srivastava, S. K., Balaji, S. J. and Kolady, D., 2016. Is there a convergence in dietary energy intake among expenditure-classes in India? *Agricultural Economics Research Review*, 29 (conf), 119-128.
Stage, J., Stage, J. and McGranahan, G., 2010. Is urbanisation contributing to higher food prices? *Environment and Urbanisation*, 22 (1), 199-215.
StataCorp, 2019. *Stata 16 Base Reference Manual*. College Station, TX: Stata Press.
Staudigel, M. and Schröck , R., 2015. Food demand in Russia: heterogeneous consumer segments over time. *Journal of Agricultural Economics*, 66 (3), 615-639.
Stańczyk, R., 2016. Convergence of health status in the European Union: a spatial econometric approach. *Athens Journal of Health*, 3 (1), 95-112.
Steenson, S. and Buttriss, J. L., 2020. The challenges of defining a healthy and 'sustainable' diet. *Nutrition Bulletin*, 45 (2), 206-222.
Stiglitz, J., 1976. The efficiency wage hypothesis, surplus labour, and the distribution of income in LDCs. *Oxford Economic Papers*, 28 (2), 185-207.
Stookey, J. D., Wang, Y., Ge, K., Lin, H. and Popkin, B. M., 2000. Measuring diet quality in China: the INFH-UNC-CH diet quality index. *European Journal of Clinical Nutrition*, 54 (11), 811.
Stuckler, D., McKee, M., Ebrahim, S. and Basu, S., 2012. Manufacturing epidemics: The role of global producers in increased consumption of unhealthy commodities including processed foods, alcohol, and tobacco. *PLoS Med*, 9 (6), e1001235.
Subramanian, S. and Deaton, A., 1996. The demand for food and calories. *Journal of Political Economy*, 104 (1), 133-162.
Sudharsanan, N., Ali, M. K., Mehta, N. K. and Narayan, K. M. V., 2015. Population aging, macroeconomic changes, and global diabetes prevalence, 1990–2008. *Population Health Metrics*, 13 (1), 33.
Sun, H., Tu, Y. and Yu, S. M., 2005. A spatio-temporal autoregressive model for multi-unit residential market analysis. *The Journal of Real Estate Finance and Economics*, 31 (2), 155-187.
Sundquist, J. and Johansson, S.-E., 1998. The influence of socioeconomic status, ethnicity and lifestyle on body mass index in a longitudinal study. *International Journal of Epidemiology*, 27 (1), 57-63.
Sununtnasuk, C. and Fiedler, J. L., 2017. Can household-based food consumption surveys be used to make inferences about nutrient intakes and inadequacies? A Bangladesh case study. *Food Policy*, 72, 121-131.
Sweitzer, M., Brown, D., Karns, S., Muth, M. K., Siegel, P. and Zhen, C., 2017. *Food-at-home expenditures: Comparing commercial household scanner data from IRI and government survey data*. US Department of Agriculture, Economic Research Service.
Swinburn, B. A., Kraak, V. I., Allender, S., Atkins, V. J., Baker, P. I., Bogard, J. R., Brinsden, H., Calvillo, A., De Schutter, O. and Devarajan, R., 2019. The global syndemic of obesity, undernutrition, and climate change: The Lancet Commission report. *The Lancet*, 393 (10173), 791-846.
Swinburn, B. A., Sacks, G., Hall, K. D., McPherson, K., Finegood, D. T., Moodie, M. L. and Gortmaker, S. L., 2011. The global obesity pandemic: Shaped by global drivers and local environments. *The Lancet*, 378 (9793), 804-814.
Tamura, K., Duncan, D. T., Athens, J. K., Bragg, M. A., Rienti Jr, M., Aldstadt, J., Scott, M. A. and Elbel, B., 2017. Geospatial clustering in sugar-sweetened beverage consumption among Boston youth. *International Journal of Food Sciences and Nutrition*, 68 (6), 719-725.

Tapsell, L. C., Neale, E. P., Satija, A. and Hu, F. B., 2016. Foods, nutrients, and dietary patterns: Interconnections and implications for dietary guidelines. *Advances in Nutrition*, 7 (3), 445-454.

The Lancet, 2019. *The Double Burden of Malnutrition* [online]. Available from: https://www.thelancet.com/series/double-burden-malnutrition [Accessed 4th March 2020].

Theil, H., 1967. *Economics and Information Theory*. North Holland.

Thow, A. M., 2009. Trade liberalisation and the nutrition transition: Mapping the pathways for public health nutritionists. *Public Health Nutrition*, 12 (11), 2150-2158.

Thow, A. M. and Hawkes, C., 2009. The implications of trade liberalisation for diet and health: A case study from Central America. *Globalisation and Health*, 5 (1), 5.

Thow, A. M., Heywood, P., Schultz, J., Quested, C., Jan, S. and Colagiuri, S., 2011. Trade and the nutrition transition: Strengthening policy for health in the Pacific. *Ecology of Food and Nutrition*, 50 (1), 18-42.

Thow, A. M., Jones, A., Hawkes, C., Ali, I. and Labonté, R., 2017. Nutrition labelling is a trade policy issue: Lessons from an analysis of specific trade concerns at the World Trade Organisation. *Health Promotion International*, 33 (4), 561-571.

Tibshirani, R., Walther, G. and Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63 (2), 411-423.

Tilman, D., Balzer, C., Hill, J. and Befort, B. L., 2011. Global food demand and the sustainable intensification of agriculture. *Proceedings of The National Academy of Sciences*, 108 (50), 20260-20264.

Timmer, C. P., Falcon, W. P., Pearson, S. R. and World Bank. Agriculture and Rural Development Dept. Economics and Policy, D., 1983. *Food policy analysis*. Baltimore: Johns Hopkins University Press.

Tobler, W., 2004. On the first law of geography: A reply. *Annals of the Association of American Geographers*, 94 (2), 304-310.

Tobler, W. R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46 (sup1), 234-240.

Toft, U., Kristoffersen, L. H., Lau, C., Borch-Johnsen, K. and Jørgensen, T., 2007. The Dietary Quality Score: validation and association with cardiovascular risk factors: the Inter99 study. *European Journal of Clinical Nutrition*, 61 (2), 270.

Tomita, A., Cuadros, D. F., Mabhaudhi, T., Sartorius, B., Ncama, B. P., Dangour, A. D., Tanser, F., Modi, A. T., Slotow, R. and Burns, J. K., 2020. Spatial clustering of food insecurity and its association with depression: A geospatial analysis of nationally representative South African data, 2008–2015. *Scientific Reports*, 10 (1), 1-11.

Torabi, M., 2016. Hierarchical multivariate mixture generalized linear models for the analysis of spatial data: An application to disease mapping. *Biometrical Journal*, 58 (5), 1138-1150.

Traill, W. B., 1997. *Convergence in US and EU food systems? The case of food consumption*. Food Marketing Policy Center, University of Connecticut and University of Massachusetts, Amherst, MA.

Traill, W. B., 2017. The role of FDI in food industries, transnational corporations and supermarkets in shifting diets. *XV European Association of Agricultural Economics (EAAE) Congress*, Palma, Italy.

Traill, W. B., Mazzocchi, M., Shankar, B. and Hallam, D., 2014. Importance of government policies and other influences in transforming global diets. *Nutrition Reviews*, 72 (9), 591-604.

Traka, M. H., Plumb, J., Berry, R., Pinchen, H. and Finglas, P. M., 2020. Maintaining and updating food composition datasets for multiple users and novel technologies: Current challenges from a UK perspective. *Nutrition Bulletin*, 45 (2), 230-240.

Trichopoulos, D. and Lagiou, P., 2004. Mediterranean diet and overall mortality differences in the European Union. *Public Health Nutrition*, 7 (7), 949-951.

Trichopoulou, A. and Benetou, V., 2019. Impact of Mediterranean diet on longevity. *Centenarians.* Springer, 161-168.

Trichopoulou, A., Costacou, T., Bamia, C. and Trichopoulos, D., 2003. Adherence to a Mediterranean diet and survival in a Greek population. *New England Journal of Medicine*, 348 (26), 2599-2608.

Trichopoulou, A., Kouris-Blazos, A., Wahlqvist, M. L., Gnardellis, C., Lagiou, P., Polychronopoulos, E., Vassilakou, T., Lipworth, L. and Trichopoulos, D., 1995. Diet and overall survival in elderly people. *Bmj*, 311 (7018), 1457-1460.
Trichopoulou, A., Orfanos, P., Norat, T., Bueno-de-Mesquita, B., Ocké, M. C., Peeters, P. H. M., van der Schouw, Y. T., Boeing, H., Hoffmann, K. and Boffetta, P., 2005. Modified Mediterranean diet and survival: EPIC-elderly prospective cohort study. *Bmj*, 330 (7498), 991.
Trijsburg, L., Talsma, E. F., de Vries, J. H. M., Kennedy, G., Kuijsten, A. and Brouwer, I. D., 2019. Diet quality indices for research in low- and middle-income countries: A systematic review. *Nutrition Reviews*, 77 (8), 515-540.
Trinh Thi, H., Simioni, M. and Thomas-Agnan, C., 2018a. Assessing the nonlinearity of the calorie-income relationship: An estimation strategy – With new insights on nutritional transition in Vietnam. *World Development*, 110, 192-204.
Trinh Thi, H., Simioni, M. and Thomas-Agnan, C., 2018b. Decomposition of changes in the consumption of macronutrients in Vietnam between 2004 and 2014. *Economics & Human Biology*, 31, 259-275.
Truswell, A. S., 1977. Diet and nutrition of hunter-gatherers. *In:* Elliott, K. and Whelan, J., eds. *Health and disease in tribal societies.* Amsterdam: Elsevier, 213-221.
Tselios, V., 2009. Growth and convergence in income per capita and income inequality in the regions of the EU. *Spatial Economic Analysis*, 4 (3), 343-370.
Tuma, M. N., Decker, R. and Scholz, S., 2011. A survey of the challenges and pitfalls of cluster analysis application in market segmentation. *International Journal of Market Research*, 53 (3), 391-414.
Tumas, N., Junyent, C. R., Aballay, L. R., Scruzzi, G. F. and Pou, S. A., 2019. Nutrition transition profiles and obesity burden in Argentina. *Public Health Nutrition*, 22 (12), 2237-2247.
Turner, C., Kalamatianou, S., Drewnowski, A., Kulkarni, B., Kinra, S. and Kadiyala, S., 2020. Food environment research in low- and middle-income countries: A systematic scoping review. *Advances in Nutrition*, 11 (2), 387-397.
Ulijaszek, S. and Schwekendiek, D., 2013. Intercontinental differences in overweight of adopted Koreans in the United States and Europe. *Economics & Human Biology*, 11 (3), 345-350.
Umberger, W. J., Rupa, J. A. and Zeng, D., 2020. Understanding food westernisation and other contemporary drivers of adult, adolescent and child nutrition quality in urban Vietnam. *Public Health Nutrition*, 23 (14), 2571-2583.
UNICEF, WHO and The World Bank, 2019. *Levels and trends in child malnutrition: key findings of the 2019 Edition of the Joint Child Malnutrition Estimates*. Geneva.
United Nations, 2020. *Millennium Development Goals* [online]. The United Nations. Available from: https://www.un.org/millenniumgoals/ [Accessed 3 March 2020].
Unnevehr, L. J., 2004. Mad cows and Bt potatoes: Global public goods in the food system. *American Journal of Agricultural Economics*, 86 (5), 1159-1166.
USDA and HHS, 2020. *Dietary Guidelines for Americans, 2020-2025*. U.S. Department of Agriculture (USDA) and U.S. Department of Health and Human Services (HHS).
Vandevijvere, S., Monteiro, C., Krebs-Smith, S. M., Lee, A., Swinburn, B., Kelly, B., Neal, B., Snowdon, W., Sacks, G. and Informas, 2013. Monitoring and benchmarking population diet quality globally: A step-wise approach. *Obesity Reviews*, 14, 135-149.
Vareiro, D., Bach-Faig, A., Raidó Quintana, B., Bertomeu, I., Buckland, G., Vaz de Almeida, M. D. and Serra-Majem, L., 2009. Availability of Mediterranean and non-Mediterranean foods during the last four decades: comparison of several geographical areas. *Public Health Nutrition*, 12 (9A), 1667-1675.
Vargas, L. A., 1990. Old and new transitions and nutrition in Mexico. *In:* Swedlund, A. C. and Armelagos, G. J., eds. *Disease in Populations in Transition.* Westport, CT: Greenwood.
Velichko, V. M. and Zagoruyko, N. G., 1970. Automatic recognition of 200 words. *International Journal of Man-Machine Studies*, 2 (3), 223-234.
Ventura Barbosa Gonçalves, H., Canella, D. S. and Bandoni, D. H., 2020. Temporal variation in food consumption of Brazilian adolescents (2009-2015). *Plos One*, 15 (9), e0239217.

Verger, E. O., Mariotti, F., Holmes, B. A., Paineau, D. and Huneau, J.-F., 2012. Evaluation of a diet quality index based on the probability of adequate nutrient intake (PANDiet) using national French and US dietary surveys. *Plos One*, 7 (8), e42155.
Vermeulen, S., Park, T., Khoury, C. K., Mockshell, J., Béné, C., Trinh Thi, H., Heard, B. and Wilson, B., 2019. *Changing diets and transforming food systems*. Wageningen, the Netherlands: CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).
Viegas, M. and Antunes, M., 2013. Convergence in the Spanish and Portuguese NUTS 3 regions: An exploratory spatial approach. *Intereconomics*, 48 (1), 59-66.
Vilar, J. A., Alonso, A. M. and Vilar, J. M., 2010. Non-linear time series clustering based on non-parametric forecast densities. *Computational Statistics & Data Analysis,* 54 (11), 2850-2865.
Vilar, J. A., Lafuente-Rego, B. and D'Urso, P., 2018. Quantile autocovariances: A powerful tool for hard and soft partitional clustering of time series. *Fuzzy Sets and Systems*, 340, 38-72.
Vilarnau, C., Stracker, D. M., Funtikov, A., da Silva, R., Estruch, R. and Bach-Faig, A., 2019. Worldwide adherence to Mediterranean Diet between 1960 and 2011. *European Journal of Clinical Nutrition*, 72, 83-91.
Villegas, R., Salim, A., Collins, M. M., Flynn, A. and Perry, I. J., 2004. Dietary patterns in middle-aged Irish men and women defined by cluster analysis. *Public Health Nutrition*, 7 (8), 1017-1024.
Viroli, C., 2011. Finite mixtures of matrix normal distributions for classifying three-way data. *Statistics and Computing*, 21 (4), 511-522.
Vogli, R. D., Kouvonen, A., Elovainio, M. and Marmot, M., 2014. Economic globalization, inequality and body mass index: A cross-national analysis of 127 countries. *Critical Public Health*, 24 (1), 7-21.
Von Ruesten, A., Illner, A. K., Buijsse, B., Heidemann, C. and Boeing, H., 2010. Adherence to recommendations of the German food pyramid and risk of chronic diseases: results from the EPIC-Potsdam study. *European Journal of Clinical Nutrition*, 64 (11), 1251.
Waijers, P. M. C. M., Feskens, E. J. M. and Ocké, M. C., 2007. A critical review of predefined diet quality scores. *British Journal of Nutrition*, 97 (2), 219-231.
Walls, H. L., Johnston, D., Mazalale, J. and Chirwa, E. W., 2018. Why we are still failing to measure the nutrition transition. *BMJ Global Health,* 3 (1), e000657.
Walthouwer, M. J. L., Oenema, A., Soetens, K., Lechner, L. and de Vries, H., 2014. Are clusters of dietary patterns and cluster membership stable over time? Results of a longitudinal cluster analysis study. *Appetite*, 82, 154-159.
Wan, G. H., 2005. Convergence in food consumption in Rural China: Evidence from household survey data. *China Economic Review*, 16 (1), 90-102.
Wang, X., Liu, Y., Chen, Y. and Liu, Y., 2016. An adaptive density-based time series clustering algorithm: A case study on rainfall patterns. *ISPRS International Journal of Geo-Information*, 5 (11), 205.
Wang, X., Wang, L. and Wirth, A., 2008a. Pattern discovery in motion time series via structure-based spectral clustering [Conference]. *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, USA. IEEE. 1. Available from: https://search.ebscohost.com/login.aspx?direct=true&db=edseee&AN=edseee.4587385&site=eds-live&scope=site [Accessed 12th June 2020].
Wang, X., Yu, F., Pedrycz, W. and Yu, L., 2019. Clustering of interval-valued time series of unequal length based on improved dynamic time warping. *Expert Systems with Applications*, 125, 293-304.
Wang, Z., Zhai, F., Du, S. and Popkin, B., 2008b. Dynamic shifts in Chinese eating behaviors. *Asia Pacific Journal of Clinical Nutrition*, 17 (1), 123-130.
Wang, Z., Zhai, F., Zhang, B. and Popkin, B. M., 2012. Trends in Chinese snacking behaviors and patterns and the social-demographic role between 1991 and 2009. *Asia Pacific Journal of Clinical Nutrition*, 21 (2), 253-262.
Webb, P. and Block, S., 2012. Support for agriculture during economic transformation: Impacts on poverty and undernutrition. *Proceedings of the National Academy of Sciences*, 109 (31), 12309-12314.
Wedel, M. and Kamakura, W. A., 2012. *Market segmentation: Conceptual and methodological foundations*. Vol. 8. Springer Science & Business Media.

Welch, N., Hunter, W., Butera, K., Willis, K., Cleland, V., Crawford, D. and Ball, K., 2009. Women's work. Maintaining a healthy body weight. *Appetite*, 53 (1), 9-15.
Well, D. N., 2007. Accounting for the effect of health on economic growth. *The Quarterly Journal of Economics*, 122 (3), 1265-1306.
Wells, J. C., Sawaya, A. L., Wibaek, R., Mwangome, M., Poullas, M. S., Yajnik, C. S. and Demaio, A., 2020. The double burden of malnutrition: aetiological pathways and consequences for health. *The Lancet*, 395 (10217), 75-88.
Wessells, K. R. and Brown, K. H., 2012. Estimating the global prevalence of zinc deficiency: Results based on zinc availability in national food supplies and the prevalence of stunting. *Plos One*, 7 (11), e50568.
WHO, 2003. *Diet, nutrition and the prevention of chronic diseases*. World Health Organisation.
WHO, 2004. *Obesity: preventing and managing the global epidemic*. World Health Organisation.
WHO, 2017a. *Obesity and overweight* [online]. World Health Organisation. Available from: http://www.who.int/mediacentre/factsheets/fs311/en/ [Accessed 23 January 2018].
WHO, 2017b. *The double burden of malnutrition: Policy Brief*. Geveva, Switzerland: World Health Organisation.
WHO, 2018. *Noncommunicable diseases* [online]. World Health Organisation. Available from: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [Accessed 24th August 2019].
WHO, 2019. *Malnutrition is a world health crisis* [online]. World Health Organisation. Available from: https://www.who.int/news-room/detail/26-09-2019-malnutrition-is-a-world-health-crisis [Accessed 25th June 2020].
WHO, 2020. *Malnutrition* [online]. World Health Organisation (WHO). Available from: https://www.who.int/news-room/fact-sheets/detail/malnutrition [Accessed 24 February 2021].
Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F. and Wood, A., 2019. Food in the Anthropocene: The EAT–Lancet Commission on healthy diets from sustainable food systems. *The Lancet*, 393 (10170), 447-492.
Wirfält, A. K. E. and Jeffery, R. W., 1997. Using cluster analysis to examine dietary patterns: nutrient intakes, gender, and weight status differ across food pattern clusters. *Journal of the American Dietetic Association*, 97 (3), 272-279.
Wirt, A. and Collins, C. E., 2009. Diet quality–what is it and does it matter? *Public health nutrition*, 12 (12), 2473-2492.
Witkamp, R. F. and van Norren, K., 2018. Let thy food be thy medicine….when possible. *European Journal of Pharmacology*, 836, 102-114.
World Bank, 2007. *Agriculture for Development*. Washington DC.
World Bank, 2020a. *Obesity: Health and Economic Consequences of an Impending Global Challenge*. Washington DC: International Bank for Reconstruction and Development/The World Bank.
World Bank, 2020b. *Trade (% of GDP)* [online]. The World Bank data. Available from: https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS [Accessed 17 January 2020].
World Bank, 2020c. *Urban population (% of total population)* [online]. The World Bank. Available from: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?end=2017&start=1960&type=shaded&view=chart [Accessed 5th March 2020].
World Bank, 2020d. *World Bank Country and Lending Groups* [online]. The World Bank. Available from: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups [Accessed 6th January 2020].
World Bank, 2020e. *World Development Indicators* [online]. Washington DC: The World Bank. Available from: https://databank.worldbank.org/source/world-development-indicators [Accessed 10th July 2020].
Wu, D. M., 1974. Alternative tests of independence between stochastic regressors and disturbances: Finite sample results. *Econometrica*, 42, 529-546.
Wuehler, S. E., Peerson, J. M. and Brown, K. H., 2005. Use of national food balance data to estimate the adequacy of zinc in national food supplies: methodology and regional estimates. *Public Health Nutrition*, 8 (7), 812-819.

Xiao, Y., 2020. The risk spillovers from the Chinese stock market to major East Asian stock markets: A MSGARCH-EVT-copula approach. *International Review of Economics & Finance*, 65, 173-186.

Xie, X. L. and Beni, G., 1991. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (8), 841-847.

Xue, L., Liu, G., Parfitt, J., Liu, X., Van Herpen, E., Stenmarck, Å., O'Connor, C., Östergren, K. and Cheng, S., 2017. Missing food, missing data? A critical review of global food losses and food waste data. *Environmental Science & Technology*, 51 (12), 6618-6633.

Yang, J., Farioli, A., Korre, M. and Kales, S. N., 2014. Modified Mediterranean diet score and cardiovascular risk in a North American working population. *PLoS One*, 9 (2), e87539.

Yang, L., Yang, L., Ho, K. C. and Hamori, S., 2020. Dependence structures and risk spillover in China's credit bond market: A copula and CoVaR approach. *Journal of Asian Economics*, 68, 101200.

Yang, M. S., 1993. A survey of fuzzy clustering. *Mathematical and Computer modelling*, 18 (11), 1-16.

Young, A. T., Higgins, M. J. and Levy, D., 2008. Sigma convergence versus beta convergence: Evidence from US county-level data. *Journal of Money, Credit and Banking*, 40 (5), 1083-1093.

Yu, D., 2014. Understanding regional development mechanisms in Greater Beijing Area, China, 1995–2001, from a spatial–temporal perspective. *GeoJournal*, 79 (2), 195-207.

Yusuf, S., Hawken, S., Ounpuu, S., Bautista, L., Franzosi, M. G., Commerford, P., Lang, C. C., Rumboldt, Z., Onen, C. L. and Lisheng, L., 2005. Obesity and the risk of myocardial infarction in 27 000 participants from 52 countries: a case-control study. *The Lancet*, 366 (9497), 1640-1649.

Zakaria, M., 2014. Effects of trade liberalisation on exports, imports and trade balance in Pakistan: A time series analysis. *Prague Economic Papers*, 23 (1), 121-139.

Zarrin, R., Ibiebele, T. I. and Marks, G. C., 2013. Development and validity assessment of a diet quality index for Australians. *Asia Pacific Journal of Clinical Nutrition*, 22 (2), 177.

Zezza, A., Carletto, C., Fiedler, J. L., Gennari, P. and Jolliffe, D., 2017. Food counts. Measuring food consumption and expenditures in household consumption and expenditure surveys (HCES). Introduction to the special issue. *Food Policy*, 72, 1-6.

Zhai, F. Y., Du, S. F., Wang, Z. H., Zhang, J. G., Du, W. W. and Popkin, B. M., 2014. Dynamics of the Chinese diet and the role of urbanicity, 1991–2011. *Obesity Reviews*, 15, 16-26.

Zhang, B. and An, B., 2018. Clustering time series based on dependence structure. *PloS one*, 13 (11), e0206753.

Zhang, N. and Ma, G., 2020. Nutritional characteristics and health effects of regional cuisines in China. *Journal of Ethnic Foods*, 7 (1), 7.

Zhang, X., Dagevos, H., He, Y., Van der Lans, I. and Zhai, F., 2008. Consumption and corpulence in China: a consumer segmentation study based on the food perspective. *Food Policy*, 33 (1), 37-47.

Zheng, Z., Henneberry, S. R., Zhao, Y. and Gao, Y., 2019. Predicting the changes in the structure of food demand in China. *Agribusiness*, 35 (3), 301-328.

Zhou, D. and Yu, X., 2015. Calorie elasticities with income dynamics: Evidence from the literature. *Applied Economic Perspectives and Policy*, 37 (4), 575-601.

Zhou, Z., Liu, H. and Cao, L., 2014. *Food consumption in China: The revolution continues*. Cheltenham: Edward Elgar Publishing.

Zimmer, D. M., 2016. Crop price comovements during extreme market downturns. *Australian Journal of Agricultural and Resource Economics*, 60 (2), 265-283.