PATTERNS OF WORLD TRADE: AN APPLICATION OF NETWORK COMPLEXITY ANALYSIS

Matthew McKay

A thesis submitted for the degree of Doctor of Philosophy of The Australian National University

October 2015

The Australian National University

Canberra, Australia

DECLARATION

I certify that this is my original work except where otherwise indicated or acknowledged in the thesis.

Matthew McKay

30 October 2015

ABSTRACT

This thesis is a contribution to network complexity analysis, which has shown a high degree of explanatory power in visualising the process of economic diversification, both across countries and over time. The thesis comprises a stage-setting introduction, a primer on the network complexity analysis and measures of economic complexity, three core chapters and a concluding chapter that summarises the key findings and makes suggestions for further research.

The first core chapter (Chapter 3) aims to place network complexity analysis, which has been criticised as a purely data driven analytical tool, within a theoretical economic context. By building a bridge between the mainstream theories of comparative advantage and the new analytical tool, the chapter demonstrates that the predictions of the theory of comparative advantage are borne out in the network analysis.

The next two chapters are novel empirical applications of network complexity analysis, using a structural construct called the product space. Chapter 4 deals with the impact of trade liberalisation on export performance, covering 96 liberalisation episodes in 129 countries over 1962–2012. The first stage of the analysis involves constructing measures of product emergence. These are then used in econometric analysis to examine the average within-country impact of trade liberalisation on product emergence and diffusion. The findings indicate that trade liberalisation is associated with an increase in the rate of product emergence and a moderate reorientation toward products that are more dissimilar to those in the pre-reform export composition.

Chapter 5 explores the implications of global production sharing on network complexity analysis. The approach taken is based on a systematic separation of production-sharing based trade (PSB trade) from the standard (reported) trade data. PSB trade is further disaggregated into parts and components and final assembly. It is found that failure to distinguish between PSB trade and standard horizontal trade tends to upwardly bias the complexity rankings of exports of some developing countries. This is because some developing countries are engaged in assembling and testing some parts and components while others engage in final assembly. It is also found that parts and components are clustered in a densely connected core of the

product space alongside relevant final products and have a higher level of average within-group proximity relative to final products, indicating higher average co-export potential. In the case of the machinery and transport equipment sector, parts and component exports tend to emerge prior to the export of final products, suggesting that global production sharing could act as a mechanism in the diffusion process from the periphery to the core of the product space.

ACKNOWLEDGMENTS

I thank my supervisor Prof. Prema-chandra Athukorala for his excellent supervision and for providing me with the opportunity to pursue this topic in applied computational development economics. His constant enthusiasm for understanding our world through the lens of economics is inspiring. I have learnt so much through our many conversations, in addition to tutoring his graduate course on development economics. I also thank Dr. Paul Burke, who provided me with excellent feedback on my drafts and always challenged me on coming up with ways to improve clarity of thought and expression.

Second I would like to thank my wife Louise McKay, who provided me with seemingly endless moments of encouragement and for providing the environment through which I could pursue these years of study. I would not have wanted to share this journey with anyone else.

I am very grateful to the Arndt-Corden Department of Economics (ACDE) at the Australian National University (ANU) and all faculty members who crafted an environment in which graduate students were an integral part of the academic community. The dedication of faculty in attending the PhD seminar series is appreciated by all graduate students and offered a valuable opportunity for feedback and further discussion and interaction. I learnt a lot from the questions that these experienced practitioners of economic development asked of us fearful presenters. My thanks also extends to the many questions and comments made by the participants of the various seminar opportunities both at ANU and the Monash PhD Conference. From the Monash conference thanks is extended particularly to Dr. Laura Puzzello and Prof. Robert Breunig. Each and every one of these experiences provided a mixture of encouragement, challenges, and insights to new avenues of thought.

I would also like to thank the following visiting professors: Ronald Findlay, Max Corden, James Riedel, and Kym Anderson. Prof. Ronald Findlay provided me with valuable feedback and encouragement at a crucial period in my topic formulation. I would also like to thank Prof. John Stachurski for the opportunity to work on the QuantEcon project. This work greatly improved the quality of my own research through the advancement of complementary python skills. It also gave me confidence knowing that the time dedicated to learning python and the associated eco-system of tools was well worth the investment.

There are many others to whom I am grateful for discussions and comments on the work included in this thesis. In particular I would like to thank Marcel Schroder and Swarnim Waglé for the many fruitful discussions and debates about economic development. They both had the patience to listen to some crazy ideas and provided a lot of valuable feedback. I would also like to thank Michele Coscia for many fruitful emails with regards to clarifications on the product space analysis and providing access to the datasets used at the Centre for International Development. A big thank you also goes to my family; mom, dad, and my two sisters. Mom and dad spent many hours providing me with feedback on my writing and drafts. My sisters gave me various moments of encouragement through spontaneous interactions over the years. Thanks to Michelle and Gabby Worthington, and Kimmy McCosker for the many spontaneous visits and coffee breaks. I would also like to thank the crew at Harvest and Hideout for providing an endless supply of amazing coffee. Canberra is a much better place with you guys around.

TABLE OF CONTENTS

ABBREVIATIONS

CEPII	Centre d'Etudes Prospectives et d'Informations Internationales
ECI	Economic Complexity Index (Indicator)
FAP	Final Assembly Product
FE	Fixed Effects
FM	Final Manufactures
FP	Final Product
GDP	Gross Domestic Product
GDPPC	Gross Domestic Product Per Capita
GMM	General Method of Moments
HO	Heckscher-Ohlin
HS	Harmonised System of international trade classification
IO	Input-Output
M_{cp}	Matrix with Country and Product Data
MNE	Multinational Enterprise
NBER	National Bureau of Economic Research
PC	Parts and Components
PCI	Product Complexity Index (Indicator)
PS	Product Space
PSB	Production Sharing Based
PSN	Product Space Network
RCA	Revealed Comparative Advantage
SITC	Standard International Trade Classification
UN	United Nations

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

> *"Ultimately, development is the expression of the total amount of productive knowledge that is embedded in a society. But the process by which this knowledge is accumulated has a structure that we are only now starting to understand." (Hausmann & Hidalgo 2014)*

International trade is one of the fundamental driving forces behind economic development and globalisation. Trade orientation is often noted as one key feature of any successful development story, as the South East Asian experience has so successfully demonstrated in recent decades. This view is supported by arguments grounded in international trade theory (Ricardo 1817; Haberler 1930; Ohlin 1933; Samuelson 1949; Vanek 1963; Dornbusch, Fischer & Samuelson 1977; Dixit & Norman 1980; Wilson 1980; Leamer 1984; Evenett & Keller 2002; Eaton & Kortum 2002; Fisher 2011; Chaney 2014), as well as a large body of empirical trade research that seeks to identify the variety of channels that link trade and development (Krueger 1997, 1998; Baldwin & Gu 2004; Winters 2004; Wacziarg & Welch 2008; Topalova & Khandelwal 2011; Costinot, Donaldson & Komunjer 2011; Costinot & Donaldson 2012). Much of the economic literature travels down the road of increasing complexity as models and analytical methods improve to allow for more realistic assumptions, or to incorporate different phenomenon that we observe from real world trade data such as intra-industry trade.

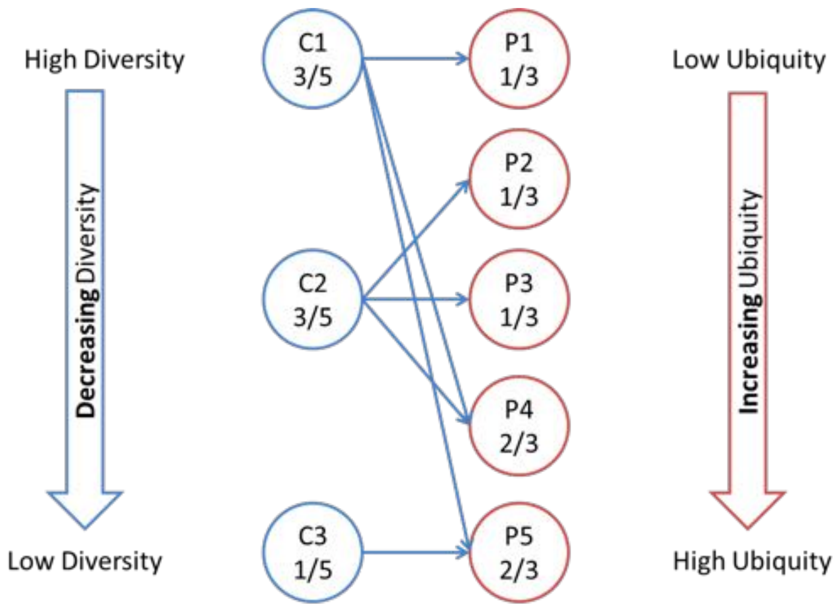
In recent years there has been renewed interest in exploring alternative methods to further our understanding of economic development (Jackson & Watts 2002; Kali, Méndez & Reyes 2007; Hidalgo et al. 2007; Reyes, Schiavo & Fagiolo 2010; Kali et al. 2013; Jackson & Nei 2015; Reyes, Wooster & Shirrell 2014; Head, Jing & Swenson 2014; Córcoles, Díaz-Mora & Gandoy 2014; Chaney 2014). The economic and product complexity approach is one such endeavour that has widened our perspective, as it adopts new analytical tools from the field of network science. This approach shows that analysing the network structure of international trade can reveal useful information about the underlying complexity of economies (Hausmann & Klinger 2006; Hidalgo et al. 2007; Hidalgo & Hausmann 2008; Hausmann & Hidalgo 2014). As an example, the economic complexity indicator (ECI) is strongly correlated with overall levels of economic development, both across countries and dynamically over time. Another finding is that clusters of highly related products found within a network that is constructed from a revealed measure of product similarities (the product space), exhibit strong correlation with more traditional product classifications that are used to group products according to type in the empirical trade literature (Leamer 1984; Lall 2000; Hausmann & Klinger 2006; Hidalgo et al. 2007). What these early findings suggest is that alternative techniques, such as the network complexity approach, offer interesting new measures that incorporate a high degree of explanatory power in differentiating between different development paths, and can be complementary to the more traditional forms of analysis.

The network complexity approach incorporates a new dimension for analysis by preserving embedded relational information and structure. This allows for more micro-level contextual analysis to be done in collaboration with broader macro-level analysis. The idea of adopting a network analysis approach is not entirely new to the study of international trade given the rich theoretical and empirical literature surrounding the gravity model (Anderson 1979; Feenstra 2001; Evenett & Keller 2002; Anderson & van Wincoop 2003; Silva & Tenreyro 2006; Helpman, Melitz & Rubinstein 2008; Chor 2010). The gravity model describes a network in which one estimates the edge values (i.e. value of exports) between any two countries (i.e. nodes) dependant on different attributes of nodes and edges. In the standard gravity model, GDP is an example of a node attribute, while the distance between two countries is an edge attribute. Network science, in recent years, has rapidly expanded to provide a rich analytical toolkit, that is not only geared towards edge value predictions, but also allows the analysis of broader structural consequences that relational data can imply. For example, network analysis provides a range of different measures of centrality that can emphasise direct network relationships, or higher-order relationships based on the connectivity structure of the network.

International trade data are a natural fit for network analysis as it describes flows of goods between countries, and has the unique property of being harmonised to allow for direct comparability. This type of data provides a unique opportunity to use network science techniques to extract new information that is embodied in the relational and structural features of this data. While there are many different ways to construct a network, the focus in this thesis is on the bipartite network, a mapping between countries and the products that they export.[1]

As demonstrated in Figure 1.1, consider a column of country nodes and a column of product nodes. The edges (that are defined in the diagram) pair countries with the products that they export with revealed comparative advantage. From the perspective of each *country node*, the sum of all edge links is a measure of diversification (the total number of exported products from a given country). Applying the same operation from the perspective of a *product node* provides a measure of ubiquity, a uniqueness rating for any given product. A ubiquitous product is one that is exported by many countries. Ubiquity and diversity, as determined by the bipartite network structure, are at the core of the recent literature on economic complexity.

Figure 1.1: Bipartite Network of 3 Countries and 5 Products



Notes: 1. The blue lines connect each country to the products that it exports. **2.** Countries are sorted by decreasing diversity (the total number of products that a country exports), and products are sorted by increasing ubiquity (total number of countries that export that product).

This thesis adds to the network complexity analysis research by first considering how these network methods relate to the standard theoretical foundation of comparative

[1] Other types of networks include undirected graphs, directed graphs, single edge graphs, and multi-edge graphs. The edges in the bipartite network mapping in this thesis are determined by the Balassa measure of revealed comparative advantage (Balassa 1965). A full discussion of the measure can be found in Appendix B.

advantage. There is a vast amount of literature regarding the different sources of comparative advantage arising from differences in technologies (Ricardo 1817; Dornbusch, Fischer & Samuelson 1977; Wilson 1980) or from differences in factor endowments (Ohlin 1933; Vanek 1963; Deardorff 1982). One of the key conclusions arising from this literature on international trade is that country heterogeneity is a fundamental determinant of comparative advantage. Any trading equilibrium that offers a different opportunity cost between any two goods to that which would otherwise be found in autarky provides an opportunity for gains from trade to be realised.[2] It is this heterogeneity that fits naturally within a network analysis framework. Network analysis techniques preserve disaggregation and can be used to analyse this heterogeneity at both the micro and macro levels.

Chapter 2 provides a primer on the network complexity analysis pioneered by Hausmann & Klinger (2006) and Hidalgo et al. (2007), and outlines the two primary strands of the foundational literature on which this thesis stands. The chapter explains how economic and product complexity indicators are computed and builds an understanding of the techniques using a small model of three countries and five products. This is then followed by the introduction of the product space, which is a network representation of revealed product similarities.

The focus of Chapter 3 is to examine the mechanics of how the product complexity indicators are computed in the context of predictions made by the mainstream theories of comparative advantage. One key outcome of comparative advantage is its emphasis on resultant patterns of specialisation that form from the international reorganisation of production. However, determination of these patterns quickly becomes confounded when considering the real world context of many-countries and many-products. General laws of comparative advantage often suffer from the presence of multiple solutions and the inability to determine the exact cross-sectional pattern of trade without imposing additional assumptions to the model.

The network complexity framework approaches identification of trade patterns directly from the data, rather than an underlying theory. The Balassa (1965) measure of revealed comparative advantage is used to identify products that each country is

[2] Assuming no associated trade or transport costs.

relatively good at producing. This serves as an input for the construction of the bipartite mapping between countries and products used in the *method of reflections* to compute economic and product complexity.[3] The network techniques are commonly discussed with a focus on two characteristics of trade: country diversity and product ubiquity (Hidalgo et al. 2007; Tacchella et al. 2012). A diverse country that exports many non-ubiquitous products will be ranked as more economically complex when compared with a country that has concentrated and ubiquitous exports. An alternative perspective is to view the *method of reflections* as a sorting heuristic that works to organise rows and columns in the cross-country trade matrix to deliver a pattern that is consistent with the notion of specialisation. Viewed in this way, ECI, PCI and the product space reveal patterns that are intrinsically linked with the predictions of both the Ricardian and Hecksher-Ohlin models.

Chapter 4 presents an empirical study that looks at trade liberalisation and the structural effects it has on patterns of export emergence across the product space network. A network of product similarity is used to characterise the structural change in export patterns by considering both the rate and distance of product emergence by comparing pre-liberalisation to post-liberalisation levels. To capture the rate of product emergence, the extensive margin is split into products that are near (probable) and far (improbable) from that country's position in the product space network. Additional measures that describe the relative distance of product diffusion are also computed to explore a continuous measure of diffusion distance. Both of these indicators are in turn used in a panel regression to estimate the impact of trade liberalisation on diffusion patterns. The findings indicate that trade liberalisation is associated with an 18% to 24% increase in the *rate* of product emergence and a moderate shift to exports that are increasingly *dissimilar* to those in the pre-reform export basket. This provides some indirect evidence that trade liberalisation is therefore supportive of both the accumulation and diversification of the underlying set of capabilities in an economy.

[3] The latest atlas of complexity uses an updated formulation for computing economic and product complexity indicators, but it delivers very similar results to the method of reflections used here to provide insight into how the network structure relates to the resultant complexity indicators.

The process of product fragmentation and its impact on the quality of information contained in the observed bipartite network structure between countries and products is the topic discussed in Chapter 5. Through improvements in production technology, coupled with large reductions in transportation costs, product fragmentation in recent decades has allowed developing countries to participate in the production of relatively complex products by providing assembly services based on their comparative advantage in labour. Products that are most susceptible to this process, called final assembly products (FAP), are identified and then differentiated by income level in the cross-country pattern of trade.

Once final assembly products are differentiated the results show that countries in East and South East Asia, in addition to Eastern Europe, drop in the economic complexity rankings due to the altered structure of the bipartite network. Low-type final assembly products, those made by less developed economies, are also typically found in the bottom quartile of the product complexity rankings. The positive bias in complexity is due to the network links between developing economies that specialise in the assembly segment of production and more advanced developed economies, such as Japan and Germany, which export products in the same SITC product code. More developed economies are both more likely to have the full suite of capabilities required to produce all segments of a product, and are likely to produce higher quality products. In this instance, network analysis inaccurately links the two types of economies, which inflates the developing economy complexity ranking due to the indirect association to relatively more advanced economies (that have lower average product ubiquity and higher average diversification) that is propagated through the network analysis.

The analysis in Chapter 5 continues by considering the positional and dynamic role that PSB products play within the product space network. It is found that these products are largely co-located with associated final products in the industrialised core of the product space and play a role in the dense connectivity that is observed. The patterns of emergence in parts and components, final assembly products, and other final products are then considered for two industries: Electronics and Machinery. In the case of the Machinery sector, evidence is found that parts and components, in addition to final assembly products, tend to act as gateway products to future participation in more sophisticated final goods. The same result was not observed in the Electronics sector, which found equal rates of diffusion between the three product groups. This is likely due to the restricted time frame of the sample to the years 1995 and 2012, or could be a reflection that it is hard to differentiate between final products and final assembly products within the already mature Electronics sector. The Electronics industry has been a frontier sector in the process of product fragmentation, in which much of the global dynamism may have occurred prior to 1995 not allowing it to be captured in this study.

A secondary outcome of the research undertaken for this thesis is the development of a Python package: *pyeconlab*. This package will be made available through github and will serve as a platform for future work with product level trade data. The package is split into sub modules for cleaning different datasets, compiling basic statistics, and building standardised data structures that can then be used with a variety of analytical tools. The heart of the package consists of python objects that make working with product level trade data much simpler. More details of this work can be found in Appendix I.

The empirical analysis of this thesis is based on data collected and compiled from a number of sources and its construction is explained in Appendix A. Chapter 3 makes use of trade data at the 4-digit level of the Standard International Trade Classification (SITC) for the year 2000. This year was selected in order to ensure comparability of the findings with Hausmann & Klinger (2006) and Hidalgo et al. (2007). The broad macro patterns that are discussed in this chapter are also reasonably stable over time, with similar observations made in other years. Chapter 4 makes use of a number of different datasets to explore export emergence. In order to incorporate as many trade liberalisation episodes as possible, SITC revision 2 at the 4-digit level was used to provide a time series from 1962 to 2012. The primary source of export data for Chapters 3 and 4 is the National Bureau of Economic Research (NBER) World Trade Flows (WTF) database between 1962 and 2000, in addition to the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII) BACI dataset for data from 1998 to 2012 (HS96). An intertemporally-consistent product classification is also constructed for Chapter 4 to test the robustness of results. Chapter 5, uses the CEPII BACI HS92 dataset for the years 1995 to 2013. The original data are based on the Harmonised

System (HS92) and is concorded to SITC revision 3 at the 5-digit level to enable the identification of parts and components at a sufficiently disaggregated level.

In summary, this thesis is a collection of essays that provide three main contributions. First it shows a complementarity between the traditional literature of comparative advantage and new network analytical techniques. These new analytical techniques can be viewed as a sorting heuristic that organises countries and products along a diagonal line of specialisation that correlates with patterns that are determined by theories of comparative advantage. Relatively complex countries tend to specialise in relatively complex exports, albeit with a large degree of observed heterogeneity. A key advantage of the new network complexity techniques is that they perform well in a many-country, many-product environment, an environment in which traditional analysis becomes complex and plagued with multiple solutions without the imposition of further assumptions on the underlying theory.

Second, the network complexity approach provides disaggregated metrics that are useful when looking at the structural evolution in country export baskets over time. This thesis uses the *product space* network as a way to characterise how country exports structurally change in light of trade liberalisation policy. This analysis shows that trade liberalisation tends to increase the *rate* of export emergence, in addition to a moderate increase in the *distance* of diffusion. These findings imply that trade liberalisation assists to increase the rate of accumulation and the level of diversity in underlying country capabilities on the export side.

Third, the limitations of these data-intensive techniques are also explored through an investigation into the process of global product fragmentation and the effect it has on the information contained in the bipartite network structure used to infer measures of complexity. This study finds that the economic complexity scores of countries that export final assembly products are biased upwards due to the presence of indirect network connections with relatively more advanced economies. When introducing differentiated final assembly products, developing countries in South Asia, South East Asia, and Eastern Europe exhibit declines in the complexity rankings. The positional and dynamic role that parts and components play in the product space was also examined and shows that these products play an important role in the densely connected core of the network. In the case of the Machinery industry, parts and components, and final assembly products tend to emerge in export baskets prior to other final goods, suggesting they have a role to play as gateway products as countries diffuse into this product cluster.

CHAPTER 2 ECONOMIC COMPLEXITY AND THE NETWORK STRUCTURE OF INTERNATIONAL TRADE

"By using data on the products exported by each country … it is possible to create measures of an economy that conserve the identity of the elements involved." (Hidalgo 2015, pg. 162)

Abstract

The network structure of trade can be used to derive revealed measures of economic and product complexity. These new metrics fundamentally rely on two characteristics of trade: country diversity and product ubiquity. A complex economy is one that is both diversified and exports many non-ubiquitous (rare) products, whereas an un-complex economy has concentrated exports in a few ubiquitous products. This chapter surveys these new economic complexity analysis techniques to provide a platform from which the remaining chapters are based. The concepts underpinning the *economic and product complexity indicators (ECI and PCI)*, and the *product space network* are explained using a small 3-country 5-product model.

2.1 INTRODUCTION

Network complexity analysis is one approach to understanding economic systems based on the multidisciplinary branch of the computational sciences, often referred to as *complexity economics* (Arthur 2013; Helbing & Kirman 2013; Krishna & Levchenko 2012). In general this set of analytical techniques embraces the inherent complexity of real-world systems that can take various forms such as high dimensionality, heterogeneity in economic actors, adaptive dynamic systems, or interactions with other non-economic systems (i.e. ecological systems). Network complexity tools are both broad and multidisciplinary and are becoming widely used in a variety of research fields. Network analysis is one such common example, in which the analytical techniques are used in computational archaeology, social network analysis, economic systems, telecommunications, and electrical systems to name a few. Network analysis has also been successfully applied to the study of international trade (Smith & White 1992; Mahutga 2006; Kali, Méndez & Reyes 2007; Reyes, Schiavo & Fagiolo 2010; He &

Deem 2010; Baskaran et al. 2011; Barigozzi, Fagiolo & Mangioni 2011; Kali et al. 2013; Benedictis et al. 2013; Chaney 2014; Dueñas & Fagiolo 2014; Shi et al. 2014; Zhu et al. 2014). In essence, adopting a network approach can be useful for analysing any form of relational data.

The network complexity approach developed by Hausmann and Hidalgo and their co-researchers, is a data-centric technique that is used as a foundation for this thesis. It casts the cross-sectional pattern of trade into a bipartite network between countries and their exported products (Hausmann & Klinger 2006; Hidalgo et al. 2007; Hidalgo & Hausmann 2008; Hausmann & Hidalgo 2014). The bipartite network is formed by filtering trade data through the Balassa (1965) definition of revealed comparative advantage (RCA) as a way to identify products that a country is relatively *"good"* at producing.[4] Another way to interpret the Balassa indicator is that it identifies products in country export baskets that have a higher share than the average expectation based on world product shares (Kunimoto 1977). While other measures of RCA are available, for the purpose of continuity with other research, the Balassa RCA indicator is used as the foundational indicator in this thesis.[5] The other measures are reviewed and discussed in Appendix B.

This measure of revealed comparative advantage (RCA) is defined:

$$RCA_{cp} = \frac{\frac{E_{cp}}{\sum_C E_{cp}}}{\frac{\sum_P E_{cp}}{\sum_{C,P} E_{cp}}}$$

where,

E_{cp} = Export of product p from country c

$\sum_C E_{cp}$= Total export from country c

[4] The use of the word *filter* is used in this context as this indicator of revealed comparative advantage is reducing noise that is contained in international trade data by excluding products that a country may export but *not* relatively well.

[5] When comparing the Balassa RCA index to other more recent indicators, the binary M_{cp} matrix that records which products each country exports is often equivalent. Many of the newer RCA measures deviate from Balassa RCA in the internal consistency and ranking of products above and below the RCA cutoff point.

$\sum_{P} E_{cp}$= Total export of product p

$\sum_{C,P} E_{cp}$= Total world export

There are two primary strands to this literature: (1) revealed indicators of economic complexity (ECI) and product complexity (PCI), and (2) the product space (PS). The complexity indicators use two primary dimensions – country diversity and product ubiquity – to rank countries and products.[6] Complex economies are classified as those that are diversified as well as having relatively non-ubiquitous products.

The product space, on the other hand, uses the cross-country pattern of trade to compute conditional probabilities of co-export as a revealed measure of product similarities. The information contained in this measure is best represented as a network in which product proximity is indicative of similarity in the factors of production. The key assumption is that two products exported by a similar set of countries, are most likely to exhibit similar underlying *capabilities* (or factors of production). It is useful to think of country capabilities as an overarching abstraction that includes the usual factors of production capital, labour, land, and other inputs such as production technology, information, institutions, or anything else that influences the capacity to produce.

The starting point for computing ECI, PCI, and the PS is the cross-country pattern of exports, as defined by the $\mathrm{M_{cp}}$ matrix.[7] Exports are used as an internationally comparable data source that indirectly provides information about what production capabilities a country has acquired through the types of products that are exported.

Using the RCA matrix, the $\mathrm{M_{cp}}$ matrix is derived by coding this information into a binary relationship between countries and products, focused on a cut-off value in which RCA is greater than or equal to one $(RCA \geq 1)$.[8] Therefore countries are deemed to be relatively good at producing a certain product when that product holds

[6] It shares some similarities with the rankings defined by Feenstra & Rose (2000) which is motivated by the product-cycle theory of international trade (Vernon 1966).

[7] Economic and product complexity indicators (ECI and PCI) and the Product Space (PS) are computed using a library of functions written in Python, which are described in Appendix H.

[8] Hidalgo et al. (2007) and Hausmann & Hidalgo (2011) consider alternative cut-off definitions and find that the overall network results are robust to these changes.

a value share in country exports that is greater than that products share in world trade. The $\mathrm{M_{cp}}$ matrix defines the set of products that are produced by all countries in any given year. For convenience the t subscript has been dropped from the following notation.

$$M_{cp} = \begin{cases} 1 & RCA \geq 1 \\ 0 & RCA < 1 \end{cases}$$

The M_{cp} matrix therefore takes the form: [9]

$$M_{cp} = \begin{bmatrix} (RCA \geq 1)_{c=1,p=1} & \cdots & (RCA \geq 1)_{c=1,p=P} \\ \vdots & \ddots & \vdots \\ (RCA \geq 1)_{c=C,p=1} & \cdots & (RCA \geq 1)_{c=C,p=P} \end{bmatrix}_t$$

This M_{cp} matrix incorporates two direct primary statistics: *country diversity* and *product ubiquity*. Diversity and Ubiquity are defined as:

$$Diversity = k_{c,0} = \sum_p M_{cp}$$

$$Ubiquity = k_{p,0} = \sum_c M_{cp}$$

Country diversity is computed by summing across the columns of the matrix, which is counting the number of products any given country exports with RCA and provides a direct measure of diversification. Summing down each column, on the other hand, gives a measure of product ubiquity. A product that is exported by many countries is deemed ubiquitous.

2.2 ECONOMIC AND PRODUCT COMPLEXITY

The *method of reflections* is an iterative procedure to compute a ranking of country complexity (ECI) and product complexity (PCI) indicators. Using country characteristics (diversity) and product characteristics (ubiquity) of exports, the method involves successive averaging over the contextual bipartite network structure between countries and products. Both country and product characteristics, as defined by the bipartite network, are the key determinants to the economic complexity approach. It is

[9] In network analysis this matrix is also called the adjacency matrix that describes the relationship between nodes in a given network.

this incorporation of context that sets this technique apart from others. The method of reflections iterates over these two dimensions, normalised by product ubiquity and country diversification, as defined by the jointly determined equations below.

$$k_{c,N} = \frac{1}{k_{c,o}} \sum_{p} M_{cp} \cdot k_{p,N-1}$$

$$k_{p,N} = \frac{1}{k_{p,0}} \sum_{c} M_{cp} \cdot k_{c,N-1}$$

where,

$k_{c,o}$ is a vector containing each countries' diversification score

$k_{p,0}$ is a vector containing each products' ubiquity score

$k_{c,N}$ is the nth iteration over the country dimension

$k_{p,N}$ is the nth iteration over the product dimension

While this procedure converges towards a global fixed-point value, iteration is terminated when a stable rank ordering of countries and products is achieved. The rank ordering of countries and products can be used as the indicator of economic and product complexity respectively. A country that exports a diversified range of products, and at the same time exports relatively unique (non-ubiquitous) products, is deemed as relatively economically complex.

These rankings can be better understood by considering the mechanics behind the method of reflections. The first iteration of this method characterises country export baskets by computing the average product ubiquity for each country.[10] The next iteration computes average country diversity of the average product ubiquity. Therefore higher order influences, based on the network structure, are propagated with subsequent iteration. A relatively less complex economy's complexity ranking would *increase* if it exported a product that is (a) relatively unique and (b) exported by relatively diversified economies that also produce other relatively unique products.

[10] This assumes one iterates over country characteristic dimension as a first step in the iteration. One can start with either country or product nodes.

These higher order influences play an important role in determining the final rank of countries and products.[11]

$$ECI = Country\ Rank\ when\ \Delta Rank\left(k_{c,N} - k_{c,N-1}\right) = 0\ for\ \forall\ c\ \in C$$

$$PCI = Product\ Rank\ when\ \Delta Rank\left(k_{p,N} - k_{p,N-1}\right) = 0\ for\ \forall\ p\ \in P$$

The recent atlas of complexity considers the eigenvector of the second largest eigenvalue of the recursive formulation of the above equations (Hausmann & Hidalgo 2014). The reason the method of reflections formulation is used here is that it allows one to more clearly understand the mechanics of the underlying computation. Both approaches yield very similar results. The recursive formulation is defined in Appendix C.

2.2.1 EXAMPLE: 3 COUNTRIES, 5 PRODUCTS

To explore the mechanics behind this iterative method, consider the case of three countries (C1, C2, and C3) and five products, as shown in Figure 2.1. The bipartite network structure in Figure 2.1 sorts' country nodes by decreasing diversity and product nodes by increasing ubiquity. The tied relationships that can be observed in the diagram are arbitrary.

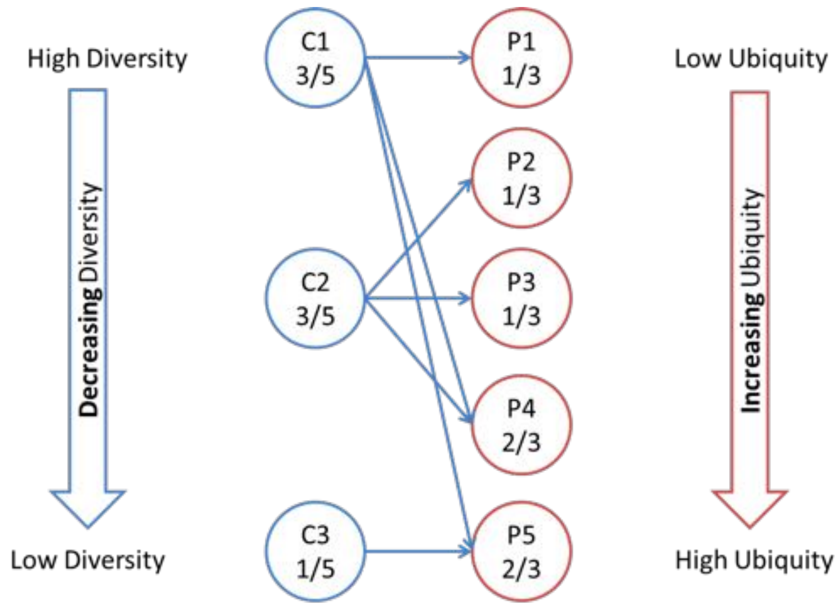
The resulting M_{cp} matrix of this bipartite network is:

$$Mcp = \begin{matrix} & P1 & P2 & P3 & P4 & P5 \\ C1 & 1 & 0 & 0 & 1 & 1 \\ C2 & 0 & 1 & 1 & 1 & 0 \\ C3 & 0 & 0 & 0 & 0 & 1 \end{matrix}$$

Now let us explore how the network information assists in ranking countries and products to compute the ECI and PCI indices. Figure 2.2 traces the logic of the method of reflections through two iterations.

[11] Using SITC Revision 2 Level 4 Product data, rank convergence is achieved in approximately 19 iterations. This is consistent with the published results (Hidalgo et al. 2007).

Figure 2.1: Bipartite Network of 3 Countries and 5 Products



Notes: 1. The blue lines connect each country to the products that it exports. **2.** Countries are sorted by decreasing diversity (the total number of products that country exports), and products are sorted by increasing ubiquity (total number of countries that export that product).

Starting with the initial bipartite network, the method of reflections computes the average product ubiquity of each country's exports.[12] Therefore, as Country 1 is connected to P1, P4, and P5, the average ubiquity score for this country is:

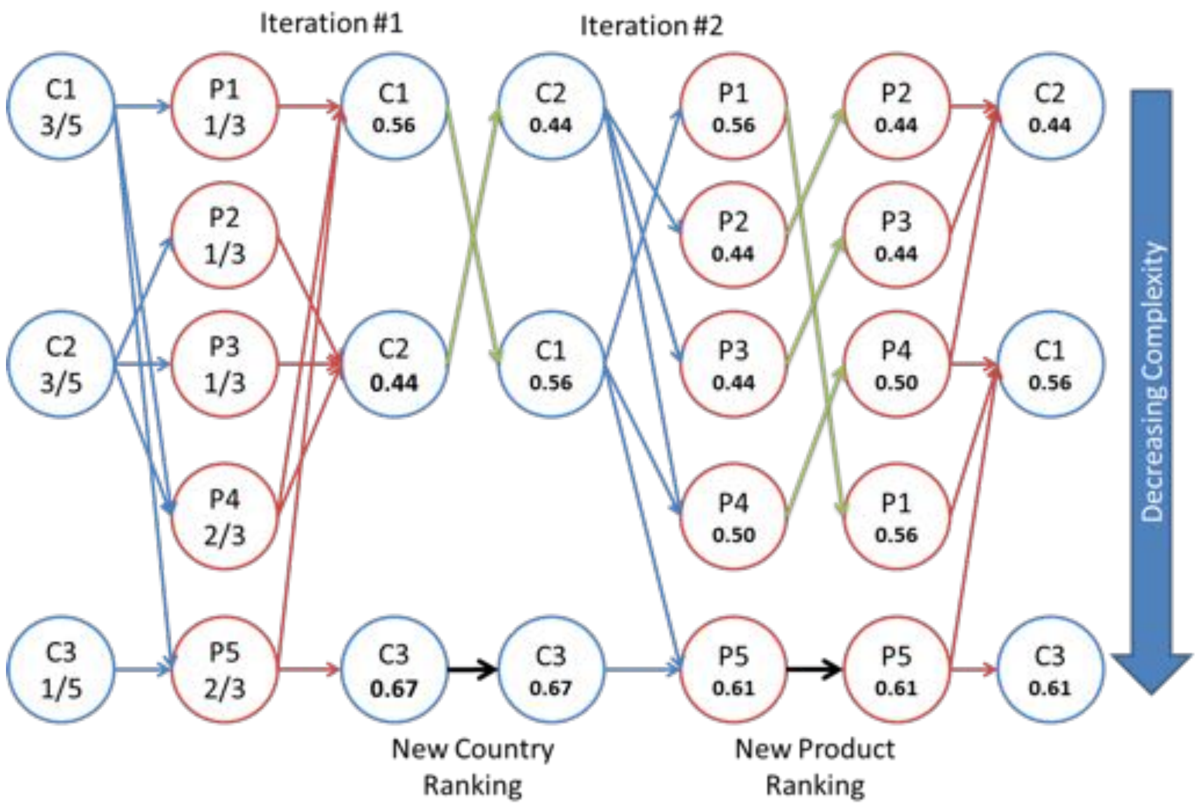
$$\left(\frac{1}{3}+\frac{2}{3}+\frac{2}{3}\right)*\left(\frac{1}{3}\right)=\frac{5}{9}\cong 0.56$$

For C2 and C3, the average product ubiquity scores are is approximately 0.44 and 0.67 respectively. Country 2 has a lower average product ubiquity, which induces a change in the country rank from C1-C2-C3 to C2-C1-C3 when ordered based on the computed average values. The reason is that Country 1 exports one non-ubiquitous product and two relatively ubiquitous products, while Country 2 exports two non-ubiquitous products and one relatively ubiquitous product. Both Country 1 and Country 2 share the same diversity score, but due to the different product ubiquity compositions, the average ubiquity is lower for Country 2 and therefore scores a lower average on the

[12] The bipartite network can be ordered in one of two ways: a column of countries followed by a column of products, or vice versa. If the bipartite network is specified as countries in the first column then country information is contained in odd iterations, and product information is contained in even iterations when using the *method of reflections*.

country complexity ranking.[13] Country 3 only exports one relatively ubiquitous product. This country is deemed less complex by the lack of diversity and the relative ubiquity of its product. The key assumption here is that on average, relative uniqueness in products is associated with a higher required levels of complexity as a precondition to export them.

Figure 2.2: Method of Reflections for 2 Iterations



Notes: 1. The first two columns of this diagram show the bipartite network observed in Figure 2.1. The diagram then demonstrates the iterative process to compute ECI and PCI on a reduced example of 3 x countries and 5 x products. **2.** Within each step of the iteration process, stage one shows the country or product node values computed through the method of reflections, and stage two considers any change in the ranking of countries or products prior to the next iteration. **3.** The final two columns show the rankings of Products and Countries with decreasing revealed measures of complexity.

Further iteration yields information relevant for product nodes, this time computing the average diversity of the average country ubiquity for each product. Here Product 1 is exported by Country 1 only, which makes it the average product ubiquity of Country 1 (~0.56). Products 2 and 3 are only exported by Country 2 and therefore exhibit

[13] The way ubiquity is coded in this example means a lower average ubiquity is when a country exports more unique products on average.

scores of (~0.44, ~0.44) respectively. Product 4 however is exported by Country 1 and Country 2 and is computed by averaging the two countries' average ubiquity scores:[14]

$$[0.44 + 0.56]\frac{1}{2} \simeq 0.50$$

Similarly for Product 5, there are two exporters and the result is averaged scores of Country 1 and Country 3. This process of iteration continues until rank convergence in both countries and products is achieved. Each odd step reveals country ranking information, and each even step reveals product-ranking information as the method reflects averages over both dimensions of the bipartite network structure. As the number of iterations gets large, the values converge towards a global fixed point. Therefore country diversity and product ubiquity information is captured in the small deviations from this global average.[15] These deviations allow for ordinal information (rank) to be captured.

2.2.2 RELATIONSHIP TO OTHER MEASURES OF DIVERSIFICATION OR CONCENTRATION

Here the method of reflections is briefly compared with the Herfindahl metric that is frequently used in economic analysis to measure country diversification. The real benefit of the method of reflections is that it integrates diversification information alongside product ubiquity information to extract more meaningful contextual information.

The Herfindahl index of concentration is defined as:

$$H = \sum_{i=1}^{P} s_i^2$$

where

s_i =share of export in product i

[14] As this method converges to a global fixed point, floating point numbers are used to distinguish rank order.

[15] A revised method has been suggested in recent literature that introduces a fitness function that does not suffer for this global convergence issue. Interested readers are referred to Tacchella et al. (2013).

While a good measure of diversification, the Herfindahl index cannot differentiate between a country that exports 20 percent wheat and 80 percent pineapples with one that exports 20 percent automobiles and 80 percent aircraft.[16] While the two aforementioned economies are equally diversified, the products that they export have vastly different characteristics requiring different levels of sophistication. This highlights a key benefit of the network approach, in that it incorporates product characteristics that are endogenously defined by the bipartite structure over all exporters.

2.2.3 RELATING TO COUNTRY CAPABILITIES

The previous section demonstrated how the network complexity measures are computed, focusing on the mechanics rather than the meaning. In this section the conceptual foundation is discussed using the dominant explanation in the literature, which uses an abstract set of country capabilities in the context of the two key parameters: diversity and ubiquity (Hausmann & Klinger 2006; Hidalgo et al. 2007; Hidalgo & Hausmann 2008; Hausmann & Hidalgo 2011, 2014). Conceptually the set of capabilities could include anything ranging from factor endowments, product-specific technological processes, supportive trade and non-trade infrastructure, institutions – in essence, it can be any factor that allows a country to be more “capable” in production.

The key assumption in the network complexity model is that countries with a larger and more diverse set of underlying capabilities are able to combine them in more complex ways, which from a combinatorial perspective, provides a greater capacity to innovate and endogenously produce other capabilities (in turn being able to produce new goods and services). From this perspective, complexity increases as countries acquire new capabilities through a process of accumulation. The unobserved set of capabilities can be measured indirectly through the analysis of the observable outcomes of this underlying process.

[16] This feature of the Herfindahl index would present more of an issue if exports exhibited a higher degree of specialisation. Given the typically observed triangular density of the M_{cp} matrix, the situation described above rarely occurs as economies that produce aircraft and automobiles tend to be more diversified on average.

So what are capabilities? Let us consider Adam Smith's famous pin factory. The factory requires a number of inputs to enable it to produce pins. First it requires land, labour, and capital. However what are really required are specific types of land, labour, and capital that are fit for purpose. A pin worker needs specific human capital to know what to do to make a pin. In addition there are many different ways one could make pins, which fundamentally relies on the way that land, labour, and capital are *organised*. Adam Smith made the case for the reorganising of production through the specialisation of labour to maximise efficiency. Capabilities can therefore include new methods of production; more modern examples include new production methods such as just in time inventory management. The atlas of complexity groups all of these factors and treats them as "knowledge that is embedded in society". Some elements in the set of capabilities are measurable, while others are not.

The literature primarily links this set of *capabilities* to country diversification and product ubiquity. These two indicators are observable measures that can be used to infer characteristics of the unobservable attributes of the set of *capabilities*.

> The diversity of a country's exports is a crude approximation of the variety of capabilities available in the country, just as the ubiquity of a product is a crude approximation of the variety of capabilities required by a product. (Hausmann & Hidalgo 2014, p. 20)

Having the capacity to produce a larger number of products is indicative of having a larger underlying set of capabilities. Ubiquity is a measure of the relative rareness of products that this capability set is able to produce as an indicator of the *type* of capabilities that are contained in each country set of capabilities. It is not these indicators in isolation that determines overall complexity, it is the *joint effect* of diversification, in addition to the type of products (as determined by ubiquity) that determine levels of complexity. The method of reflections averages over both dimensions using the bipartite network structure of trade, where diversified countries that export more unique (non-ubiquitous) products are ranked as more complex than countries that exhibit relatively concentrated exports in ubiquitous products.

The atlas of complexity explains this concept well with a few key examples that are worth repeating here to form a solid understanding of the importance of viewing diversity and ubiquity together (Hausmann & Hidalgo 2014). Consider the following two products and countries: [17]

Medical imaging devices are only made by a few countries such as Germany and the United States, which are both highly diversified exporters. Medical imaging devices are therefore complex because they are both non-ubiquitous (inferred to mean they require a novel set of capabilities to produce) and in addition are exported by countries that are inferred to have large underlying sets of capabilities due to a high degree of diversification.

Diamonds are only made by a few countries such as Botswana and Sierra Leone. Given this rarity one may conclude that diamonds' are difficult to produce, but the fundamental reason for this rarity is due to geographic endowments. Once the structures of Botswana and Sierra Leone's exports are taken into account these countries are, however, relatively concentrated in the export of resources and lack diversification. These average country characteristics can be used to discount the observed rareness.

As these examples demonstrate, the attributes of country diversification and product ubiquity in concert provide useful information regarding underlying capabilities based on context.

2.3 THE PRODUCT SPACE: A STRUCTURAL VIEW OF PRODUCT SIMILARITIES

Another strand of the economic complexity literature considers a structural view of exports by constructing a network of product relatedness called the product space. The same M_{cp} matrix can be used to define a proximity $\phi_{pp'}$ matrix that considers the cross-sectional conditional probability of co-exporting any two products $P(P_2 \mid P_1)$. Proximity is a measure of product relatedness, and is a useful transformation of the data contained in the M_{cp} matrix when exploring the links to the HO model (Chapter 3) and providing structure to study diffusion characteristics (Chapter 4).

[17] The atlas of complexity is a freely available website (https://atlas.media.mit.edu) which allows users to explore different country profiles and explore the outcome of the complexity techniques.

Proximity is computed from the minimum joint probabilities of co-exporting any two goods, Product 1 (P1) and Product 2 (P2):[18]

$$\phi_{p_1p_2} = \frac{\sum_c M_{cp_1} \bullet M_{cp_2}}{max(k_{p_1,0}, k_{p_2,0})}$$

As conditional probability is not a symmetric function (i.e. $P(P_1 \mid P_2) \neq P(P_2 \mid P_1)$), the denominator is taken to be the maximum ubiquity between Product 1 (P_1) or Product 2 (P_2).[19] This uses the most conservative conditional probability and forces the relationships to be symmetric, an important property when finding a minimum spanning tree for visualising the network.[20] Symmetry is a useful property in the network context, as it enables the construction of an undirected network, where the relationship shared between any two nodes (also known as an edge) is the same. Put more simply, proximity is derived as the number of countries that export both P_1 and P_2, divided by the total number of countries that export P_1 or P_2 (whichever is higher). The resultant square matrix takes the form:[21]

$$\phi_{pp'} = \overset{\qquad\qquad P'}{P \begin{bmatrix} 1 & \cdots & \frac{\sum_c M_{cp_1} M_{cp'_N}}{max(k_{p_1,0}, k_{p'_N,0})} \\ \vdots & \ddots & \vdots \\ \frac{\sum_c M_{cp_N} M_{cp'_N}}{max(k_{p_N,0}, k_{p'_N,0})} & \cdots & 1 \end{bmatrix}}$$

In practice proximity quantifies the main idea that if any country exports t-shirts it is more likely to export trousers than it is to export automobiles. Framed using the capabilities approach, given that t-shirts and trousers share more similar factors of

[18] Minimum conditional probabilities are used so the network remains undirected. This means that when two column vectors of cross-country product patterns are multiplied elementally the sum of joint products is divided by the maximum of either product's ubiquity (the sum of each column vector in the M_{cp} matrix).

[19] The idea of using asymmetric probabilities is discussed in Appendix G. The primary disadvantage is that it generates a directed network.

[20] In network science terminology, a symmetric similarity matrix forms an undirected network, which is advantageous in that it allows the use of many network algorithms that depend on this characteristic (such as Minimum Spanning Tree). A minimum spanning tree is a sub graph that has a minimum number of edges but ensures a complete network (where all nodes are connected).

[21] It is this matrix that is used in classifying the type of product emergence in the context of trade liberalisation in Chapter 4.

production such as labour skills (i.e. sewing), they are much more likely to be co-exported. T-shirts and automobiles, on the other hand, do not share many similar factors or capabilities. While the two products may have some overlapping factors (i.e. production of seats requires some similar capabilities as T-shirts in that stitching and sewing are both inputs to both products), automobiles are more complex products and thus require more capabilities than would be required of producing t-shirts. In this case, automobiles require both a different set of inputs and a more diverse range of required capabilities (such as higher levels of human and physical capital) to produce. Proximity is therefore a measure of similarity in the unobserved set of capabilities.

The key to *symmetric* conditional probability is that it produces high scores for products that have relatively similar ubiquity characteristics as well as being exported by similar countries. Proximity therefore serves as an indirect measure of overlap in the underlying set of capabilities. In the product space, two products that are near to each other are likely to share capabilities that are required to make each product. Products that are far away have very different sets of capabilities that are required in their underlying production processes. This idea is often best visualised as a network which intuitively captures the idea that some products are more similar to each other according to their relative positions to one another.

This constructed product space has already delivered some interesting findings. Countries that specialise in the periphery of the product space typically have lower gross domestic product per capita (GDPPC) relative to a country that is spread across the product space (Hidalgo et al. 2007).[22] In addition countries tend to diffuse locally to products that are close to other products that it currently exports.

2.3.1 EXAMPLE: 3 COUNTRIES, 5 PRODUCTS (CONTINUED)

Continuing with the three-country, five-product example used to explore measures of economic and product complexity, we can compute the proximity matrix and observe how it relates to a network representation of a product space matrix. A network diagram is adopted to explore the notion of a core and periphery structure that is observed and why it forms.

[22] The atlas of complexity provides access to numerous country product space maps. http://atlas.media.mit.edu

Recall the sorted M_{cp} matrix:

$$M_{cp} = \begin{matrix} \\ C2 \\ C1 \\ C3 \end{matrix} \begin{matrix} P2 & P3 & P4 & P1 & P5 \\ \end{matrix} \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

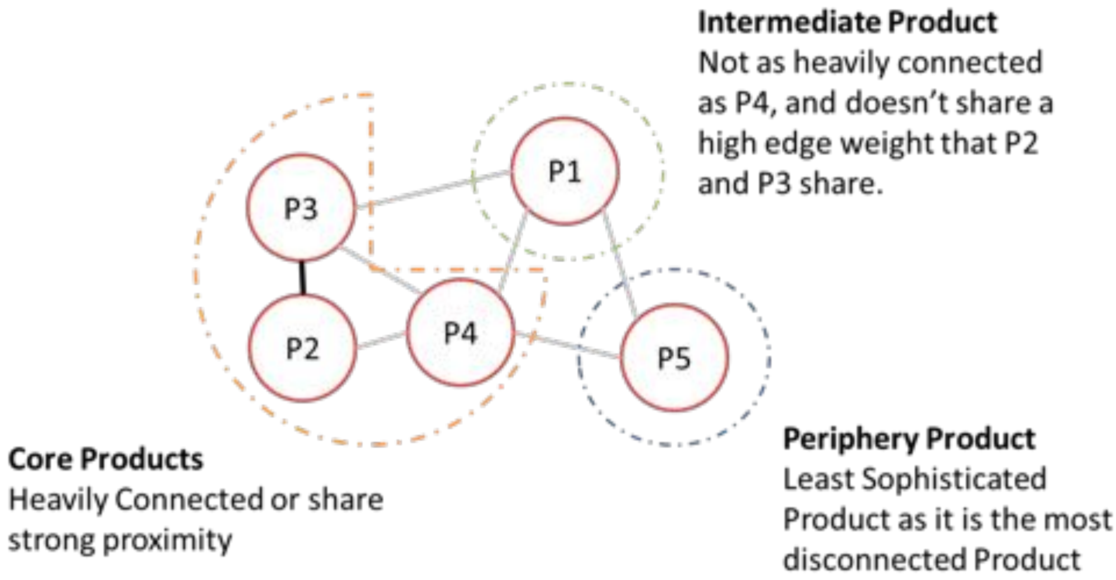
In this matrix rows are sorted by the economic complexity indicator and columns are sorted by the product complexity indicator. The proximity matrix is computed by comparing all pairwise combinations of products (columns in the matrix). Each column is pairwise multiplied and then summed to compute the total number of countries that export both products. This sum is subsequently normalised with the maximum ubiquity characteristics of either column, which is equivalent to the minimum conditional probability of co-export. The resultant matrix is:

$$\phi_{pp'} = \begin{matrix} P2 \\ P3 \\ P4 \\ P1 \\ P5 \end{matrix} \overset{\begin{matrix} P2 & P3 & P4 & P1 & P5 \end{matrix}}{\begin{bmatrix} 1 & 1 & 0.5 & 0 & 0 \\ 1 & 1 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 1 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 1 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 1 \end{bmatrix}}$$

Each product, by definition, has a perfect relationship with itself as is evident from the diagonal of the matrix consisting solely of the value one. Now consider the relationship between product 1 and product 2. The relationship is zero as when both column vectors are multiplied in the Mcp matrix there are no countries that export both products resulting in a 0 in the numerator. In the bipartite network diagram this is akin to saying that product 1 and product 2 are not connected directly through any single country traversals in the network. Product 2 and Product 3, on the other hand, share a perfect relationship. This occurs when two products share exactly the same cross-country export profile, and have the same product ubiquity characteristics. In this case, it is driven by the fact that these products are exported by a single country – a condition that is rare when considering real world data.

Figure 2.3 shows one of many possible clustered network layouts of the proximity relationships between products. [23]

Figure 2.3: Network representation of the product space



Notes: 1. Nodes in the network represent Products. **2.** Edges in the network are proximity values computed through conditional probability of co-export. **3.** This diagram only presents one possible clustering and relative layout of nodes.

What is evident is that P2, P3, and P4 are more heavily connected (based on edge weights and the total number of edges that connect them to one another) than P1 and P5. P5 is a less connected periphery product. P1 could be considered an intermediate product with relatively high connectedness but relatively low edge weights. Other clustering outcomes can also be considered and depend on the algorithms used to analyse the network. In the product space literature, random walk clustering is employed to find the 34 product space communities (Hidalgo et al. 2007).[24]

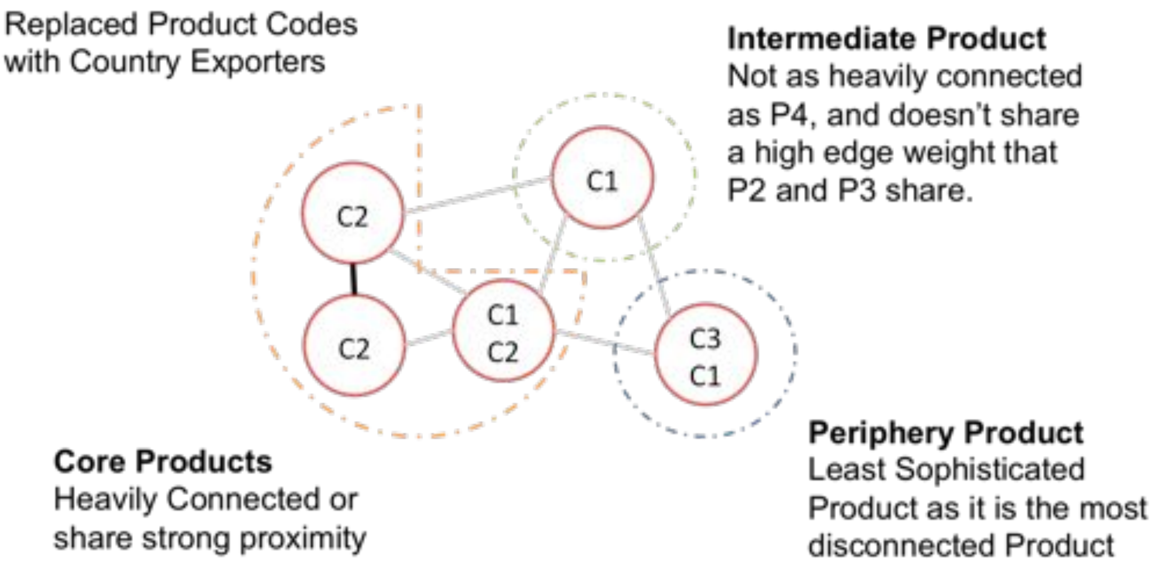
Overlaying country information within each product node identifies a country's relative position in the product space. Countries that are rated as more complex

[23] Given a network consists of relational data, there are many ways to view the actual layout of nodes. Network science uses many visualisation techniques, such as force-directed layouts to view the structure of relationships. However many of these techniques will never produce the same exact layout due to random starting conditions.

[24] Dropping random agents onto nodes in the network and allowing these agents to explore the neighbourhood of connected nodes, over time, start to form clusters that represent relatively strongly connected neighbourhoods over many repetitions.

typically occupy the densely-connected core of the product space, as illustrated in Figure 2.4. In this instance Country 3 occupies a single periphery product node (marked in blue dotted lines). Country 3 exports one product P5 that has a direct relationship with P4 and P1. The connections suggest that countries that export P5 tend to also export P1 and P4. However P5 has relatively low proximity values that connect it to P1 and P4 due to the large ubiquity effect operating through the denominator of the conditional probability. Country 2 on the other hand occupies all of the densely connected core products P2, P3, and P4 in the network and is classified as the most complex economy.

Figure 2.4: Country Information in Nodes of the Product Space



Notes: 1. This is the same diagram as Figure 2.3 except country information has been placed on each node in the network. **2.** This presents only one possible clustering and network layout.

As is demonstrated by this simple example, the product space offers a view of the data in terms of inferred product similarities. Country positions within this network are strongly correlated with varying levels of development, and allows one to consider a disaggregated structural view of the data (Hidalgo et al. 2007). As an alternative to overlaying country data on top of the product space, a country space can be computed over the country dimension of the Mcp matrix. For interested readers this is explored further in Appendix D using real world trade data.

2.4 LIMITATIONS OF FOCUSING ON EXPORTS

To conclude this discussion it is important to note some limitations of the economic complexity and product space approach, particularly with respect to its focus on export data. These limitations are summarised in the following three points (Hausmann & Klinger 2006):

1. *Uses export data rather than country production data*. This is a limitation as exports are not necessarily a full representation of what a country has the ability to produce (domestic market production and non-tradeables).

2. *Collected export data captures goods but not services*. Services are an increasingly important component of world trade, particularly as developed countries transition from production-based economies to service based economies.

3. *Export patterns are observed ex-post* and are influenced by various distortions such as tariffs and other protectionist policies, which could distort the pattern of exports which serves as an input to these data intensive techniques. (e.g. domestic agricultural subsidies in the USA and Europe have the capacity to distort export patterns).

While acknowledging the limitations there are some significant advantages to using export data. A key benefit is the ability to emphasise products that an economy is relatively “good” at producing. This is accomplished using RCA which serves as a filter of export data to focus on products that a country exports with a greater share than the average expected level (product share in world trade). This goes somewhat towards addressing issue 3. In addition, if production data were available it is not immediately apparent that it would suffer any less from the issue of domestic distortions such as state based industry retention policies or domestic subsidies. Due to the common classification of goods, international trade is the only dataset that allows for large cross-sectional comparison of countries based on the products that they export.

2.5 CONCLUSION

The purpose of this chapter is to serve as a primer for the network complexity approach that is used in the remainder of this thesis. The key benefit of these new analytical techniques is that they operate at high levels of disaggregation, which allows for a larger degree of heterogeneity when compared to aggregative approaches such as Leamer's ten types of products or Lall's technological intensity classification. The disaggregated analysis, made possible by the network complexity approach, caters to the great richness of variation we often observe between countries and allows for more detailed contextual analysis.

Through the use of a small-scale example of three countries and five products, the reasoning behind these methods was discussed in the hope to convey that many results using these techniques stem from the joint dependency on levels in country diversification and the respective product ubiquity characteristics. The method of reflections averages information contained in these two dimensions to produce a complexity ranking of countries and products. In Chapter 3 we examine the links between these complexity indicators and move our focus from these two country and product characteristics to demonstrate that these methods also form a strong complementarity with patterns of specialisation that are observed from the theory of comparative advantage. It is also shown that the *method of reflections* can be viewed as a sorting heuristic that orders countries and products in line with these macro-patterns of specialisation.

CHAPTER 3 THE THEORY OF COMPARATIVE ADVANTAGE AND PRODUCT COMPLEXITY ANALYSIS

"Economists ought to abandon the idea that models are either true or false in favour of the notion that models are sometimes useful and sometimes misleading" (Leamer, 1993)

Abstract

This chapter relates the measurement of economic complexity back to the traditional literature on international trade to show that the two approaches are complementary. The *method of reflections* can be viewed as a sorting heuristic that provides a ranking of countries and products in a many-country, many-product setting that reorders the information contained in the global pattern of exports to be in line with macro patterns of specialisation. These observed patterns of specialisation are broadly comparable to those predicted by both the Ricardian and Heckscher-Ohlin models of international trade.

3.1 INTRODUCTION

Comparative advantage is a fundamental determinant of the overall pattern of international trade. This chapter surveys the economic literature to understand the implications of this theory for cross-country patterns of exports in a many-country, many-product context. The links between the theory of comparative advantage and the recent network complexity analysis is then studied to show that they are complementary rather than competing tools. The network complexity approach is typically focused on the central idea that diversified countries which export unique products have higher levels of economic complexity while less complex economies only produce a few ubiquitous products (Hausmann & Klinger 2006; Hidalgo et al. 2007; Hausmann & Hidalgo 2014). This two-dimensional approach not only considers how many products countries make but also accounts for the type of products countries produce that is captured through the notion of product ubiquity. The mechanics of how these methods produce rankings of economic and product complexity is highly related to matrix reordering, in which columns and rows are organised in line with a pattern that is observable in the matrix data. When considered from this viewpoint, the rankings of economic and product complexity are correlated to the patterns of specialisation we would expect to observe based on the two fundamental theories of comparative advantage.

The chapter is structured as follows. The theory of comparative advantage is discussed in Section 3.2, with a focus on how it relates to the real-world context of many countries, many products, and many factors. Both the Ricardian and Heckscher-Ohlin frameworks are used to set expectations regarding the structure of the M_{cp} matrix in this high dimensionality setting. The economic complexity and product space approach is then discussed in Section 3.3. The data driven analytical techniques used in computing revealed measures of the economic complexity index (ECI) and the product complexity index (PCI) are briefly reviewed. These measurements are then framed in the primary explanatory framework of country capabilities and compared to the Ricardian and Hecksher-Ohlin models. This section then brings the two theoretical discussions together by considering their intersection at the cross-country pattern (M_{cp}) of international export data for the year 2000.[25] A key outcome of Section 3.3 is that it demonstrates how effective the *method of reflections* technique is at producing the multi-dimensional indicators of revealed economic and product complexity. In Section 3.4 a new measure of trade distortion is proposed and computed, using real-world data, to demonstrate this observed complementarity. Section 3.5 provides the main conclusions from this chapter.

3.2 THE THEORY OF COMPARATIVE ADVANTAGE

> *"Comparative advantage is the best example of an economic principle that is undeniably true yet not obvious to intelligent people" (Samuelson 1969)*

Comparative advantage, in its most basic form, is driven by country heterogeneity that results in differences in relative prices when comparing autarkic conditions with those that would prevail under free trade. This heterogeneity can be driven from a range of sources such as differences in technology, endowments, institutions, geography, market design, and the structure of demand (Ricardo 1817; Ohlin 1933; Vanek 1963;

[25] The year 2000 has been chosen as it appears in the key literature and therefore valuable for direct comparability with the key literature (Hidalgo et al. 2007). Similar diagrams for the year 2012 are available in Appendix C.

Helpman & Krugman 1985; Davis 1995; Eaton & Kortum 2002; Costinot 2009; Morrow 2010; Chor 2010; Markusen 2013).

This chapter places its focus on the two fundamental frameworks of comparative advantage, the Ricardian and Heckscher-Ohlin models of international trade. The Ricardian model formalised the concept of comparative advantage by introducing a model in which labour is the sole factor in production and heterogeneity arose due to differences in labour productivity. Other sources of heterogeneity were subsequently explored in the Heckscher-Ohlin model in which factor endowments are the prime motivations to trade. These two fundamental models have laid a rich foundation that has shaped the way we think about both the gains from trade and the generation of overall trade patterns.

The network complexity framework of analysis has been predominantly focused on the idea of country *capabilities*, a construct that ultimately provides a more general level of abstraction when considering the factors that drive comparative advantage. For example, the *set of capabilities* could represent a discrete range of different production processes, in which some processes are more efficient then others. This can be aligned with a Ricardian view of international trade. Alternatively the *set of capabilities* could include different discrete factor endowments providing a natural Hecksher-Ohlin perspective.

What is novel in framing ones thinking about *country capabilities* is the ability to incorporate more complex origins of comparative advantage. For example, capabilities as presented by network complexity analysis moves beyond thinking in aggregate constructs such as capital, land, and labour and moves towards a framework that may include many different factors of production such as product specific capital (i.e. a factory designed to make batteries and not clothing) or any other factor that may influence differences between countries on the supply side. Once the aggregate measures of capital, land, and labour become more disaggregated, this introduces the notion that some countries will have access to capabilities that other countries do not. This leads to a combinatorial perspective when thinking about factors of production and the many ways they can be combined to produce output.

3.2.1 RICARDIAN MODELS

The basic Ricardian model is motivated by differences in labour productivity across countries as the primary determinant that induces trade between countries. The setup involves two countries, two goods, and only one factor (labour) [2x2x1]. The model provides the well-known result that at least one country specialises in the production of good p, depending on relative labour productivity.[26] If country 1 is relatively more productive in product 1, case 'A' will be observed which drives a line of specialisation through the cross country pattern of trade as captured by the M_{cp} matrix. In the alternative, matrix 'B' would be observed. In matrix form, where countries are rows and products are columns, the line of specialisation is either downward (Case A) or upward (B) sloping depending on the ordering of countries and products.

Case A: $M_{cp} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ *or* **Case B:** $M_{cp} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

This model is readily expandable to a discrete many-goods case by constructing the relative demand and relative supply curves that jointly determine the pattern of specialisation between two countries [2xPx1]. Given there is only one factor the goods can be placed into an ordered set based on labour productivity in each of the two countries. With the assumption of full employment (implicitly assuming that the number of work hours is unrelated to a change in wage), the pattern of specialisation is determined by the intersection of the relative wage and the relative demand curve. Products on one side of this cut-off will be exported by one country, and the alternative set of products by the other country. Introducing trade costs into this framework can also drive a wedge between the products that border the cut-off point, rendering the possibility of some products being non-traded and produced domestically by both countries for domestic consumption.[27]

---

[26] Country 1 will export good 1 if $\frac{a_1^1}{a_2^1} < \frac{a_1^2}{a_2^2}$ where a_p^c is labour productivity.

[27] These models typically assume that all products are demanded in all countries. This is akin to saying demand is homothetic across countries. This ensures that the ratios of relative prices determine consumption patterns and is not complicated by changes in consumption patterns due to income etc.

In the two-country case, Ricardian forces lend themselves to M_{cp} matrices with the following structure (where product 3 is a non-traded example and could ultimately be removed from this export matrix):[28]

$$M_{cp}= \begin{bmatrix} 1 & 1 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 1 & 1 \end{bmatrix}$$

This framework is often used in the analysis of comparative two-country cases or aggregative studies such as North-South models.[29] Further extensions have also been applied that introduce higher dimensionality in either countries or goods through the idea of chaining relative technological intensity. A model that incorporates a continuum of goods allows one to focus on determinants of trade such as tariffs and transport costs and provides strong comparative static results in a two-country setting (Dornbusch, Fischer & Samuelson 1977). The Dornbusch model provides a good basis for understanding trade in a multi-commodity world but the results do not generalise to a many-country case. A more generalised extension to the model allows for many countries by assuming a simpler production structure (Wilson 1980). However in a multi-country, multi-product setting it might not be possible to represent the set of exported goods on a single interval, which introduces significant complexity without imposing further structure on the model. The Eaton and Kortum (2002) (EK) model provides one way forward by using a Frechet distribution (an extreme value distribution) to describe differences in productivity across countries. The EK model is a successful generalisation of a many-country, many-product Ricardian model of bilateral trade flows, incorporating a role for technology, in addition to a role for geography that impacts trade by imposing country-specific costs.[30] The use of the Frechet distribution to describe differences in cross-country productivity builds a pattern of cross-country specialisation that is in-line with the notion of perfect competition (i.e. lowest price). The emphasis is not typically placed on the specific

[28] Products are indexed by columns, and countries are indexed by rows.

[29] However country comparative studies should acknowledge that due to weak comparative advantage, the relationship across goods is correlative due to the presence of other countries in the world trading system.

[30] Geographic effects are correlated with impacts on trade derived from items such as transport costs and tariffs and quotas. This allows for country-specific influences on trade such that countries may choose to trade with closer countries due to potential costs of trading further afield. Infrastructure weighs heavily on this relationship.

cross-country pattern of trade, but rather on gains from trade through the mechanism of comparative advantage.

Full empirical tests of the Ricardian model are difficult due to the fact that ex-post differences in productivity are not observed (Deardorff 1984; Costinot & Donaldson 2012). Given the strong emphasis on specialisation in the Ricardian framework, the implicit fact that one country will specialise produces an intrinsic problem of knowing what labour productivity would be if a country were to produce the imported good that exists in the trading equilibrium. Two primary strategies have been employed in the literature to overcome these empirical limitations in testing the Ricardian model of international trade:

1. Impose structure on productivity across countries and industries (or products) to allow one to estimate the unobservable productivity levels. This is essentially the technique employed by Eaton and Kortum (2002), who use random productivity draws from the Frechet distribution.
2. Consider a single industry with known absolute productivity information. Constinot et al. (2012) consider the agriculture sector to test the Ricardian model and find that this extended Ricardian model exhibits significant explanatory power.

What all of these models have in common is the strong emphasis on specialisation. Therefore one would expect to observe a diagonal structure in the data if the M_{cp} matrix were sorted in line with country technological capability and product technological intensity.

$$M_{cp}=\begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 1 \end{bmatrix}$$

As expressed by this matrix, a strong degree of specialisation is expected, as indicated by the country-product relationship being dominated by ones along the main quasi-diagonal of the matrix.[31] One explanation for the possible presence of overlaps is due

[31] Based on the ordering of countries and products, the line of specialisation could also be upward sloping.

to differences in country size, a situation where two or more countries, with similar technology, would supply exports of a particular product to world markets. Given the Ricardian model's emphasis on a single factor, typically labour, extreme patterns of specialisation tend to be observed.[32]

This concept of *capability* can be recast into a Ricardian framework. It can be assumed that countries that are complex have accumulated a large and diverse set of capabilities. By being able to re-combine these capabilities, a combinatorial process suggests that this will lead to increasingly more sophisticated production methods. Assume, for a moment, there is a given product that has three known methods of production. A relatively less advanced country may only have acquired method 1 and it is the only method available in its set of capabilities. A relatively more complex economy however could choose to produce the good with a variety of methods 1, 2, or 3 for example.[33] Assuming that methods are sorted according to efficiency, method 3 provides a maximising choice for producers. Viewed in this way, more complex economies use relatively sophisticated production techniques but have, in theory, access to other less sophisticated methods that are not optimal production choices. In this way technological advancement can be viewed as a discrete process, where more advanced techniques dominate less sophisticated options, and the *size of capabilities* increases as countries' acquire access to new production methods.

3.2.2 HECKSCHER-OHLIN MODELS

The Hecksher-Ohlin Model places the primary focus on difference in factor endowments among countries. The fundamental HO model with two countries, two factors, and two goods (2x2x2) delivers the well-known result that a country that is abundant in a factor, exports the good whose production is relatively intensive in that factor. Therefore a capital-intensive country will tend to export capital-intensive

[32] Substitutability of factors can be explored in the Hecksher-Ohlin framework and are reviewed in the next section.

[33] This assumes that if a country is able to produce using method 2 or 3 that it has the capacity to produce using method 1, but would choose not to as it is a less productive method. It is interesting to consider that while the more complex economy may only use method 3 in production, it may be the case that method 1 and 2 are implicit in the set of capabilities (adding to the effective set size). This also lays down a potential foundation to investigate path dependence in the discovery of method 3.

products, and labour-intensive countries will tend to export labour-intensive products. Consider two countries ranked by relative endowment in capital ($(K/L)_1 > (K/L)_2$) and two products ranked by relative capital intensity (relatively high capital intensity: P_1 and relatively low capital intensity: P_2). Then C_1 will export the good that is relatively capital intensive (P_1), while C_2 will export the good that is relatively labour intensive (P_2). This is expressed in the M_{cp} matrix, where specialisation is observed along the diagonal of the matrix:

$$M_{cp} = \begin{bmatrix} (K/L)_1 \\ (K/L)_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

In the 2 Countries, 3 Goods, 3 Factors Case, assuming Country 1 is relatively labour and land abundant and Country 2 is relatively capital abundant, there will exist a sorted ordering of products by country-specific factor abundance which renders a quasi-diagonal nature to the M_{cp} Matrix. These results are possible due to the number of goods being equal to the number of factors delivering a solvable linear system of equations and the inclusion of only two countries. This is expressed in the M_{cp} matrices:

Case A: $M_{cp} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ *or* Case **B:** $M_{cp} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

A strict quasi-diagonal is not necessarily globally enforced (as is demonstrated by case A) based on the ordering of products by factor intensities. In this two-country case there will, however, exist some ordering pattern that produces a quasi-diagonal line of specialisation (as is demonstrated by case B).

Generalising the HO model to many goods, many countries, and many factors introduces predictive barriers. Starting with a three-country, two-product, and two-factor setting, exports follow a similar pattern as in the basic two-country model, with the exception that some exports can now be exported by more than one country and complete international specialisation is no longer required (Leamer 1984).[34] This approach renders a M_{cp} matrix with an ordering that produces a quasi-diagonal line with crossover as more than one country can export boundary goods.

[34] This possibility becomes fundamental when considering the Product Space, a construct that considers the conditional probability of co-exporting products. Conditional probability requires non-complete specialisation to yield results.

$$M_{cp}=\begin{bmatrix}1 & 1 & 0 & 0 & 0\\ 0 & 1 & 1 & 1 & 0\\ 0 & 0 & 0 & 1 & 1\end{bmatrix}$$

The above matrix has been sorted to demonstrate a clear specialisation pattern. More generally, sorting products by labour, capital, and land intensity could conceptually produce groups of products that exhibit local clusters with intra-group specialisation. The matrix below contains the same information as the previous matrix but exhibits block clustering with quasi-diagonal lines of specialisation by grouping capital, labour, and land intensiveness as primary factors.

$$M_{cp}=\begin{bmatrix}1 & 0 & 1 & 0 & 0\\ 0 & 1 & 1 & 0 & 1\\ 0 & 0 & 0 & 1 & 1\end{bmatrix}$$

More generalised results can be inferred from the main findings of the three-country case (Leamer 1984). The cross-country trade pattern (M_{cp}) for a generalised number of countries and products therefore may exhibit global patterns of specialisation such as:

$$M_{cp} = \begin{bmatrix}1 & 1 & 0 & \cdots & 0 & 0 & 0\cdots & 0 & 0\\ 0 & 1 & 1 & \cdots & 0 & 0 & 0\ldots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & 1 & 0 & \vdots\ \ddots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & 1 & 1 & 0\ldots & 1 & 0\\ 0 & 0 & 0 & \cdots & 0 & 0 & 0\ldots & 1 & 1\end{bmatrix}$$

However, global patterns of specialisation are dependent on the ordering of factors capital, labour, and land. Other clustered outcomes are possible that incorporate only local intra-block diagonals *without* a necessary macro semi-diagonal shape:

$$M_{cp} = \begin{bmatrix}1 & 1 & 0 & \cdots & 0 & 0 & 0\cdots & 0 & 0\\ 0 & 1 & 1 & \cdots & 0 & 0 & 0\ldots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & 0 & 0 & \vdots\ \ddots & 1 & 1\\ 0 & 0 & 0 & \cdots & 1 & 0 & 0\ldots & 1 & 1\\ 0 & 0 & 0 & \cdots & 1 & 1 & 0\ldots & 1 & 0\end{bmatrix}$$

Extending this further by incorporating factor substitution generates the possibility of inter-block relationships, as products that are relatively capital-intensive in one country may be relatively labour-intensive in another.[35] This produces relationships *between* clusters such that relatively capital-intensive countries using a different mix of factors may produce some products that are categorised as labour-intensive in other

[35] This becomes increasingly more important at higher relative levels of product aggregation

countries. Wheat, for example, is produced by a diverse range of countries in which some farmers produce much higher quantities due to capital intensity such as tractors. This product can be produced in different ways, which drives off-diagonal terms in the cross-country trade pattern.

The capacity for substitution therefore reduces the tendency for specialisation to generate smooth diagonal landscapes by incorporating potentially distant off-diagonal elements such as:[36]

$$M_{cp} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 & 0 & \ldots & [\mathbf{1}] & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 & 0 & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 & 0 & \ldots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \ldots & 1 & 1 \end{bmatrix}$$
$$\qquad capital \qquad\qquad land \qquad\qquad labour$$

The introduction of off-diagonal elements in the M_{cp} matrix is also supported by the specific factor model (Neary 1978; Hausmann & Klinger 2006). A key assumption in much of the theory discussed thus far is that all countries have the capacity to produce all goods in autarky. Production varies due to differences in the size of endowments and the effect that it has on prices. Therefore the Hecksher-Ohlin-Samuelson (H.O.S) model is effectively a long-run model. The specific factors model, on the other hand, imposes restrictions on the mobility of capital in the short and medium terms (Neary 1978). This drives different rental rates and price responses when comparing short run with long run outcomes.

Taking the specific factor model to the extreme, some countries lack access to a specific factor entirely. For example, oil production first requires the existence of oil endowments, but in addition requires complementary sector-specific capital to extract oil from its reserves. The capital used to extract oil is not the same as the capital required in the production of footwear, imposing stickiness in the mobility of capital. Short- to medium-term effects therefore can drive heterogeneity of country export vectors as it takes time for production to be reorganised. This also provides a role for international capital mobility in forming the observed heterogeneity in country

[36] The actual degree of off-diagonal relationship depends on the sorted ordering of the matrix according to different factor intensity

exports. Countries with oil endowments but which lack the capital to extract oil may engage with foreign firms to acquire this capital.

The main implications of this discussion around specific factors, is that it takes time for trade patterns to be reorganised in line with comparative advantage. Products that are in the off-diagonal positions in any given year may be exposed to a dynamic process as underlying resources in the economy are reallocated through the forces of general equilibrium. Romalis (2004) also demonstrates that a dynamic process of rapid factor accumulation systematically causes a shift towards industries that make intensive use of that factor. International capital mobility may drive off-diagonal elements by providing product specific capital to countries that would otherwise lack the capacity to produce these products.

It is clear that the HO framework quickly becomes complex in the many-country, many-product, many-factor setting. Thompson (2001) provides a clear study of the case of three factors, goods or countries. The case is made that the intuition provided by the two-country case is convenient but “leads to incomplete intuition”, and similarly to Leamer, makes the case that many complexities introduced by higher dimensional models can be learnt from studying the 3 country, product and factor case. Thompson’s analysis then focuses on a combinatorial perspective when thinking about the excess supply matrix.

$$\begin{bmatrix} X_{1A} & X_{2A} & X_{3A} \\ X_{1B} & X_{2B} & X_{3B} \\ X_{1C} & X_{2C} & X_{3C} \end{bmatrix}$$

Thompson shows that if all countries participate in trade, then we know that each country must export at least one good, which generates a main diagonal that is positive. Further, if one assumes that no two countries share the same trade pattern (excluding cohorts), there are four fundamental trade patterns. While each pattern (a) to (d) contains six possible arrangements, they can all be arranged in such a way that positive signs are clustered around the diagonal.

$$\begin{matrix} + & - & - \\ - & + & - \\ - & - & + \\ & (a) & \end{matrix} \qquad \begin{matrix} + & - & - \\ - & + & - \\ + & - & + \\ & (b) & \end{matrix} \qquad \begin{matrix} + & + & - \\ - & + & - \\ + & - & + \\ & (c) & \end{matrix} \qquad \begin{matrix} + & + & - \\ - & + & + \\ - & - & + \\ & (d) & \end{matrix}$$

In summary, the HO framework predominantly predicts that there is an ordering of countries and products such that a macro line of specialisation will exist. However it does introduce the potential for block-diagonal clusters in addition to distant off-diagonal relationships to be observed in the M_{cp} matrix between relative factor groups. Some products can span multiple factors based on different production functions, and sector specific endowments can drive heterogeneity in export outcomes. As is demonstrated in the 3 country, goods, and factors case, expectations of cross-country patterns tend to be clustered around a diagonal line of specialisation.

By considering a discrete range of factors of production, the interpretation of the HO can apply directly to the country *capabilities* framework, where factors of production are entries in the *capability* set (i.e. an oil endowment is a first step requirement for the production of oil). The capabilities framework implies that factors are more disaggregated into industry or even product specific factors that can encompass more than just the aggregate inputs into production. While all countries have some level of capital, land, and labour, having a more disaggregate factor space will lead to the scenario whereby one country has the capital that allows for the production of a certain good (i.e. oil drilling platforms in the production of oil), while another country will not necessarily have access to that capital to extract their reserves.

3.2.3 KEY OBSERVATIONS

The main implications from this discussion are that the Ricardian model predicts strong patterns of specialisation across countries with distinct diagonal, or quasi-diagonal, patterns observed in the M_{cp} matrix.[37] This outcome occurs due to the implicit sorting of countries and products of the underlying model, whereby countries tend to specialise.

Patterns that are observed in the HO model, in part, depend on the sorting of the M_{cp} matrix. There is however an ordering of countries and products that forms along a macro diagonal line of specialisation, with the possibility of off-diagonal components due to factor substitution or short run dynamics. Block clustering is also possible

[37] A quasi-diagonal pattern forms when the number of countries is not equal to the number of products rendering a rectangular matrix rather than a square matrix.

driven by groups of products that share a given primary factor intensity. Aggregation, product differentiation, and quality differences in products within a specific SITC or HS product group can also offer an additional explanation of off-diagonal patterns in the M_{cp} matrix.

3.3 THE THEORY OF COMPARATIVE ADVANTAGE AND THE NETWORK STRUCTURE OF TRADE

The section considers the top down view and uses international trade data to empirically observe patterns of specialisation across countries using a number of various core economic concepts such as physical capital, land, and labour. This is achieved by considering a recent UNCTAD study which uses a method similar to the method of reflections to compute revealed measures of physical capital intensity for each traded product from a dataset of physical capital endowments in each country. This UNCTAD dataset is computed for the SITC rev 2 classification.

Attention is then turned to the measures of economic (ECI) and product (PCI) complexity that serves as a useful multi-dimensional measure. The data used in this chapter are briefly outlined, and then different country and product characteristics are used to sort the M_{cp} matrix to observe overall patterns.

3.3.1 DATA

Export data for the year 2000 are provided by the NBER World Trade Flows dataset (Feenstra et al. 2005). The year 2000 is used for two primary reasons. First, the broad macro patterns are fairly consistent for all other years between 1962 and 2012; second, this year is the one that is used in the primary literature and therefore the results can be directly compared.

Additional data sources include:

1. GDP per capita from the World Development Indicators (World Bank 2015).
2. UNCTAD revealed dataset on Physical Capital, Human Capital, and Land Intensity (Shirotori, Tumurchudur & Cadot 2010).

3. Product classifications that group products based on technological intensity (Lall 2000), as well as the classification by Leamer categorising ten primary categories of products (Leamer 1984).

3.3.2 THE PATTERN OF SPECIALISATION

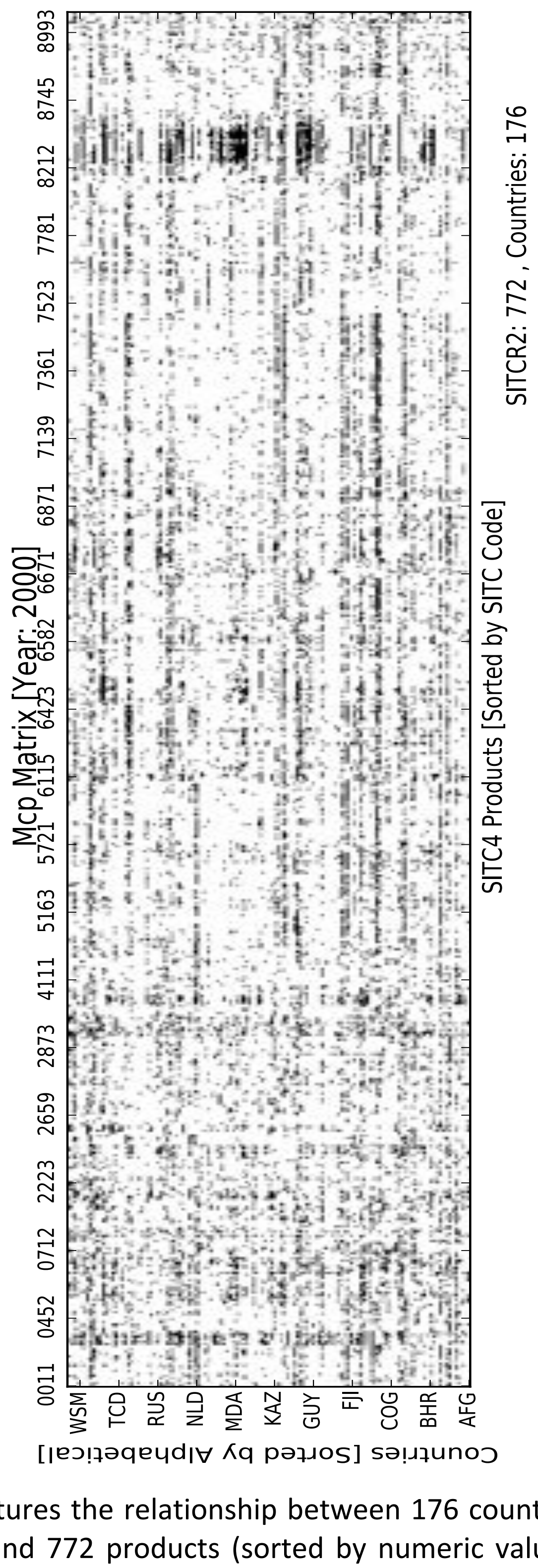
The structure of the M_{cp} matrix is interesting to consider in its own right. In Figure 3.1, the countries are sorted alphabetically and products are ordered numerically according to the SITC revision 2 level 4 product codes. This does not paint a particularly meaningful picture as countries and products are ordered in relatively arbitrary ways. Despite this one could start to observe some basic patterns in the data. For instance, one can start to observe that some products are exported by many countries, such as SITC chapter 84 (apparel and clothing). Countries that export one product within Chapter 84 tend to also export other products within the same category.

Sorting the M_{cp} matrix rows by country diversification and columns by product ubiquity characteristics, results in an M_{cp} matrix that exhibits a striking macro pattern, as shown in Figure 3.2.[38] First, diversified countries on average tend to be more developed than the less diversified countries in the bottom quartile of the rows.[39] Second, diversified countries (top of the diagram) tend to also export less ubiquitous products (right hand side of the diagram) on average. Third, countries that are more diversified (top of the diagram) also appear to export many of the same ubiquitous products (left hand side of the diagram) that less diversified countries export.

[38] The countries and product codes that are labelled in the diagram are representative product codes chosen at random with even spacing. The dimension of this diagram is 176 countries and 772 products. As you move up the y-axis countries are becoming more diversified while moving along the x-axis products are becoming less ubiquitous.

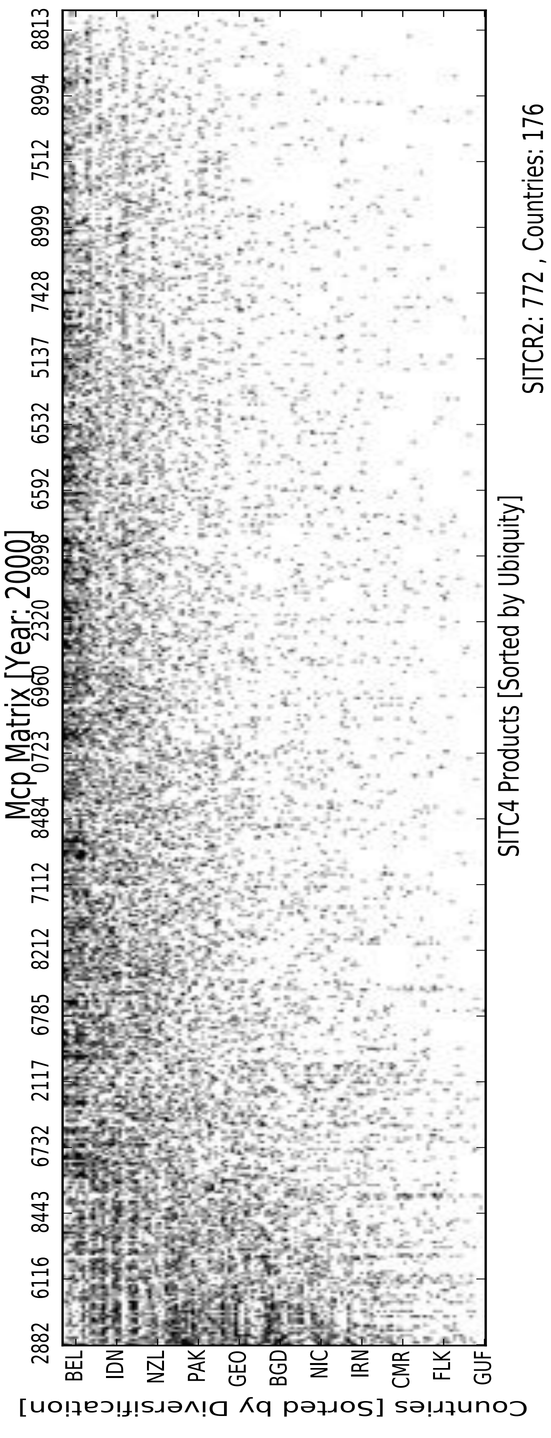
[39] Upon closer inspection country size clearly plays a role, as Indonesia (IDN) is more diversified than New Zealand (NZL).

Figure 3.1: Basic M_{cp} Matrix



Notes: 1. This captures the relationship between 176 countries (alphabetically sorted along the y-axis) and 772 products (sorted by numeric value along the x-axis). Black dots are products exported by that country with Balassa RCA index values greater than or equal to one. **2.** The data contained in this matrix are computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. **3.** Only a few countries and products are identified based on a specified gap to allow a sub sample of labels to be visible.

Figure 3.2: M_{cp} for the Year 2000 sorted by country diversity and product ubiquity



Notes: 1. This captures the relationship between 176 countries (sorted by the *increasing level of country diversification along the y-axis*) and 772 products (sorted by the *decreasing level of product ubiquity along the x-axis*). Black dots are products exported by that country with Balassa RCA index values greater than or equal to one. **2.** The data contained in this matrix is computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. **3.** Only a few countries and products are identified based on a specified gap to allow a sub sample of labels to be visible.

The observed triangular density shape is similarly observed in all other years and forms the following stylised fact.

> ***Stylised Fact:*** The M_{cp} matrix is typically comprised of a dense triangular shape when sorted by country diversity and product ubiquity. Diversified countries tend to export both common (high ubiquity) products and more unique products (low ubiquity).[40]

It is this emphasis on diversification and ubiquity that is most often considered in the literature to date (Hausmann & Hidalgo 2014; Cristelli, Tacchella & Pietronero 2014). By sorting the matrix in this way, trade data appears to strongly deviate from the diagonal patterns of specialisation that are predicted by the traditional theories of comparative advantage.

One plausible explanation for this pattern could be that as countries diversify the set of capabilities grows, leading to newer production techniques that enables the continued production of less complex products in increasingly more efficient ways (i.e. through capital substitution for labour). Improvements in productivity release resources that can be then be used for other production tasks. Therefore, this process allows for the manufacture of new goods, while retaining the capacity to manufacture existing goods. This is a *competition* view of trade that is not consistent with the theory of comparative advantage, and raises the importance of the dynamic dimension. A country with an absolute advantage in the production of all goods still has an opportunity to benefit from trade through the global reallocation of resources. A country should move out of products it holds with a *relative disadvantage*.

The dynamic element, highlighted by the specific-factors model, is also an important consideration. Not only are there restrictions on specific factor mobility, firms invest with potentially large fixed costs and will retain an incentive to remain in business over time, despite the forces of comparative advantage potentially working against them.[41]

---

[40] Highly diversified countries (at the top-left of the matrix view) continue to export goods that are highly ubiquitous, in addition to goods that are unique.

[41] This endogenous process is not only influenced by productivity considerations, but also through domestic and international protection mechanisms, and the political economy of lobbying.

This dynamic story provides incentive for firms to innovate and change production methods as underlying economic forces change.

Yet another alternative is that improvements in productivity may differentiate products within a single product category, in which higher quality products are exported by more advanced economies. Aggregation in the reported data may make differentiated products appear to be the same when in-fact they are not.

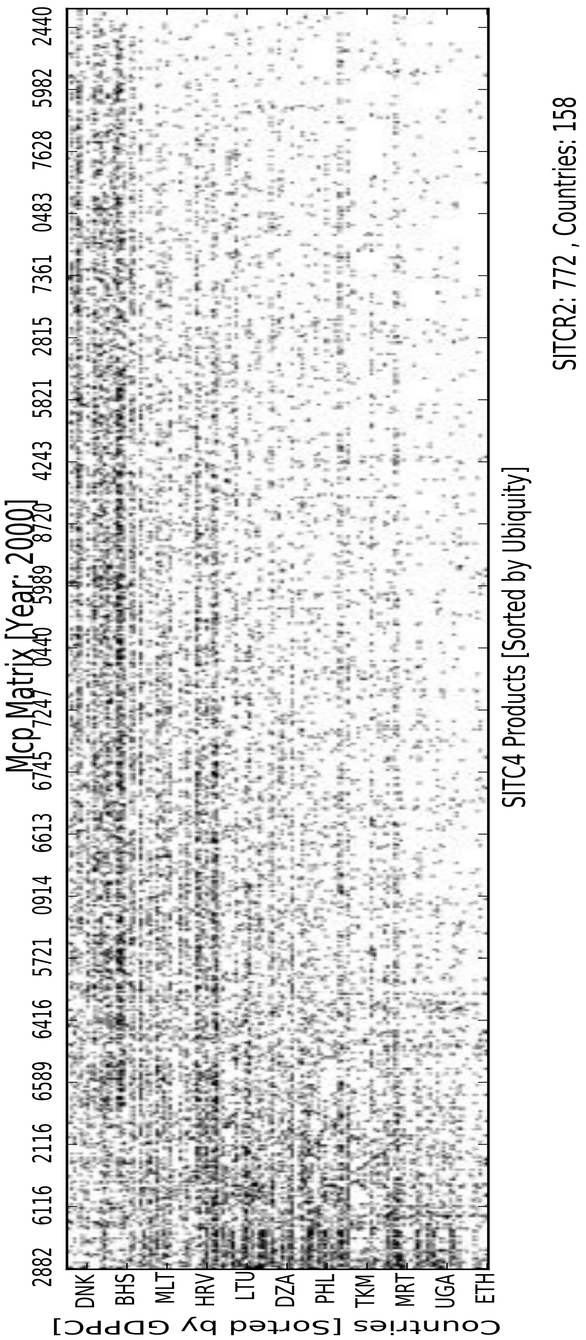
It is also possible that the observed triangular density is a result of the chosen ordering of countries and products, which provides an incomplete picture. For example, country size is going to distort the measure of diversification. Indonesia (IDN) has a higher diversification level than New Zealand (NZL), despite NZL having a much larger income level. Rows can be sorted by country-specific attributes and columns can be sorted by product-specific attributes of which there are many possible dimensions one could consider.

To compare more directly with the expected theoretical pattern, a number of different sorting arrangements are now presented in the following sections with respect to some of the conventional determinants of international trade (such as physical capital endowment and product capital intensity). By taking one dimension at a time, we can observe how well the information within the matrix conforms to a pattern of specialisation. From both the Ricardian and Heckscher-Ohlin viewpoints, one would have expected an ordering of countries and products to be centred on the diagonal of the M_{cp} matrix of specialisation. The HO model offers some explanation for off diagonal elements, but not to the degree that is observed in the data.

3.3.2.1 LEVEL OF DEVELOPMENT (GDPPC) AS A PROXY FOR ECONOMIC CAPABILITIES

Figure 3.3 takes an initial step by sorting the countries by GDP per capita (GDPPC). Output per capita, considered as a proxy for the level of economic development, is closely associated with other trade determinants such as the size of the physical and human capital stock. The effect of this ordering of countries is to demote a number of countries in the vertical axes, which tends to decrease the triangular density that has been observed as a stylised fact. High GDPPC countries continue to export both ubiquitous and less ubiquitous products, while middle and low-income countries tend to export only ubiquitous products, on average.

Figure 3.3: M_{cp} for the Year 2000, Sorted by GDPPC and Product Ubiquity



Notes: 1. This captures the relationship between 158 countries (sorted by the *increasing level of GDP per capita along the y-axis*) and 772 products (sorted by the *decreasing level of product ubiquity along the x-axis*). Black dots are products exported by that country with Balassa RCA index values greater than or equal to one. **2.** The data contained in this matrix are computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. **3.** Only a few countries and products are identified based on a specified gap to allow a sub sample of labels to be visible.

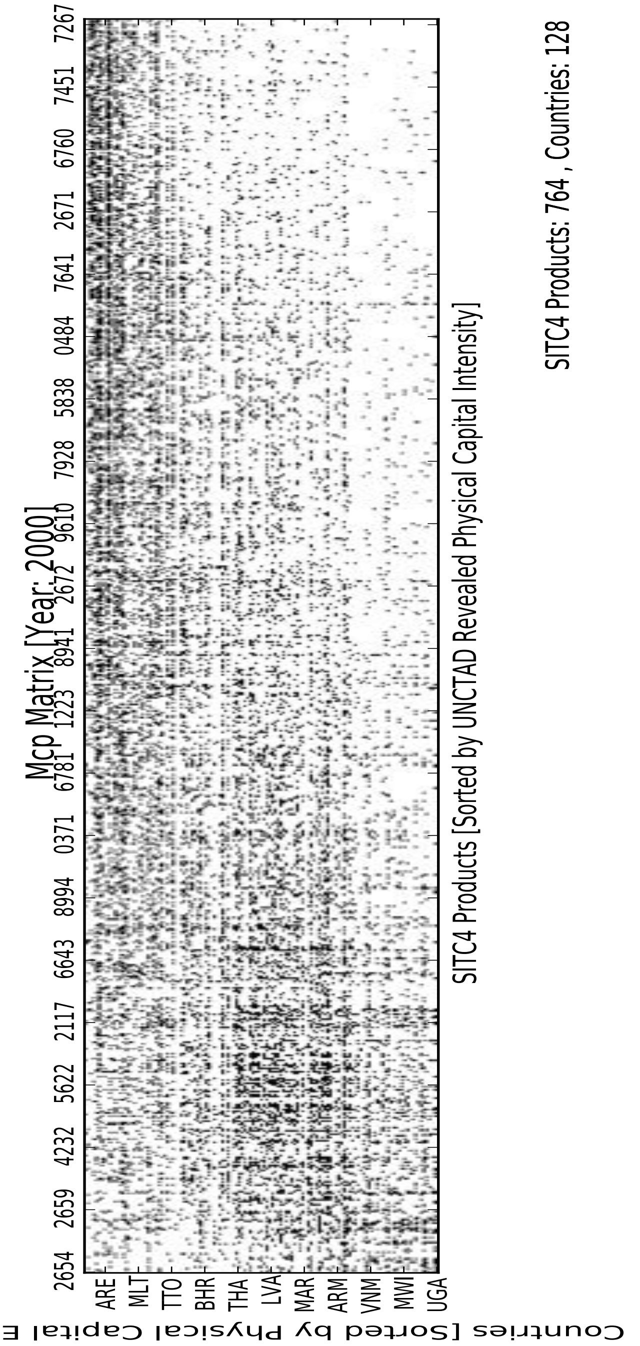
3.3.2.2 REVEALED FACTOR ENDOWMENTS

UNCTAD released a dataset that considers country endowments of physical capital, human capital, and arable land, which have been computed at the product level, and followed a similar procedure to the method of reflections. Products that exhibit high physical capital intensity are products exported by countries with high endowments of physical capital, averaged over the bipartite network structure. The revealed factor intensity dataset is used, in this sub-section, to sort the countries by endowment stock and products by revealed product intensity. Figure 3.4 considers the physical capital dimension. Countries are sorted by *increasing* physical capital stock and products are sorted by *increasing* revealed capital intensity. The pattern of trade moves closer to the direction of clustered exports around a quasi-diagonal, albeit with a lot of noise. Two dominant clusters can be observed where, on average, high physical capital countries tend to export products with relatively high physical capital intensity, a proposition predicted by the Hecksher-Ohlin model of trade. There are still however a lot of high capital intensity countries exporting products of relatively low capital intensity.

Figure 3.5 tells a similar but less convincing story when sorting countries solely by human capital endowment and products by revealed human capital intensity. Figure 3.6 considers the land endowment dimension. Land is evidently not a strong determinant on the macro pattern of trade with an even dispersion and relatively few clusters observable in the diagram. This is not overly surprising as land is not a binding constraint except perhaps in agricultural and some resource production. Land intensity would be a better predictor within that single industry.

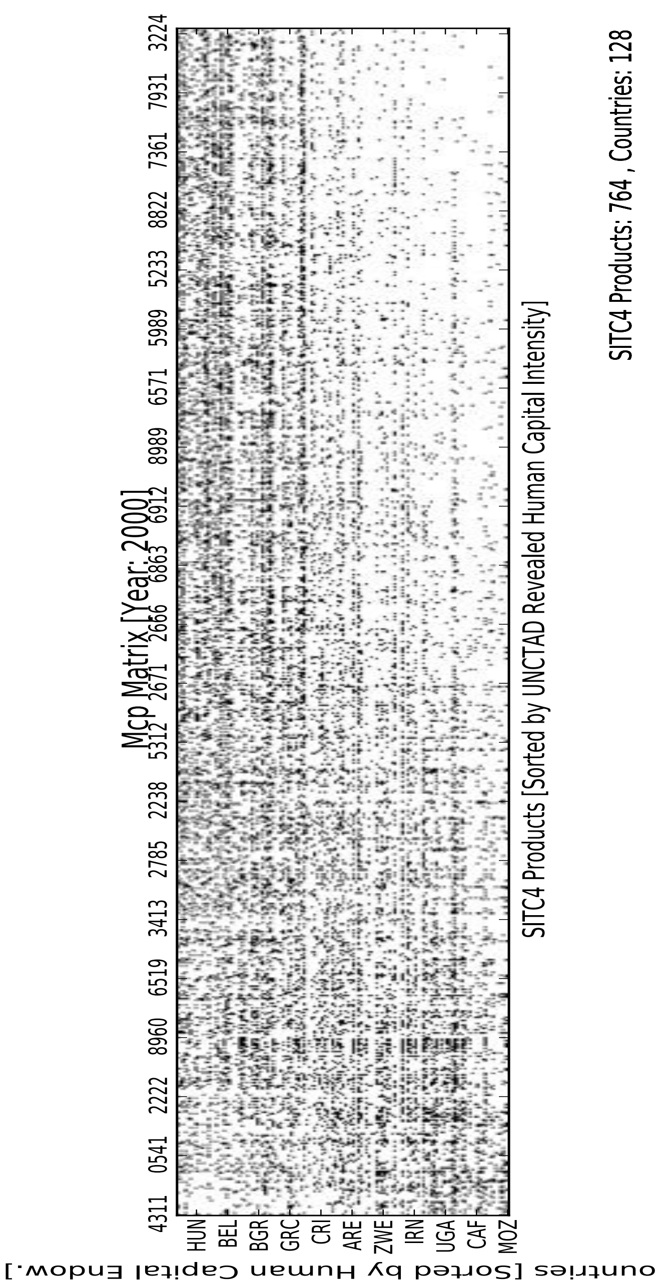
The key limitation to visualising this data is that only two dimensions of the factor space can be considered at a time. Physical and human capital have broad, but noisy, patterns of specialisation, while land provides little in terms of overall macro-level trade patterns.

Figure 3.4: M_{cp} for the Year 2000, Sorted by physical capital stock and capital intensity



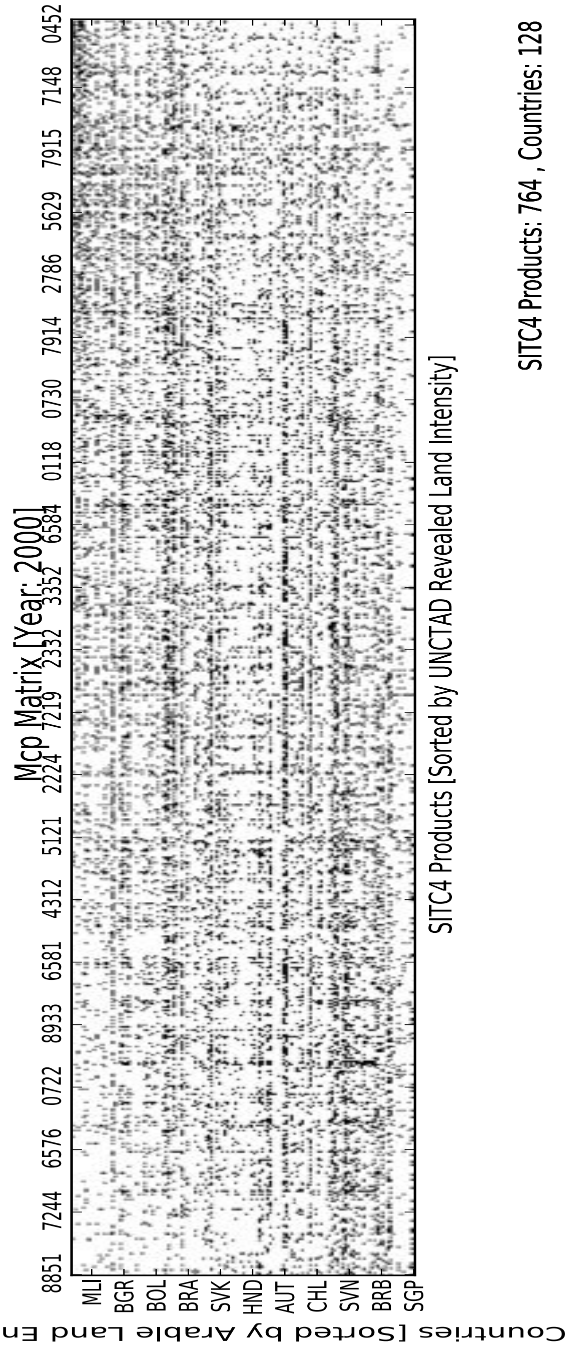
Notes: 1. This captures the relationship between 128 countries (sorted by the increasing level of physical capital endowment along the y-axis) and 764 products (sorted by the decreasing level of revealed physical capital intensity along the x-axis), Black dots are products exported by that country with Balassa RCA index values greater than or equal to one. **2.** The data contained in this matrix are computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. The revealed measures of physical capital endowments and intensity are from Shirotori et al. (2010). **3.** Only a few countries and products are identified based on a specified gap to allow a sub sample of labels to be visible.

Figure 3.5: M_{cp} for the Year 2000 Sorted by Human Capital Stock and Human Capital Intensity



Notes: 1. This captures the relationship between 128 countries (sorted by the increasing level of human capital endowment along the y-axis) and 764 products (sorted by the decreasing level of revealed human capital intensity along the x-axis), Black dots are products exported by that country with Balassa RCA index values greater than or equal to one. **2.** The data contained in this matrix is computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. The revealed measures of human capital endowments and intensity are from Shirotori et al. (2010). **3.** Only a few countries and products are identified based on a specified gap to allow a sub sample of labels to be visible.

Figure 3.6: M_{cp} for the Year 2000 Sorted by Land Endowment and Land Intensity



Notes: 1. This captures the relationship between 128 countries (sorted by the *increasing level of arable land endowment along the y-axis*) and 764 products (sorted by the *decreasing level of revealed arable land intensity along the x-axis*). **2**. The data contained in this matrix is computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. The revealed measures of physical capital endowments and intensity are from Shirotori et al. (2010). **3.** Only a few countries and products are identified based on a specified gap to allow a sub sample of labels to be visible.

3.3.2.3 REVEALED ECONOMIC COMPLEXITY (ECI) AND PRODUCT COMPLEXITY (PCI)

Now we consider the ECI and PCI measures of economic complexity. Figure 3.7 shows the M_{cp} matrix with countries sorted in terms of their revealed economic complexity, and products sorted by their revealed level of complexity. What is observed is a degree of decompression moving away from the stylised fact of a dense triangular relationship towards a thick but highly skewed band along a quasi-diagonal. There is comparatively more white space visible in the upper left hand quadrant. A line of best fit would bend from the lower left of the figure to the upper right.

In the diagrams considered so far, the height and width of each row and column is somewhat arbitrary given that countries vary significantly by size and products vary significantly by value. Currently each country and product receives the same weight in observing each element of the matrix. To obtain a clearer perspective, the row height can be normalised by country size (using country share in world trade), and the row width can be normalised by the product size (using product share in world trade).[42] This procedure renders a non-dimensional square matrix, where each axis varies between zero and one, and represents the cumulative shares of world trade. This is shown in Figure 3.8.

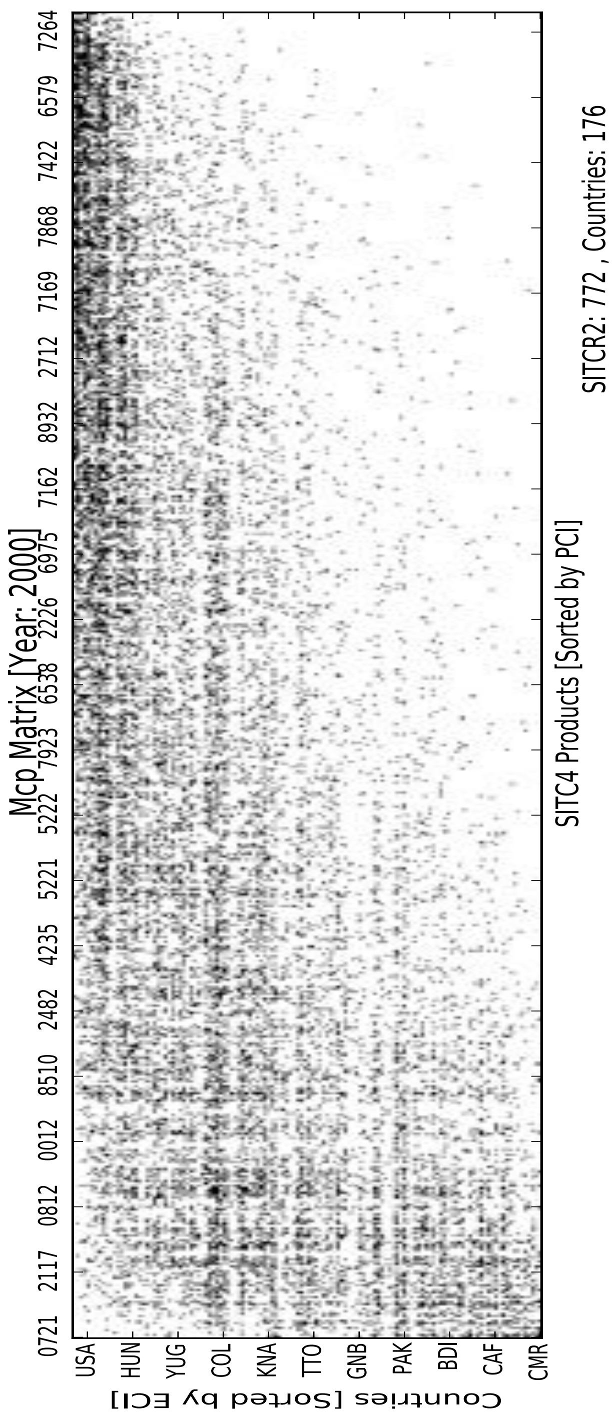
The rows of this matrix are now sorted by ECI and scaled by total country export shares. The columns are sorted by PCI and scaled by total product export shares. Through this non-dimensional transformation, a clustering around the diagonal of the matrix can now be observed, coupled with a very large degree of within and between-country heterogeneity.[43] Countries deemed economically complex, such as Japan (JPN), tend to export products that are on average relatively more complex. Such products include Automobiles (7810) and Specialty Measuring Equipment (8748). Products that are deemed relatively less complex include items such as Crude Oil

[42] The recent release of trade in value added data from the OECD may be another interesting avenue to explore particularly for the last 10 years given the growth in product fragmentation. It is suggested that other features, such as population size, are positively correlated with export share. Value scaling adjusts the column size due to variation in both prices and quantities.

[43] This observed heterogeneity within country vectors may in part be contributed by the constant fixed width scaling applied to the row heights and column widths.

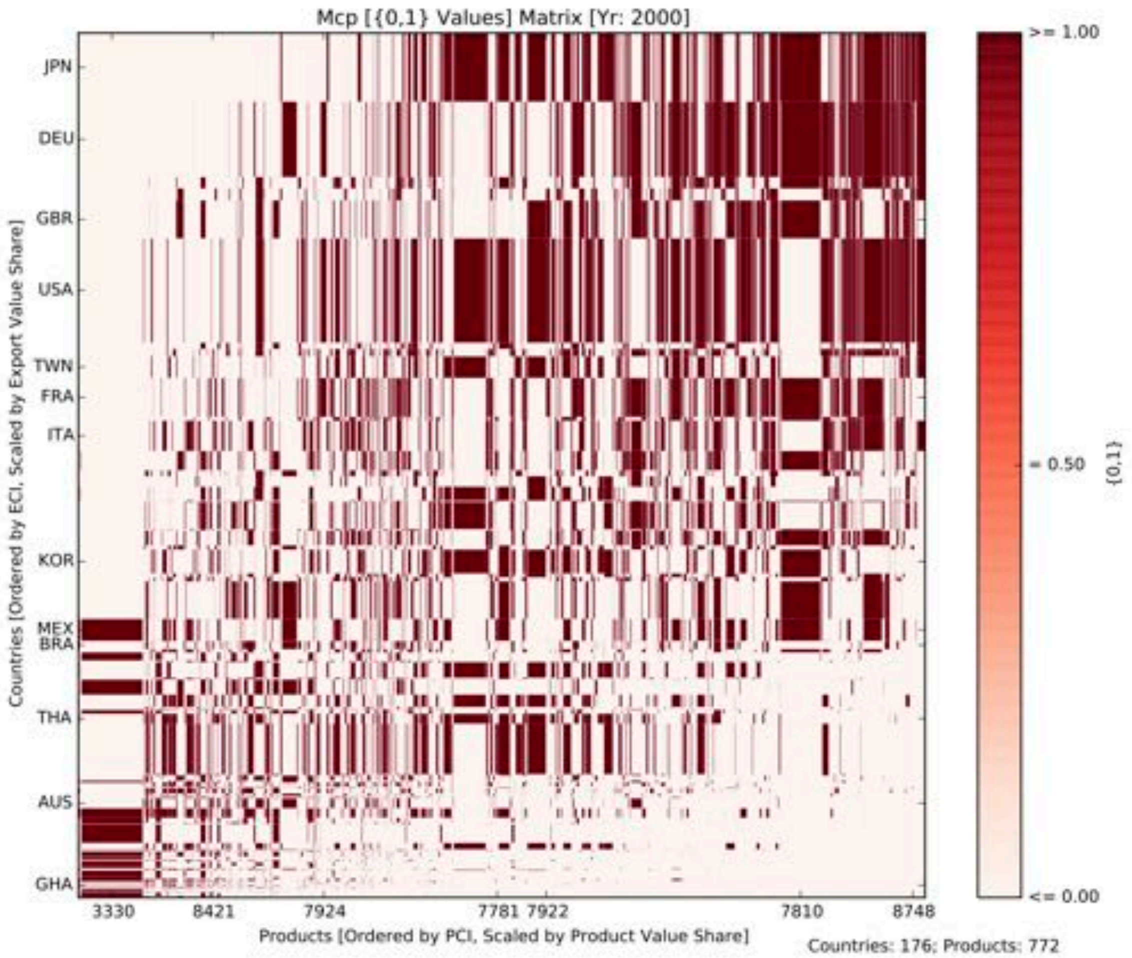
(3330) and Overcoats (8421). A broad macro pattern of specialisation can now be observed.

Figure 3.7: M_{cp} for the Year 2000 Sorted by ECI and PCI



Notes: 1. This captures the relationship between 176 countries (sorted by the increasing rank of economic complexity indicator along the y-axis) and 772 products (sorted by the decreasing rank of product complexity indicator along the x-axis). **2.** The data contained in this matrix are computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. **3.** Only a few countries and products are identified based on a specified gap to allow a sub sample of labels to be visible.

Figure 3.8: M_{cp} for the Year 2000, Scaled by Country Share in World Trade and Product Share in World Trade



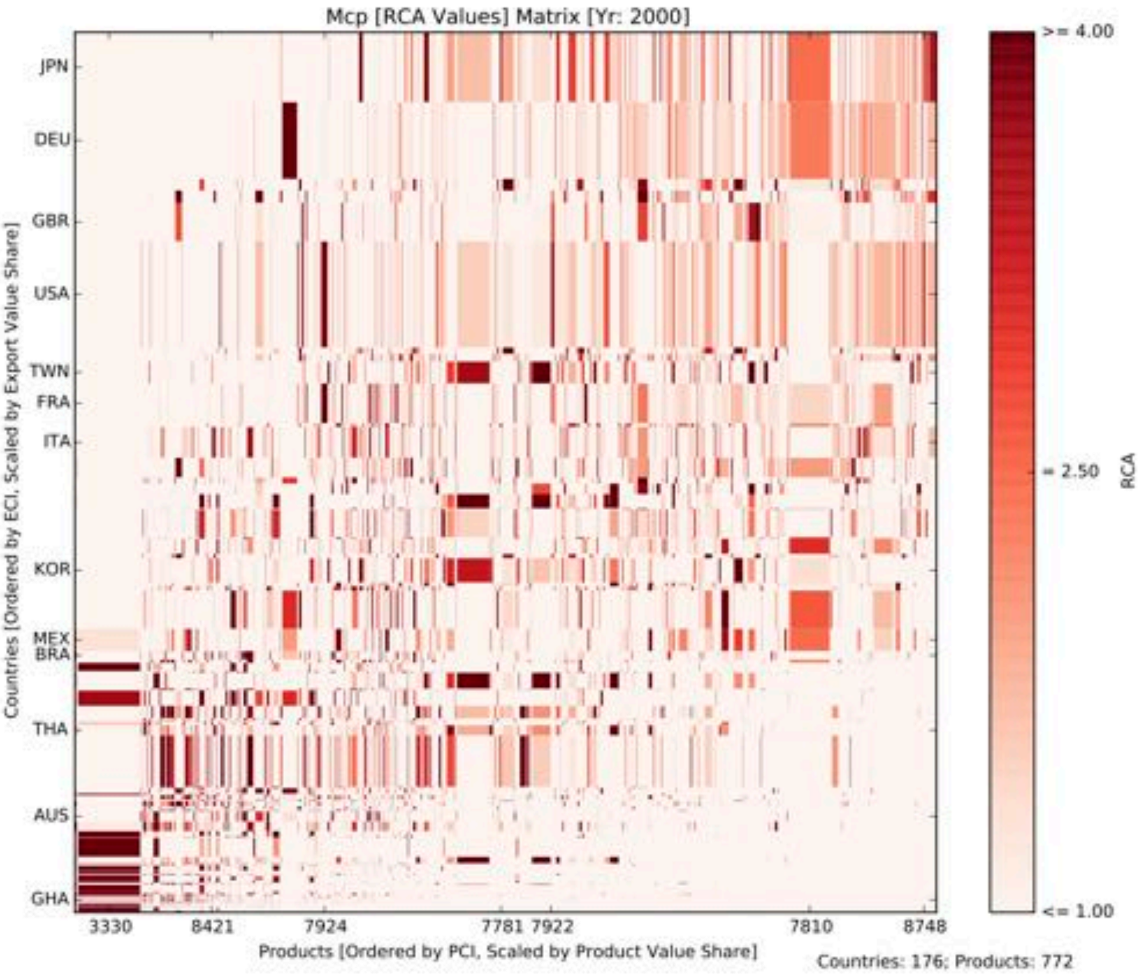
Notes: 1. Using ECI and PCI along with value scaling, a line of specialisation can be observed in the data, albeit with a very large degree of heterogeneity. **2.** M_{cp} is the {0,1} relationship between countries and products that countries export with Balassa RCA. **3.** This captures the relationship between 176 countries (sorted by the *increasing rank of economic complexity indicator along the y-axis, and each row is scaled by country export share*) and 772 products (sorted by the *decreasing rank of product complexity indicator along the x-axis, and each column is scaled by the product share*). **4.** The data contained in this matrix are computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. **5.** Only a few countries and products are identified based on a specified gap to allow a sub sample of labels to be visible.

The same diagram can be overlayed with the intensity information provided by RCA values rather than the simple binary relationships of the M_{cp} matrix. In Figure 3.9 the heat map is constructed using an upper bound cut-off value of 4, which is represented by the dark red RCA intensity. There are many stories to be told from this diagram. Oil (3330) and Automobiles (7810) together hold large shares in total world trade by value as evident by the width of these columns. Many less sophisticated countries have high RCA values in Oil (3330). This is in part because countries that export oil tend to have highly concentrated exports, which demotes them, along with the product in the complexity rankings. Some countries have darker RCA blocks that are near to the diagonal line of specialisation, but this is not always the case with large RCA values appearing, in some cases, a significant distance from a country's relative position.

Taking a horizontal slice through the matrix allows the inspection of how RCA values vary when products are ordered by PCI in a given country. As an example, Germany's RCA product vector, which has been sorted by PCI, is shown in Figure 3.10. Germany's rank (2) in the ECI index indicates that it has a relative advantage in the export of relatively *complex* products. This figure shows that Germany tends to not only export relatively more complex products, as is evident by the right-skewed density distribution of RCA values, but Germany also tends to export more complex products with *increasing* RCA intensity values on average.[44]

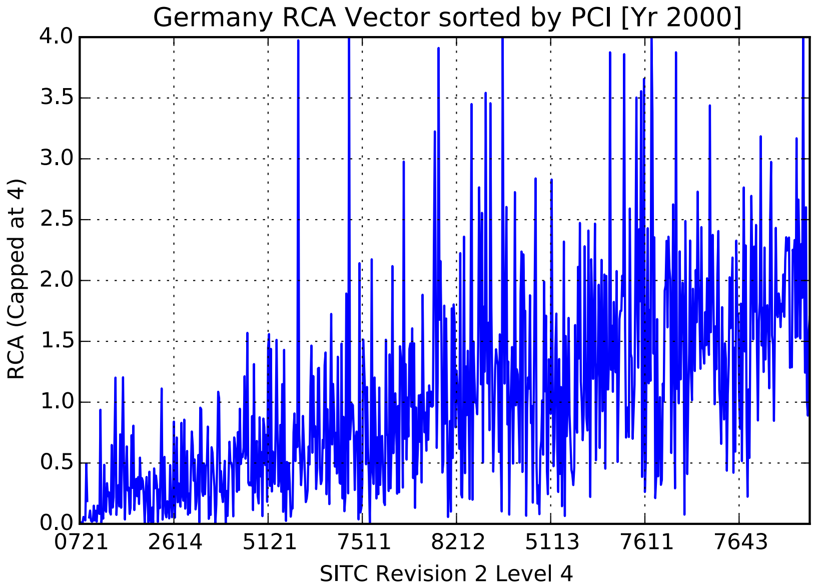
[44] The main outlier for DEU consisting of a large rectangular cell that is a long way to the left of its relative position along the diagonal is 9310 – Special transactions, nes.

Figure 3.9: RCA for the Year 2000, Scaled by Country Share in World Trade and Product Share in World Trade



Notes: 1. This tells a similar story to Figure 3.8, however is overlayed with intensity of RCA. It is notable that while many dark red cells are observed close to the line of specialisation, there are also other dark red cells that fall a long distance from that country's relative position along the line of specialisation. **2.** RCA is the Balassa RCA index. **3.** The RCA measure is truncated at 4 due to the presence of some very large values. **4.** This captures the relationship between 176 countries (sorted by the *increasing rank of economic complexity indicator along the y-axis, and each row is scaled by country export share in world trade*) and 772 products (sorted by the *decreasing rank of product complexity indicator along the x-axis, and each column is scaled by the product share in world trade*). **5.** The data contained in this matrix is computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. **6.** Only a few countries and products are identified to allow a sub sample of labels to be visible.

Figure 3.10: Germany's (DEU) RCA vector sorted by PCI



Notes: 1. Germany is not only diversified, as shown by the large number of exports across many different products, but on average exports more complex products with increasingly higher RCA intensities, as shown by the positive upward slope. **2.** RCA values are the Balassa RCA index. **3.** RCA values are truncated at 4 due to the presence of some very large values. **4.** Products along the x-axis are sorted by the product complexity indicator (PCI). **5.** The graph shows data from the NBER world trade flows dataset for the year 2000.

3.3.3 ECONOMIC COMPLEXITY AND COMPARATIVE ADVANTAGE: A SORTING HEURISTIC

The picture painted so far demonstrates that appropriate scaling provides a clearer synergy between the network complexity approach and the expected patterns set by the traditional theories of international trade. In this section we examine the method of reflections at various intermediate steps to demonstrate how it behaves with respect to sorting information within the M_{cp} to compute a ranked ordering of countries and products. To do this we briefly revisit the 3 x country, 5 x product model, that was presented in Chapter 2, followed by a closer look at discrete steps in the iterative process of computing ECI and PCI.

The *method of reflections* works by reflecting information that is contained in the bipartite network structure of trade over diversification and ubiquity to capture a joint effect of both country and product characteristics. This method can alternatively be shown to act as a sorting heuristic that sorts countries (rows) and products (columns) into a cross-country pattern that works by clustering the country-product relationships contained in the M_{cp} matrix along a diagonal line of specialisation.[45] Recall from Chapter 2, the method of reflections effectively maximised the clustering around the quasi-diagonal of the M_{cp} matrix in the reduced example containing 3 countries and 5 products. The following matrix was observed:

$$M_{cp} = \begin{matrix} \\ C2 \\ C1 \\ C3 \end{matrix} \begin{matrix} P2 & P3 & P4 & P1 & P5 \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{matrix}$$

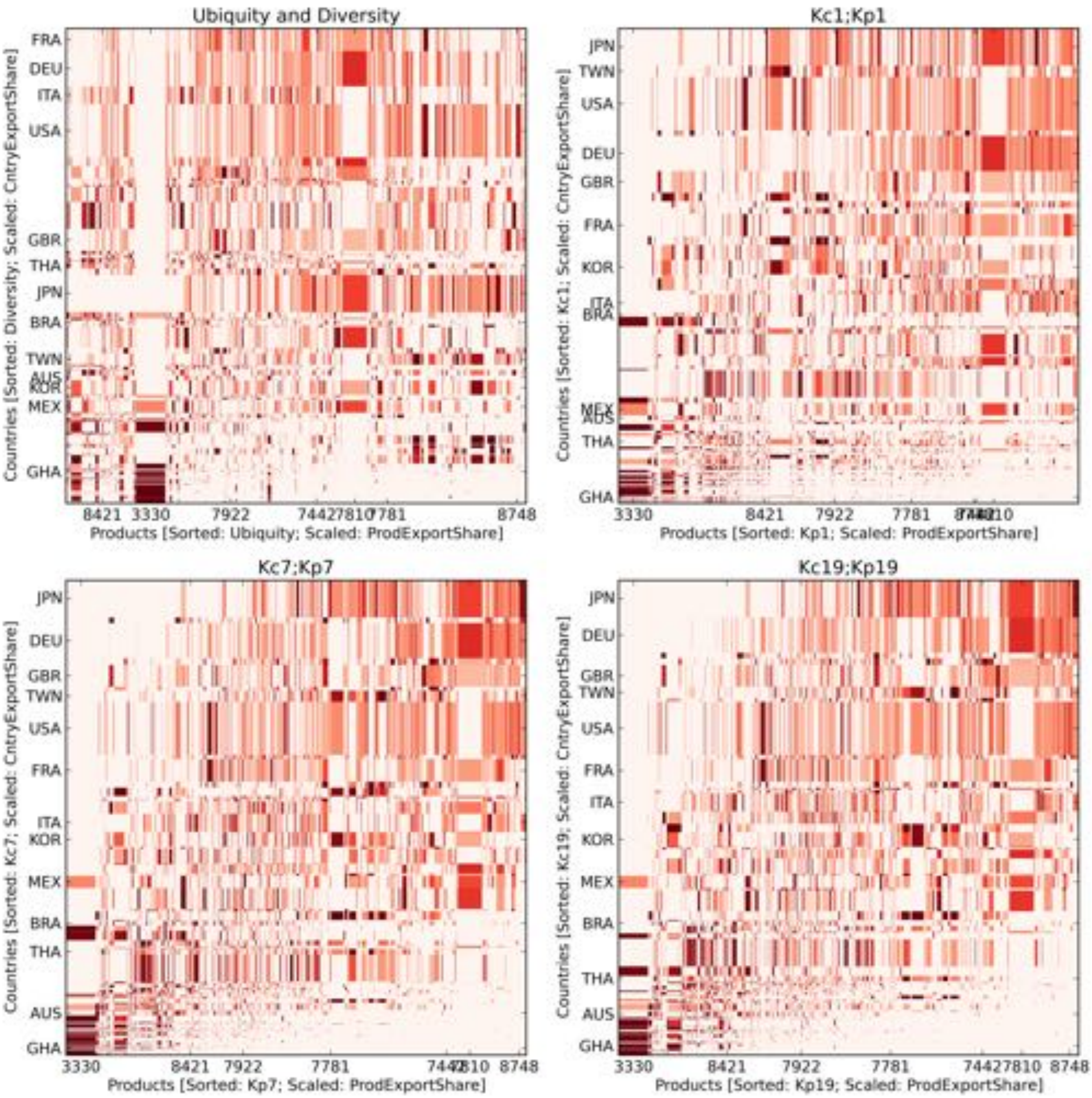
This generates a broad pattern of specialisation that is closely linked to the notion of comparative advantage informed by the macro patterns discussed in Section 3.1. This suggests that countries that have relatively high endowments of *capabilities* produce products that are relatively intensive in the *required number of capabilities* (i.e. higher complexity).

The M_{cp} matrix is a natural meeting point between the bottom-up theories of international trade and the top-down analytics of the network complexity approach. When thinking in terms of theoretical models, the information contained within this matrix is implicitly sorted, which produces the expected patterns of specialisation. Aligning the within-matrix information in accordance with patterns of specialisation could therefore *reveal* useful rank information through the resultant ordered list of countries and products. Part of the explanatory power of this method stems from conducting matrix reordering that is broadly in line with the theory of comparative advantage. To demonstrate how well the method of reflections performs as a sorting heuristic of the high dimensional trade data, a *seriation* exercise is conducted in Appendix E.

To observe the sorting behaviour, various stages of iteration are presented in Figure 3.11, using the same data that was defined in Section 3.3.1.

[45] Another matrix reordering technique, called *seriation* is briefly discussed in Appendix E. A short exercise was undertaken to show that the method of reflections is an efficient heuristic in achieving broad patterns aligned with those set by the theory of comparative advantage in Section 3.1. It performed better than Bond-Energy Algorithms (BEA) due to BEA’s focus on immediate neighbours rather than broader macro relationships contained in these matrices.

Figure 3.11: Iterations of the Method of Reflections for the Year 2000



Notes: 1. This diagram shows the different stages of iteration through the method of reflections starting with a matrix initially sorted by ubiquity and diversity. It shows how the formation of ECI and PCI is organising the within matrix information around a diagonal line of specialisation. **2.** RCA is the Balassa RCA index. **3.** The RCA measure is truncated at 4 due to the presence of some very large values. **4.** This captures the relationship between 176 countries (sorted by *different stages of iteration when computing the rank of economic complexity indicator along the y-axis, and each row is scaled by country export share in world trade*) and 772 products (sorted by different stages of iteration when computing the *rank of product complexity indicator along the x-axis, and each column is scaled by the product share in world trade*). **5.** The data contained in this matrix is computed from NBER world trade flows SITC revision 2 Level 4 for the year 2000. **6.** A consistent set of countries and products are identified to allow a sub sample of labels to be visible and changes in rank.

Starting from a matrix that is initially sorted by country diversity and product ubiquity, the same selection of countries and products are consistently displayed in each of the four panels to trace the relative changes in rank.

Japan (JPN) is promoted in the complexity ranking despite its initial lower overall level of diversification. The main reason that this occurs is that Japan, on average tends to export less ubiquitous products, than other countries with similar levels of diversification. Thailand (THA), while reasonably diversified, exports products that are more ubiquitous on average, and therefore gets demoted in the complexity ranking.

An interesting outlier in this method is Australia (AUS). It is revealed to have a relatively low overall complexity ranking due to its export structure consisting of many resources that tend to be exported by countries that have low average diversification. This is also in part a size effect, as a country like Australia is unlikely to every be as diverse as the United States.[46] The non-dimensionalised view as is presented here also compresses information in the bottom of the chart. Australia is ranked 69th (out of 115) in the ECI index. Therefore 40% of countries are compressed in rows below Australia due to relatively small export shares.

Consideration of the *method of reflections* as a sorting heuristic aligns particularly well with the Ricardian model, albeit with a very thick observed line of specialisation. This approach highlights the vast degree of heterogeneity that is observed in the data that is not well represented in current Ricardian models.[47] Section 3.4 exploits this heterogeneity, by computing relative product deviations within country export baskets as an exercise that demonstrates the complementarity between these new data-centric methods and underlying theory.

Prior to this analytical exercise, it is important to consider the HO model and how it relates to the network complexity approach. It is helpful to use the product space as a measure of revealed product similarities. This links in with an HO framework as the

[46] Studying resource rich countries would form an interesting side study in the economic complexity framework.

[47] This also suggests that there are many determinants of international trade, and the recent literature that integrates the Ricardian and Heckscher-Ohlin type models could be a fruitful avenue for further research.

product space is constructed from the proximity matrix, which serves as a proxy for the overall similarity in the underlying factors of production.

3.3.4 THE PRODUCT SPACE ($\phi_{pp'}$) AND COMPARATIVE ADVANTAGE

The $\boldsymbol{\phi_{pp'}}$ matrix is essentially a transformation of the M_{cp} matrix to compute revealed proximity measures based on the conditional probability of co-export. For a moment let us consider the strict mechanistic relationship between alternative M_{cp} shapes and the overall impact on the $\phi_{pp'}$ matrix. Figure 3.12 compares four possible outcomes: (1) Perfect specialisation, (2) a thick-skewed line of specialisation, (3) a macro line of specialisation with intra-cluster overlap of exports, and (4) the extreme case of a triangular density shape linearly ranging from perfectly diversified country to perfectly specialised (with one perfectly ubiquitous product exported).

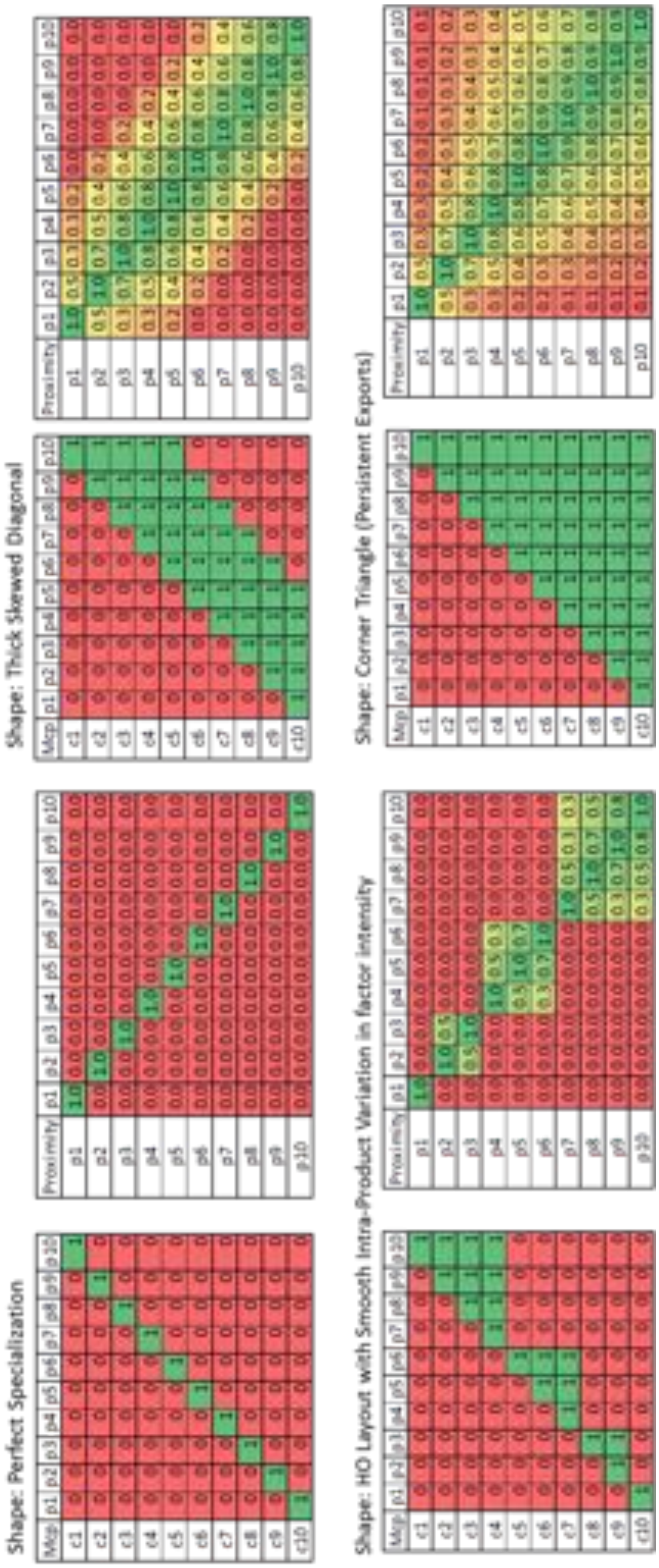
By construction, all proximity matrices are square and symmetric. Symmetry is due to the normalisation of conditional probability using the maximum product ubiquity. All of the cases considered here produce either smooth diagonals with decaying values, or alternatively block diagonals with intra-group values decreasing as one moves away from the diagonal.

The most basic Ricardian case renders a proximity matrix that is similar to the M_{cp} matrix due to the lack of product overlaps. As there is no co-export in the cross-sectional pattern of trade, under perfect specialisation, no products have revealed product similarities. The Ricardian case with “thick” lines of specialisation therefore renders a proximity matrix with decaying proximity values as you move away from the diagonal of the matrix.

$$\phi_{pp'} = \begin{bmatrix} 1 & c & d & \cdots & 0 \\ c & 1 & \ddots & \ddots & \vdots \\ d & \ddots & \ddots & \ddots & d \\ \vdots & \ddots & \ddots & 1 & c \\ 0 & \cdots & d & c & 1 \end{bmatrix}$$

where d < c < 1.

Figure 3.12: Patterns of M_{cp} matrices and impact on the $\phi_{pp'}$ matrix



Notes: 1. This diagram relates four macro-patterns that could be observed in the $\mathrm{M_{cp}}$ matrix and the resultant pattern that would be observed in the $\phi_{\mathrm{pp'}}$ matrix. **2.** In these four examples there are ten countries and ten products.

In the HO framework, cohorts of countries would form based on their relative factor endowments. An example of this would be to sort the matrix by grouping capital abundant countries together, creating clusters of exporters that produce relatively capital-intensive products.

This leads to block-clustering in the $\boldsymbol{\phi_{pp'}}$ matrix.

$$\phi_{pp'} = \begin{bmatrix} \begin{bmatrix} 1 & l \\ l & 1 \end{bmatrix} & 0 & 0 \\ 0 & \begin{bmatrix} 1 & k \\ k & 1 \end{bmatrix} & 0 \\ 0 & 0 & \begin{bmatrix} 1 & h \\ h & 1 \end{bmatrix} \end{bmatrix}$$

As discussed in Section 3.1, clustering by factor endowment is complicated by the possible substitutability between factors.[48] This may produce smooth diagonals with decaying proximity values within highly correlated clusters of products along the diagonal of the proximity matrix and more limited intra-group correlation as some countries may export in different cones of diversification (Deardorff 1999).

$$\phi_{pp'} = \begin{bmatrix} \begin{bmatrix} 1 & l \\ l & 1 \end{bmatrix} & \begin{matrix} \cdots & \cdots \\ lk_1 & lk_2 \end{matrix} & \begin{matrix} lh_1 & 0 \\ lh_2 & \vdots \end{matrix} \\ \begin{matrix} \vdots & lk_1 \\ \vdots & lk_2 \end{matrix} & \begin{bmatrix} 1 & k \\ k & 1 \end{bmatrix} & \begin{matrix} kh_2 & \vdots \\ kh_1 & \vdots \end{matrix} \\ \begin{matrix} \vdots & lh_1 \\ 0 & lh_2 \end{matrix} & \begin{matrix} kh_2 & kh_1 \\ \cdots & \cdots \end{matrix} & \begin{bmatrix} 1 & h \\ h & 1 \end{bmatrix} \end{bmatrix}$$

The main lesson learnt from this exercise is that by sorting the proximity matrix, we would expect either smooth gradients, observed from a Ricardian perspective or blocks of related products that are aligned along the diagonal of the matrix, which is more similar to a Heckscher-Ohlin pattern.

In the same way as the gravity model can describe a number of different trade theories due to its general functional form, there is also no reason to believe that the clustered ordering of the proximity matrix, taking an HO perspective, and the smooth

[48] Addition issues to consider are trade in factors and product intensity reversals in the production process (i.e. different methods of production).

gradients of the Ricardian viewpoint could not be expressed using the same matrix that is sorted differently.[49]

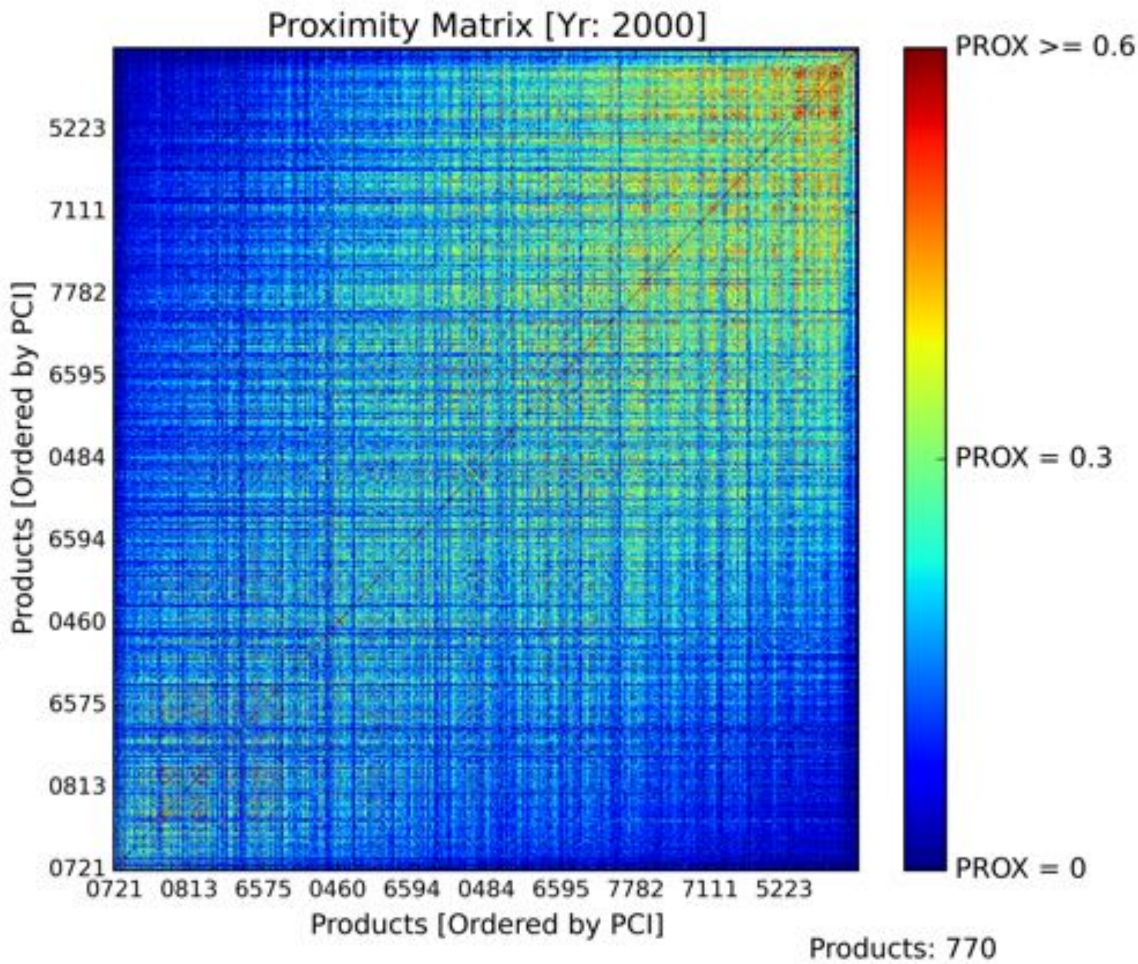
The Ricardian viewpoint is explored first by considering the PCI indicator given its nature to cluster exports around the line of specialisation in the $\mathrm{M_{cp}}$ matrix. Sorting the Proximity Matrix by PCI along both axes is shown in Figure 3.13. The colour BLUE points to product similarity relationships that are very low in proximity, while RED indicates product pairs that share a high level of proximity. From a macro perspective there is evidence of a Ricardian pattern, observing on average, decreasing values as one moves away from the diagonal. Products that are far apart share dark blue regions while products that are close together share higher proximity values when ordered this way. It is also interesting to note that there tends to be increasingly broad levels of relatedness as product complexity increases. This may be a feature of the SITC classification system, as manufactured goods tend to be reported in greater detail, in effect creating more opportunities for co-exported relationships than raw materials, for instance.

To consider the HO viewpoint, the K-means clustering algorithm is used to identify highly correlated groups of products. These clusters are then plotted as shown in Figure 3.14. A group parameter of 10 has been chosen to coincide investigation with the Leamer classification of products.[50] Hausmann & Klinger (2006) find that the Leamer classification is highly correlated with the K-means result. K-means is a similarity-based clustering algorithm that partitions the data into groups based on minimising the within cluster sum of squares (where μ_i is the mean of points in set S_i). The main reason K-means is used in this visualisation is to demonstrate that there are groups of products that tend to be much more inter-related to one another than to other products.

[49] It can be argued that subtle differences in the parameters can be related back to identification of different theoretical models, using the gravity model as is done in Feenstra (2001).

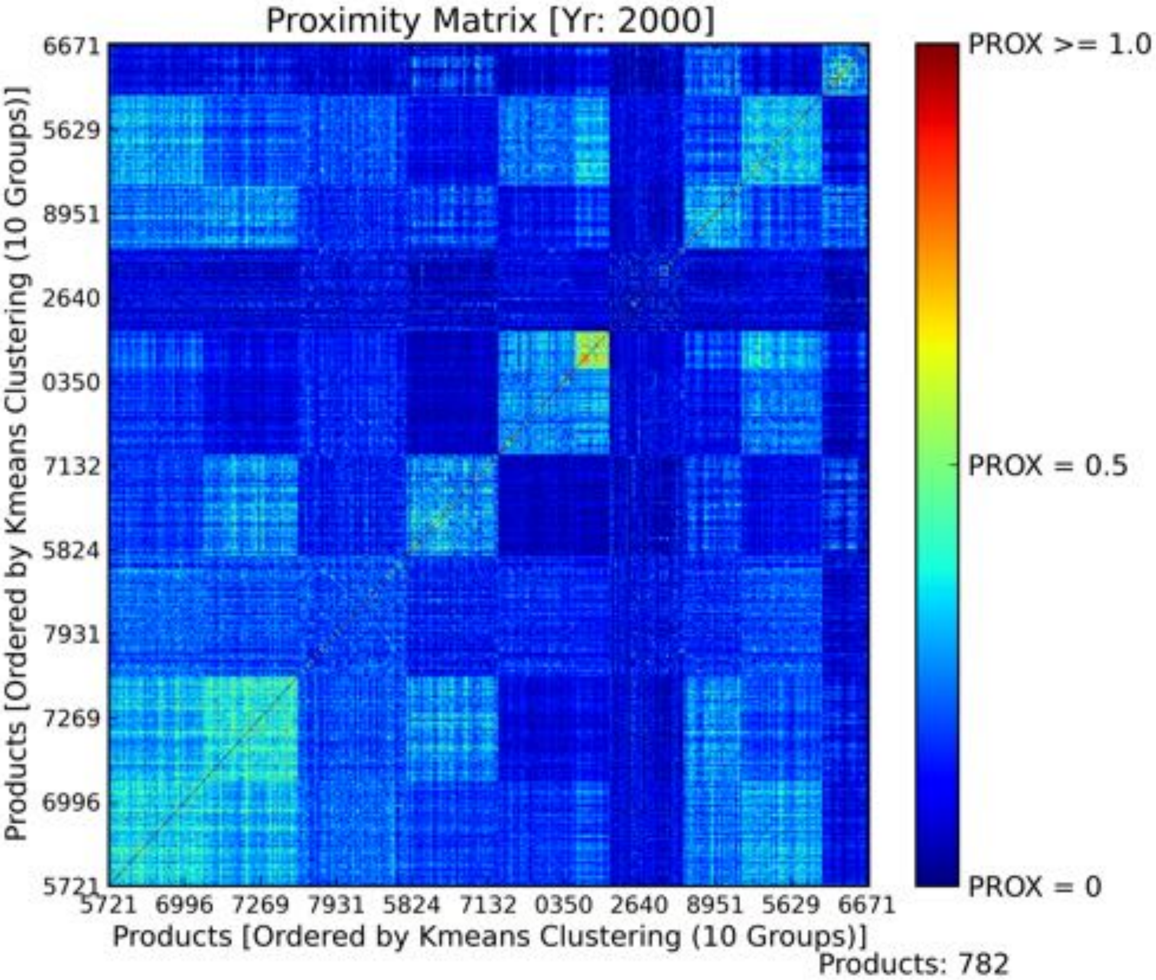
[50] K-means clustering algorithm requires the specification of a known number of groups. The disadvantage with this requirement is that it may force products into groups that may not naturally fit if the underlying data should be grouped into more disaggregated groups.

Figure 3.13: Proximity Matrix for the Year 2000 sorted by PCI



Notes: 1. This graph shows the product similarity matrix that has been sorted by the product complexity indicator (PCI). This shows that products tend to have higher similarity scores with products that are more similarly ranked when compared to products that are very different in the complexity ranking. In addition relatively more complex products tend to form a more densely connected region of the matrix. **2.** The smooth gradient view of the product space proximity matrix aligning with a Ricardian motivation to trade through specialisation. **3.** Proximity values are truncated at values > 0.6 to display greater colour variation in the relationships between product. **4.** Both rows and columns are sorted by increasing rank of the product complexity indicator (PCI). **5.** Data used to compute this matrix is from the NBER world trade flows dataset for the year 2000.

Figure 3.14: Proximity sorted by K-means Clustering (Group parameter of 10)



Notes: 1. This matrix is the result of K-means clustering, which groups products with other similar products. This shows that the highest product proximity is within K-means groups, with less frequent proximity associations between groups. **2.** Rows and columns have been sorted into clusters that were identified by the K-means algorithm with a group parameter of 10. **3.** Data used to compute this matrix are from the NBER world trade flows dataset for the year 2000. **4.** While the K-mean algorithm determines groups of products, the ordering of each group is presented in an arbitrary manner. Therefore there may exist a better arrangement of the data to maximise clustering along the diagonal than is shown in this figure.

Computing K-means is accomplished using the formula:

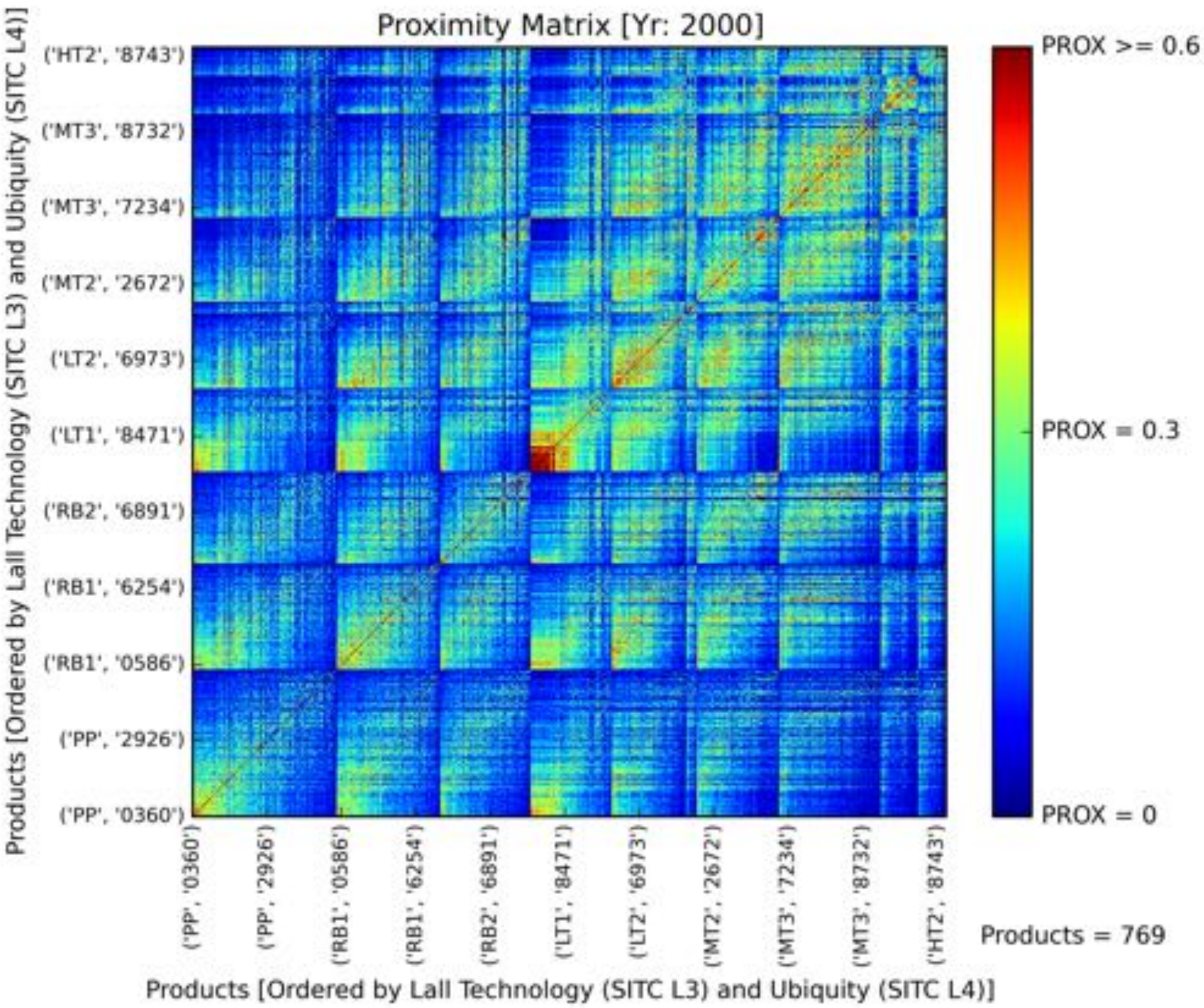
$$\underset{S}{argmin}\sum_{i=1}^{k}\sum_{x_j \varepsilon S_i}(x_j-\mu_i)^2$$

What this process enables us to observe is a level of revealed relatedness between products, which groups them by the degree of overlap in *capabilities* that are required to produce them. Products within each group have high proximity as shown by the colour gradients (Red = highly related, Blue = unrelated). Inter-group relationships also vary from product group to product group. Some products in group one, the left hand column, are also heavily related to the second group. Tubes and piping (6996) [Group 1] and parts of printing and book binding machinery (7269) [Group 2] have been classified into different macro-groups but are still exhibit reasonably high intra-group proximity values.

Another classification developed by Lall (2000) provides an alternative grouping based on clustering products based on technological content. SITC 3-digit products were categorised into primary products (PP), resource based products (RB), light technological products (LT1 and LT2), medium technological products (MT1 and MT2), and high technology products (HT1 and HT2). Figure 3.15, sorts the proximity matrix by the Lall category at the 3-digit level, and then sorts within each group by product ubiquity to demonstrate.

The proximity matrix provides a measure of revealed product similarities that is determined by co-export potentials from sets of similar countries. Figure 3.15 shows that the highest average proximity between products is observed in light manufactures (LT1), primarily driven by apparel production. Apparel is highly disaggregated in recorded trade data and many apparel-producing countries participate in more than one type of apparel product. The dense red cluster that is located in the middle of Figure 3.15 demonstrates that apparel products tend to be exported by sets of similar countries. This same group of products also has a much higher level of proximity to ubiquitous products that are located in both the primary product (PP) and resource based (RB1) categories, and next to no proximity relationship with medium technology (MT3), or high technology (HT1 or HT2) products.

Figure 3.15: Proximity sorted by Lall (SITC L3) and Ubiquity (within groups)



Notes: 1. This figure shows the proximity between products when grouped by the Lall technology intensity product classification at the 3-digit level and then sorted by product ubiquity within each product cluster. **2.** Proximity values are truncated at values > 0.6 to display greater colour variation in the relationships between products. **3.** Both rows and columns are sorted at the 3-digit level by Lall Technology classification (Lall 2000), and then sorted by ubiquity within each 3 digit category. **4.** Data used to compute this matrix are from the NBER world trade flows dataset for the year 2000.

In summary, the proximity matrix broadly contains both of the expected patterns, set by the Ricardian and HO frameworks, depending on the way the rows and columns are sorted. While it is also possible to form product groups directly based on the conventional classifications used in international trade, such as the Leamer and Lall classification system. What the K-means cluster analysis shows is that there is still a considerable amount of variation hierarchically within the factor proportions groups and therefore a high level of heterogeneity exists below the broad Leamer and Lall categories (Hausmann & Klinger 2006). This is reflected in the fact that the product space community algorithm defines 34 product clusters (Hidalgo 2009).

3.4 A MEASURE OF TRADE DISTORTION

To demonstrate the usefulness of the links identified between Section 3.2 (theory) and Section 3.3 (data analytics) it is possible to identify the products that sit a long way from any country's relative position along the diagonal of the M_{cp} matrix. As has been shown, underlying theories of comparative advantage suggest exports should be largely clustered around a diagonal line of specialisation. The deviations from a country's relative position along the diagonal of the sorted M_{cp} matrix therefore reveal information about the degree of departure from that country's relative point of specialisation in complexity. In the case of the USA, the following analysis identifies many agricultural products that are the recipients of large domestic subsidies.

The proposed metric is the negative percent deviation from a country's relative position along the line of specialisation. This is computed by cumulatively summing the product shares that are between the relative positions along the diagonal to the products that sit to the left in the country's export vector.

$$d_{cp} = \sum_{p=\rho_{cl}}^{p=\rho_{cd}} \rho_{cp} - \rho_{c,p=d}$$

where,

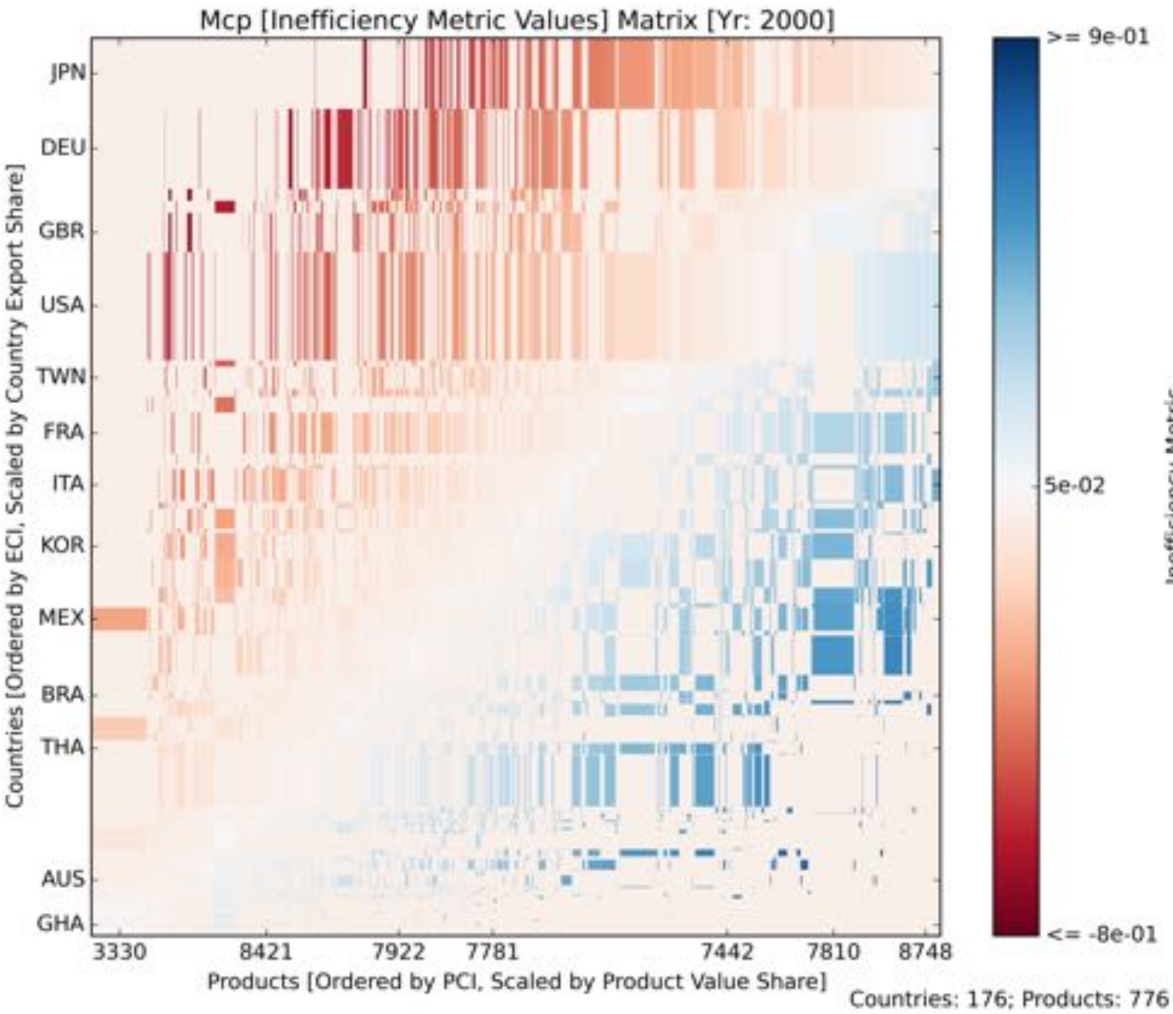
$\rho = share\ of\ export\ p\ in\ world\ trade$

$\rho_{cl} = left\ bounded\ product\ horizontal\ position$

$\rho_{cd} = country\ c\ relative\ horizontal\ position\ along\ diagonal$

The d_{cp} measure varies between 0 and 1 and represents the percent deviation of the value-weighted square cross-sectional trade matrix in which rows have been sorted by the ECI, and columns have been sorted by the PCI.

Figure 3.16: A measure of trade deviation



Notes: 1. Data contained within this heat map are the horizontal deviations from a country's relative position along the line of specialisation. **2.** Rows are sorted by the economic complexity indicator (ECI) and scaled by country export share in world trade. Columns are sorted by product complexity indicator (PCI) and scaled by product export share in world trade. **3.** Red indicates relatively un-complex products that are exported by relatively complex countries. (i.e. Rice exports from the USA). Blue indicates relatively complex products that are exported by relatively un-complex economies (Automobile exports from Australia). **4.** Data used to compute this matrix is from the NBER world trade flows dataset for the year 2000.

This serves as a novel measure that identifies products that are most likely to be a result of trade distortion.[51] Figure 3.16 shows a transformation of the M_{cp} matrix that computes deviations from a country's position along the diagonal. This information is presented in the form of the heat map. Red indicates products that are found in relatively complex economies but are relatively un-complex products. While these products have been identified to have RCA, they sit a long way from that economy's relative point of specialisation.

Consider the United States (USA), which exports a number of products that sit a long way to the left of its relative position along the diagonal. The top twenty items with the largest deviations are listed in Table 3.1.

Table 3.1: Top 20 Negative Deviations from the Line of Specialisation for the USA in the Year 2000

SITC Code	DESCRIPTION	Deviation Metric
2631	Cotton (other than linters), not carded or combed	-0.69
2223	Cotton Seeds	-0.69
2221	Groundnuts (peanuts)	-0.69
0577	Edible Nuts (excluding nuts chiefly used for extraction of oil)	-0.68
0542	Beans, peas, lentils and other leguminous vegetables	-0.67
0342	Fish, frozen (excluding Fillets)	-0.67
2911	Bones, horns, ivory, hooves, claws, coral, shells and similar …	-0.67
0812	Bran, sharps and residues derived from sifting, milling, working …	-0.67
0459	Buckwheat, millet, canary seed, grain sorghum and other …	-0.67
2632	Cotton Linters	-0.67
2483	Wood of non-coniferous species, sawn, planed, tongued, grooved …	-0.67
4233	Cotton seed oil	-0.67
2633	Cotton waste (including pulled or garneted rags) …	-0.66
1212	Tobacco, not stripped	-0.66
2119	Hides and Skins, nes	-0.66
1213	Tobacco, refuse	-0.66
2222	Soya Beans	-0.65
0440	Maize (corn), unmilled	-0.64
0813	Oil-cake and other residues (except dregs) resulting from the …	-0.63
0421	Rice, in the husk or husked	-0.63

Notes: 1. The above measures are computed using SITC revision 2 Level 4 export data for the year 2000. **2.** The inefficiency metric is a measure of product deviation from the United States' relative position along the diagonal line of specialisation. Cotton, for example, falls 69 percent of the width of the entire matrix away from its position at 0.8 of the distance from the left hand axes.

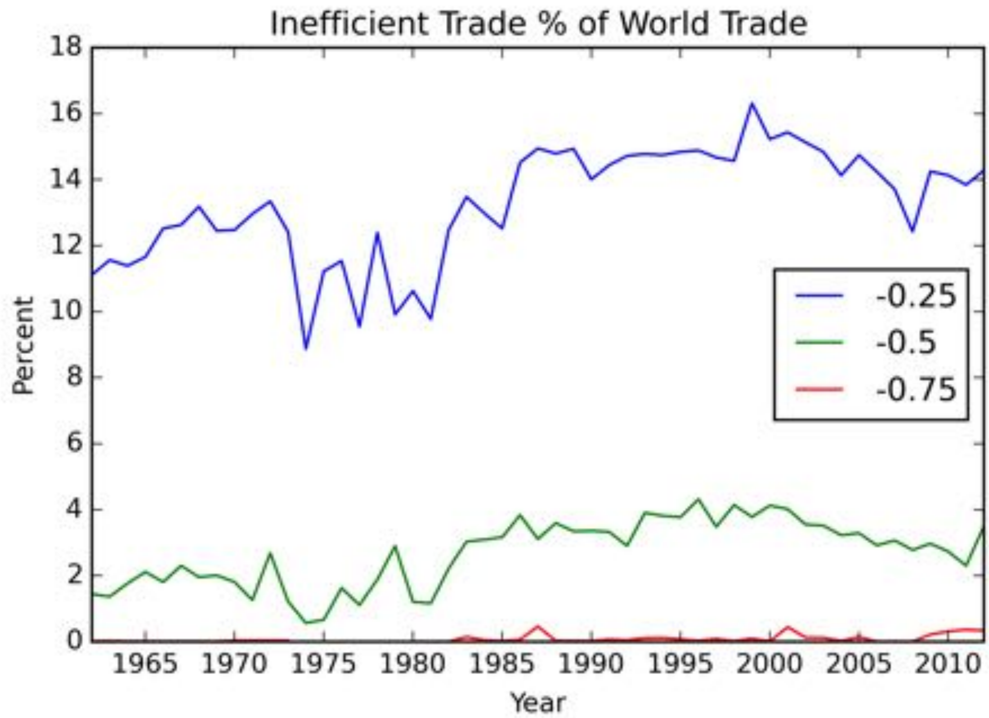
[51] What remains to be explored is a theory-informed cutoff point.

The majority of these negatively deviated items are agricultural products that are heavily supported by domestic U.S. agricultural subsidy programs. In the OECD (2014) agricultural monitoring and evaluation report, the US provided $1.2 billion in 2013 to wheat (-0.62 deviation score), $2.8 billion to maize (-0.64 deviation score), $177 million to sorghum (-0.67 deviation score), $1.5 billion to soybean (-0.65 deviation score), and $473 million to cotton producers (-0.69 deviation score). All of these products are identified as items that are the most negatively deviated relative to the expected level of sophistication of the USA. These results are encouraging, as in the case of the United States, the matrix reordering information using ECI and PCI has correctly identified products that are typically considered least in line with its comparative advantage.

Another way to use this metric is to aggregate the most deviated products across all country export baskets in a triangular fashion. As was previously demonstrated, in the case of Germany (Figure 3.10), the fact that RCA values on average increase towards the line of specialisation suggests that the majority of export *value* is clustered around the diagonal. A global aggregate indicator was computed to assess the total export value contained in these highly deviated products across all countries. The process effectively sums across the top left triangular region of the matrix, where the hypotenuse is parallel to the diagonal. A range of deviation cut-off values was considered to understand the value of global of trade that is contained in the triangular quadrants as shown in Figure 3.17.

This aggregation supports the notion that products positioned along the diagonal capture the majority of export value. Aggregating all product values that fall outside a 25 percent deviation line from the diagonal suggests that 10 to 15 percent of total world export value is contained in the top left triangular quadrant. While, this is by no means a small number, it does confirm that trade *values* have a tendency to fall in line with specialisation. This value non-linearly declines as you increase the cut-off distance. When summing the value of products that fall further than a 50 percent deviation contains between only 1 and 4 percent of export value between 1962 and 2012. In the US case the 50 percent cut-off point differentiates many of the primary product agricultural exports that receive large domestic subsidies with some of the relatively more sophisticated products such as frozen beef exports which falls just short of the 50 percent mark at (-0.49) and clay and refractory minerals (-0.49).[52]

Figure 3.17: Value (% of World Trade) of products that are negatively deviated from country relative positions along the diagonal line of specialisation



Notes: 1. This graph shows the percent of world trade that is contained in products that are negatively deviated from the line of specialisation. This is equivalent to drawing a thick diagonal equal to 25 percent of the width of the matrix and aggregating product values that are contained in the top left triangular quadrant. This shows that the majority of trade, by value, is not negatively deviated. Complex countries earn the largest export revenue from complex products, on average. **2.** Data are from 1962 to 2012 and sourced from NBER-BACI dataset D at the SITC 4 digit level of disaggregated. More information on the dataset can be found in Appendix A.

As is widely known, developing economies would stand to benefit greatly from reduced exports in these less sophisticated, mainly agricultural, products from countries such as the United States and Western Europe. According to OECD (2014), total net official development assistance in 2012 was US$ 125.6 billion. Taking the cut-off value of 0.5 or 50% of the width of the adjusted M_{cp} matrix, this analysis suggests it

[52] Further work should be conducted to identify a more rigorous cutoff point across countries. The current cutoff point is simply defined to be illustrative and mainly based on the US case.

would open up export opportunities worth US $550 billion or almost four and half times larger than official development assistance.

3.5 CONCLUSION

The purpose of this chapter was to demonstrate a complementarity between the network complexity framework and the more traditional literature on comparative advantage. This was achieved by considering the information contained within the $\mathrm{M_{cp}}$ matrix. As a first step expected patterns were constructed by considering many-country, many-product cases of the Ricardian and Heckscher-Ohlin theories of comparative advantage. While exact cross-sectional patterns are typically difficult to obtain due to the high dimensionality setting, the broader expectation demonstrated by Section 3.2, is that comparative advantage works to generate either an overall line of specialisation, or block-diagonal clusters, in the $\mathrm{M_{cp}}$ matrix depending on a Ricardian or Heckscher-Ohlin perspective.

These patterns were then compared to those produced when using the *method of reflections* to compute ECI and PCI. It is found that based on matrix reordering, patterns set by both theories can be observed in the $\mathrm{M_{cp}}$ and $\phi_{\mathrm{pp'}}$ matrix. While much of the research on the network complexity approach has focused on diversification and ubiquity, this chapter considers the method of reflections as a sorting heuristic which sorts' countries and products in a way that clusters exports around a diagonal line of specialisation.[53] The strong explanatory power contained in this new measure of complexity, is in part, driven by aligning countries and their exports around a line of specialisation in a many-country, many-product context.

Where this research deviates from the underlying theory of comparative advantage is in the large degree of heterogeneity that is observed in the data. The data-centric network methods allow economists to study the rich heterogeneity that persists in the

[53] This idea is strongly related to another analytical technique used in computational archaeology known as *seriation*. These matrix-reordering methods are used to discover underling patterns in archaeological data. Heuristics are an important component of this literature due to the computational difficulties in this discrete combinatorial problem. The method of reflections is considered in light of the basic seriation techniques and ECI and PCI are shown to be very effective at reordering countries and products compared to a technique known as BEA as is discussed in the Appendix E.

high dimensional setting of the real world. The analysis in Section 3.4 showed that while relatively complex countries tend to obtain a majority of export value through the export of relatively complex products, in many cases these same countries also export relatively less complex products that deviate a long way from their relative position along the line of specialisation. The percent deviation from a country's relative position along the diagonal line of specialisation was proposed as a measure of trade distortion. When looking at the United States, this measure identified a large number of agricultural products that are subject to large distortionary domestic subsidies.

CHAPTER 4 PROBABLE AND IMPROBABLE PRODUCT EMERGENCE DURING TRADE LIBERALISATION

"It is far easier to show why, especially over time, liberalising a restrictive trade regime is conducive to more rapid growth than it is to show why outer-oriented trade strategies have been so highly successful." (Krueger, 1998)

Abstract

This chapter examines the impact of trade liberalisation on export performance, covering 96 liberalisation episodes in 129 countries over 1962–2012. Product emergence and diffusion metrics are constructed and used in an econometric analysis to examine the average within-country impact of trade liberalisation on product emergence and diffusion characteristics through the product space. The findings indicate that trade liberalisation is associated with: a) an increase of 18% to 24% the rate of product emergence that emerge with high proximity to current exports, and b) a moderate 8% increase in the distance of diffusion to products that are more dissimilar to those in the pre-reform export basket. Overall, trade liberalisation appears to contribute to both accumulation and diversification of underlying country capabilities.

4.1 INTRODUCTION

Export orientation is now widely acknowledged as an important source of economic growth. This view is based on both economic theory and the experience of East Asian countries since the early 1970s. Consequently, over the past few decades the consensus in policy thinking has shifted from protectionist-import substitution policies towards outward-oriented trade policies. However, despite decades of research, there remains conflicting evidence in support of trade liberalisation. Direct evidence is lacking primarily due to the central difficulty in establishing causality running from trade policy reforms through to economic growth (Papageorgiou, Choksi & Michaely 1990; Krueger 1997; Greenaway, Morgan & Wright 1997, 1998, 2002; Winters 2004; Wacziarg & Welch 2008).

Winters (2004) provides the best summary:

> *While there are serious methodological challenges and disagreements about the strength of the evidence, the most plausible conclusion is that trade liberalisation generally induces a temporary (but possibly long-lived) increase in growth*

Finding empirical evidence is also hampered by: difficulties in measuring and identifying trade liberalisation episodes, the multidimensional nature of complementary and non-complementary policies that act in parallel to trade liberalisation, and the observed heterogeneity of country experiences.

This chapter focuses on the direct relationship between trade liberalisation and export performance, rather than explicitly considering the link between trade liberalisation and economic growth. The *a priori* view is that trade liberalisation promotes export growth by reinforcing current export products (intensive margin) in addition to enabling the emergence of new products that are in line with a country's comparative advantage (extensive margin). The productivity improvements that are observed in the empirical literature, via improved access to foreign intermediates, provides a good basis for an increase in export response (Amiti & Konings 2007; Goldberg et al. 2009; Topalova & Khandelwal 2011).[54] In addition, economic theory is supportive of the emergence of new products, as access to new intermediates can remove one of the potential barriers for firms to become future exporters (Puga & Venables 1999). Although trade liberalisation does not solely operate through the export channel, it is the focus in this chapter as it integrates cleanly with the recent literature on the product space network.[55]

The product complexity analysis, developed by Hausmann and his co-researchers, allows for a disaggregated structural approach to be used for the purpose of characterising product emergence (Hausmann & Klinger 2006; Hidalgo et al. 2007; Hidalgo 2009; Hausmann & Hidalgo 2011, 2014). This research has shown that

[54] This literature is reviewed in more detail in section 4.2.

[55] In addition to an export response, Santos-Paulino (2005) also identifies a strong import response, which can lead to a worsening in the balance of trade. This evidence provides a strong case for the sequencing of trade liberalisation and careful consideration for how it is implemented.

relatively poor countries occupy the periphery of the product space network while developed countries tend to export products located in the densely connected core (Hidalgo et al. 2007). To date much of the focus has been on the interaction between country diversification and product ubiquity and there has not been a significant focus on how diffusion patterns change in relation to public policy. This chapter addresses part of this gap by considering how trade liberalisation policy alters these diffusion patterns as countries migrate from the periphery to the core of the product space network.

A key benefit of network analysis is that it allows for the contextual starting position of each country, and captures the nature of each product that a country exports. The product space is best visualised as a network and measures the similarity between products as determined by their co-export potential observed across all countries. This network consists of nodes that represent products and edges that represent the degree of similarity. This proximity metric provides a way to measure the distance between products and is used in this chapter to classify product emergence (extensive margin) into probable emergence (emergence of new products that are close to a country's position in the network) and improbable emergence (emergence of new products that are distant from a country's position in the network).

Given this network is defined by the proximity matrix, its technical definition is briefly reviewed to serve as a brief reminder about the information that is contained within it. This is followed by definitions of the diffusion measures used in this study, which are broken into two different approaches. A cut-off procedure was used to split the extensive margin into probable and improbable product emergence, in addition to two continuous measures of the distance of diffusion: the variance and width of proximity. These measures are then used in a cross-country panel regression framework to understand how trade liberalisation affects the rate of emergence, as well as the distance of emergence of new products for the years 1962 to 2012.

The analysis of this chapter yields two novel findings. First, trade liberalisation is positively associated with higher rates of probable product emergence, that is emergence of products that are in high proximity to at least one other product in a countries export basket. This finding, that products emerge close to existing products, is consistent with the findings by Hidalgo et al. (2007). What is new is the finding that there is an associated 18% to 24% increase in the *rate* of local diffusion observed under periods of trade liberalisation. In a country capability context (discussed in Chapter 2) this indicates that countries tend to accumulate production *capabilities* at a faster rate under trade liberalisation. Countries increase the rate of diffusion within current product clusters or to product clusters that are close to the existing export structure of the country. For example, clothing producers that liberalise would more quickly diffuse to other clothing categories, as they have already acquired similar underlying production capabilities in these product types.

The second finding is that trade liberalisation is associated with a *moderate* increase in the width of diffusion through the product space. While no evidence was found of an increase in the rate of extreme improbable product emergence, analysis using a continuous measure of the distance of diffusion suggests a moderate 8% increase in diffusion distance. By supporting larger diffusion jumps in the product space, trade liberalisation has a role to play in the accumulation of more *diverse* underlying *capabilities* for a country. For example, a clothing manufacturer may start to produce other light manufactures such as textiles and footwear but tends not to make large improbable jumps to dissimilar industries such as automobile manufacturing.

Taken together, both of these findings lend support to trade liberalisation having a positive effect on increasing the underlying *size* and *diversity* of country *capabilities*.[56]

The chapter is organised as follows. Section 4.2 provides a more detailed discussion of the literature on the links between trade liberalisation and export response. Section 4.3 defines the product space and the measurement of product proximity. Section 4.4 details the derived measures of product diffusion in the context of the product space. Section 4.5 discusses the data sources. Section 4.6 provides country case studies focusing on the derived measures of export diffusion, followed by Section 4.7 which performs a statistical analysis of these metrics in relation to episodes of trade liberalisation using a panel fixed effects estimation strategy and discusses the results. Section 4.8 concludes with the main findings and suggestions for future research.

---

[56] Diversity in the underlying country *capabilities* vector is strongly linked to a country's level of development, and therefore may also have indirect implications for economic growth.

4.2 TRADE LIBERALISATION, GROWTH, AND EXPORT RESPONSE

While there are many potential channels determining the relationship between trade liberalisation and export (and economic) growth, productivity appears to be one key explanatory factor. The pro-competitiveness aspects of trade liberalisation induce discipline on local firms to become and remain efficient (Amiti & Konings 2007; Topalova & Khandelwal 2011). Models have been constructed to show that more integrated economies, through market size and trade effects, exhibit higher aggregate productivity (Melitz & Ottaviano 2008). Additional channels include effects on the evolution of institutional quality. Deconstructing protectionist intervention in the economy improves the institutional setting by removing incentives for rent seeking opportunities that come from protection (Krueger 1998). Despite these arguments the findings of multi-country econometric analyses of the relationship between trade liberalisation and economic growth are mixed. However in-depth country case-studies have generally come up with evidence that is consistent with the postulate that trade liberalisation promotes growth.[57]

Empirical evidence at the country level has also shed light on more specific channels through which trade liberalisation affects export performance (and growth) through productivity improvements. While trade liberalisation not only acts to improve domestic productivity through a competitiveness effect, empirical evidence suggests that access to foreign intermediates is an equally, if not more significant channel for productivity improvements. Amiti & Konings (2007) find evidence in the Indonesian context that a 10% fall in input tariffs results in a productivity gain of around 12% for the firms that import inputs from abroad. This is almost twice as large as the gains that they estimate from reducing the tariffs on final goods.[58] These observed productivity gains are largely attributed to learning, variety, and quality effects. This evidence is also supported in the Indian case where Topalova & Khandelwal (2011) find firm level productivity improvements through the reduction of both final and intermediate

[57] See Krueger (1997) for a survey of this literature.

[58] This does not suggest that countries should reduce tariffs on intermediates while retaining tariffs on final goods, as both channels are still estimated to provide productivity improvements. In addition, altering the structure of tariffs in this way will drive up the effective rate of protection, providing higher protection for import competing goods (Corden 1971).

tariffs, with reduction in tariffs applied to imported intermediates associated with a larger productivity improvements.

Considering a dynamic setting, access to lower cost intermediates also has the potential to remove one of the barriers to future industrial development opportunities (Puga & Venables 1999). Having access to foreign intermediates, such as high quality textile and fabrics, may allow a country to compete in apparel production without the need to develop relatively capital-intensive production technology to supply the required textile inputs domestically. Modern sporting goods, for example, require relatively sophisticated materials that have properties conducive to sporting activities (i.e. flexibility, breathability, and durability). This can require relatively sophisticated production techniques to produce fabrics with the desired properties that are of high quality – and may present as a technical barrier for countries to participate effectively in the production of this type of apparel and clothing products.[59]

An alternative strand of literature that is more critical of broad-based trade liberalisation argues that import protection, substituting imported goods with domestically produced goods, has the potential to provide an environment for domestic industrial development (Lee 1997; Melitz 2005). The primary focus of this argument is either: (1) on the formation of industry and not on the efficient allocation of resources used in the production of goods, or (2) a dynamic setting where short term losses can move an industry to a new equilibrium involving long term sustainable gains. It is the dynamic gains that most interest the advocates of this strategy, preparing a country for the future with a domestically thriving modern sector. What is not apparent however is how the information problem surrounding the predictability of economic success can be overcome (Dornbusch 1992), in addition to the negative impact selective protection can have on institutional quality and rent seeking. While Rodriguez & Rodrik (2001) do not know of a case where trade restrictions are systematically associated with higher growth rates, they advise caution around the general enthusiasm towards trade policy. They make the case that economies are multidimensional and trade policy may not induce the economic rewards that are

[59] This is also intrinsically linked with the product fragmentation process that has been observed in recent decades. As transportation and service link costs decline, distance becomes less of a barrier for participation in different segments of the value chain.

expected, and that other determinants of growth may be more rewarding (such as focusing on other institutional determinants).

The lack of consensus in this debate continues to provide a rich environment for additional research to more fully understand the effect trade liberalisation has on the development process. The next section of this chapter details a new branch of literature called the product space, which enables a *structural network view* of how new products emerge in light of trade liberalisation.

4.3 THE PRODUCT SPACE

This section briefly revisits the product space, as introduced in Hidalgo et al. (2007), and provides a short discussion on the various definitions and components used to identify the de-novo ("new") goods margin.

The position of a country within the product space is highly correlated with the level of development. In the product space developed countries tend to export products across the network, including exports that occupy the densely connected core, while developing economies tend to only export products that are found in the periphery. The product space is constructed from the conditional joint probabilities of co-exporting any two products. This forms a disaggregated similarity matrix ($\phi_{pp'}$) over the product dimension by comparing all pairwise product relationships across the countries that export them. This provides a measurement of the main idea that if a country exports t-shirts it is more likely to export trousers than it is to export automobiles.

The proximity matrix encapsulates this notion of similarity (or alternatively a measure of distance) between products, and provides a structured landscape over which one can use to characterise the type of product emergence. This *structural* perspective can provide information not only concerning the change in country diversification levels, but in the type of product changes that are occurring through the diffusion process.[60] If a country were observed to jump to more distant products due to some policy

[60] The product space is often best visualized as a network of products. Diffusion therefore is the process of jumping from already occupied product nodes to new previously unoccupied nodes.

intervention, this would be indicative of the country acquiring a more *diverse* set of underlying *capabilities*.

In summary, products that emerge close to existing products are more in line with that country's current set of capabilities. Products that emerge further away from current products suggest the acquisition of more *diverse capabilities*. The product space network provides a method to observe these structural changes in export patterns, as the proximity measure reveals useful information about how similar any two products are. High product proximity implies overlaps in the unobserved set of country capabilities required to produce each product. [61]

As discussed in Chapter 2, the basic unit of data required to compute the product space is the M_{cp} matrix:

$$M_{cp} = \begin{cases} 1 & RCA_{cp} \geq 1 \\ 0 & RCA_{cp} < 1 \end{cases}$$

This matrix describes the assignment between country *c* and product *p*, in which that country exports an RCA value which is greater than or equal to one.[62] The definition of RCA used in this study is the Balassa (1965) measure, which is the value share of a product in a country's export basket over the value share that that product holds in world exports defined as:[63]

$$RCA_{cp} = \frac{\frac{E_{cp}}{\sum_{C} E_{cp}}}{\frac{\sum_{P} E_{cp}}{\sum_{C,P} E_{cp}}}$$

[61] When considering the formulae for computing proximity it can be observed that to attain a high conditional co-export probability, not only do the products need to be exported by a similar set of countries, they also needs to have similar ubiquity characteristics. If either product is relatively more ubiquitous, then the proximity score is decreased due the use of maximum ubiquity in the denominator of the relationship.
[62] Other cut-off definitions of RCA have been considered in the literature, and find the structure of the product space to be robust to changes in the RCA cut-off value. An in-depth discussion of revealed comparative advantage can be found in Appendix C.
[63] A discussion on the alternative measures of revealed comparative advantage (RCA) can be found in Appendix B. The Balassa measure compares favourably with more recent measures of revealed comparative advantage particularly when considering the construction of the binary M_{cp} matrix.

where,

E_{cp} = Export of product p from country c

The proximity matrix $\phi_{pp'}$ is the derived measure of product relatedness and is computed from the minimum conditional probabilities of co-exporting any two goods; Product 1 (P_1) and Product 2 (P_2) and is defined as:

$$\phi_{p_1p_2} = \frac{\sum_c M_{cp_1} \bullet M_{cp_2}}{max(k_{p_1,0}, k_{p_2,0})}$$

where,

$M_{c\mathrm{p_n}}$ is the country-product export assignment matrix

$k_{p_1,0}$ is the ubiquity of product 1, and

$k_{p_2,0}$ is the ubiquity of product 2

Within this context, proximity provides a measure of similarity based on the observed similarities across the set of countries that export these goods. Products, which are commonly co-exported by the same set of countries, as well as sharing very similar ubiquity characteristics, will score highly. Consider the simple case:

$$M_{cp} = \begin{bmatrix} & P & P' & \\ \cdots & 0 & 0 & \cdots \\ \cdots & 1 & 1 & \cdots \\ \cdots & 1 & 1 & \cdots \\ \cdots & 0 & 0 & \cdots \end{bmatrix}$$

Regardless of any other columns, the multiplication and summing of both column vectors will generate a numerator of 2. The ubiquity scores (the sum of each column) of both P and P’ are both 2. That indicates that P and P’ are similar products in terms of their ubiquity characteristics, and are also exported by exactly the same two countries. Consider an alternative case where the two products are not exported by exactly the same set of countries:

$$M_{cp} = \begin{bmatrix} & P & P' & \\ \cdots & 0 & 0 & \cdots \\ \cdots & 1 & 0 & \cdots \\ \cdots & 1 & 1 & \cdots \\ \cdots & 0 & 1 & \cdots \end{bmatrix}$$

While both products have equivalent ubiquity scores the numerator is now only equal to the overlap, which equals 1. In this case the proximity measure is reduced to 0.5, as the measure is penalised due to differences in the set of common exporters.

A final case to consider is the perfect overlap in one product:

$$M_{cp} = \begin{bmatrix} & P & P' & \\ \cdots & 0 & 0 & \cdots \\ \cdots & 1 & 0 & \cdots \\ \cdots & 1 & 1 & \cdots \\ \cdots & 0 & 0 & \cdots \end{bmatrix}$$

In this case *P'* is produced by the same set of countries, while *P* is not. Due to the maximum function in the denominator, the ubiquity score of *P* is used to produce a symmetric pairwise relationship due to the asymmetry in the column vectors and the ubiquity scores.[64]

To illustrate what the values are in practice, Table 4.1 provides proximity values that are computed for mens and boys trousers (SITC4 8423) as well as green and roasted coffee (SITC4 0711) from the NBER world trade flows (WTF) dataset for the year 2000.[65] These are compared with a small selection of partner products across a spectrum of types. What this table shows is that this measure does a good job of ranking products that are of similar type with high proximity values, such as trousers and shirts. Products that have dissimilar factors of production have very low proximity scores. The overall distribution of proximity, for the year 2000, is shown in Figure 4.1.

[64] Appendix G compares this symmetric formulation with the asymmetric proximity matrix that allows for both conditional probabilities to be computed. The use of symmetric conditional probability values is in alignment with the current literature on economic and product complexity.

[65] The dataset used to compute these values is the NBER SITC Revision 2 Level 4 dataset for WTF (Feenstra et al. 2005) .

Table 4.1: Examples of Proximity Values for the Year 2000

Product 1	Product 2	Proximity
Mens & Boys Trousers (8423)	Mens and Boys Shirts (8441)	0.81
	Jerseys / Cardigans (8451)	0.77
	Linens / Textile Coverings (6584)	0.36
	Motor Vehicles (7810)	0.09
	Commercial Aircraft (7924)	0.06
Coffee (Green, Roasted) (0711)	Sugars, Raw (0611)	0.5
	Cut Flowers (2927)	0.45
	Jerseys / Cardigans (8451)	0.33
	Motor Vehicles (7810)	0.03

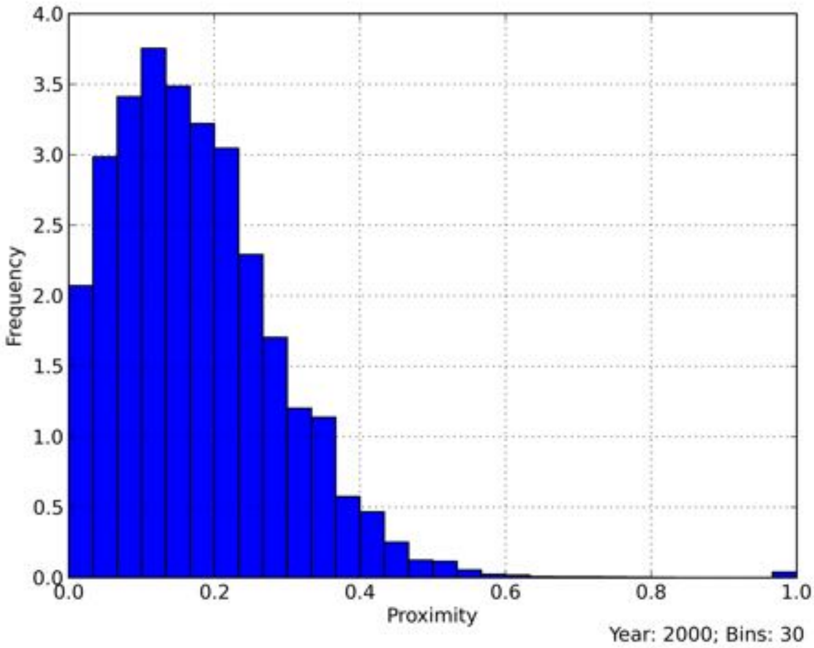
It is this $\phi_{pp'}$ matrix that provides structure for characterising the extensive margin into probable and improbable product emergence relative to each country's export basket. The overall distribution of proximity values are relatively consistent over time and portrays a relatively sparse proximity matrix with a majority of low values between product pairs.

The product space network is constructed by focusing on the strong bonds between products through the use of a minimum spanning tree along with adding edge values that are greater than 0.55.[66] Consideration of a minimum spanning tree ensures that the product space is a fully connected network by ensuring that a minimum of at least one link connects all products. Groups of products that share large proximity values will form *communities* in the product space network.[67] This community structure is discussed in Hidalgo et al. (2007).

[66] These strong connections are the right hand side tail of the proximity distribution as shown in Figure 4.1.

[67] It is possible to argue that if products are filtered through the Balassa index, then by definition, products that emerge in a country's export basket have comparative advantage due to the way M_{cp} is constructed. This view can be challenged on the grounds that the Balassa index is only an *indicator* of comparative advantage. This indicator is a measure of a country's export participation that is excess of the expected world average (as judged by that products share in world trade). Considering the US case, based on the Balassa definition it has RCA in a wide variety of goods including many unexpected products such as agricultural products.

Figure 4.1: Proximity Histogram for the Year 2000



Notes: 1. This diagram is the distribution of proximity values constructed in the year 2000. It is a highly skewed distribution that shows that strong proximity values are relatively rare. **2.** This diagram excludes a large number of zero relationships to consider only +ve proximity values.

4.4 MEASUREMENT OF PRODUCT EMERGENCE AND DIFFUSION

Before commencing with an econometric study, the independent variables to describe the rate and distance of diffusion to be used in the analysis need to be computed. In section 4.4.1, to characterise how a country diffuses through the product space, the extensive margin ("new" goods) is decomposed into probable (products that emerge near to current country exports) and improbable emergence (products that emerge far from current country exports). Continuing with the network description of the product space, this is akin to classifying products in terms of how close they are to a country's current collection of exported products. In section 4.4.2 the size of jumps within the product space are computed to provide a continuous measure of the distance of diffusion. This is accomplished by measuring the width and variance of all pairwise distance relationships in each country's export basket between any two periods.

4.4.1 VECTORS OF IMPROBABLE AND PROBABLE EMERGENCE ("JUMPS")

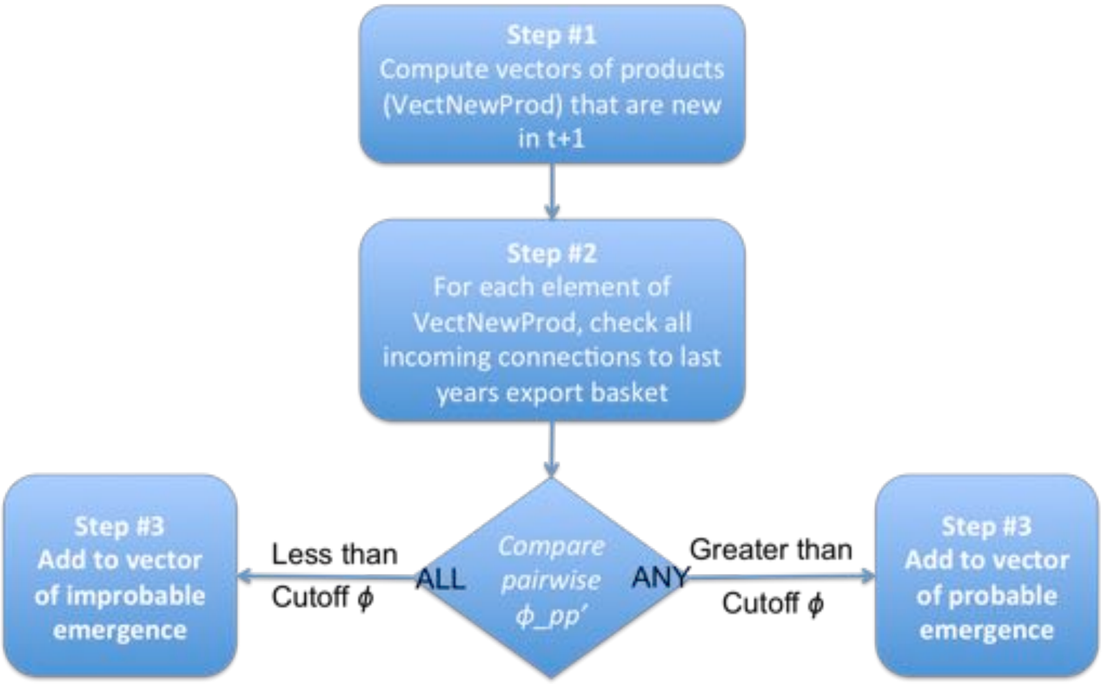
This measure constructed first, is a decomposition of the extensive margin to provide a structural measure of the *rate* of product emergence. Probable emergence is defined as any new products a country produces with RCA that has links to *at least one* existing export that has proximity greater than the median proximity value.[68] Improbable emergence is therefore when *all* connections between the new product and other existing products have values that are less than the median value.[69] Given the observed skewness in the distribution of proximity, median values are used as a better measure of the centrality of values.

Using the network analogy this operation classifies products in accordance with the type of diffusive jump between or within product clusters. If the jump is within a current export neighbourhood of products, this new product is classified as a near and probable emergence. If the jump is between two distant communities, and the new community does not contain any current exports (with RCA), this new product is classified as a far and improbable emergence. The algorithm used to compute the two emergence vectors, for each country, is presented in Figure 4.2.[70] The aggregation of the classified products provides a measure of the rate of emergence, or the number of new products that emerge close to existing products, in the case of probable product emergence.

To ensure the data is not subject to noisy single-year events, a persistence test is added that requires the new product to be exported in the following year immediately following the emergent event. If the same product is not exported in the subsequent year it is then dropped as an emergent product. Therefore, the emergence is classified as probable-persistent emergence and improbable-persistent emergence.

[68] Median values vary from year to year. But as a rough guide across all years this is approximately equal to 0.24.
[69] The median value used to compute these transitions is the median proximity value of both years of pooled proximity values.
[70] Given the nature of this cut-off style of analysis, sensitivity tests were conducted based on various values observed in the skewed distribution of proximity as presented in Section 4 when the measures are computing with product level export data.

Figure 4.2: Probable and Improbable Classification Algorithm



4.4.2 DIFFUSION MEASURES

To characterise the degree of diffusion, continuous time-series measures are computed at the country level, for each year-to-year transition period. Three such measures were calculated: the *average* of proximity values through diffusion (AvgProx); the *variance* of proximity through diffusion (VarProx); and the *width* of proximity through diffusion (WidthProx). Given $\rho = \frac{1}{2}(\phi_{pp',t} + \phi_{pp',t+1})$, a time averaged proximity value, and $n = \sum(P_t, P_{t+1} > 0)$ is the number of connections between new and existing products, each measure is defined as: [71]

Average proximity diffusion

$$\phi_{avg} = \frac{1}{n}\sum_{i=1}^{n}\rho_i$$

[71] Alternatives were considered such as using the base year and the next year. The average is used as it uses information contained in the proximity values of both years. Given that proximity values are found to change slowly over time, as measured by Pearson's correlation measure, this decision makes little difference to the metrics.

Variance proximity diffusion

$$\phi_{var} = \frac{1}{n}\sum_{i=1}^{n}(\rho_i - \mu)^2$$

Width proximity diffusion

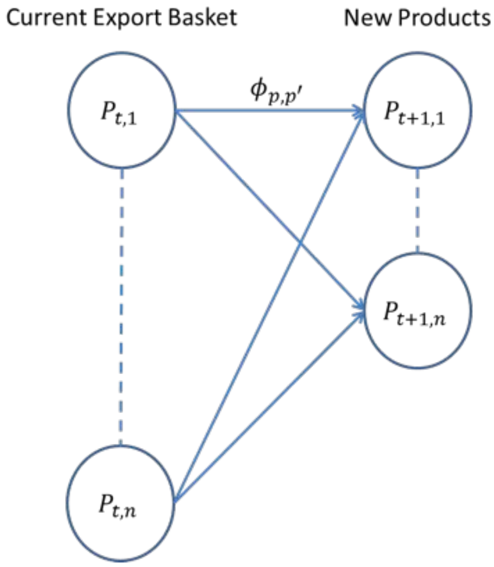
$$\phi_{width} = max(\rho_{1..N}) - min(\rho_{1..N})$$

Of the three measures, focus is placed on the variance and the width of the proximity values. A high *variance* represents a wider distribution of proximity connections between year t and year t+1. The *width* is the most direct measure of diffusion by taking the difference between the maximum and minimum proximity value when comparing adjacent year export baskets. A key benefit to the diffusion width measure is that it is insensitive to the overall distribution and is therefore a measure that will pick up relatively rare distant emergence events. Diffusion width also enables more direct interpretation when used in statistical regressions.

Averaging the $\phi_{pp'}$ values in this context does not perform well due to the large number of connections observed between emergent products and those that currently exist. If *all* new products emerged close to the existing products in the product space, you would expect a high average proximity. Conversely if *all* products emerged far away from the current export base, a low average proximity would be expected. As is documented in Hidalgo et al. (2007), in most cases products emerge with relatively close proximity to current products, which crowds out any rare jump in the product space, having a very small impact on the overall average.

To illustrate how each measure is computed, Figure 4.3 shows all pairwise connections that exist between a country's current export basket (at time t) and the same country's export basket the following year (at time t+1). Product similarities are captured by comparing $\phi_{pp'}$ values from each yearly cross section. Given $\phi_{pp'}$ values are computed cross-sectionally – an average value is considered as cross year link relationships.

Figure 4.3: Product relationships between new exports and current exports



Notes: 1. This diagram describes the product proximity values that need to be computed across adjacent years. **2.** Given $\phi_{pp'}$ is computed for every year (cross-sectionally) a transition $\phi_{pp'}$ needs to be computed between a country's export basket in time t and its export basket at time t+1.

From a country *capabilities* point of view, these metrics provide useful information in regard to changes in the unobserved *diversity* of capabilities being accumulated over time in a given country. New product emergence suggests that a country has acquired a new capability required to produce that good for international markets. By imposing a structural viewpoint for emergence, one can then indirectly comment on the degree of *diversity* of the newly acquired capability. Products that emerge close to current products are indicative of gains in capabilities that are relatively similar to capabilities a country has already acquired. If countries diffuse more *widely* then this is indicative of a *more diverse process* of accumulation in country capabilities, as the product that emerged is less similar relative to current exports. This is important as a more diverse set of capabilities is strongly correlated with future economic growth rates, as has been empirically demonstrated by Hausmann & Hidalgo (2014).

4.5 DATA AND COMPUTED METRICS

To conduct an analysis this chapter uses international trade data at the 4-digit level of the Standard International Trade Classification (SITC) Revision 2. To obtain the longest time series possible, two datasets were combined: (1) the National Bureau of

Economic Research (NBER) dataset, as developed by Feenstra et al. (2005), which covers the years 1962 to 2000; and (2) the Base pour l'Analyse du Commerce International (BACI) dataset, as developed by the Institute for Research on the International Economy (CEPII), for years 1998 to 2012. [72] Both NBER and BACI datasets are based on the UN Comtrade data and combine data from exporter and importer records to improve the overall coverage of data on bilateral trade flows. Due to the fact that data on exports from some developing countries can be extracted only from importer reports, these methods expand the set of countries available for analysis. Given there are three years of overlap from 1998 to 2000, the matched products are averaged to link the two datasets. The additional countries that are only contained in the BACI dataset, which are generally very small trading nations, are dropped in combining the two datasets for continuity.[73] The construction of the export dataset is documented in more detail in Appendix A.

As a first step, export data at the 4-digit SITC level are used to construct country level indicators of product diffusion. All country-product differences are computed for each year and the proximity relationship from each new product to each existing product is analysed to classify the products according to the type of emergence.[74] This analysis is then aggregated and combined with country-level data from the world development indicators published by the World Bank (2015) and the Wacziarg and Welch (2008) trade liberalisation indicators, to form datasets suitable for cross-country regression analysis. Two country groups are considered. The first contains all countries for which data is available, and the second focuses only on developing countries, as defined at the start of the dataset period in 1962. The full lists of countries are provided in Appendix F.

Three product classifications were considered for stage one computations. Dataset C refers to a dataset that is most similar to the raw NBER data. This consists of a mixture of official and unofficial SITC revision 2 product codes. Dataset D tracks official SITC Revision 2 data codes only, which provides a more consistent dataset across time, but

[72] Centre d'Etudes Prospectives et d'Informations Internationales (CEPII).

[73] This largely drops many small countries that would otherwise be dropped when the datasets are joined with WDI and other country economic variables.

[74] Functions that conduct this analysis have been written using the python programming language and are available on request.

drops 20 to 25% of trade by value between 1962 and 1984. Given proximity is computed cross-sectionally this does not pose a direct challenge to estimation, but it does impact on the computed level of product emergence in the transition years 1974, 1984, and 1998.[75] In these transition years, a number of new products are introduced due to changes in the underlying product classification system within the NBER dataset. Dataset E undertakes an algorithmic approach to harmonising the data through selective aggregation. This dataset is used as an important check of the robustness of the results to ensure product emergence is not simply a figment of the underlying transition dynamics contained in datasets C and D.[76] Summary statistics of the data used in the regression analysis are given in Table 4.2.[77]

Table 4.2: Export dataset statistics

	All Countries			Developing Countries		
Dataset ID	C	D	E	C	D	E
Number of Products	1066	783	283	1066	783	283
Number of Countries	129			107		
Years	1963 to 2012			1963 to 2012		
Observations	5469			4441		
Number of Countries under a Liberal Trade Regime						
Year 1963	21			6		
Year 2012	96			75		
Period 1963-1969	28			12		
Period 1970-1979	33			15		
Period 1980-1989	54			34		
Period 1990-1999	97			76		
Period 2000-2009	97			76		
Period 2010-2012	96			75		
Countries with Multiple Trade Liberalisation Episodes (#)	3 KEN (2), LKA (2), VEN (2)			3 KEN (2), LKA (2), VEN (2)		
Average Length of Trade Episodes	22.8 years (std. dev. = 17.0)			18.4 years (std. dev. = 14.2)		

Notes: 1. The number of products in this table is the total number of products in the underlying trade dataset that is used to compute the country level statistics of product emergence and each diffusion metric. The regression datasets are cross-country. **2.** List of countries is available in Appendix F. **3.** Product emergence data is computed from year-to-year transitions and therefore starts from 1963.

[75] The transition years are controlled for in the estimation through the use of year fixed effects as specified in the following section.

[76] As part of this research, additional datasets A and B were computed but are only intermediate datasets that were prepared to construct datasets C and D.

[77] See Appendix A for explanations regarding the construction of the underlying export datasets that are used to compute product emergence and diffusion statistics. The table presents the statistics relating to the regression dataset used in estimation.

The disaggregation level of the product classification naturally influences the level of product diffusion. More disaggregated datasets tend to produce higher overall levels of product emergence due to the availability of a greater number of products available for diffusion. Therefore each regression dataset C, D, and E is directly associated with the underlying export dataset that is used to compile the product emergence statistics in stage one. The number of products listed in Table 4.2 document the number of product codes available in each export dataset. For example, dataset C derives emergence statistics from an export dataset that contains 1066 mixed SITC rev 1 and SITC rev 2 product codes which has over 2.9 million export flow observations.

Other economic data used in this analysis (such as GDP and GDP per capita) are sourced from the World Development Indicators (WDI) dataset. Price and exchange rate data used in robustness checks are sourced from Penn World Tables version 8.1. The Sachs & Warner (1995) trade liberalisation measure used here, was updated to the year 2000 along with a wider country coverage by Wacziarg and Welch (2008). For the purpose of this study, given the time coverage of the export dataset, the trade liberalisation indicators were extended to 2012 under the assumption of the status quo.[78] The validity of this assumption is checked on a restricted dataset between 1962 and 2000.

The main explanatory variable is a Sachs-Warner measure of trade liberalisation that uses a binary value to classify if a country and open or closed economy. An economy is open if the following five criteria are not met for the entire duration of a give time period: (i) Non-tariff barrier coverage of intermediate and capital goods imports of 40 per cent or more; (ii) Average tariff on intermediate and capital goods imports of 40 per cent or more; (iii) A black market exchange rate that is depreciated by 20 per cent or more relative to the official exchange rate; (iv) A socialist economic system (as defined by Kornai 1992) and (v) A state monopoly on major exports (Sachs & Warner 1995).

[78] Robustness checks were undertaken by applying a year filter, running the regressions over the NBER dataset as an independent exercise. See Appendix F for further details.

4.6 TRADE LIBERALISATION AND PRODUCT EMERGENCE

Given the *a priori* view that trade liberalisation promotes export growth, a structural view of product emergence is now considered to understand both the type of new products that emerge, in addition to the relative distance of de-novo emergent products. This section first provides a short analytical narrative for the derived measures of export emergence and diffusion at both the global and country specific level. This is then followed by a global panel regression analysis to understand how the rate of export emergence and the diffusion patterns of products change in relation to episodes of trade liberalisation.

At the global level there is an increasing trend in the *rate* of product emergence from the 1960s through to 2012 as more countries integrate with world trade. As is evident in Figure 4.4, rapid growth in product emergence can be observed from the mid-1980s and through the 1990s, which overlaps with a well-recognised period of rapid globalisation. Spikes in the data for the years 1974, 1984, and 1998 correspond to transitions between differences in product classification. These spikes in the data will need to be controlled for with the inclusion of year fixed effects.[79]

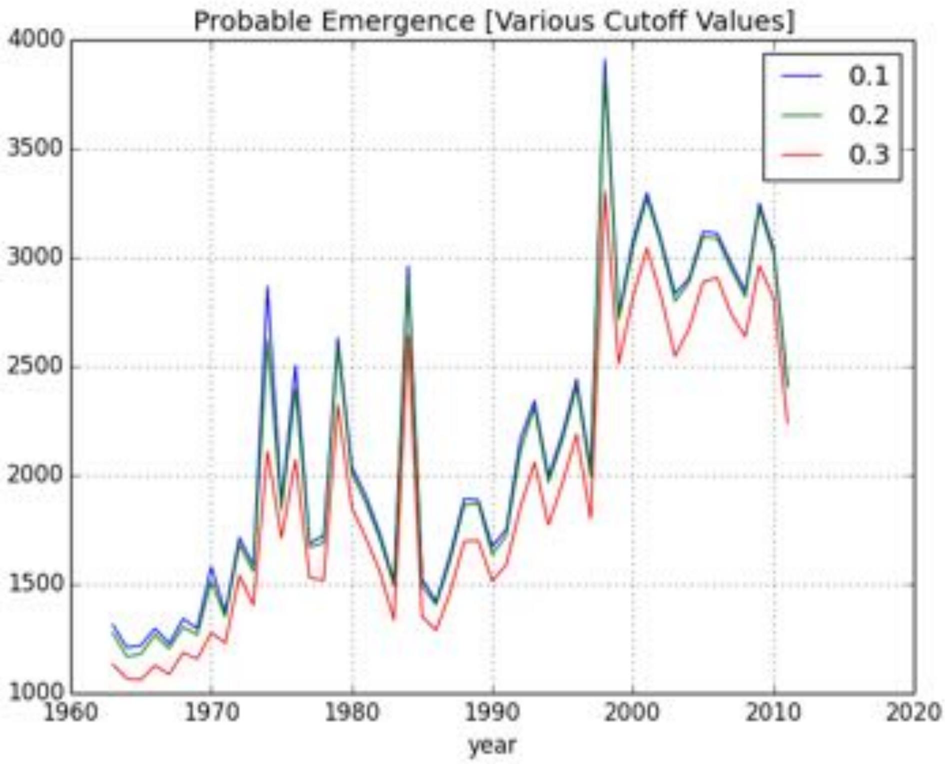
With a particular focus on the measures of probable and improbable product emergence, a sensitivity analysis of this first data-generating step is warranted due to the use of a cut-off in classifying between the two types of product emergent events.[80] Figures 4.4 and 4.5 demonstrate the aggregate affect on the metric using proximity cut-off values of 0.1, 0.2, and 0.3. These values represent a wide spread in the densest region of the proximity histogram. Improbable emergence is naturally more sensitive to the cut-off values due to the very nature of these events occurring infrequently. Reductions in probable emergence are simply increases in improbable emergence as one product is moved from one classification to the other. The dynamics of probable emergence are however remarkably stable over time, with only moderate changes in the level, given the large variance in cut-off values.

[79] Despite being particularly important controls for transition years, year fixed effects will be applied to all years in the regression to control for unobserved year specific heterogeneity.

[80] This sensitivity analysis to product cut-off definitions was conducted on Dataset D with SITC revision 2 codes at the 4-digit level.

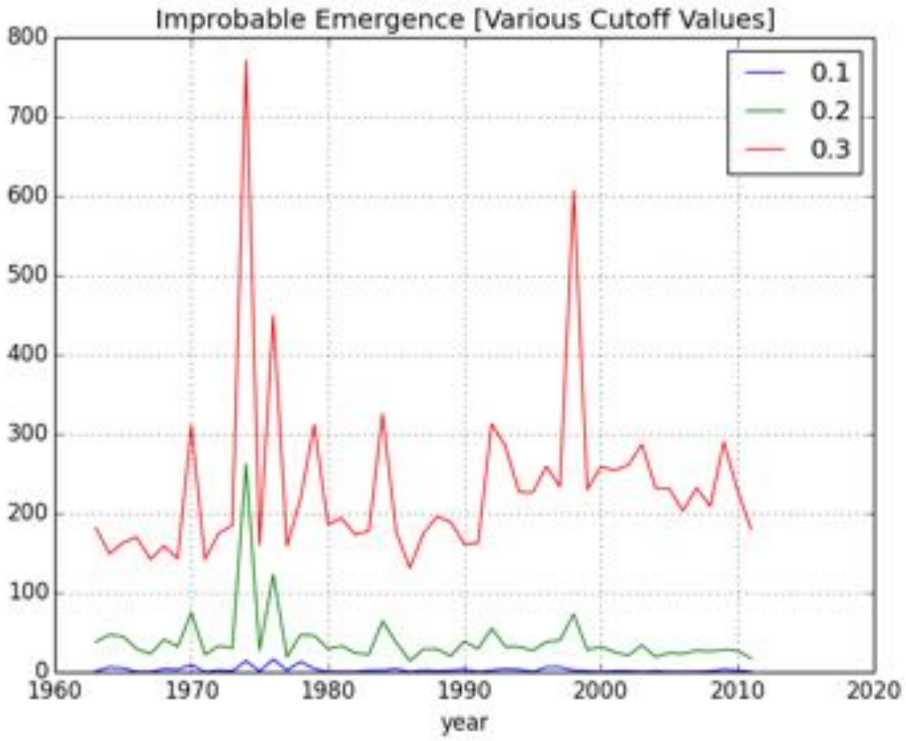
For example, if we were to use a cut-off value of 0.3 that sits to the far right of the proximity distribution, the results show that the vast majority of products that emerge have at least one close relationship with a current export within that country's export basket. It is noteworthy that connections greater than 0.3 constitute only 7.5 percent of all computed proximity values. This supports the finding of Hidalgo et al. (2007) which found that products emerge close to product nodes that a country already occupies in the product space network. Using this highly skewed cut-off point still only classifies 8 percent of emergent products as improbable in the year 2000.

Figure 4.4: Total probable product emergence for the Years 1963 to 2012



Notes: 1. This figure plots total probable product emergence from 1962 to 2012. Probable emergence is equivalent to the number of new products that emerge in country export baskets that shares at least one close proximity connection with an existing export. **2.** The cutoff values correspond the value used in classifying products as probable or improbable export emergence (algorithm defined in Figure 4.2).

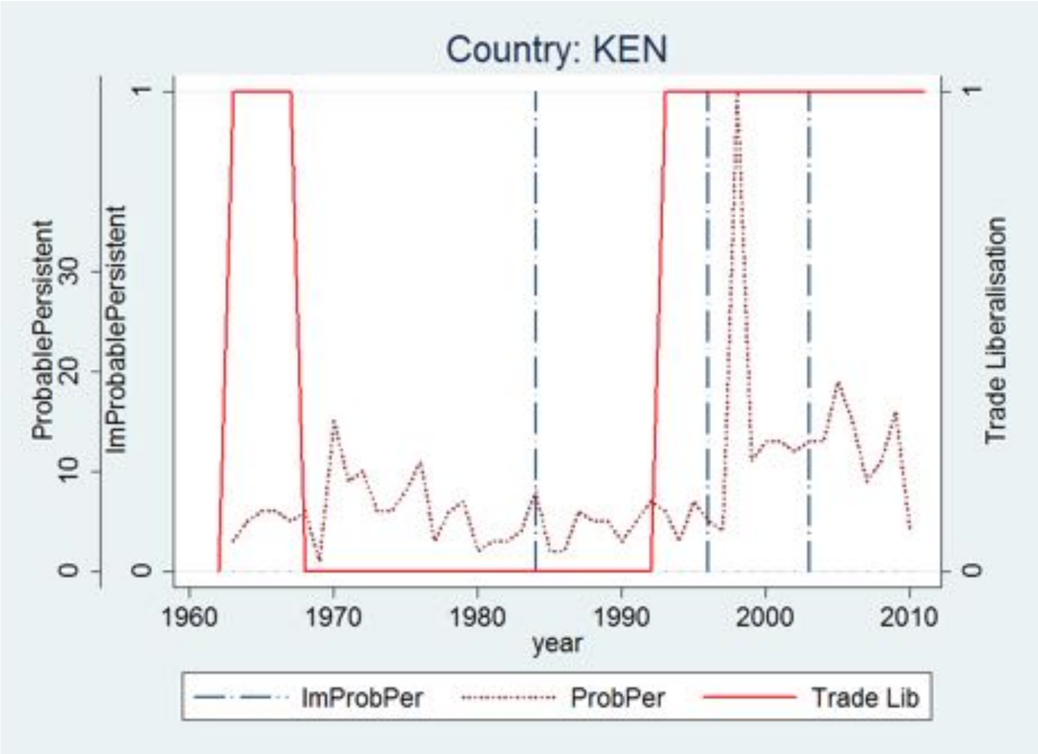
Figure 4.5: Total Improbable Product Emergence for the Years 1963 to 2012

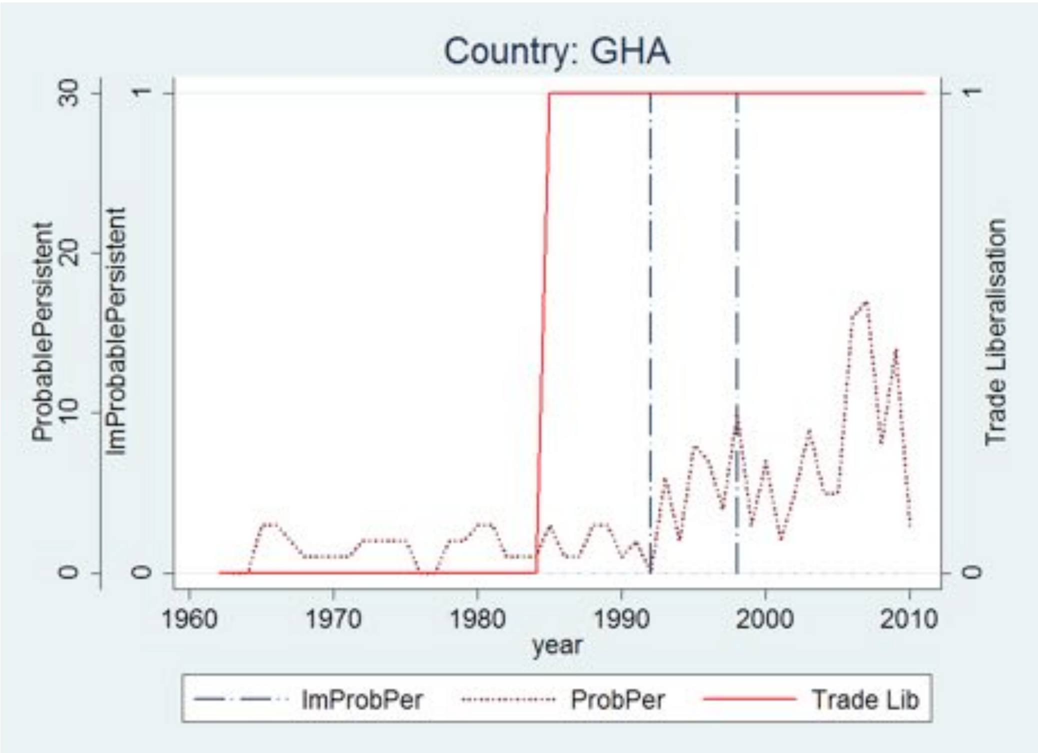


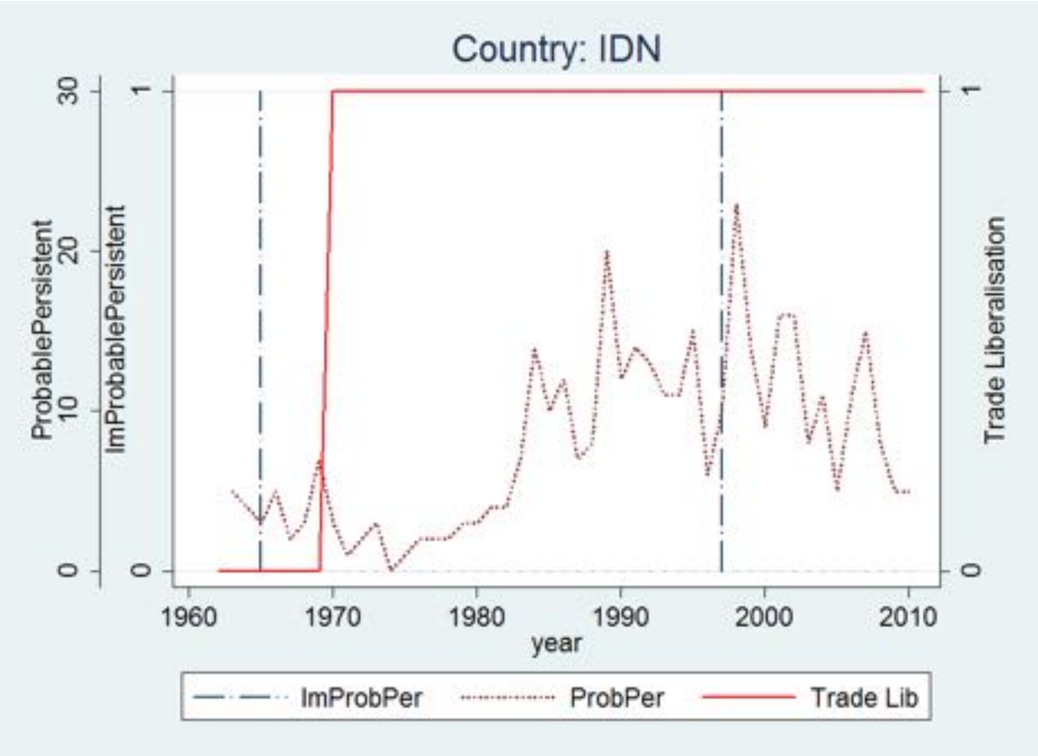
Notes: 1. This figure plots total improbable product emergence from 1962 to 2012. Improbable emergence is equivalent to the number of new products that emerge in country export baskets that does not share any close proximity with an existing export. **2.** The cutoff values correspond the value used in classifying products as probable or improbable export emergence (algorithm defined in Figure 4.2).

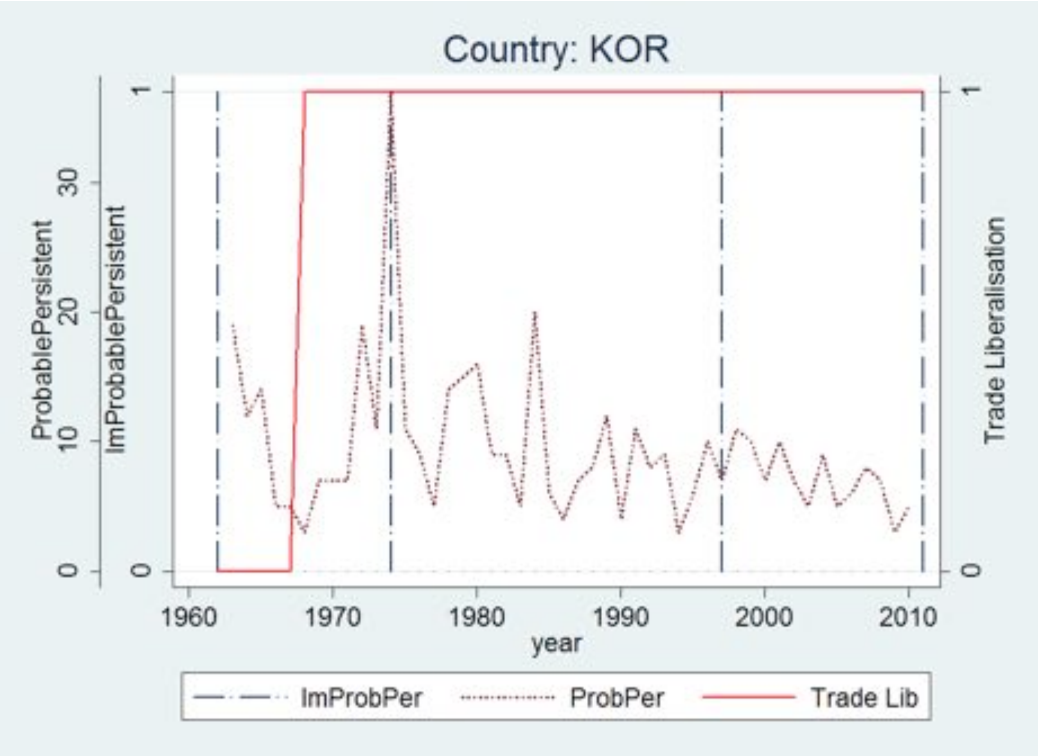
As a second step, four country case studies are presented in Figure 4.6. These graphs show the relationship between periods of trade liberalisation and the rate of persistent probable and improbable product emergence events. The four countries considered are Kenya (KEN), Ghana (GHA), Indonesia (IDN), and South Korea (KOR) and were selected as they represent two very different regions and have a variety of economic development histories. Asia is an important consideration given its broadly recognised emphasis on outward oriented trade policies. Africa provides insight into a region that has historically not benefited from the trade booms that have been observed in East and South-East Asia. Exploring these two contexts is important to more fully appreciate the links between different country contexts and the observable response from trade liberalisation. Persistent probable emergence is reported here as it captures only those products that persist for longer than a single year transition. The full statistical analysis considers both raw and persistent measures of export emergence.

Figure 4.6: Country examples of Probable and Improbable Product Emergence and Trade Liberalisation Episodes









Notes: 1. Countries are Kenya (KEN), Ghana (GHA), Indonesia (IDN), and South Korea (KOR). **2.** The vertical dotted bar is an indicator of an improbable product emergence event and the doted red line indicates the number of new probable products that emerge in a given year. **3.** The solid red line varies between zero and one and indicates periods of trade liberalisation.

The graphs show that for Ghana, Indonesia, and Kenya, periods of trade liberalisation appear to be associated with an increase in probable and persistent export emergence. Kenya and Indonesia rise to an average rate of 10 emergent products per year relative to a pre-trade liberalisation rate of approximately 5 emergent products. The pattern for Ghana is similar, with a more gradual rise in the rate of emergent products after trade liberalisation.

In the case of South Korea (KOR), the most developed of the four countries, one can observe a downward trend over time, after an initial increase in product emergence. This is likely to be due to a saturation effect. As countries become increasingly diversified in the product space there are fewer diffusion opportunities. This trend is also observed in other more developed countries such as the United States and Great Britain.

Another possibility for this trend is the rise of tradeable services that are not captured in this analysis. This analysis only captures trade in physical goods. Despite the overall trend, the level of South Korea’s product emergence is still a relatively large number of emergent products, at a similar level to developing countries of about 10 emergent and persistent products a year. Overall this paints a picture of interesting dynamics as exports come and go across the wide range of products.

The overall descriptive statistics of probable and persistent export emergence can be observed in Table 4.3.

Table 4.3: Descriptive Statistics of Probable and Persistent Export Emergence for the Years 1962 to 2012

All Countries	**Dataset C**	**Dataset D**	**Dataset E**
Mean	7.42	6.77	2.92
Std	9.49	8.77	3.31
Developing Countries	**Dataset C**	**Dataset D**	**Dataset E**
Mean	6.43	5.93	2.74
Std	8.18	7.72	3.35

The dynamic patterns of trade liberalisation are also interesting to observe. In the four cases that are presented, the increase in probable and persistent product emergence tends to be observed with a lag response after the declared classification of having liberalised trade. This observation is most apparent in the case of Ghana and

Indonesia. The lag in export response may be due to the time required for the general reorganisation of the underlying economy to take effect (through general equilibrium), or could be indicative of *effective* compared to *paper* liberalisation.

Motivated by these country-level observations, an econometric model is used to analyse these trends more rigorously. Two samples are considered, one that contains all countries, and a second that focuses on developing countries only (as declared by the UN at the beginning of the sample in 1962).

The econometric estimation strategy used is similar to that specified by Wacziarg and Welch (2008). The following model is estimated:

$$ProbableEmergence_{\mathrm{ct}} = \alpha + \beta_1 TradeLib_{\mathrm{ct}} + GDPPC_{ct} + Y_t + C_c + \varepsilon$$

where,

$TradeLib_{\mathrm{ct}}$ is the Sachs-Warner measure of trade liberalisation regime.

$GDPPC_{ct}$ is GDP per Capita (Constant US$ 2005) from WDI.

Y_t are year fixed effects.

C_c are country fixed effects.

ε is the standard error term.

Year and country fixed effects are added to control for unobserved year and country specific shocks respectively. The estimated coefficients for fixed effects are not reported in the regression tables. Year-specific shocks are particularly important controls to account for product emergence shocks that are observed in some years for datasets C and D.[81] Real GDP per capita (GDPPC) is included as a control for a country's level of development.

Additional control variables were also considered. World GDP growth (to proxy change in world demand) and world GDP levels as a proxy for the level of world demand were introduced to the model. Both were significant but did not impact on the estimated coefficient for trade liberalisation in any significant way. Additional robustness checks

[81] Year specific shocks are observed to be larger in some transition years due to changes in the underlying product classification system.

were also undertaken by introducing a linear time trend and 5-year average panels were also constructed. Both robustness checks delivered similar results to those reported. These have been left out of the regression reporting, as they are not a central component to the story.

This analysis focuses on the average within-country relationship between product emergence and trade liberalisation. While the fixed effects model only partially addresses the potential issue of endogeneity between product emergence and trade liberalisation, there are many country case studies that document the increase in *de novo* products following the effective implementation of trade liberalisation policies (Papageorgiou, Choksi & Michaely 1990; Athukorala & Rajapatirana 2000).[82] From an estimation perspective Santos-Paulino & Thirlwall (2004), studies the effect of trade liberalisation on exports and shows that regression results using Fixed Effects (FE) and General Method of Moments (GMM) estimation do not significantly differ, except that the GMM results provide statistically stronger relationships. Kehoe & Ruhl (2013) also finds that growth in the new good (extensive) margin of international trade coincides with trade liberalisation. Therefore there is both theoretical and empirical evidence that effective trade liberalisation leads to the emergence of new export products.[83][84] While GMM and Arellano-Bond dynamic estimation approaches were considered, fixed effects (FE) panel methods were used due to the large time span of the panel, in addition to the dummy variable nature of the trade liberalisation variable (Arellano & Bond 1991).

4.7 RESULTS AND DISCUSSION

The results for datasets C, D and E in a sample with all countries (developed and developing) are reported in Table 4.4. Trade liberalisation is associated with an

[82] Given country fixed effects only addresses time invariant unobserved country effects, estimating the model over different time periods (as presented in Appendix F) provides some reassurance that this assumption holds.
[83] The death of products is also a potential result of structural change and is an interesting avenue for future research.
[84] Reverse causality in which product emergence causes trade liberalisation is not directly addressed by the fixed effects model used in this chapter and is an avenue worthy of exploration to ensure the robustness of the results. The fixed effects model does however address concerns over omitted variable bias.

additional emergence of 1.8 probable and persistent products and approximately 2.6 probable emergent products on average.[85] Considering a sample of developing countries only (results found in Table 4.5), the estimate coefficients drop to an increase in 1.2 probable and persistent emergent products and 1.8 probable emergent products. This translates to a 24% increase relative to average rates of emergence across all countries, and an 18% increase relative to average rates of emergence across developing countries. Across all regressions, GDPPC is estimated to have a negative impact -0.1e-4 to -0.5e-4 and is statistically significant in all cases. As countries become more developed, the rate of product emergence tends to decline, which is likely due to fewer available opportunities for diffusion given developed economies already exhibit higher levels of export diversification.

Table 4.4: Fixed Effects Panel Regression over All Countries

	All Countries					
Export Dataset	C	D	E	C	D	E
SITC Product (#)	1066	783	283	1066	783	283
	Probable			Probable and Persistent		
Trade Liberalisation (0=NotLib, 1=Lib)	2.76*** (1.01)	2.64*** (0.92)	0.84** (0.39)	1.84*** (0.50)	1.89*** (0.45)	0.58*** (0.21)
Observations	5469	5469	5469	5342	5342	5342
Country Clusters	129	129	129	129	129	129
R-Squared						
Within	0.19	0.21	0.12	0.19	0.19	0.10
Between	0.09	0.02	0.02	0.13	0.03	0.00
Overall	0.01	0.07	0.07	0.03	0.08	0.05
Other Controls	GDPPC (Constant US$ 2005)***					
Fixed Effects	Country Year					
Years	1963-2012					

Notes: 1. This model has been estimated using stata and xtreg [options: fe and vce(cluster country)] with the addition of year fixed effects. **2.** The reported standard errors in this table have been adjusted for country clusters. **3.** Standard errors are reported in parenthesis. **4.** Levels of significance are indicated by pvalues ***= <0.01, **=<0.05, *=<0.1. **5.** GDPPC in constant US$ 2005 is sourced from WDI and is a control for the level of development. This is significant at the 1 percent level across all regressions. The estimated coefficients are –ve and are not reported. **6.** Additional controls such as XR/PPP were considered but did not change the estimate significantly. **7.** The country list for both datasets can be found in Appendix F. **8.** The lower coefficients for dataset 'E' is expected due to significant reduction in the number of available products in the underlying export dataset from which measures of emergence were computed.

[85] These results should be taken in consideration of an SITC level 4 dataset with 786 products.

To address the potential impact of underlying product reclassification over time, the regressions are replicated using dataset E as a robustness check. This dataset has been compiled as a dynamically consistent dataset (dataset E, Table 4.4 and Table 4.5). The large decline in the number of products is due to aggregation in the underlying export dataset from which emergence is computed, which naturally leads to an estimated coefficient lower in magnitude given the lower average number of products that are available for diffusion. However, the estimated coefficient is relatively proportionate when scaled by the number of available product candidates for diffusion (i.e. 0.8 * 786/283 = 2.2). The scaled value of 2.2 is similar in magnitude to estimates obtained from dataset C and D).[86]

Table 4.5: Fixed Effects Regression over Developing Countries Only

	Developing Countries Only					
	C	**D**	**E**	**C**	**D**	**E**
SITC Products (#)	1070	786	283	1070	786	283
	Probable			Probable and Persistent		
Trade Liberalisation (0=NotLib, 1=Lib)	1.73*** (1.02)	1.91** (0.95)	0.80** (0.41)	1.21*** (0.48)	1.36*** (0.45)	0.59*** (0.21)
Observations	4441	4441	4441	4336	4336	4336
Country Clusters	107	107	107	107	107	107
R-Squared						
Within	0.17	0.20	0.14	0.16	0.17	0.11
Between	0.08	0.00	0.00	0.13	0.02	0.00
Overall	0.03	0.10	0.09	0.03	0.08	0.07
Other Controls	GDPPC (Constant US$ 2005)					
Fixed Effects	Country Year					
Years	1962-2012					

Notes: 1. This model has been estimated using stata and xtreg [options: fe and vce(cluster cid)] with the addition of year fixed effect. **2.** The reported standard errors in this table have been adjusted for country clusters. **3.** Standard errors are reported in parenthesis. **4.** Levels of significance are indicated by pvalues ***= <0.01, **=<0.05, *=<0.1. **5.** GDPPC in constant US$ 2005 is sourced from WDI and is a control for the level of development. This is significant at the 1 percent level across all regressions. The estimated coefficients are not reported. **6.** Additional controls such as XR/PPP were considered and discussed but did not change the estimate significantly. **7.** The country list can be found in Appendix F.

[86] This simple calculation implicitly assumes that all products are equally likely candidates for diffusion. This is in all likelihood an invalid assumption, but useful in comparing the two estimated coefficient estimates.

Both sets of regression results are interesting on two fronts. First, they highlight that trade liberalisation is positively associated with a moderate increase in the rate of emergence in probable products. Second, by comparing estimates for *probable* with *probable and persistent* emergence, the results indicate that there are a significant number of products that do not persist and are not immediately observed in two adjacent time periods, which deflates the overall measure of product emergence.[87] Products that are persistent are more likely to provide a higher quality estimate by ensuring survivability of products for at least one additional year. This additional requirement filters out products that are close to the RCA cut-off and are simply fluctuating across this boundary.

An equivalent fixed effects panel analysis was conducted to explore the likelihood of an improbable product emergence event. The estimation strategy adopted focused on emergent events, where any improbable emergent outcome was recoded to be a categorical variable, and then likelihoods were directly estimated using the logistic model.[88]

The following model was estimated:

$$ImProbableEmergence_{\mathrm{ct}} = \beta_1 TradeLib_{\mathrm{ct}} + GDPPC_{ct} + Y_t + C_c + \varepsilon$$

where,

$TradeLib_{\mathrm{ct}}$ is the Wacziarg (2008) indicator for trade liberalisation regime.

$GDPPC_{ct}$ is the GDP per Capita (Constant US$ 2005) from WDI.

Y_t are year fixed effects.

C_c are country fixed effects

[87] This may be due to the measurement of RCA as a product close to the cut-off of 1.0 is sensitive to changes in country composition, in addition to changes in product composition in world exports. The effect of persistence considers emergence with sustained RCA values above the cut-off value for two years

[88] Given the small reduction in variation when converting the improbable product emergence to a categorical variable an OLS estimate was also conducted, which provides the non-significant results.

Both a fixed effects panel OLS over the pre-recoded data, in addition to the logistic model detailed above found no statistical relationship between improbable product emergence, or the likelihood of an improbable emergent event occurring more frequently under trade liberalisation. The estimates show no statistical relationship between trade liberalisation and improbable product emergence. At first this result was unexpected. Large jumps would be indicative of a larger increase in the underlying diversity of country capabilities, as revealed by increased diversity in exports. In retrospect, however, this is likely due to the extreme characterisation of improbable emergence. Trade liberalisation is more likely, if anything, to have a more moderate effect on the diffusion distance of newly emergent products.

A weakness of a cut-off-style of analysis is that it provides only one of many possible characterisations of distance that depends directly on the cut-off that is used. Consequently, intermediate levels of diffusion are missed entirely with the current definition of improbable emergence. Instead of large and improbable jumps, the attention is now turned to the continuous measures of diffusion distance, to test if trade liberalisation is associated with more moderate levels of diffusion distance to more proximal parts of the product space.

Of the three proposed metrics (average, variances and width) the width of proximity is the main focus, as it is considered to be the best indicator for the distance of diffusion given its insensitivity to the overall sampling distribution of $\phi_{pp'}$ values.

All three diffusion indicators are reported in similar panel fixed effects regressions:

$$AverageProximity_{\text{ct}} = \beta_1 TradeLib_{\text{ct}} + GDPPC_{ct} + Y_t + C_c + \varepsilon$$

$$VarianceProximity_{\text{ct}} = \beta_1 TradeLib_{\text{ct}} + GDPPC_{ct} + Y_t + C_c + \varepsilon$$

$$WidthProximity_{\text{ct}} = \beta_1 TradeLib_{\text{ct}} + GDPPC_{ct} + Y_t + C_c + \varepsilon$$

where,

$TradeLib_{\text{ct}}$ is the Wacziarg (2008) indicator for trade liberalisation regime

$GDPPC_{ct}$ is the GDP per Capita (Constant US$ 2005) from WDI

Y_t are year fixed effects, and

C_c are country fixed effect.

As is shown in Table 4.6, trade liberalisation is positively associated with all three measures. Notably evident are positive coefficients for both the variance and the width of diffusion distance. Taking these results together, this provides evidence that trade liberalisation is associated with increase in a country's participation in more moderately distant products.

The positive coefficient on trade liberalisation for an increase in average proximity lends support of an increased *rate* of probable products. This would increase the within-cluster density of exports in existing product neighbourhoods, which increases the overall density of high proximity values, moving the average proximity value in a positive direction.[89] For this reason it is not a good indicator of overall diffusion patterns as more distant products could still be emerging, but not at a rate that alters the average density of proximity values due to the relative rarity of these events.

Table 4.6: Trade Liberalisation and the Average, Variance, and Width of Proximity (FE estimates) [Regression Dataset 'D']

	Average Proximity	**Variance Proximity**	**Width Proximity**
Trade Lib (0=NotLib; 1=Lib)	0.012*** (0.004)	0.001** (0.000)	0.04** (0.017)
R-Squared (within)	0.15	0.09	0.16
Fixed Effects	Country (129) Year (50)		
Observations	5069	5069	5069
Years	1963-2012		

Due to ease of interpretation, particular focus is placed on the width of proximity relationships between new exports and current exports, as is presented in Table 4.7. Given the width of proximity, where $\phi'_{pp} \in \{0,1\}$ is a skewed distribution with the majority of values between 0 and 0.5, trade liberalisation is associated with a moderate increase in width of approximately 0.04 units on average. This constitutes an

[89] Throughout this thesis the words *community* and *neighborhood* are used interchangeably to signify a cluster of products that are more related to one another than they are to other products.

average widening in proximity values of approximately 8 percent of the effective region of the histogram.

These results, provide evidence that trade liberalisation is associated with a moderate increase in the underlying diversity of products exported from developing countries through wider diffusion in the product space. This increase in diversity is not only through the addition of new exports (which increases a country's diversification) but also through an increase in products that share less direct similarity in the post-reform export basket. A greater diversity in exported products is indicative of an increase in the *diversity* of the underlying set of country capabilities. As countries acquire more diverse capabilities, in line with the current literature, this suggests that they have access to new diffusion opportunities in addition to higher levels of observed economic complexity.

Table 4.7: Width of Diffusion for All Countries

SITC Products (#)	**All Countries**			**Developing Countries**		
	C	D	E	C	D	E
	1070	786	283	1070	786	283
	Width of Diffusion			Width of Diffusion		
Trade Liberalisation (0=NotLib, 1=Lib)	0.037** (0.018)	0.041** (0.017)	0.033** (0.014)	0.041** (0.018)	0.038** (0.018)	0.032** (0.015)
Observations	4534	5069	5458	3899	4191	4432
Country Clusters	129	129	129	107	107	107
R-Squared						
Within	0.13	0.15	0.10	0.13	0.17	0.11
Between	0.06	0.07	0.10	0.01	0.02	0.04
Overall	0.01	0.01	0.00	0.05	0.06	0.03
Other Controls	GDPPC (Constant US$ 2005)					
Fixed Effects	Country (107) Year (49)					
Years	1963-2012					

Notes: 1. This model has been estimated using stata and xtreg [options: fe and vce(cluster cid)] with the addition of year fixed effects. **2.** The number of observations tends to decrease with more disaggregated export data due to the higher presence of missing data. When trade data is classified at higher levels of disaggregation individual products find it more difficult to meet the RCA test, which produces more years with no emergent events.

These results should be viewed in light of some important data considerations. In the 1962 to 2012 dataset, there are some countries clustered at the beginning and the end of the panel, which may not offer enough within-country variation when comparing pre and post trade liberalisation periods. For example, South Korea (as shown in Figure 4.6) liberalised in 1968, which means there is only 6 years of available data pre-liberalisation. To account for this, robustness tests are performed that consider different periods of time. The results are robust to restricting the dataset to 1984 to 2000, a period where the NBER world trade flow dataset is most consistent in product classification at the highest level of product disaggregation. These regression results are reported in Appendix F.

To check the effect on extending the Wacziarg and Welch (2008) trade liberalisation indicator to 2012, a robustness check was conducted for the years between 1962 and 2000, which aligns naturally with the NBER WTF dataset. The results remain robust to this restriction, albeit with some observed increases in the estimated standard error, rendering less significant results. In one instance using dataset E renders an estimated coefficient for trade liberalisation that is just larger than the 10 percent level. While discussed further in Appendix F, one possible reason is the reduced product variation found in dataset E with 63 percent fewer products available for diffusion than Datasets C and D.

Additional control variables were also considered from the PENN world tables, such as the exchange rate (XR), in addition to the price level (XR/PPP). The addition of these controls tends to increase the significance of the trade liberalisation estimate with little change to the estimated coefficient value.

As a final check, the full 1962 to 2012 Dataset was converted to a 5-year average panel and estimated using the same regression specifications where the year fixed effects are replaced with period fixed effects. Within any given period in which the trade liberalisation indicator was found in more than three years it was deemed liberalised for the entire 5-year period. All estimated coefficients were found to be similar under these conditions.

4.8 CONCLUSION

The *a priori* view is that trade liberalisation promotes export growth by reinforcing current export products and enabling the emergence of new products in line with a country's comparative advantage. In this chapter, new tools from the field of network complexity analysis have been used to conduct a systematic analysis of the structure of export emergence. New metrics were constructed to characterise how exports emerge over the period 1962–2012 to observe how they change following trade liberalisation.

This research was carried out in two stages. The initial construction of the measures of export emergence from disaggregated country-product SITC level 4 export data; followed by the consideration of the links between both the *rate* of product emergence and the *distance of diffusion* in relation to periods of trade liberalisation. While trade liberalisation has no statistical relationship with improbable emergence, it was positively associated with an increase of between 1 and 1.8 additional emergent probable and persistent products on average under a liberal trade regime. This faster rate of product emergence is suggestive of an accumulation in the underlying country capabilities required to produce these goods. In addition to the rate of emergence, this study also finds that trade liberalisation is associated with a moderate increase in the distance of diffusion through the product space. A larger distance of diffusion is indicative of greater diversity in the attainment of underlying country capabilities through the process of diffusing to less similar products in the product space network. Therefore the results indicate that trade liberalisation has a positive role to play in the accumulation of country capabilities as measured through this analysis on the export channel.

Understanding factors that encourage countries to diffuse into new clusters may shed light on new policies for accelerating the development process. There are many avenues to consider for future research: (1) A richer bipartite network representation of proximity that uses non-symmetric conditional probabilities, which in effect produces a directed graph to study directed diffusion patterns, (2) Inclusion of a dynamic model to understand if trade liberalisation generates a temporary short term increase or sustained long term increase in new emergent products, (3) Alternative characterisations of jumps could be investigated to gain further insight into what clusters are important. Are jumps into “core” clusters (those found in the densest part of the network) the only growth inducing “jumps” or is it that economic diversity is the important factor?, (4) Studying typical cross-country diffusion patterns to explore whether there are consistent emergent product paths, as evidence for path dependencies which may shed further light on effectiveness of guided industrial policy, and finally (5) Understanding how the import structure changes during trade liberalisation (i.e. capital-intensive imports (with embodied technology) versus a larger variety of products).

CHAPTER 5 PRODUCT FRAGMENTATION AND TRADE PATTERNS

"The very nature of the process of global production sharing is the continuous shaking up of industry through the emergence of new products and production processes in place of the old ones. Engagement in a manufacturing process involving a variety of goods and inputs could contribute more to growth than perpetual specialisation in a narrow range of products" (Athukorala, 2014)

Abstract

This chapter explores the implications of global production sharing for the measurement of economic complexity. The analysis is based on the separation of production sharing based (PSB) products from the standard (reported) trade data. Production sharing trade is further disaggregated into parts and components (PC) and final assembly products (FAP). It is found that failure to distinguish between product sharing based trade and standard horizontal trade tends to upwardly bias the complexity rankings of exports of some developing countries. It is also found that PC tend to be located in the densely connected core of the product space alongside associated final products. In the case of the machinery and transport equipment sector, PC and FAP tend to emerge in country export baskets prior to the export of final machinery products. This suggests that PSB products act as gateway products in the diffusion process from the periphery to the core of the product space.

5.1 INTRODUCTION

The manner in which goods and services are produced has become increasingly internationalised as part of the ongoing process of economic globalisation. Through the convergence of new production technology coupled with the emergence of low transport costs, production processes are increasingly split up and different slices of the process are undertaken in countries that are most efficient in their production. The electronics industry offers a clear example of this phenomenon. The Apple iPod was designed in the USA, with components manufactured in Japan, South Korea, Taiwan, and the EU with final assembly located in China prior to being shipped to target retail markets such as the USA (Linden, Kraemer & Dedrick 2009; Shin, Kraemer & Dedrick 2012). Though most commonly discussed in the context of electronics and automotive industries, 'global production fragmentation' has over the past four decades extended to many types of industries such as the apparel, footwear, and furniture industries (Yeats 2001).[90]

This chapter explores this product fragmentation phenomenon in trade by considering parts and components (PC) and final assembly products (FAP) in the context of the recent literature on economic and product complexity and the product space (Hidalgo et al. 2007; Hausmann & Hidalgo 2014). Product fragmentation has broader implications for the analysis of international trade. The recent literature on economic complexity is based on analytical methods that determine measures of *revealed* economic and product complexity based on the assignment relationship between countries and their exports. This assignment relationship is what determines the bipartite network structure used to compute the complexity measures. The main issue is that the process of product fragmentation distorts the information that is contained in these assignments as some exporters of complex products are engaged only in their assembly.

This chapter examines the implications of product fragmentation for the measurement of economic complexity using the new network analytical techniques. The chapter is structured into three primary contributions. First, trends and patterns of product fragmentation are examined using disaggregated international trade data for the years 1995 to 2013.[91] Second, the link between final assembly products and revealed measures of economic and product complexity is considered by altering the bipartite mapping of country-product relationships and introducing differentiated final assembly products based on income levels. As an approximation, low GDP per capita countries are assumed to engage primarily in the labour-intensive final assembly stage of production, while high GDP per capita countries have typically acquired the full suite of capabilities to produce a much larger portion of the value chain. An alternative

[90] Alternative terms used in the literature include international production fragmentation, offshoring, vertical specialisation, and slicing up the value chain.

[91] Data are sourced from the recent BACI dataset Guillaume & Zignago (2010).

justification for this characterisation is that high GDP per capita countries may also produce high quality specialised goods that are differentiated within the same SITC category.[92] This adjustment leads to lower economic complexity rankings for countries such as China, Malaysia, and Thailand – in addition to reduced product complexity rankings for the low-type differentiated product. Finally, the analysis finds that PC and FAP play a key role in producing a dense, highly connected core in product space network. Industry-specific PC and FAP are highly related to final products within their respective product space communities. In the case of the machinery sector, PC and FAP serve as gateway products as they appear in country export baskets prior to the emergence of more complex final products within that sector.

The chapter is organised as follows. Section 5.2 provides an overview of the ongoing process of product fragmentation and trade patterns to provide the context for the ensuing analysis. Section 5.3 details the dataset used, which is constructed from the BACI HS92 dataset for the years 1995 to 2013. Section 5.4 then considers a traditional analysis of this new dataset to supplement the current literature on product fragmentation. Section 5.5 explores the linkages between product fragmentation and the observable role that PC and FAP play in revealed measures of economic complexity. This is discussed in two sub-sections, 5.5.1 introduces differentiated FAP products to disrupt the bipartite network structure, and 5.5.2 conducts an dynamic analysis to understand what role PC and FAP products play as countries diffuse through the product space network over time. Section 5.6 provides conclusions and suggestions for further research.

5.2 THE PROCESS OF PRODUCTION FRAGMENTATION

The expansion of production sharing has primarily been driven by four mutually reinforcing determinants. First, improvements in manufacturing technology have allowed the production of some final goods to be split into smaller discrete “mobile” segments. As a consequence, these production segments are produced by firms in the

[92] This assumption stems from country case studies in which may countries in developing Asia have primarily participated in the assembly services of a product chain. This cutoff was chosen as a differentiator between the advanced Asian economies (i.e. Japan, Korea, and Taiwan) and developing Asian economies (i.e. China, Indonesia, Malaysia, Philippines)

most efficient (least cost) way before being brought together for final assembly. Second, improvements in global communication, transportation, and trade infrastructure have reduced the barriers that previously inhibited production to be located in many geographic locations. Communications crucially assist the organisation of production, while falling transport costs and improvements in trade-related infrastructure have significantly reduced the effective distance between different regions of the globe. Third, institutional environments have evolved in host nations, to a point that ensures timely delivery and reasonable levels of contract enforceability forming enduring commercial partnerships. Fourth, trade liberalisation policies over the past four decades have greatly reduced the economic barriers to trade through broad based reductions in tariff rates and the removal of import quotas, stimulating an era of outward orientation and globalisation. While not exhaustive, this list captures many of the key determinants driving the change.[93]

Production fragmentation is not a new phenomenon, however the nature and spread of production sharing is continuously evolving as new enabling technologies continue to emerge. In the 1960s product fragmentation largely consisted of simple two-way relationships between a multinational enterprise (MNE) and a subsidiary located in a host country (Brown & Linden 2005). Parts and components were shipped to subsidiary firms that assembled them and returned final goods to the parent country for any final processing and retail. As these operations became more firmly embedded in the host country, technological spillovers and production know-how spread with the emergence of local industries – often leveraging initial MNE networks and enterprise activities (Swenson 2008; Ahn, Khandelwal & Wei 2011; Head, Jing & Swenson 2014). Firms in developed countries then began moving final assembly of a wide range of goods, not just to subsidiaries, but also to contracted firms in developing countries for final assembly (Krugman 2008). Within these product segments, firms in developed economies started to reshape themselves to specialise in research and development, and design tasks while dismantling their domestic manufacturing capabilities due to the emergence of product assembly networks.

[93] Additional determinants such as smaller minimum efficient size of scale for global corporations, more efficient capital markets, and agglomeration are discussed further in Price (2001).

While this process was initially motivated by component assembly and testing, it has now led to the creation of complex global production networks, whereby specialisation occurs in different segments of the value chain. Firms, across many industries, have migrated from building a product from start to finish in one geographic location in favour of leveraging vertical specialisers to produce different segments of a product in a location determined by cost, quality and intra-firm networks. Coupled with industry "standards" and interchangeability of parts, firms now have the capacity to not only engage other businesses for final assembly tasks but to also leverage a wide range of high quality parts and components developed to common industry specification.

The empirical evidence suggests that trade in parts and components has accounted for a growing share of total manufacturing trade over the past few decades (Hummels, Ishii & Kei-Mu 2001; Athukorala 2005, 2010). Starting in the 1960s the deepening of product fragmentation has resulted in rapid industrial expansion, predominantly throughout Asia, supplying both parts and components and providing final assembly services. Two main approaches have been applied in the empirical literature to measure and describe the evolution of product fragmentation. The first approach makes use of input-output tables in an effort to track how traded and domestic goods flow into final products (implicitly treating each country as a closed manufacturing system). Hummels et al. (2001) use this technique with 10 OECD and four emerging market countries and find that product fragmentation accounts for 21% of these countries' exports. In addition Hummels et al. (2001) finds this type of trade grew almost 30% between 1970 and 1990. The main limitation with this approach is the extensive data that is required to construct input-output (IO) tables for a large sample of countries. As a result IO tables are generally aggregated to the sector level, rather than the product level, and have more limited country coverage.

A second approach is to consider decompositions of the SITC product classification, made possible in more recent decades with the inclusion of parts and components product codes in international trade statistics. Yeats (2001) uses this approach to measure the rise in parts and components in trade statistics.[94] While it is well documented that this approach most likely underestimates the true magnitude of

[94] This survey was undertaken using SITC revision 2 trade data (Yeats 2001)

product fragmentation, it provides a detailed view into those industries that are reported, such as electronics and machinery. The expected underestimate is due to the limited coverage in some industries, such as textile and apparel, that have been shown to be susceptible to the forces of fragmentation but contain fewer parts and components details in the SITC product classification (Yeats 2001; Athukorala 2014). Naturally, industries that are more conducive to fragmentation have a broad array of parts and components defined within the SITC classification system and are the focus in these empirical papers. Yeats (2001) found parts and components accounted for 30 percent of machinery and transport equipment trade and were growing at a rate faster than the overall SITC chapter 7 total growth rate. An increasing share of parts and components in manufactures provides convincing evidence of a deepening trend towards product fragmentation. Given the limitations in the data, Yeats (2001) uses alternative OECD data and finds supporting evidence that product sharing is also a significant phenomenon in other industries, such as the textiles and clothing, leather goods, footwear, and other labour-intensive manufactures.

In a more recent study, Athukorala (2011) uses an updated classification of parts and components and shows that this trend of increasing product fragmentation has continued unabated through the 1990s and the first half of the 2000s with parts and components accounting for 29.7 percent of manufacturing exports globally. This study also highlights how East Asia operates at the very core of the product fragmentation story and has emerged as a dominant beneficiary. Parts and components have not only become a significant share of total exports, intra-regional trade statistics also show a vibrant and significant production sharing network between countries within East and South East Asia (Athukorala 2014). Overall product fragmentation has played an important role in opening up opportunities for countries in East Asia to participate in global manufacturing (Brown & Linden 2005; Athukorala 2005; Swenson 2005, 2006; Athukorala & Waglé 2014).

This evolutionary process towards increasing international product fragmentation introduces new challenges for the study of international trade flows. Parts and components typically cross multiple country borders, causing double counting, as they flow through the supply chain towards final destination markets. As highlighted by "teardown" studies, a product is assigned the full value to country export baskets, despite only participating in some fraction of value added (Dedrick, Kraemer & Linden 2009). For example, the iPod's full value is exported from China while only a small fraction of value added was actually provided by China in the task of final assembly. As a result product fragmentation produces a complex web of interactions between exports, imports, and domestic production, and therefore increases the difficulty of understanding true measures of value added. This in turn distorts the information provided by current trade datasets that acts as an input to the analytical methods used in the literature on economic complexity.

5.3 DATA

For the purpose of this study, the empirical analysis makes use of the BACI (HS 1992) dataset compiled by CEPII using the UN Comtrade dataset. This dataset provides time coverage from 1995 to 2013 and is classified according to the HS 1992 system. The data are concorded to the SITC revision 3 product classification in order to match with the latest available parts and components list.[95] The details of the construction and description of this dataset are given in Appendix H. The basic statistics are detailed in Table 5.1.

Table 5.1: Basic Statistics of Trade Dataset

	Trade	Export	Import
Number of exporters	204	204	-
Number of importers	204	-	204
Number of total products	3112		
Number of manufactured products	2329		
Number of final manufactured products	1975		
Number of parts and components	354		
Number of trade flows	87,684,655	5,208,779	8,500,433
Years	1995 to 2013		

Notes: 1. Full details of the construction and description of this dataset can be found in Appendix H. **2.** This dataset is concorded to SITC revision 3 level 5 using the standard UN conversion tables. **3.** Parts and Components are provided by Athukorala (2014)

In addition to the trade data, various product classifications are also used to obtain a deeper understanding about the categories of products impacted most by the process of product fragmentation. Both the Leamer product classification, based on a factor

[95] The parts and components list can be found in Appendix H and is an updated version from Athukorala (2014)

content perspective, and the Lall technology product classification are used to characterise trends in product fragmentation for products within these different categories (Leamer 1984; Lall 2000). In addition, the product space communities sourced from Hausmann & Hidalgo (2014), which have been shown to be broadly consistent with the Lall and Leamer classifications, are used to explore a more disaggregated perspective.[96, 97] The product space *communities* are originally constructed in SITC revision 2, which required them to be matched to the SITC revision 3 data through the UN SITC concordance.

The parts and components classification is from Athukorala (2014), and final assembly products are defined as SITC chapters 75, 76, 78, 87, and 88 net of any sub-chapter parts and components. Other standard economic development data are provided by the World Development Indicators.

One important consideration to note is that trade share estimates are likely to be lower, when using this dataset, as BACI provides greater country coverage through the harmonisation of both exporter and importer reports available through UN Comtrade. The effect of this harmonisation process increases the total value of trade, due to the increased number of trade flows, which effectively increases the denominator (total manufacturing trade) when compared with raw UN Comtrade data.[98] Also, taking a direct analysis based on product groups that are defined by the international trade classification, favours industries that have detailed coverage defined within the classification. An example of this limitation is the clothing and apparel industry. While this industry is heavily involved in product fragmentation, it is not effectively captured by the SITC classification due to fewer specific parts and components product codes. While acknowledging the constraints of the data, there is value to be gained by focusing on PC and FAP product categories, in industries that are defined, as it

[96] Product space communities are defined by a clustering technique in which statistical relationships in the product space network are uncovered using random walkers, which over time discover the clusters of related notes.
[97] The actual data were downloaded from Michele Coscia's dataset repository located at: http://www.michelecoscia.com/?page_id=223.
[98] Total manufacturing trade is defined as SITC chapters 5+6-68+7+8. The parts and components list can be found in Appendix H

provides insight into a large component of trade that is subject to the forces of fragmentation, such as Machinery and Electronics.

5.4 PRODUCTION FRAGMENTATION AND WORLD TRADE PATTERNS

Prior to considering how product fragmentation impacts the recent measures of economic complexity, this section reviews the dataset compiled here in relation to the existing literature for two primary reasons. First, is to ensure that the BACI dataset provides a consistent story with reference to previous literature that is primarily based on raw UN Comtrade data. Second, to provide an updated view of the product fragmentation process through to 2013.

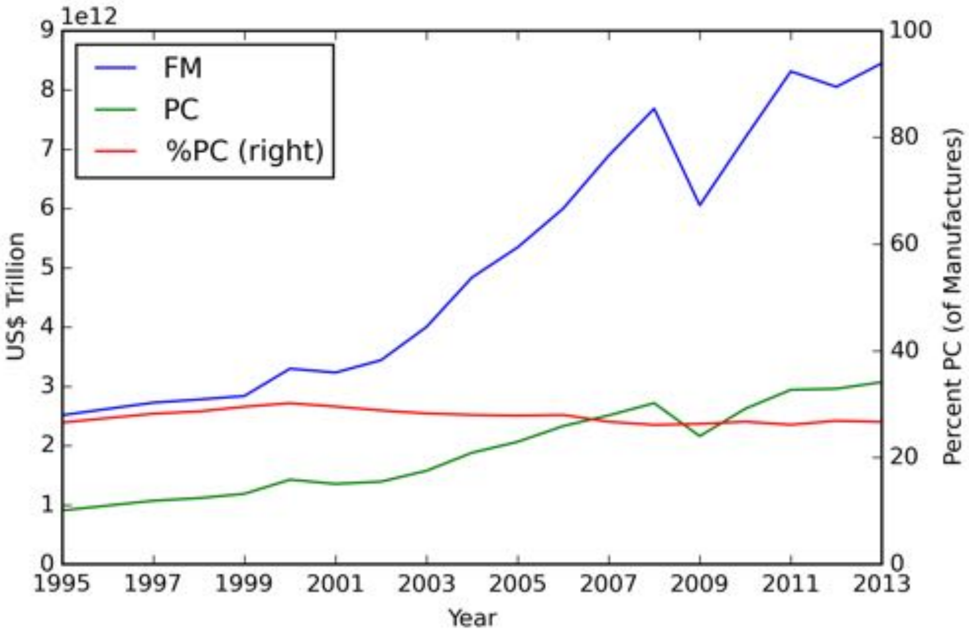
The existing literature documents the rise of product fragmentation over the past four decades, showing that parts and components have risen as a share of manufactures from 15% in the 1980s to 30% by the mid-1990s (Yeats 2001). These estimates are also similar to more recent studies that extend the analysis from the 1990s through the 2000s (Athukorala 2005, 2014; Athukorala & Yamashita 2006; Athukorala 2012).[99] This study uses a different data source, the BACI dataset, and finds broad agreement with the previous findings, albeit at slightly lower shares uniformly spread across time. At the global level the share in manufactures increases to a peak of 30% in 2001 with a softening to 26.4% in 2013, as shown in Figure 5.1.

To examine the rise of parts and components, disaggregated manufacturing trade data were grouped into three primary groups: parts and components (PC), final assembly products (FAP), and other final manufactures. Final assembly products are those that are considered more conducive to product fragmentation, and consist of electronics, automobiles, machinery, and specialist equipment. These particular product categories have detailed parts and components defined within each SITC chapter, and are frequently studied in the product fragmentation literature. Throughout this chapter both parts and components, in addition to final assembly products, are collectively referred to as production sharing based (PSB) products.

---

[99] Some of the most recent studies use SITC revision 3 as it provides greater resolution of parts and components.

Figure 5.1: World Exports of Parts and Components (PC) and Final Manufactures (FM)



Notes: 1. Percent Parts and Components (PC) is the share of parts and components within manufacturing exports.

The geographic landscape of PSB trade paints a much richer picture of the underlying dynamics of trade, as global production shifts between different countries and regions. What is observed are declining world shares in PC among many developed regions and rising shares in many developing countries, particularly those located in East and South East Asia. East Asia now accounts for almost half of world trade in PSB products as shown in Table 5.2 (Export) and Table 5.3 (Import). [100] In 1995 China produced 2.9% of the world's parts and components, by 2013, this share has risen to 21.9%. Asia as a region now accounts for 48.5% of total PSB products, with 38% coming from East Asia alone. On a global scale, as part of this widespread geographic shift due to product fragmentation, the following declines in share of world PSB trade in developed countries have been observed. North America, show declines from 16.6% in 1995 to 11.0% in 2013; Western Europe, from 26.8% in 1995 to 19.3% in 2013; and Japan, from 17.0% to 7.3% in 2013 in shares of world PSB trade.

[100] East Asia consists of China, Hong Kong, Japan, South Korea, Macau, Mongolia and Taiwan

Table 5.2: Geographic profile of Total Export and PSB Exports (%)

Export		Total			Manufacturing			Production Sharing Based (PSB) Products: Total PSB Products			Parts & Components			Final Assembly			Share PC in PSB Products (%)		
		1995	2005	2013	1995	2005	2013	1995	2005	2013	1995	2005	2013	1995	2005	2013	1995	2005	2013
Asia		29.8	35.5	40.7	31.8	37.6	43.8	34.4	42.1	48.5	34.6	40.3	47.0	34.4	42.1	48.5	57.8	53.0	53.0
Eastern Asia		19.9	21.6	23.3	24.5	27.5	32.3	27.3	31.9	38.4	26.5	30.2	37.0	27.3	31.9	38.4	55.8	52.4	52.6
	JPN	9.5	6.2	4.2	12.1	8.1	5.9	17.3	10.6	7.8	17.0	9.3	7.4	17.3	10.6	7.8	56.6	48.7	51.6
	CHN	4.4	9.2	12.8	5.1	11.7	18.3	2.9	11.8	20.3	2.9	12.6	21.9	2.9	11.8	20.3	57.5	59.0	59.2
	HKG	1.2	0.8	0.9	1.5	1.0	0.6	1.1	1.0	0.6	1.2	1.0	0.7	1.1	1.0	0.6	59.8	55.5	56.7
	TWN	2.0	2.4	2.2	2.4	3.1	3.1	2.3	4.0	4.5	2.7	3.1	2.4	2.3	4.0	4.5	68.5	43.2	29.5
	KOR	2.7	2.9	3.3	3.3	3.7	4.3	3.7	4.5	5.2	2.6	4.2	4.6	3.7	4.5	5.2	41.5	51.1	47.9
South-Eastern Asia		5.8	6.6	6.9	5.1	6.5	6.8	6.4	8.5	7.9	7.3	8.4	7.7	6.4	8.5	7.9	66.0	54.9	53.3
	IDN	1.1	1.0	1.2	0.8	0.7	0.7	0.3	0.5	0.5	0.3	0.7	0.5	0.3	0.5	0.5	58.0	69.1	61.0
	MYS	1.4	1.6	1.5	1.3	1.8	1.4	1.9	2.7	1.9	1.9	2.5	1.6	1.9	2.7	1.9	56.5	50.7	46.4
	PHL	0.3	0.6	0.4	0.3	0.7	0.5	0.3	1.3	0.9	0.4	1.0	0.9	0.3	1.3	0.9	69.8	43.3	53.6
	SGP	1.6	1.6	1.5	1.6	1.7	1.5	2.6	2.3	1.8	3.2	2.4	1.5	2.6	2.3	1.8	70.6	58.3	44.9
	THA	1.2	1.2	1.4	1.1	1.3	1.5	1.2	1.6	1.9	1.5	1.7	1.9	1.2	1.6	1.9	72.8	59.4	55.5
	VNM	0.1	0.4	0.8	0.1	0.3	1.0	0.0	0.1	1.0	0.0	0.2	1.3	0.0	0.1	1.0	67.9	83.4	73.6
Southern Asia		1.3	1.9	2.3	1.0	1.4	2.1	0.2	0.3	0.7	0.2	0.4	0.9	0.2	0.3	0.7	68.8	68.7	68.1
	IND	0.7	1.0	1.6	0.6	0.9	1.5	0.2	0.3	0.7	0.2	0.4	0.8	0.2	0.3	0.7	68.4	69.4	68.1
	BGD	0.1	0.1	0.2	0.1	0.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	82.2	63.8	54.9
Northern America		14.8	12.1	10.6	14.3	12.8	10.7	16.7	13.7	11.0	16.4	14.3	11.0	16.7	13.7	11.0	56.5	57.8	54.2
	CAN	3.4	3.3	2.5	2.7	2.8	1.8	3.1	3.0	1.9	1.3	2.4	1.5	3.1	3.0	1.9	23.9	44.8	43.2
	USA	11.4	8.8	8.1	11.6	10.0	8.9	13.5	10.7	9.2	15.1	11.9	9.5	13.5	10.7	9.2	64.1	61.4	56.5
	MEX	1.7	2.1	2.2	1.8	2.2	2.5	2.6	3.2	3.9	2.5	3.3	3.6	2.6	3.2	3.9	54.2	56.2	50.4
Europe		47.1	42.0	37.1	49.4	44.2	39.8	45.0	39.1	34.5	45.1	40.2	36.5	45.0	39.1	34.5	57.6	57.0	57.8
Western Europe		26.1	21.5	18.8	28.3	24.1	21.9	26.9	22.4	19.3	26.0	21.5	19.4	26.9	22.4	19.3	55.6	53.1	54.9
	GBR	4.8	3.5	2.6	5.1	3.7	2.8	5.6	3.6	2.8	5.8	3.9	2.9	5.6	3.6	2.8	60.2	59.5	55.2
	DEU	10.8	9.2	7.8	12.7	11.1	10.3	13.4	11.7	10.7	12.8	11.1	10.5	13.4	11.7	10.7	54.8	52.6	53.5
	FRA	5.8	4.3	3.2	6.2	4.8	3.9	5.8	4.2	2.9	6.0	4.3	3.4	5.8	4.2	2.9	59.0	56.5	64.5
	ITA	4.8	3.5	2.8	5.6	4.2	3.6	4.1	2.8	2.5	4.5	3.5	3.2	4.1	2.8	2.5	62.9	68.0	70.6
Eastern Europe		2.8	5.4	6.4	2.1	4.1	5.2	1.1	3.7	5.1	1.4	4.5	6.0	1.1	3.7	5.1	70.5	68.5	64.0
	CZE	0.5	0.7	0.9	0.5	0.9	1.2	0.4	1.0	1.5	0.5	1.3	1.8	0.4	1.0	1.5	74.0	69.2	66.7
Latin America and the Caribbean		5.0	5.9	6.5	3.3	4.1	4.3	3.4	4.4	5.2	3.3	4.4	4.6	3.4	4.4	5.2	56.4	56.2	49.2
Africa		2.0	3.1	3.4	0.7	0.9	1.0	0.2	0.4	0.6	0.3	0.5	0.6	0.2	0.4	0.6	71.9	62.7	56.4
Oceania		1.4	1.3	1.8	0.5	0.4	0.4	0.3	0.3	0.3	0.4	0.3	0.3	0.3	0.3	0.3	67.0	53.1	59.0

Notes: 1. Total provides the total export shares for selected regions and countries. **2.** Manufacturing provides to total manufacturing shares (SITC 5+6-68+7+8). **3.** PSB Products are Parts and Components and Final Assembly Products. **4.** Parts and Components are defined in Appendix H. **5.** Final Assembly Products are SITC 75+76+77+87+88 with relevant parts and components removed from each chapter.

Table 5.3: Geographic profile of Total Import and PSB Imports (%)

Import		Total			Manufacturing			Production Sharing Based (PSB) Products: Total PSB Products			Parts & Components			Final Assembly			Share PC in PSB Products (%)		
		1995	2005	2013	1995	2005	2013	1995	2005	2013	1995	2005	2013	1995	2005	2013	1995	2005	2013
Asia		27.6	29.3	37.7	26.0	27.8	33.2	23.9	29.7	35.9	26.2	29.8	34.4	23.9	29.7	35.9	63.1	55.8	52.3
Eastern Asia		17.0	17.4	22.0	15.2	16.2	18.1	13.5	18.1	22.0	14.1	17.3	19.8	13.5	18.1	22.0	60.3	53.1	49.3
	JPN	6.6	4.8	4.4	4.6	3.6	3.3	4.4	3.7	3.4	4.2	3.6	3.4	4.4	3.7	3.4	54.9	54.0	54.8
	CHN	2.6	5.3	8.8	2.7	5.2	7.3	2.0	5.7	8.8	2.7	5.5	6.9	2.0	5.7	8.8	74.8	53.4	43.0
	HKG	3.4	2.9	3.5	3.7	3.5	4.0	3.0	4.4	6.0	2.8	4.5	6.0	3.0	4.4	6.0	54.4	56.5	54.9
	TWN	1.6	1.8	2.4	1.7	1.7	1.2	1.6	1.9	1.3	1.8	1.6	1.1	1.6	1.9	1.3	64.7	47.5	46.4
	KOR	2.8	2.5	2.8	2.5	2.1	2.2	2.4	2.2	2.4	2.6	1.9	2.3	2.4	2.2	2.4	62.6	49.0	52.7
South-Eastern Asia		6.3	5.6	6.8	6.5	5.5	6.7	7.0	6.7	7.2	8.5	7.4	7.8	7.0	6.7	7.2	69.9	61.5	58.6
	IDN	0.9	0.6	1.0	0.9	0.5	1.0	0.7	0.4	0.8	1.0	0.5	1.1	0.7	0.4	0.8	80.6	74.0	72.2
	MYS	1.2	1.0	1.1	1.3	1.1	1.1	1.4	1.5	1.2	1.7	1.6	1.2	1.4	1.5	1.2	70.2	59.5	54.8
	PHL	0.4	0.5	0.4	0.4	0.5	0.4	0.4	0.6	0.5	0.5	0.8	0.6	0.4	0.6	0.5	68.6	71.5	61.6
	SGP	2.1	1.9	2.0	2.1	2.0	1.9	2.8	2.9	2.5	3.2	2.9	2.2	2.8	2.9	2.5	65.4	55.2	48.0
	THA	1.4	1.1	1.3	1.5	1.0	1.3	1.5	1.0	1.3	2.0	1.3	1.6	1.5	1.0	1.3	77.4	71.5	71.2
	VNM	0.2	0.4	0.7	0.2	0.4	0.8	0.1	0.2	0.7	0.1	0.2	0.9	0.1	0.2	0.7	45.0	66.8	65.3
Southern Asia		1.2	2.1	3.3	1.1	1.8	2.3	0.8	1.3	1.6	1.0	1.5	2.0	0.8	1.3	1.6	74.5	67.6	68.0
	IND	0.7	1.1	2.4	0.6	0.9	1.5	0.4	0.6	1.1	0.6	0.9	1.5	0.4	0.6	1.1	80.2	74.8	71.3
	BGD	0.1	0.1	0.2	0.1	0.1	0.2	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.1	0.1	63.6	69.1	65.5
Northern America		16.9	18.9	14.7	17.9	19.4	16.5	21.7	21.2	18.8	19.5	20.0	17.8	21.7	21.2	18.8	51.7	52.2	51.7
	CAN	2.7	3.0	2.5	2.9	3.3	2.9	3.4	3.7	3.1	3.5	3.9	3.0	3.4	3.7	3.1	58.9	58.9	53.8
	USA	14.1	15.8	12.1	15.0	16.0	13.6	18.2	17.5	15.7	16.0	16.1	14.7	18.2	17.5	15.7	50.3	50.8	51.3
	MEX	1.4	2.1	2.0	1.6	2.4	2.5	1.8	2.8	2.9	2.4	3.5	3.8	1.8	2.8	2.9	77.6	69.6	70.1
Europe		46.4	42.5	36.4	46.6	43.1	38.0	45.8	40.3	34.1	45.7	40.8	36.3	45.8	40.3	34.1	57.3	56.2	58.2
Western Europe		25.1	20.7	18.1	25.4	21.1	19.0	24.5	19.6	17.0	24.2	20.7	18.0	24.5	19.6	17.0	56.8	58.4	57.9
	GBR	5.5	4.8	3.6	5.8	5.1	3.8	6.5	5.1	3.7	6.6	4.8	3.6	6.5	5.1	3.7	58.2	52.3	53.8
	DEU	9.4	7.1	6.2	9.6	7.5	6.9	9.5	7.7	6.8	9.4	8.6	7.9	9.5	7.7	6.8	56.8	61.9	63.5
	FRA	6.0	4.7	3.7	6.1	4.8	4.1	6.0	4.5	3.7	5.8	4.6	3.8	6.0	4.5	3.7	55.4	56.6	56.7
	ITA	4.2	3.7	2.6	3.9	3.4	2.5	3.7	3.0	1.9	3.2	2.5	2.0	3.7	3.0	1.9	50.6	46.3	55.7
Eastern Europe		2.8	4.6	5.7	2.8	4.8	6.6	2.3	4.4	6.3	2.2	4.9	7.3	2.3	4.4	6.3	56.0	61.3	63.2
	CZE	0.5	0.7	0.8	0.6	0.8	1.0	0.5	0.8	1.1	0.6	1.0	1.4	0.5	0.8	1.1	62.7	70.2	72.9
Latin America and the Caribbean		5.3	5.3	6.3	5.6	5.6	7.2	5.2	5.4	6.9	5.5	6.1	7.8	5.2	5.4	6.9	61.0	63.6	61.2
Africa		2.5	2.6	3.4	2.4	2.4	3.3	1.9	1.9	2.5	1.9	1.9	2.4	1.9	1.9	2.5	58.3	56.2	52.1
Oceania		1.4	1.5	1.6	1.6	1.7	1.8	1.5	1.6	1.7	1.3	1.2	1.3	1.5	1.6	1.7	47.3	43.9	41.8

Notes: 1. Total provides the total import shares for selected regions and countries. **2.** Manufacturing provides to total manufacturing shares (SITC 5+6-68+7+8). **3.** PSB Products are Parts and Components and Final Assembly Products. **4.** Parts and Components are defined in Appendix H. **5.** Final Assembly Products are SITC 75+76+77+87+88 with relevant parts and components removed from each chapter.

The most notable production fragmentation story outside of the China and East Asia region is Eastern Europe, which saw an increase from 1.4% to 6.0% of world trade in parts and components and a 1.1% to 5% of final assembly products between 1995 and 2013. Given one key determinant of production fragmentation is transport costs, proximity to Western Europe is undoubtedly a significant contributing factor as Germany and France outsource production to Eastern European countries. This trade is also enhanced through low barriers to trade within the European Union.

Other regions showing signs of growth include India, rising from 0.17% to 0.67% share of world export, and Africa, which has more than doubled its world share from 0.21% to 0.58%. While small in magnitude, this demonstrates that product fragmentation is not simply a localised phenomenon in Asia as other regions show early signs of growth in this trade category. This change in the geographic composition of the production of parts and components is shown in Figure 5.2. East Asia is clearly the major winner capturing over 35% of world parts and components trade, up from 25% in 1995. Figure 5.3 shows the six largest export countries, which collectively account for approximately 60 percent of total world exports, with China observed to rise rapidly, capturing 22% of world trade alone.

Figure 5.2: World Share (%) in Parts and Component Exports by Select Regions

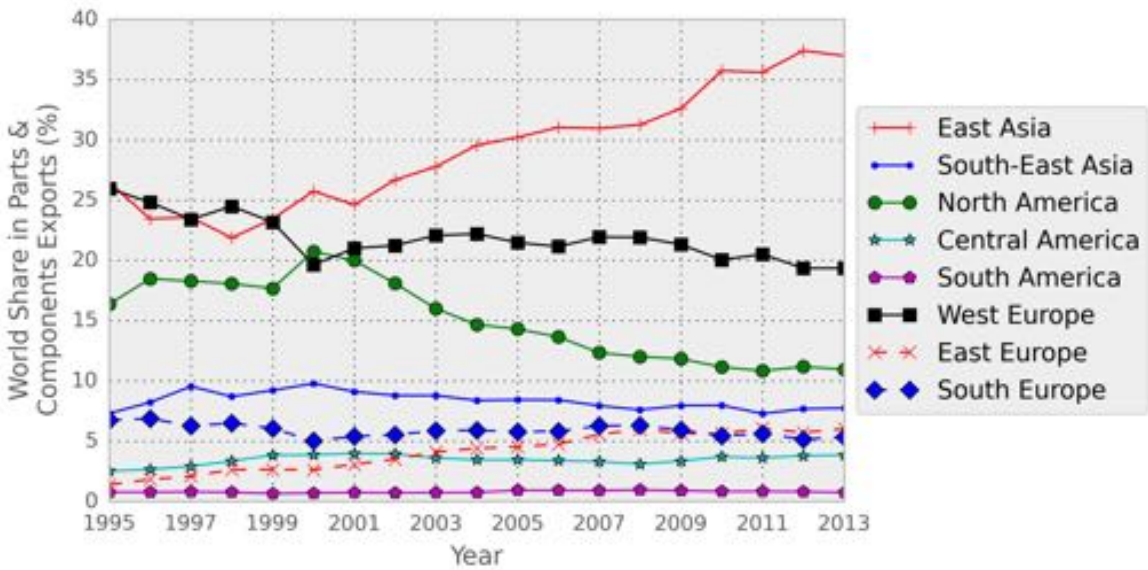
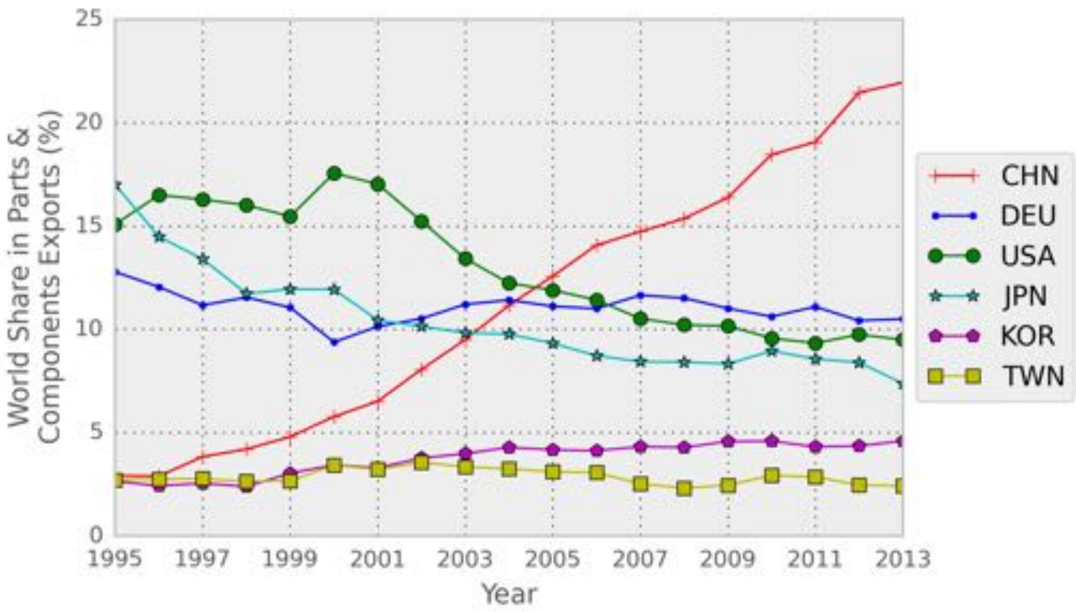


Figure 5.3: World Share (%) in Parts and Components Exports



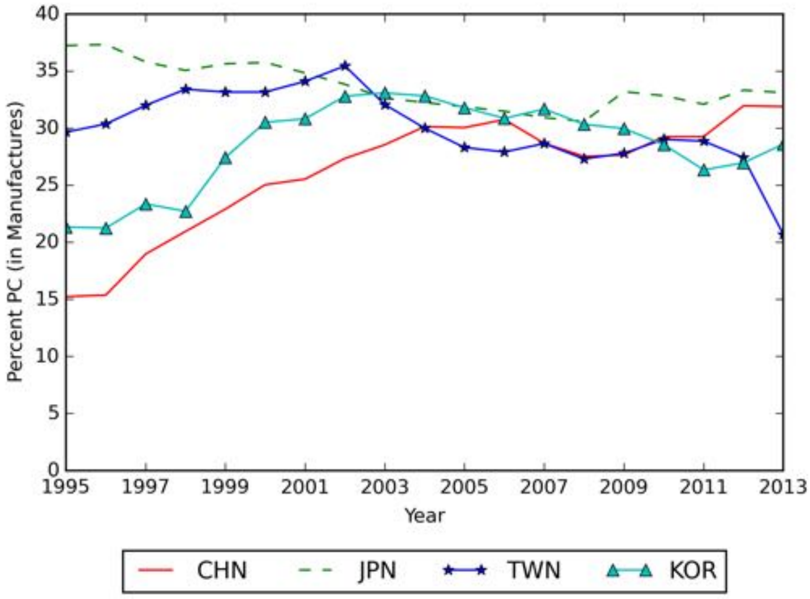
The rise of this segment of international trade is also evident when considering the composition of within-country shares of parts and components in total manufacturing exports (provided in Figure 5.4 and Table 5.4).[101] Total PSB products make up 58% of total manufacturing exports in East Asia for 2013, up from 51% in 1995. These products also account for 57% of total manufacturing exports from South-East Asia and notably now account for 48% of Eastern European manufactures exports, up from 24% in 1995. China's total manufactured exports in 2013 now comprise a 31% share in parts and components and a 22% share in final assembly goods, increasing from 15% and 11% since 1995 respectively. PC continues to remain a significant component of manufactures exports in countries such as Japan, and South Korea, and Taiwan, albeit with some degree of long-term downward trends observed for Japan and Taiwan. Figure 5.4 (B) shows long-term declines in shares of PC of manufactures for Singapore and Malaysia, from a high level of 50% in Malaysia's case to a similar share of 30% that is observed in other developed countries in the Asian region.

What is evident from these statistics is that PSB trade plays a significant role in the dynamics of global trade. Developing countries continue to acquire larger shares in parts and components manufacturing, in addition to final assembly products, specialising in one segment of the production process rather than requiring large sets

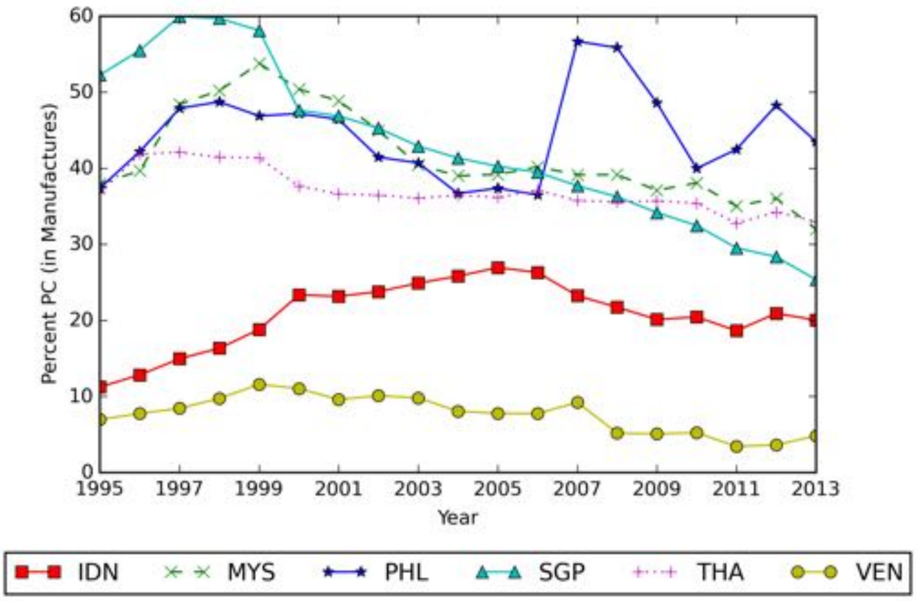
[101] The composition of imports is available in Figure 5.5.

of capabilities to produce a good from start to finish. This phenomenon has opened new opportunities for countries to participate in within-product specialisation patterns.

Figure 5.4: Percentage of parts and components in manufacturing exports of A) China, Japan, Taiwan, and South Korea, and B) Indonesia, Malaysia, Philippines, Singapore, Thailand and Vietnam



(A)



(B)

Table 5.4: Shares of PSB Products in Manufacturing Exports (%)

Export		Parts & Components				Final Assembly				Total PSB Products			
		1995	2000	2005	2013	1995	2000	2005	2013	1995	2000	2005	2013
Asia		28.87	32.08	29.83	28.60	21.10	25.38	26.47	25.34	49.97	57.46	56.30	53.93
Eastern Asia		28.74	31.16	30.60	30.54	22.80	27.04	27.81	27.53	51.53	58.20	58.40	58.07
	JPN	37.23	35.74	31.87	33.10	28.60	34.53	33.61	31.04	65.82	70.27	65.48	64.15
	CHN	15.23	25.04	30.04	31.88	11.27	16.11	20.87	22.02	26.50	41.15	50.91	53.90
	HKG	21.51	26.80	30.41	29.93	14.45	25.30	24.39	22.87	35.96	52.11	54.80	52.80
	TWN	29.65	33.16	28.30	20.69	13.63	26.70	37.14	49.42	43.27	59.85	65.44	70.11
	KOR	21.33	30.51	31.79	28.59	30.13	28.67	30.39	31.07	51.46	59.18	62.17	59.66
South-Eastern Asia		37.71	42.17	35.93	30.41	19.39	25.65	29.51	26.64	57.10	67.82	65.44	57.06
	IDN	11.23	23.34	26.92	20.04	8.15	9.74	12.06	12.81	19.38	33.08	38.98	32.84
	MYS	37.96	50.37	39.17	31.96	29.23	27.72	38.02	36.91	67.19	78.09	77.18	68.88
	PHL	37.40	47.18	37.36	43.49	16.17	37.22	48.91	37.61	53.56	84.40	86.27	81.10
	SGP	52.25	47.60	40.25	25.34	21.72	31.45	28.81	31.09	73.96	79.05	69.05	56.44
	THA	36.71	37.64	36.16	33.00	13.74	21.20	24.73	26.41	50.45	58.84	60.89	59.41
	VNM	2.69	14.43	17.16	36.25	1.27	2.90	3.42	12.98	3.96	17.34	20.57	49.23
Southern Asia		5.91	7.08	8.50	11.25	2.68	2.55	3.88	5.28	8.59	9.63	12.38	16.52
	IND	8.71	9.99	11.04	14.86	4.03	3.39	4.86	6.95	12.74	13.39	15.90	21.81
	BGD	1.42	1.32	1.51	0.46	0.31	0.45	0.86	0.38	1.73	1.77	2.37	0.83
	PAK	0.69	0.87	1.70	1.50	0.77	1.27	1.59	1.57	1.46	2.15	3.29	3.08
Northern America		30.32	34.96	31.18	27.24	23.30	25.33	22.80	22.99	53.62	60.29	53.98	50.23
	CAN	12.75	25.57	24.15	21.68	40.60	35.17	29.73	28.53	53.35	60.74	53.88	50.21
	USA	34.41	37.47	33.16	28.37	19.28	22.72	20.88	21.87	53.69	60.19	54.04	50.24
	MEX	36.79	39.13	41.71	38.83	31.04	36.26	32.54	38.20	67.83	75.39	74.24	77.04
Latin America and the Caribbean		26.31	31.25	30.47	28.94	20.32	27.07	23.74	29.91	46.63	58.32	54.21	58.85
	ARG	17.07	17.44	15.69	12.99	14.49	18.76	18.56	35.42	31.56	36.20	34.25	48.41
	BRA	21.32	22.23	22.84	19.01	8.32	12.43	14.84	12.57	29.64	34.66	37.68	31.59
	CHL	7.73	9.31	8.57	10.27	6.41	7.30	6.60	8.33	14.14	16.61	15.17	18.60
Europe		24.23	26.90	25.36	24.47	17.80	19.51	19.14	17.85	42.03	46.41	44.50	42.32
Western Europe		24.32	26.26	24.83	23.61	19.45	22.09	21.90	19.38	43.76	48.35	46.73	42.99
	GBR	30.64	34.37	28.94	27.36	20.27	20.07	19.72	22.21	50.91	54.44	48.65	49.57
	DEU	26.74	28.50	27.81	27.14	22.05	25.02	25.09	23.58	48.79	53.52	52.91	50.72
	FRA	25.71	27.55	24.76	23.72	17.89	18.95	19.09	13.05	43.60	46.50	43.85	36.77
	ITA	21.19	21.77	23.33	24.03	12.48	12.13	10.96	10.00	33.67	33.90	34.29	34.03
Eastern Europe		17.04	27.65	30.48	31.00	7.14	12.57	14.02	17.43	24.18	40.22	44.50	48.43
	RUS	5.14	12.85	9.83	11.51	3.53	5.07	5.31	6.98	8.67	17.93	15.14	18.49
Africa		10.09	12.23	14.29	15.20	3.94	6.28	8.51	11.74	14.03	18.51	22.80	26.94
Oceania		22.85	21.80	19.65	21.12	11.24	15.45	17.32	14.69	34.09	37.25	36.97	35.81

Notes: 1. Shares are constructed from baci92 dataset concorded to SITC rev 3 level 5 data. **2.** PSB products are parts and components and final assembly products.

Table 5.5: Shares of PSB Products in Manufacturing Imports (%)

Import		Parts & Components				Final Assembly				Total PSB Products			
		1995	2000	2005	2013	1995	2000	2005	2013	1995	2000	2005	2013
Asia		26.78	32.28	29.94	27.65	15.68	21.90	23.73	25.26	42.46	54.18	53.67	52.91
Eastern Asia		24.64	31.13	29.77	29.22	16.23	22.62	26.29	30.06	40.87	53.75	56.06	59.29
	JPN	24.33	28.80	28.13	27.09	20.01	25.86	24.00	22.38	44.34	54.66	52.13	49.47
	CHN	26.30	32.52	29.68	25.53	8.88	12.98	25.91	33.81	35.18	45.50	55.59	59.34
	HKG	20.12	32.87	35.55	40.53	16.89	20.94	27.40	33.23	37.02	53.82	62.96	73.76
	TWN	28.82	32.34	27.19	25.08	15.75	27.77	30.05	29.03	44.57	60.11	57.24	54.11
	KOR	27.82	30.23	25.56	27.36	16.59	27.64	26.60	24.59	44.40	57.87	52.16	51.96
South-Eastern Asia		34.59	40.69	37.24	30.79	14.87	22.45	23.27	21.77	49.46	63.15	60.51	52.56
	IDN	30.98	27.84	28.99	29.07	7.44	8.99	10.17	11.19	38.42	36.83	39.16	40.25
	MYS	33.75	41.71	40.37	30.02	14.33	28.02	27.49	24.81	48.07	69.73	67.86	54.83
	PHL	31.48	47.82	47.89	34.50	14.40	18.80	19.09	21.52	45.88	66.61	66.98	56.02
	SGP	39.19	44.92	40.55	32.04	20.77	27.64	32.91	34.77	59.95	72.56	73.47	66.81
	THA	34.96	39.47	34.83	34.13	10.21	15.09	13.87	13.80	45.17	54.56	48.70	47.93
	VNM	16.35	20.66	19.45	28.48	19.97	14.45	9.69	15.16	36.32	35.11	29.13	43.64
Southern Asia		25.10	25.74	24.27	23.55	8.60	10.00	11.65	11.10	33.70	35.74	35.92	34.65
	IND	26.02	30.52	27.96	26.87	6.42	10.40	9.42	10.79	32.43	40.93	37.38	37.65
	BGD	12.70	14.03	17.27	15.05	7.26	7.18	7.70	7.93	19.97	21.21	24.97	22.98
	PAK	30.93	25.45	20.46	19.63	10.54	13.59	15.85	10.23	41.47	39.04	36.31	29.86
Northern America		28.85	30.99	28.76	28.66	26.97	28.79	26.33	26.74	55.82	59.77	55.09	55.40
	CAN	32.02	37.84	33.08	28.31	22.34	23.27	23.04	24.32	54.35	61.11	56.12	52.63
	USA	28.26	29.62	27.90	28.79	27.88	29.89	27.03	27.32	56.15	59.52	54.93	56.11
	MEX	39.93	41.51	40.36	40.35	11.50	17.08	17.61	17.22	51.44	58.59	57.97	57.56
Latin America and the Caribbean		26.05	32.83	30.60	29.01	16.67	16.87	17.52	18.40	42.72	49.70	48.12	47.41
	ARG	26.84	30.41	29.03	30.81	17.05	16.71	17.93	24.18	43.89	47.12	46.96	54.99
	BRA	27.12	35.42	34.73	29.56	21.22	17.81	15.39	17.38	48.34	53.23	50.13	46.94
	CHL	18.74	21.16	20.13	18.30	23.01	20.19	23.90	25.12	41.75	41.35	44.03	43.41
Europe		26.00	28.51	26.37	25.52	19.34	21.01	20.55	18.31	45.34	49.52	46.91	43.83
Western Europe		25.27	28.64	27.38	25.35	19.22	20.67	19.47	18.43	44.49	49.31	46.85	43.78
	GBR	30.38	30.97	26.33	25.67	21.83	23.62	23.98	22.08	52.21	54.58	50.32	47.75
	DEU	26.13	30.79	31.91	30.62	19.84	21.02	19.67	17.63	45.97	51.81	51.58	48.25
	FRA	25.12	27.39	26.74	24.98	20.20	21.37	20.52	19.11	45.32	48.76	47.26	44.09
	ITA	22.07	22.72	20.61	21.10	21.54	23.92	23.90	16.80	43.61	46.64	44.51	37.90
Eastern Europe		21.33	29.52	28.55	29.42	16.74	14.76	18.03	17.10	38.06	44.29	46.57	46.53
	RUS	15.74	18.34	20.69	23.58	22.18	15.88	21.97	19.98	37.93	34.22	42.67	43.56
Africa		21.12	22.38	22.28	19.49	15.11	15.11	17.36	17.89	36.23	37.49	39.63	37.38
Oceania		21.68	26.27	20.85	19.10	24.13	25.11	26.69	26.57	45.81	51.39	47.54	45.68

Notes: 1. Shares are constructed from BACI92 dataset concorded to SITC rev 3 level 5 data. **2.** PSB products are parts and components and final assembly products.

5.4.1 PARTS AND COMPONENTS AND PRODUCT SPACE COMMUNITIES

In this section parts and components are identified in each of the product space communities as defined by Hidalgo et al. (2007) and shown in Figure 5.5. Consideration of these *communities* helps to relate the analytical narrative thus far to later sections that look specifically at the product space network. The clustering of products into communities is defined based on their relative locations in the product space network, which have been determined by a clustering algorithm.[102] Similarities between the product space communities and the Leamer and Lall product classifications have been demonstrated by Hausmann & Klinger (2006). For example, products that are found in the "Electronics" community predominantly fall within the Lall HT1 classification. The key benefit is a broader classification from 10 to 34 product groups offering a higher degree of disaggregation, grouping products into relevant types. Overall, product space communities that tend to have high shares in parts and components are Aircraft, Electronics, and Machinery.

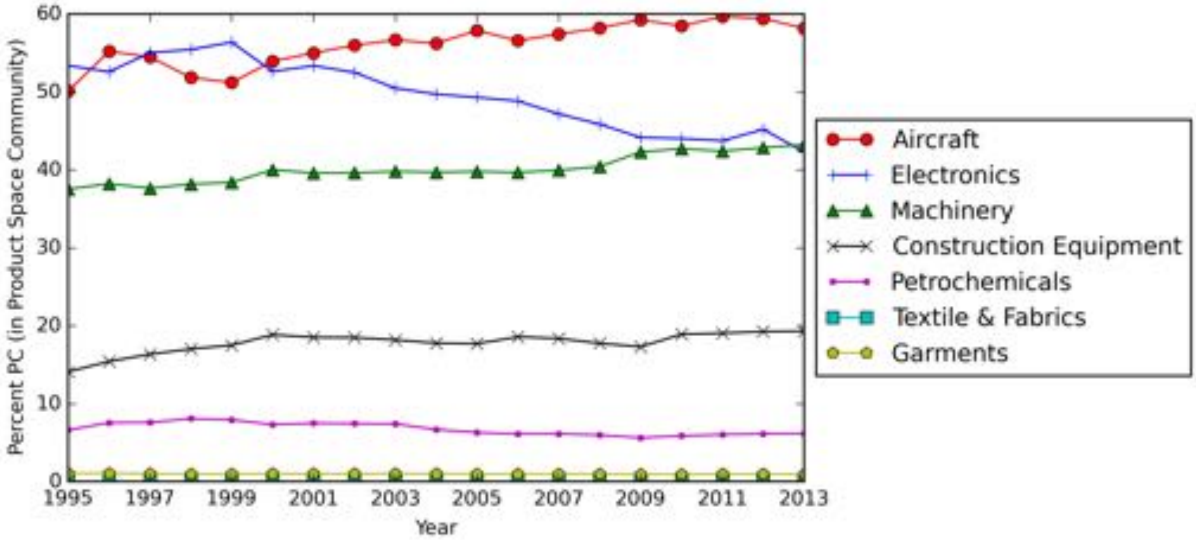
One advantage to using product groupings, such as Lall, Leamer, and the product space communities, is that it removes the need for parts and components to fall hierarchically within the SITC classification. Take for example the product space community identified as "Aircraft". As shown in Figure 5.5, parts and components within the aircraft industry account for a 55 percent share (by value) within this clustering of products. The equivalent analysis of the share of parts and components within SITC chapter 79 would underestimate the true share within the industry as jet engines, a high cost component in aircraft assembly, are located in chapter 71. The "Aircraft" product space community, as defined in Table 5.6, consists of product codes from chapters 62, 71, and 79. Clustering techniques can therefore provide cross cutting groups of products that are highly related and allow for deviations from the hierarchical structure of the SITC classification. The next highest shares within the product space communities are contained in the well-known Electronics and Machinery industries.[103]

---

[102] As has been noted previously, the product space communities are identified using random walk clustering techniques.

[103] While these industries are still quite broad in the product space, further work may identify within community clusters of different sectors of the electronics industry.

Figure 5.5: Percentage of parts and components in select Product Space Communities (%)



Notes: 1. Product space communities are defined at SITC rev 2. Level 4. This has been concorded to the current dataset using the UN SITC rev 3. to SITC rev. 2 conversion table.

Table 5.6: Aircraft Community in the Product Space

SITC Rev2 Level 4	Parts and Components	SITC Description
6253	Yes	Tyres, Pneumatic, New, of a kind used on aircraft
7131	Yes	Internal Combustion Piston Engines for Aircraft
7144	Yes	Reaction Engines
7148	Yes	Gas Turbines, NES
7149	Yes	Parts of the Engines & Motors of 714- and 718.8-
7922	No	Aircraft Not Exceeding an Unladen Weight of 2000 KG
7923	No	Aircraft Not Exceeding an Unladen Weight of 15000 KG
7924	No	Aircraft Exceeding an Unladen Weight of 15000 KG
7928	No	Aircraft, N.E.S Ballons, Gliders Etc and Equipment
7929	Yes	Parts of Heading 792--, Excl Tyres and Engines

When working with the product space, a network diagram is a more intuitive representation and has become the standard when presenting the product space data. Following the same methodology as found in the network complexity literature, the product space network was recomputed for the SITC revision 3 level 5 data used in this chapter.[104] Proximity is computed using the same metric described in the primary economic complexity literature.

$$\phi_{p_1p_2} = \frac{\sum_c M_{cp_1} \bullet M_{cp_2}}{max(k_{p_1,0}, k_{p_2,0})}$$

The network view is constructed from the minimum spanning tree, with the inclusion of high proximity connections in excess of 0.55. The parameters used in the construction of this product space are the same as those used in Hidalgo et al. (2007). The same communities of products that are identified in the *Atlas of Complexity* are used to highlight the different regions of the product space (Hausmann & Hidalgo 2014). Given the published product space communities are defined at the SITC rev 2 4-digit level, these were concorded to the 4-digit component of the 5-digit SITC rev 3 product codes.[105] To account for the higher level of disaggregation of data used here, nodes that occupied the same location in the diagram were moved to prevent overlaps, which explain the differences observed between the diagrams in this chapter with those found in the literature.

The relationship with the Lall classification can be observed by comparing product space communities and Lall technology clusters which are highlighted in Figure 5.6 and Figure 5.7 respectively. In both figures the location of product nodes are fixed (as per the layout defined by Figure 5.6), which allows for direct comparisons to be made. Strong similarities can be observed between the two product groupings, as is found in Hausmann and Klinger (2006).[106] For example, the Lall category MT3 (Medium Technology 3) contains relatively complex engineering industry based products such as engines and pumps. This cluster as shown in blue in Figure 5.7 and largely corresponds with the machinery product space community in Figure 5.6.

---

[104] This work is carried out using the python tools developed to support this research and is available through github. These tools can be used to compute ECI and PCI for any trade dataset. See Appendix G for further details.

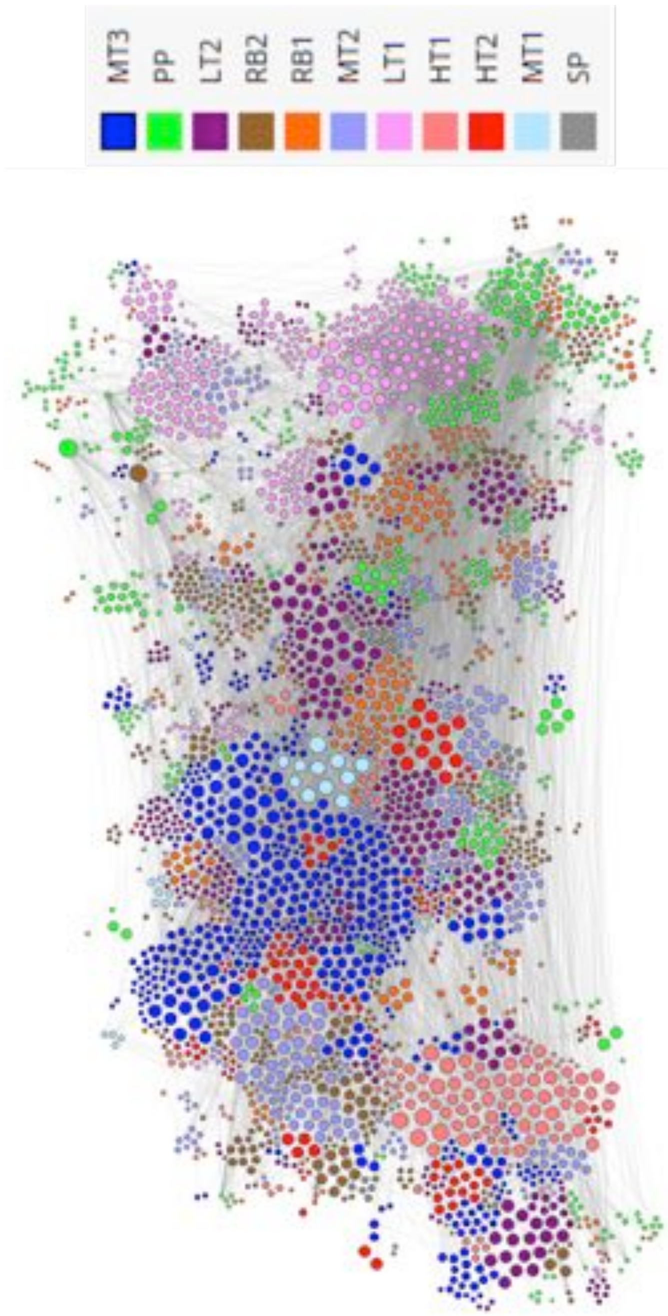
[105] Due to the higher level of disaggregation of the underlying trade data there are a few cases where a product would move to another cluster. Rather than reconstructing the product space community, the product space communities as previously defined in the literature at the 4 digit level presents a similar narrative and only minor differences are observed when conducting a separate clustering exercise at the 5 digit level.

[106] The Leamer (1984) version of this diagram can be found in Appendix H.

Figure 5.6: Product Space (SITC Rev 3. L5) in Atlas of Complexity Colours

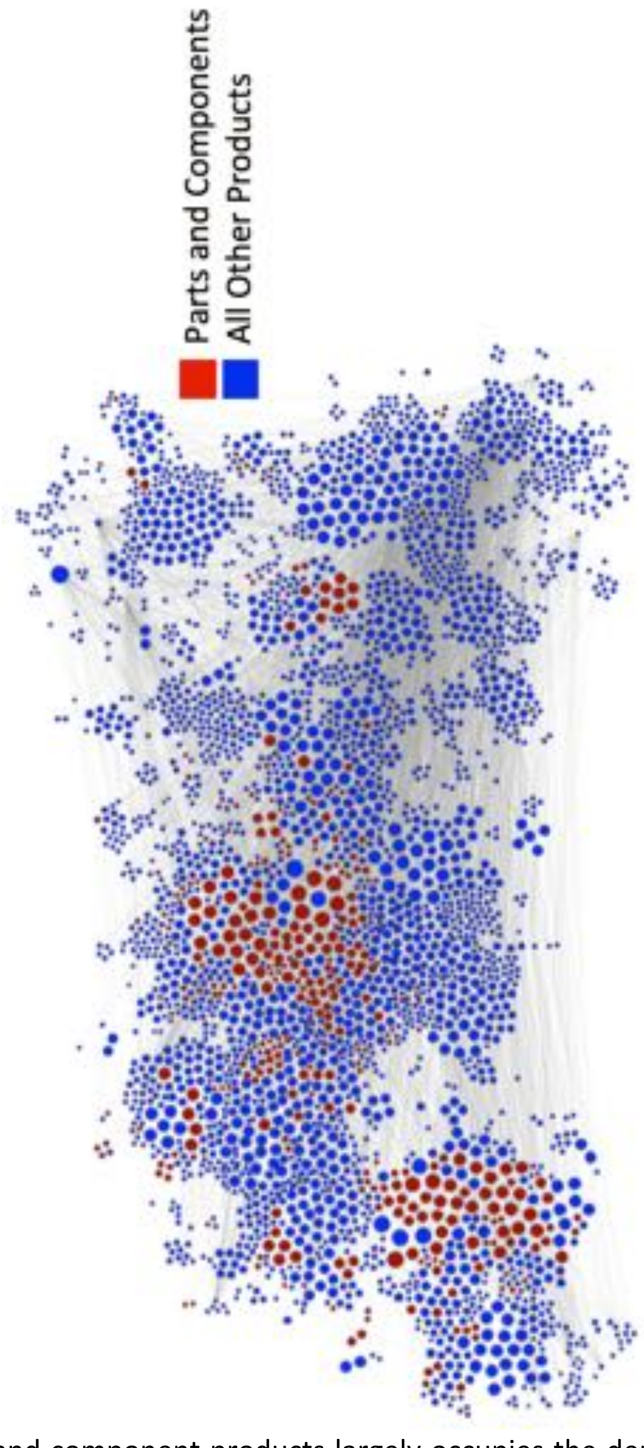
Notes: 1. This diagram provides a network view of product similarities, which form communities of products that are identified by different colours. **2.** The product space communities used are the ones defined in the Atlas of Complexity, defined at the four-digit level. There is a slightly larger degree of mixed clusters due to the use of data at a higher level of disaggregation, which includes a greater number of nodes than the product space defined in the literature. **3.** Legend is organised by decreasing cluster size based on the number of within cluster nodes. **4.** This diagram contains 3111 products.

Figure 5.7: Product Space (SITC Rev 3 L5) Atlas Layout in Lall (2000) Categories



Notes: 1. Lall Categories correlate strongly with the identified product space communities. ***HT***-High Technology (HT1–electronics, HT2-aerospace, pharmaceuticals, and specialist equipment), ***MT***-Medium Technology (MT1-automotive, MT2-process industries such as chemicals, fertilizers, and plastics, MT3-engineering industries such as engines, pumps and ships), ***LT***-Low Technology (LT1-textile and fashion, LT2-pottery, jewellery and furniture), ***RB***-Resource Based (RB1-agro based, RB2-mining based), ***PP***-Primary Products, and **SP**-Other. **2.** Legend is organised by decreasing cluster size based on the number of within cluster nodes; **3.** This diagram contains 3111 products.

Figure 5.8: Product Space (SITC Rev 3 L5) Atlas Layout with Parts and Components



Notes: 1. Parts and component products largely occupies the densely connected core of the product space network. Parts and components are also collocated within industry communities suggesting a strong co-export relationship exists between parts and components, in addition to final assembly products, and other final goods. **2.** This diagram contains 3111 products.

The placement of Parts and Components can be viewed in Figure 5.8, and are observed to largely inhabit the densely connected industrial core. Upon more disaggregated levels of inspection, it is also observed that parts and components tend to be clustered with the relevant final goods. The aircraft community includes industry-specific parts and components that are embedded within the same cluster as final aircraft exports. A similar pattern is also observed for Electronics and Machinery. This has two main implications. First, as the process of product fragmentation continues and enables higher degrees of within-product specialisation, the relationship between countries and PSB products may change over time, with the potential for the formation of new clusters. Aircraft parts and components may become detached from the related final products that they serve. Second, product emergence typically occurs in close proximity to other products that a country already exports (Hidalgo et al. 2007). Therefore, it is possible that parts and components, or final assembly products, may serve as gateway products that enable countries to diffuse from the periphery to the core of the product space. This diffusion pattern towards clusters that contain sophisticated final products is supportive of a process of accumulation in country capabilities towards a higher degree of overall complexity. This is explored further in section 5.5.2.

5.5 PRODUCT FRAGMENTATION AND ECONOMIC COMPLEXITY

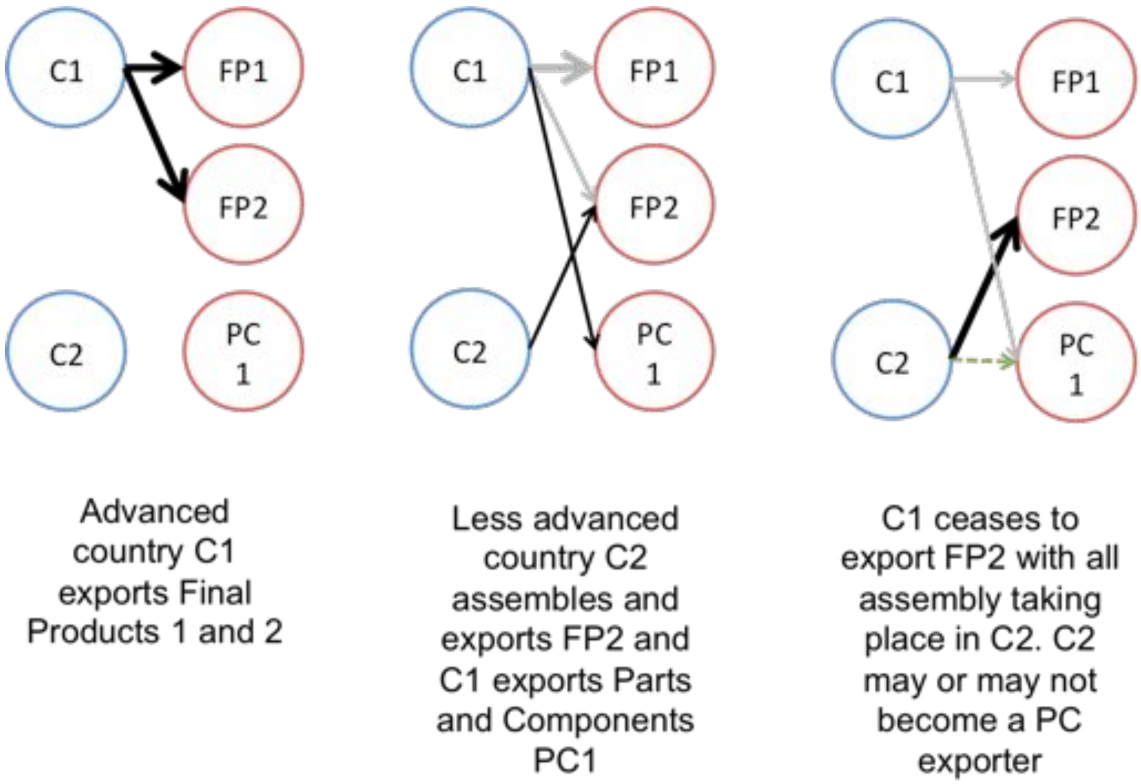
The next step is to examine how product fragmentation affects the analytical techniques used in producing the economic complexity rankings. As has been discussed, economic complexity is fundamentally determined by the cross-country patterns of exports in any given year. The key determinant of complexity measurement is the underlying structure of the bilateral network mapping between countries and products, which links country diversification with product ubiquity characteristics. The observed network structure is determined by measurements taken from international trade data statistics.

Product fragmentation introduces a new challenge for these new data-driven techniques that measure the complexity of economies. Due to specialisation within the vertical value chain, countries no longer require the *full suite of capabilities* to produce the entire good, and in some cases only engages in the final assembly process that is labour intensive. In reality, the process of product fragmentation unfolds as a dynamic process that links countries that have the full suite of *capabilities* to produce a given complex product from start to finish, with countries that have only have a relatively *small sub-set of capabilities* required.

A well-studied example is the Apple iPhone (SITC Chapter 76) that is produced by China. The data-centric technique infers that China has acquired the capabilities to manufacture complex electronics, when in reality it provides assembly services to produce large quantities of iPhones from complex imported intermediates such as motherboards, screens, cases etc. The information used by the empirically driven metrics, such as product ubiquity characteristics, are therefore distorted, as trade statistics attribute the full export value of the iPhone export to China. The dynamic evolution of the bipartite network, that links between countries and products, needs to be considered in light of the process of product fragmentation. The impact of product fragmentation on the measurement of economic complexity is explored by introducing differentiated products for the family of final assembly products.

The evolution of product fragmentation is expressed in Figure 5.9. This diagram contains two countries (C1, C2) and three products (FP1, FP2, and PC1). This figure illustrates the following: (1) Final product 2 (FP2) happens to be a final assembly product (FAP), which has characteristics that make it susceptible to the forces of product fragmentation, (2) Comparing the process from start to finish suggests that the production of FP2 moves from C1 to C2, while C1 provides complex intermediates to C2 (as represented by the grey line), (3) If there were no intermediate step, and dynamic adjustment was instantaneous, the economic complexity approach would then endogenously adjust to the information contained in this new environment.

Figure 5.9: Process of Product Fragmentation in Bipartite Network Structure



Now consider C1 and C2 to be groups of countries. If the shift in production occurred instantaneously, and no dynamic intermediate step was considered, then the product characteristics of FP2 would endogenously change as the average country characteristics are now derived from C2-type countries only. In the case of final assembly products, this would demote the complexity ranking of FP2, as C2 countries are likely to be less sophisticated, on average, providing only assembly services with relatively lower labour rates. C1 countries remain diversified as they export complex intermediates to nations that assemble the products. This has described a step process of internal adjustment that is built into the method of reflections.

The intermediate step however highlights the potential for the assignment of inaccurate relationships to form indirectly between countries through products. The complexity measure is, in part, constructed by averaging product characteristics over all exporters. The intermediate adjustment step links many less advanced economies (C2) to the highly diversified production structures of more advanced economies (C1). The formation of this link will bias the less advanced country's complexity ranking upwards due to the joint assignment of countries that export FP2. In this case, product fragmentation has allowed the less developed economy to give the illusion that it exports sophisticated products, while in reality it is vertically specialised in a small segment of the production chain. This renders the possibility that two exporters of the same product may have vastly different sets of *capability*.

The process of product fragmentation has enabled the geographic reassignment of exports, through the mobility of intermediates, which alters the information that is contained in the bipartite network. In summary, countries participating in a vertical segment of a product will get larger than expected complexity rankings due to ties with relatively complex economies, while products undergoing this process will get lower than expected complexity rankings due to declining country sophistication, as new assemblers emerge over time.

5.5.1 DIFFERENTIATED FINAL ASSEMBLY PRODUCTS

Based on the discussion in the previous section, the intermediate dynamics that occur due to product fragmentation may link countries through the bipartite network structure in ways that may distort the measurement of complexity. There are two primary implications of these reassignment dynamics for economic complexity analysis. First, the product characteristics are changing over time as new exporters are able to participate in final goods assembly, potentially rendering relatively sophisticated products more ubiquitous over time. Second, the assignment relationship between developing economies that participate in final assembly from imported parts and components may inaccurately represent the true level of *capability* within each given economy. This is driven, through the *method of reflections*, by linking the export structure of developing economies with that of more developed countries through the joint relationship of a final assembly product. A country that exports a product, made domestically from start to finish, would require a larger set of capabilities than a country that only participates in a small segment of that product's value chain.

To test these implications, final assembly products (final goods that correspond to industries that heavily participate in product fragmentation) are differentiated into two categories based on the exporting country's level of GDP per capita (PPP adjusted). A high category is when a final assembly product is exported from a high GDP per capita country and a low category is when a final assembly product is exported from a low GDP per capita country. A cut-off value of US$ 12,736 was used to differentiate between countries such as China, Thailand, and Indonesia from countries such as South Korea, Japan, and Taiwan.

The main assumption of this approach is that products coming from low-type economies are fundamentally different to products coming from high-type economies, which serves as a proxy for the differences in unobserved participation in different stages of the production process.[107] It is implicitly assumed that low-type economies have a propensity to engage in assembly operations, importing the majority of parts and components, while high-type economies are more likely engaged in higher quality products, or products that are specialised within the same product category. While the focus is on product fragmentation, product differentiation also has the advantage of capturing the idea that low-type and high-type economies may produce goods in the same category but are differentiated based on quality. Differences in quality may also imply differences in the underlying set of available capabilities.

It is important to note that ECI and PCI rankings are a result of the network structure of the bipartite network, and does not itself incorporate the income level of countries.[108] The use of income to differentiate products does not influence the computation of economic complexity directly, but rather through the remapping of bipartite network connections between countries and products. This change in the bipartite network structure has three primary implications for the computation of economic and product complexity ranks. First, low type economies will export final assembly products that are more similar to final assembly products of other low type economies. This will reduce the complexity score of these products as they are now associated with economies that are less diversified and export less ubiquitous products, on average. Second, this change in structure will also alter country complexity scores by linking these economies with other exporters of final assembly products that are, on average,

[107] This method of demarcation based on GDP per capita (PPP) is used at the level of development correlates with a number of economic variables such as the size of the capital stock at the country level. Many other differentiating data sources are typically not available for the large numbers of countries considered in this analysis.
[108] This is done in some early papers, which make use of PRODY and EXPY. These measures define levels of export complexity based on products that are exported from certain countries based on income levels. In the formulation used here it is only the bipartite network defined by the M_{cp} matrix, which is used to compute complexity.

more similar in productive capabilities. Finally, the ubiquity characteristics of these products will decrease due to splitting them into two categories. While initially this may suggest that these products would become relatively more sophisticated making them less ubiquitous across countries, the averaging over the network structure drives the final result between these new product characteristics and countries to which they are paired.

To progress in this analysis, the ECI and PCI are computed from an M_{cp} matrix that consists only of countries that have GDP per capita (PPP) information. This reduces the number of countries from 204 to 173. The columns of final assembly products are then split into the two columns, and products are reassigned in accordance with the level of GDP per capita. The total number of final assembly products is 199, which increases the number of products from 3112 to 3311. The ECI and PCI rankings are then recomputed and compared to the original rankings prior to the enforcement of differentiated products. The year 2000 is used as it serves as a baseline year in the literature. Similar results are obtained using other yearly cross-sections due to relatively static proximity matrices over time.

Allowing for differentiated final assembly products, the complexity rankings of countries in South-East Asia, Central America, and Eastern Europe are observed to decline as shown in Table 5.7. These regions are the same regions that participate heavily in final assembly products. Exporting low-type final assembly products decreases their overall complexity ranking by decoupling them from more advanced economies such as Taiwan, South Korea, and Japan that export similar but differentiated goods.

Table 5.7: Economic Complexity Indicator (ECI) Ranks with Final Assembly Product Adjustments for the Year 2000

Region	Country	ECI Rank	ECI Rank FAP Adjusted	Change in ECI Rank
South-East Asia	MYS	56	76	-20
	PHL	91	108	-17
	THA	83	86	-3
	IDN	101	103	-2
	VNM	139	141	-2
East Asia	CHN	74	79	-5
	JPN	1	1	0
	TWN	21	19	2
	KOR	26	22	4
	HKG	73	47	26
South Asia	IND	58	62	-4
Central America	MEX	33	46	-13
Eastern Europe	SVK	18	27	-9
	HUN	25	29	-4
	CZE	12	15	-3
	UKR	28	30	-2
North America	USA	6	6	0
Western Europe	DEU	2	2	0

Notes: 1. The ECI ranks have been computing using SITC rev 3 Level 5 data for the year 2000. **2.** FAP adjusted has split the final assembly products into two columns according to different levels of country GDPPC. Low type is defined as a final assembly product exported by countries with a GDP per capita < US$12,736 PPP adjusted.

Similar evidence of the constraints imposed by the bipartite mapping is revealed when considering individual product complexity rankings. Table 5.8 compares the PCI rankings of a selection of products that are observed before and after the addition of differentiated goods. Once the products are differentiated, low-type telephonic apparatus (76419) falls from a ranking of 443 to 2759 (out of 3311 products), while high type telephonic apparatus remains in the top quartile. Decoupling these products into two groups by country characteristics leads to changes in the ranking of revealed measures of product complexity.

This analysis suggests that product fragmentation tends to bias the results in the economic complexity literature by inappropriately linking countries where the forces of fragmentation increasingly allow less sophisticated economies to specialise in less sophisticated production segments of relatively sophisticated products. As a consequence, this inaccurately pairs countries to relatively sophisticated products, and indirectly links to countries that are relatively diversified. These links lead to an overestimation of their true complexity ranking. The introduction of differentiated products, linked with low and high type economies, demonstrates the direction of this bias.[109]

Table 5.8: Product Complexity Indicator (PCI) Ranks with Final Assembly Product Adjustments for the Year 2000

SITCR3L5	Description	Type	PCI Rank	PCI Rank FAP Adjusted	Change in PCI Rank
75210	Analogue or hybrid (analogue-digital) data processing machines	**L**	834	2913	-2080
		H	834	192	642
75220	Digital automatic data-processing machines, containing in the same housing at least a central processing unit and an input and output unit, whether or not combined	**L**	826	2663	-1838
		H	826	81	745
76419	Other telephonic or telegraphic apparatus	**L**	443	2759	-2317
		H	443	37	406
87413	Surveying (including photogrammetrical surveying), hydrographic, oceanographic, hydrological or geophysical instruments and appliances (excluding compasses); rangefinders	**L**	1833	3191	-1359
		H	1833	68	1765

Notes: 1. The PCI ranks have been computing using SITC rev 3 Level 5 data for the year 2000. **2.** FAP adjusted has split the final assembly products into two columns according to different levels of country GDPPC. Low type is defined as a final assembly product exported by countries with a GDP per capita < US$12,736 PPP adjusted.

5.5.2 PATTERNS OF EMERGENCE IN THE PRODUCT SPACE

While the previous section focused on the implications of product fragmentation on ECI and PCI, this section will explore the structural features of production sharing based (PSB) products within the context of the product space network. The product space exhibits a distinct periphery-core network structure, where developing economies tend to occupy the periphery and diversified advanced economies are spread through the densely connected core of the network (Hidalgo et al. 2007). The analysis first considers the properties of parts and components (PC) and final assembly products (FAP) in the proximity matrix, and then considers the role these products play

[109] Alternative methods for differentiating these overlaps in the network structure could be an interesting avenue for further research.

as gateway products through the process of country diffusion over the product space network.

Proximity is defined as the minimum conditional probability of co-export.

$$\phi_{p_1 p_2} = \frac{\sum_c M_{cp_1} \bullet M_{cp_2}}{max(k_{p_1,0}, k_{p_2,0})}$$

A summary of the proximity data, by product types, is presented in Table 5.9

Table 5.9: Proximity Description for Different Product Types

Proximity	All Products	Final Manufactures	Parts and Components
Product Pairs	9,678,832	5,419,584	124,316
Mean ($\phi_{pp}{}'$)	0.15	0.17	0.23
Standard Deviation	0.101	0.105	0.123
25 Percentile	0.075	0.091	0.142
50 Percentile	0.136	0.158	0.222
75 Percentile	0.211	0.235	0.310

Table 5.10 then presents the average proximity between and within PC and final products (FP). It shows that parts and components have relatively high average proximity values (0.23) when compared to all other products (0.16) as well as compared to final manufactures (0.17). This means that parts and components therefore have higher average levels of co-export potential and as a group share more similar revealed *capabilities* required to produce them. This outcome seems reasonable, as parts and components reside within industry clusters of related products that are relatively more similar to each other. For example sub-assemblies of iPhone's are more related to each other than they are to milk production.

As countries tend to diffuse locally within the product space, participation within clusters with high average proximity should also lead to further exports of other parts and components as the country becomes embedded within a supply chain. High average proximity implies that a country's capability to export one PC product suggests that it also has potentially acquired the capabilities to export many types of similar parts and components with only margin changes required in the set of capabilities. This should increase the set of available future export options into other PC products.

Table 5.10: Proximity between and within Final Products and Parts and Component Manufactures

	Final Manufactures	Parts and Components
Final Manufactures	0.17 (0.105)	0.16 (0.104)
Parts and Components		0.23 (0.12)

While these global averages show that PC have a high average within cluster proximity (meaning they are closely linked), PC also share a large number of strong connections with other final manufactured products. Of the 354 parts and components, the highest proximity values between parts and component products consist of 48 final assembly products, 146 final manufactures, 150 other parts and component products, and 10 other traded products. Analysis of the top ten connections for each part and component product code yields an average composition of 19.5% with final assembly products, 49% with other final manufactured goods, 40% with other parts and components, and only 14% are shared with other non-manufactured goods. To illustrate this, table 5.11 shows the top 5 proximity relationships (by value) between PC and all other types of products (P2) to demonstrate the heterogeneity in observed partnership types.

Table 5.11: Top 5 Highest Proximity values for Parts and Components for the Year 2000

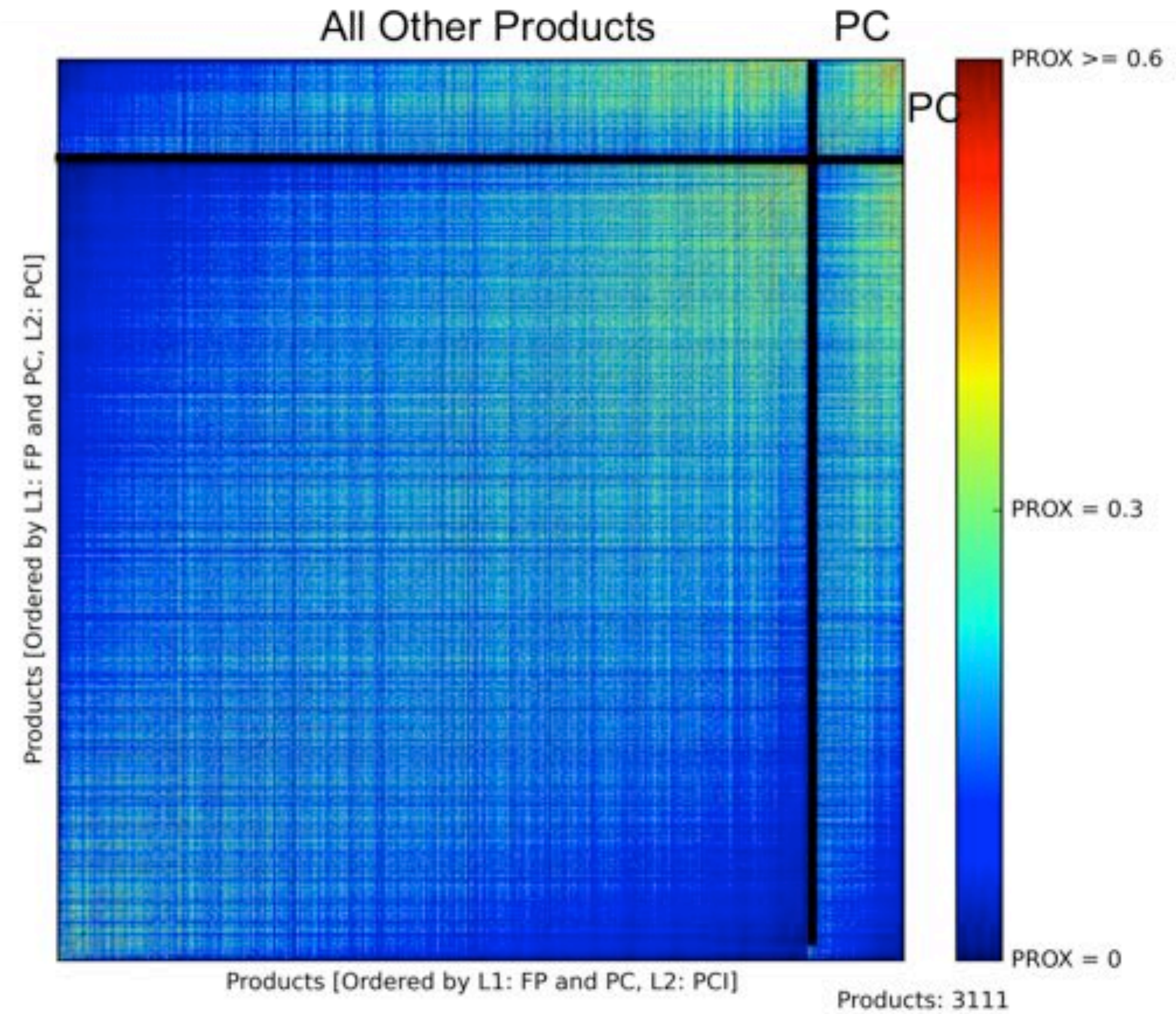
Part and Component	Description	Product 2	Description	P2 Category	Proximity
84552	Girdles, corsets, braces …	84551	Brassieres	Other Manufactures	0.771
78432	Automobile bodies (incl. cab)	77834	Windscreen wipers, defrosters of kind used for motor vehicles	Parts and Components	0.769
77834	Windscreen wipers, defrosters of kind used for motor vehicles	78436	Non-driving axles, and parts thereof	Final Assembly Product	0.769
88114	Parts of photographic cameras	88552	Watch movements, neither battery nor accumulator powered	Final Assembly Product	0.75
77632	Diodes and Transistors	77863	Aluminium capacitors	Parts and Components	0.75

What these observations suggest is that parts and components play a role in connecting the dense industrial core of the product space together, not only to each other, but also to other product types such as FAP and FP. This can also be observed graphically in Figure 5.10, which is the proximity matrix for the year 2000. In this matrix, parts and components (PC) have been grouped together in the top right hand corner. The PC cluster is also shown in greater detail in Figure 5.11. These plots demonstrate the high average level of connectedness that has been discussed, within parts and components. But it also shows the extensive connecting role PC play in the off-diagonal quadrants, which compares PC with all other products (see top left quadrant of Figure 5.10).

As these figures are sorted by the product complexity indicator (PC) within the four quadrants, parts and components also primarily correspond with relatively sophisticated products, with much of the high valued proximity connections, indicated by the heat map colours, present in the top right hand of the upper left quadrant of Figure 5.10.

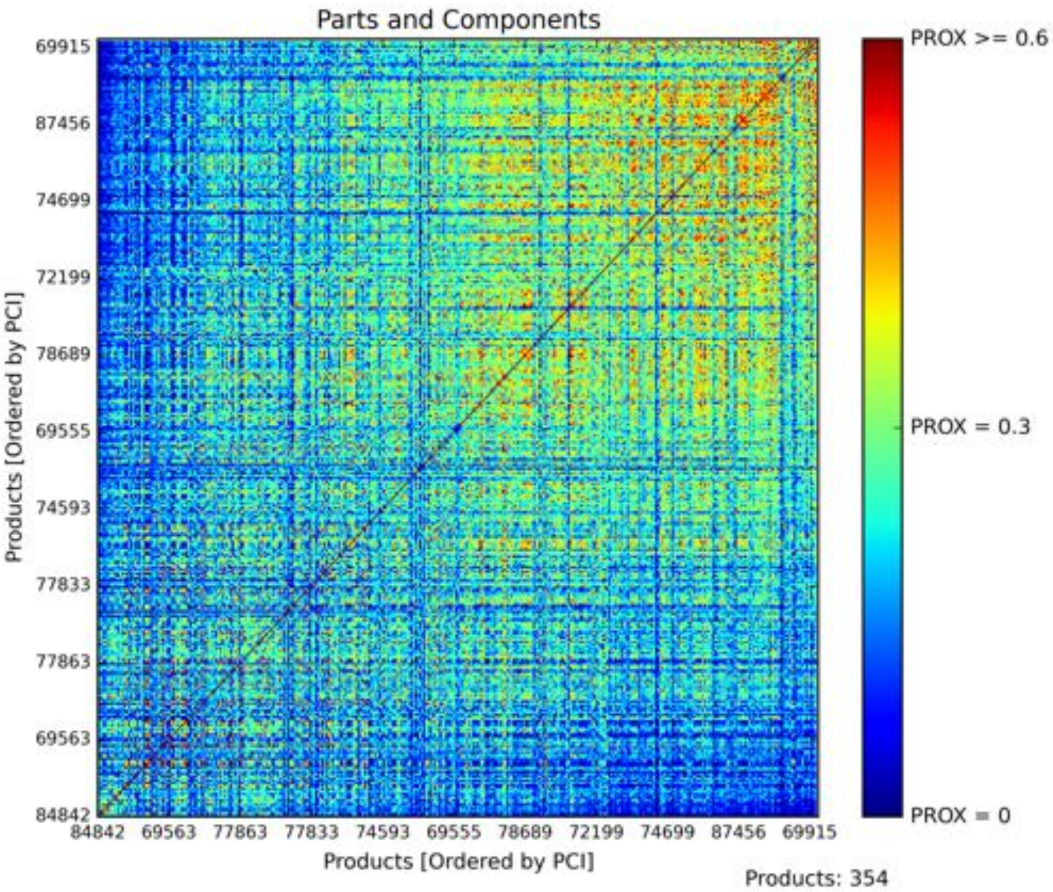
Given the high degree of network connectedness, parts and components and final assembly products therefore have the potential to act as gateway products into these industrial clusters through the process of product fragmentation. To assess the possibility of this hypothesis, a sector-level emergence analysis was undertaken to see if this could be observed in the data by computing the order of emergence across all countries. The Electronics and Machinery industries were considered for further analysis, given the well-defined nature of PSB products in these industries. In undertaking this analysis, we are able to see if there are any regularities in whether producing parts and components, or final assembly goods, on average leads to future exports in final assembly products, or vice versa.

Figure 5.10: Proximity Matrix grouped into PSB and final goods, sorted by product complexity indicator for the Year 2000



Notes: 1. Top right hand quadrant shows the within parts and components proximity relationships, and the bottom left hand quadrant show proximity relationships for all other goods. The other quadrants contain the same information and compares proximity relationship between parts and components with all other traded products. **2.** Proximity, which varies between zero and one, is an indicator of product similarity in the product space.

Figure 5.11: Proximity matrix of Parts and Components Only in the Year 2000



Notes: 1. This contains the same information as is contained in the upper right quadrant of Figure 5.10. **2.** Parts and components, as a group, are more co-exported, and have higher average proximity values than many other clusters of goods.

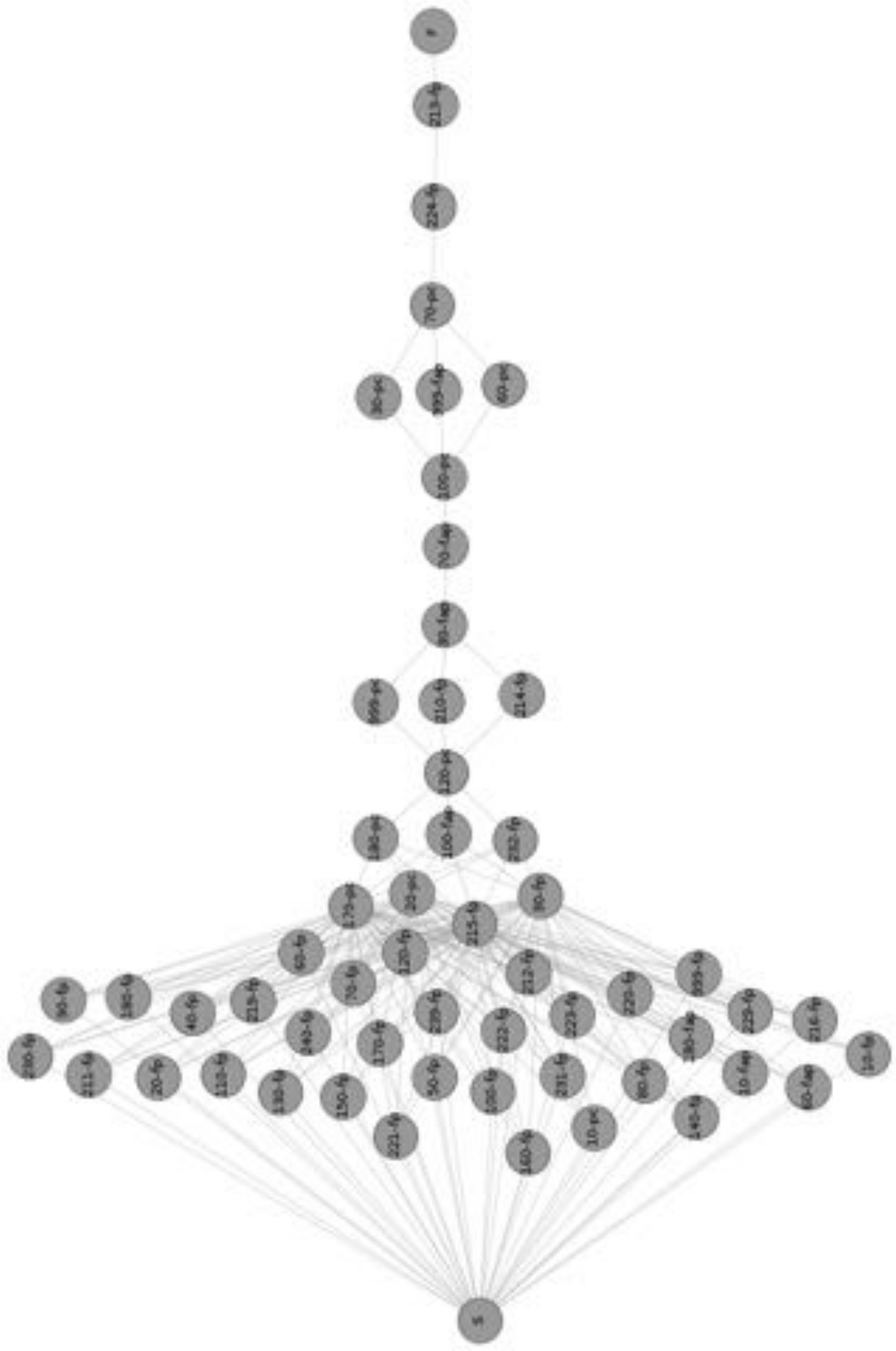
To account for the computational challenge of comparing all product pairs, for each country over time, the network analysis was done at a higher level of aggregation using product space communities that have been disaggregated into the three components, parts and components, final assembly products, and other final product clusters.

Comparing the average proximity between these different product nodes within the product space communities over time, it becomes apparent that clustering relationships in the product space are reasonably stable over time. Using the Electronics industry community as an example, the mean proximity between electronics PC and electronics FAP remains reasonably constant between 1995 and 2013 with a relatively high average proximity value of 0.22. Therefore, on average, the co-export potential between these two categories of products, over this period of time has remained constant. When considering the relationship between PC and FP, the average proximity decreases slowly over time from 0.19 in 1995 to 0.16 in 2013. Given these values fall in the densest part of the proximity histogram, for this average to change, a large number of inter-cluster proximity values need to change, which suggests that these two clusters are becoming more distant. Parts and components moving further away from final products is indicative of increased specialisation, and highlights the possibility that the within electronics cluster of PC could be slowly moving further away from the associated electronics FP. Intra-cluster PC proximity values in this industry varied between 0.26 and 0.31 over the timeframe, showing strong average co-export potentials of PC within the Electronics industry.

It has long been recognised that product fragmentation should increase the set of available opportunities by allowing countries to jump into new clusters by participating in only one segment of the value chain. To consider emergence patterns, a product emergence network was computed for all 204 countries. As an example, the diffusion network for Kenya (KEN) is shown in Figure 5.12. The figure orders sectors as they emerge in Kenya's export basket, with an RCA value that is greater than one. The 'S' node represents the year 1995 and each transition between years is used to identify new sectors that a country has not previously participated in. The large number of comparisons requires a relatively high-level analysis to be conducted which considered 58 product types, made up of 38 product space communities along with the three sub-cluster types: final products, final assembly products, and parts and components for each relevant product space community.[110]

[110] A higher level of disaggregation could offer much richer insight into these diffusion dynamics. This is an avenue for future work.

Figure 5.12: Kenya's product emergence diagram from 1995 to 2013



Notes: 1. The data used to construct this diagram is the BACI HS92 dataset that has been converted to SITC Rev 3 L5 data. **2.** The product nodes are encoded as the product space community number followed by the type of product: *FP*=final product, *FAP*=final assembly product, *PC*=parts and components. **3.** Emergence is constrained by no repeat instances; therefore product emergence is recorded as the first instance a country exports one of the products found in the associated product cluster. **4.** A large number of products emerge in 1996, as this is the start of the dataset used.

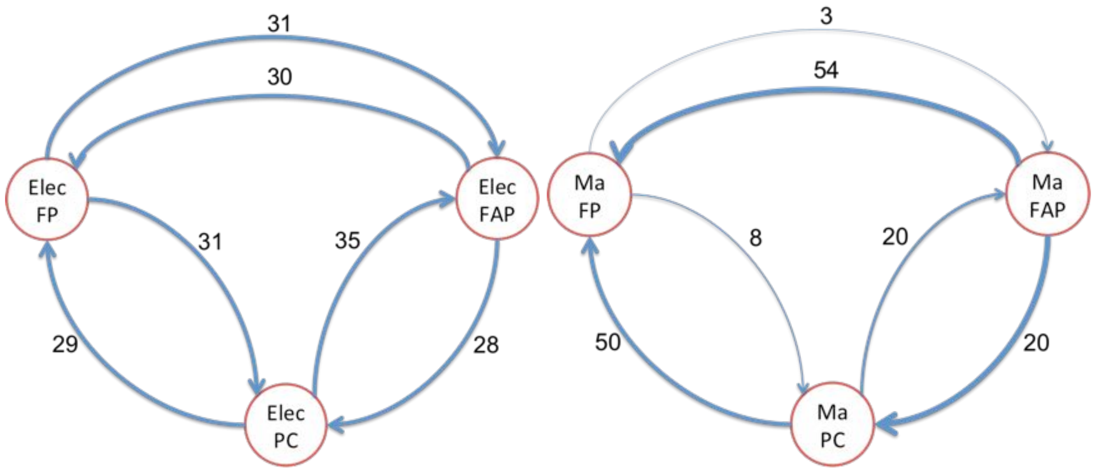
By computing the product emergence networks for each country, regularities in the pattern of diffusion can be computed by considering all pairwise sequences between the start node 'S' and the finish node 'F'. As countries vary in size, products are not expected to always emerge in adjacent years and therefore each level of the product emergence network is iterated over with full sequences considered with decaying influence $\frac{1}{\Delta t}$. Therefore the further away each comparative product pair is along the intertemporal chain of products, the weaker the influence on the estimated transition matrix.

As an example, if a country exports in the machinery final products (10-fp) cluster in 1999 and participated in machinery parts and components (10-pc) in 1998, a transition event is recorded between 10-fp and 10-pc with a value of 1. If, however, participation in the PC category was in 1996 then a weaker value of 0.5 is recorded. The main assumption is that products that emerge closer together in a sequence are given stronger weight when computing the overall transition dynamics, which recognises that country size may impact on emergence timing and therefore it is important to consider more than simple adjacent year transitions.

The computed transition diagrams for both the Electronic and Machinery product space communities are shown in Figure 5.13, computed over the years 1995 to 2013. The Electronics (Elec) sector offers a story in which transition scores found between all sub-sectors of the industry cluster are reasonably similar. Countries tended to diffuse between PC, FAP, and FP at similar rates. Each edge can be converted to relative probabilities by dividing each edge value by the total number of transition events between any two sub-industry clusters. It is observed that between 1995 and 2013, countries that diffuse between intra Electronics industry clusters do so at much the same probability. Countries tend to move from final products to the parts and component cluster with 52 percent chance compared to 48 percent from parts and components to final products. Therefore countries diffusing within the Electronics industry occur at similar frequencies between the three different product types.

One explanation of this finding is that for an industry that has participated heavily in production fragmentation since the late 1960's, it may have already reached a state of maturity that makes differentiating between final assembly products (FAP and final products (FP) difficult. The fact that intra-community clusters emerge at the same rate could be interpreted as a strong signal in support of product fragmentation in this industry. Countries have the capacity to participate in final goods products via specialising in a small segment of that product's value chain driving similar probabilities that final goods and/or parts and components are the first entry points into the Electronics cluster. An alternative explanation is that these dynamics may be observed due to the restricted time sample of this analysis. The industry between 1995 and 2013 is already mature and production networks are now well established in different segments of the value chain prior to 1995.

Figure 5.13: Product emergence dynamics within Electronics and Machinery export sectors [1995 to 2013]



Notes: 1. *Elec* = Electronics Industry, and *Ma* = Machinery Industry as identified by the product space community cluster. *FP* = Final Products, *FAP* = Final Assembly Products, *PC* = Parts and Components. **2.** Edge weights are relative frequencies of moving from one node to another. **3.** While transitions between FP, FAP, and PC in the electronics sector are similar between 1995 and 2013, the Machinery sector shows a strong tendency to move from PC to FP and from FAP to FP with similar transition frequencies between PC and FAP products.

Diffusion in the Machinery industry presents a more dynamic picture, as it appears to follow an asymmetric diffusion pattern. Within the Machinery product community, countries tend to participate in PC and FAP prior to the export of FP. Final machinery products tend to emerge after participation in machinery parts and components in 86 percent of cases, and from final assembly machinery products to final products in 95 percent of cases. However, participation between final assembly products and parts and components show equal rates of emergence, which is an interesting finding given final assembly products are those that have been identified to have favourable characteristics for production fragmentation.

This analysis provides evidence that specialisation in PC and FAP play an equally important role as gateway products in the Machinery industry between the year 1995 and 2013. Countries were observed to diffuse from parts and components and final assembly products to other final goods.

5.6 CONCLUSION

This chapter has considered the process of product fragmentation and the challenge it imposes on the data-centric techniques used to infer economic complexity. By altering product types of final assembly products (FAP), it is shown that relative rankings of some developing countries may be biased upwards due to the indirect linking of these economies to relatively more advanced economies through the shared relationship with final assembly products in the bipartite network structure of exports.

This bias was demonstrated by differentiating final assembly products into separate product groups within the M_{cp} matrix. As a result China, other South-East Asian, and Eastern European countries' complexity rankings decrease as one key bond linking them to more sophisticated economies such as Japan and Germany is severed. With the introduction of differentiated final assembly products (FAP) developing countries that exported FAP became predominantly linked with countries that possessed less sophisticated export structures on average. When comparing low-type and high-type final assembly products, this alteration in the bipartite network structure resulted in low-type products being demoted in the complexity rankings, and high-type final assembly products promoted relative to their original ranked positions when both types were pooled across countries. Further consideration of value added data would be an important avenue for additional research to examine the affect that this mismatch between countries and products has on these new measures of revealed rankings of complexity.

The structural role and diffusion characteristics of PC and FAP type products in the product space were also studied. Parts and components play a clear role in connecting the dense core of the product space. They not only share strong proximity relationships with other PC but also share high proximity to other final products. The dynamic characteristics of emergence were then studied by estimating a transition matrix based on the order of product emergence computed for each country. In the case of the Machinery sector, parts and components or final assembly products tended to emerge prior to other final machinery products. This suggests that parts and components and final assembly products may act as gateway products prior to diffusion to more sophisticated machinery products. This pattern was not observed for the Electronics sector, presumably because it is a relatively mature industry that has been subject to the forces of product fragmentation for many decades. These transition dynamics may no longer be observable for the period 1995 to 2013 sample examined here. Alternatively, this result may be interpreted as strong evidence of product fragmentation in this industry as final products emerge at similar rates in country exports baskets to parts and components or final assembly products.

This chapter lends itself to a number of extensions. First, the method of constructing differentiated products based on levels of GDP per capita is a crude approximation for splitting final assembly products into two groups. The key assumption is that less developed economies are most likely to participate in a small component of the value chain, such as assembly services from imported parts and components. Closer country-specific examination, and the use of more disaggregated HS6 data may provide more precise estimates of the upward complexity bias.

Second, the estimation of dynamic product emergence paths lends itself to many possible avenues for future work. Using more disaggregated product level emergence to estimate Markov transition matrices may assist with exploring the heterogeneity and more detailed complex patterns of product emergence. Using trade data in this way could assist in our understanding of whether the product space can be used as an industrial policy tool, or alternatively exploring whether there exists a regular path dependant outcome through the development process.

CHAPTER 6 CONCLUSION

> *"However, in international trade and international relations, there are fundamental unanswered questions that require modelling beyond a two-at-a-time or a market approach, as the actors involved are inherently networked and face large externalities"* (Jackson 2014)

The study of economic development is a challenging task given the great variety in historical development paths among countries. The diversity in underlying factors leads to a corresponding diversity in the theoretical and analytical tools that we require to dissect this complex multidimensional system we call the economy. In recent years there has been a broader interest in new analytical techniques that have the potential to provide alternative avenues for research and new insight. The recent network complexity analysis techniques, that form the foundation for this thesis, is one example of how adopting methods from the field of network science can provide a new set of tools available for research in economic development. This multidisciplinary approach in collaboration with advancements in computational capacity and infrastructure make for exciting research opportunities ahead.

The network complexity approach draws on the relational structure of international trade data to infer information about underlying levels of economic complexity. Exploiting the relational component of this data allows for highly disaggregated measures to be computed, opening up new opportunities to study more detailed patterns of economic development. Relational data sits at the core of complex network analysis that forms a bipartite network between countries and the products that each country exports with RCA. This network provides immediate information regarding the level of country diversification, and a characterisation of the rareness of products in terms of ubiquity. The higher order relationships that form as you reflect over this bipartite data structure yield the economic and product complexity indicators (ECI and PCI). Complex economies are those that are *both* highly diversified and export relatively non-ubiquitous products. The novelty of this technique is using information that is contained in two dimensions, diversity and ubiquity, to contextualise country export structure. This allows for the differentiation between unique exports that are largely endowment driven from relatively undiversified economies, with unique products that tend to be more complex by nature and are exported by diversified countries. This preserves the identity of data, rather than using aggregate forms that hide a richness of underlying heterogeneity.

The primary contributions of this thesis are embedded in Chapters 3, 4, and 5. Chapter 3 looked to build a bridge between the traditional literature on comparative advantage and the measures of ECI and PCI. Theories of comparative advantage in the many-country, many-product, many-factor context are used to build expectations on the broad patterns of specialisations that are predicted by the fundamental Ricardian and Heckscher-Ohlin theories. Theoretical models tend to cluster products around a diagonal line of specialisation by implicitly ordering the rows and columns of a cross-country record of exports, such as the M_{cp} matrix. While there are many possible reasons for exports to be located away from the diagonal, such as factor intensity reversals, a dominant pattern of specialisation should still be observable based on the ordering of rows and columns in this matrix.

However this research shows that when one plots data on revealed comparative advantage, the pattern is initially a striking departure from specialisation towards a triangular density shape presented in the M_{cp} matrix. This shape tells us that diversified economies tend to export the both ubiquitous and non-ubiquitous products, while less diversified economies primarily export only ubiquitous products.[111] Chapter 3 recognises that country diversification and product ubiquity is only one way to organise the information that is contained within this matrix. When inspecting the mechanics behind the *method of reflections* it can alternatively be viewed as a sorting heuristic that implicitly sorts rows and columns of the matrix in a way that clusters the pattern of exports around a diagonal line of specialisation, and achieves this in in the high-dimensional context of many countries and many products. While the results contain a high degree of observed heterogeneity, transforming the matrix using country and product value shares brings the data closer towards the theoretical patterns of specialisation predicted by the theory of comparative advantage.

The connection between these two strands of literature was then illustrated in an empirical analysis, which identified products that most negatively deviated from a

[111] In network science this also suggests that this data has a hierarchical structure.

country's position along this diagonal line of specialisation. This deviation represents a measure, which identifies products that are *least* associated with that country's comparative advantage in *complexity*. In the case of the United States, the top twenty most negatively deviating products correlate strongly with industries that received large domestic subsidies including Wheat, Soybeans and other agricultural products. It was also shown that Germany's RCA vector within the M_{cp} vector was upward sloping from least complex to more complex products. Germany tends to export more complex products with a greater intensity of RCA as one moves toward its relative position of specialisation along the diagonal. This short exercise demonstrates how the existing literature on comparative advantage and these new network complexity methods can be used as complementary tools for further research.

The network complexity approach also paves the way for new forms of empirical analytics using disaggregated measures such as proximity. Chapter 4 examined a structural approach by using the product space network as a disaggregated measure of product similarity. This structure that is provided by the product space network, is then used to investigate the *rate* and *distance* of product emergence in relation to periods of trade liberalisation. The work in chapter 4 suggests that trade liberalisation is supportive of both an increase of 18% to 24% in the *rate* of product emergence, in addition to a moderate 8% increase in the *distance* of diffusion. In light of a country capabilities viewpoint, this provides indirect evidence that both the *size* and the *diversity* of underlying country *capabilities* tend to increase under periods of trade liberalisation.

The remaining work in Chapter 5 reflects on a limitation of the new measures of complexity by considering how the process of product fragmentation impacts the information that is contained in the bipartite network mapping between countries and products that is used to infer complexity. By considering final assembly products, it is suggested that indirect links are formed between developing and developed economies that encapsulate incorrect information. Many developing economies export final assembly products by importing parts and components and providing final assembly services based on their comparative advantage in labour. Developed economies on the other hand, are more likely to have the full suite of capabilities required to produce these goods from start to finish, or alternatively tend to produce higher quality differentiated products. These indirect links between these two types of production environments will bias developing country complexity scores upwards. This bias is demonstrated by differentiating final assembly products (FAP) into low-type and high-type products. In doing so the complexity ranks of many countries in East Asia, South East Asia, and Eastern Europe decline. Low-type products tend to fall to the bottom quartile of the product complexity ranking, while the equivalent high-type products remain in the top quartile of the product complexity ranks.

This research on product fragmentation concludes with an investigation into the role of parts and components in the product space network. Parts and components (PC) are found to be embedded within the industrial core alongside associated industry-specific final goods. As a group they exhibit much higher levels of co-export potential relative to all other products. However despite higher average proximity within parts and components, this product type also shares high proximity connections with other associated final goods. This heterogeneity in proximity relationships, suggests that parts and components may play a gateway role for countries to diffuse from the periphery to the densely connected core of the product space network. This was explored using a sector-level dynamic analysis of product emergence for the machinery and the electronics sectors. In the case of the machinery sector, PC and FAP tended to emerge in country export vectors prior to the export of associated final exports. This provides some evidence of path dependent emergence patterns in which the fragmentation of the production is providing new opportunities to participate in manufacturing. This pattern is not observed however, for the electronics sector, presumably because it is a relatively mature industry that has been evolving since the late 1960s and hence may have passed the point at which these transition dynamics are broadly observable in the 1995 to 2013 sample examined here.

There has always been a feedback loop between the development of economic theory and the analysis of real world data. With more prevalent availability of data, coupled with new computational capacity, and a vast array of new analytical methods and techniques, there are more opportunities than ever to conduct research at this interesting junction point.

6.1 FURTHER RESEARCH

This is an exciting time to be studying economic development. By applying advanced multidisciplinary programming tools to large data sets constructed from many difference sources, this thesis has explored the possibility of using the network complexity analysis to broaden our understand of international dimensions of the process of economic development. Ideas for future research have already been discussed in the concluding sections of each chapter. It is pertinent to end this thesis with my listing of two key areas, which deserve specific attention in further research.

The product space of imports

Is it possible to identify capabilities that countries may lack due to import intensities? More specifically, can the interaction between exports and imports reveal patterns that more succinctly identify country strengths and weaknesses in the traded goods space?

Industrial policy and the predictability of export patterns

Much has been written on the potential for the product space work to act as an identification tool for future country participation. Can we use the recent advances in statistical pattern recognition through machine learning methods to accurately predict the emergence of new products?

REFERENCES

Ahn, J, Khandelwal, AK & Wei, SJ 2011, 'The role of intermediaries in facilitating trade', *Journal of International Economics*, vol. 84, no. 1, pp. 73–85.

Alessandrini, M, Fattouh, B, Ferrarini, B & Scaramozzino, P 2011, 'Tariff liberalization and trade specialization: Lessons from India', *Journal of Comparative Economics*, vol. 39, no. 4, pp. 499–513.

Amiti, M & Konings, J 2007, 'Trade Intermediate and Productivity: Liberalization, Inputs, from Indonesia Evidence', *The American Economic Review*, vol. 97, no. 5, pp. 1611–1638.

Anderson, JE 1979, 'A Theoretical Foundation For the Gravity Equation', *The American Economic Review*, vol. 69, no. 1, pp. 106–116.

Anderson, JE & van Wincoop, E 2003, 'Gravity with Gravitas : A Solution to the Border Puzzle', *The American Economic Review*, vol. 93, no. 1, pp. 170–192.

Arellano, M & Bond, S 1991, 'Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations', *Review of Economic Studies*, vol. 58, no. 2, pp. 277–297.

Arkolakis, C, Costinot, A & Rodriguez-Clare, A 2012, 'New trade models, same old gains?', *American Economic Review*, vol. 102, no. 1, pp. 94–130.

Arthur, WB 2013, *Complexity Economics : A Different Framework for Economic Thought*, SFI Working Paper (Santa Fe Institute).

Athukorala, P 2005, 'Product Fragmentation and Trade Patterns in East Asia', *Asian Economic Papers*, vol. 4, no. 3, pp. 1–27.

Athukorala, P 2011, 'Trade Liberalisation and The Poverty of Nations: A Review Article', *The Journal of Development Studies*, vol. 47, no. 3, pp. 533–543.

Athukorala, P 2014, 'How India Fits into Global Production Sharing: Experience, Prospects, and Policy Options', in S Shekhar, B Barry, & P Arvind (eds), *India Policy Forum*, Sage Publications, pp. 57–108.

Athukorala, P & Rajapatirana, S 2000, *Liberalization and Industrial Transformation: Sri Lanka in International Perspective*, Oxford University Press.

Athukorala, P & Waglé, S 2014, 'Trade Liberalisation and Export Performance in Transition: The Case of Georgia', *The World Economy*, vol. 37, no. 12, pp. 1669–1691.

Athukorala, PC 2010, *Production Networks and Trade Patterns in East Asia: Regionalization or Globalization?*, ADB - Working Paper.

Athukorala, PC 2012, 'Asian trade flows: Trends, patterns and prospects', *Japan and the World Economy*, vol. 24, no. 2, pp. 150–162.

Athukorala, PC & Yamashita, N 2006, 'Production fragmentation and trade integration: East Asia in a global context', *North American Journal of Economics and Finance*, vol. 17, no. 3, pp. 233–256.

Balassa, B 1965, 'Trade Liberalisation and Revealed Comparative Advantage', *The Manchester School*, vol. 33, no. 2, pp. 99–123.

Baldwin, JR & Gu, W 2004, 'Trade Liberalization: Export-market Participation, Productivity Growth, and Innovation', *Oxford Review of Economic Policy*, vol. 20, no. 3, pp. 372–392.

Barigozzi, M, Fagiolo, G & Mangioni, G 2011, 'Identifying the community structure of the international-trade multi-network', *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 11, pp. 2051–2066.

Baskaran, T, Blöchl, F, Brück, T & Theis, FJ 2011, 'The Heckscher–Ohlin model and the network structure of international trade', *International Review of Economics & Finance*, vol. 20, no. 2, pp. 135–145.

Benedictis, L De, Nenci, S, Santoni, G, Tajoli, L & Vicarelli, C 2013, *Network Analysis of World Trade using the BACI-CEPII dataset*, Centre D'Etudes Prospectives Et D'Informations Internationales (Working Paper).

Benedictis, L De & Tamberi, M 2001, *A note on the Balassa Index of Revealed Comparative Advantage*, Mimeo.

De Benedictis, L & Tamberi, M 2004, 'Overall specialization empirics: Techniques and applications', *Open Economies Review*, vol. 15, no. 4, pp. 323–346.

Bowen, H 1983, 'On the theoretical interpretation of indices of trade intensity and revealed comparative advantage', *Review of World Economics:*

Weltwirtschaftliches Archiv, vol. 119, no. 3, pp. 464–472.

Brown, C & Linden, G 2005, 'Offshoring in the Semiconductor Industry: A Historical Perspective', *Brookings Trade Forum*, vol. 2005, no. 1, pp. 279–322.

Brusco, MJ & Stahl, S 2001, 'An interactive multiobjective programming approach to combinatorial data analysis', *Psychometrika*, vol. 66, no. 1, pp. 5–24.

Bugamelli, M 2001, *International Specialization Patterns of the Euro Area*, Italian Central Bank (Working Paper).

Chaney, T 2014, 'The Network Structure of International Trade', *American Economic Review*, vol. 104, no. 11, pp. 3600–3634.

Chor, D 2010, 'Unpacking sources of comparative advantage: A quantitative approach', *Journal of International Economics*, vol. 82, no. 2, pp. 152–167.

Córcoles, D, Díaz-Mora, C & Gandoy, R 2014, 'Product sophistication: A tie that binds partners in international trade', *Economic Modelling*, vol. 44, no. Supplement 1, pp. S33–S41.

Corden, MW 1971, *The Theory of Protection*, Oxford University Press.

Costinot, A 2009, 'On the origins of comparative advantage', *Journal of International Economics*, vol. 77, no. 2, pp. 255–264.

Costinot, A & Donaldson, D 2012, 'Ricardo's theory of comparative advantage: old idea, new evidence', *The American Economic Review*, vol. 102, no. 3, pp. 453–458.

Costinot, A, Donaldson, D & Komunjer, I 2011, 'What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas', *The Review of Economic Studies*, vol. 79, no. 2, pp. 581–608.

Cristelli, M, Tacchella, A & Pietronero, L 2014, *Economic complexity Measuring the Intangibles A Consumer's Guide*, Working Paper.

Davis, DR 1995, 'Intra-industry trade: A Heckscher-Ohlin-Ricardo approach', *Journal of International Economics*, vol. 39, no. 3-4, pp. 201–226.

Deardorff, A 1982, 'The general validity of the Heckscher-Ohlin theorem', *The American Economic Review*, vol. 72, no. 4, pp. 683–694.

Deardorff, A 1984, 'Testing Trade Theories and Predicting Trade Flows', in RW Kenen &

PB Jones (eds), *Handbook of International Economics*, North Holland, Amsterdam.

Deardorff, A V 1999, *Patterns of Trade and Growth across Cones Patterns of Trade and Growth across Cones*, Working Paper No 443 (University of Michigan).

Dedrick, J, Kraemer, KL & Linden, G 2009, 'Who profits from innovation in global value chains?: A study of the iPod and notebook PCs', *Industrial and Corporate Change*, vol. 19, no. 1, pp. 81–116.

Dixit, AK & Norman, V 1980, *Theory of International Trade*, Cambridge University Press.

van Dongen, S 2000, *Graph clustering by flow simulation*, University of Utrecht (Phd Thesis).

Dornbusch, R 1992, 'The Case for Trade Liberalization in Developing Countries', *Journal of Economic Perspectives*, vol. 6, no. 1, pp. 69–85.

Dornbusch, R, Fischer, S & Samuelson, PA 1977, 'Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods', *The American Economic Review*, vol. 67, no. 5, pp. 823–839.

Dueñas, M & Fagiolo, G 2014, 'Global Trade Imbalances: a Network Approach', *Advances in Complex Systems*, vol. 17, no. 3&4, p. 1450014 (29 pages).

Eaton, J & Kortum, S 2002, 'Technology, geography, and trade', *Econometrica*, vol. 70, no. 5, pp. 1741–1779.

Evenett, SJ & Keller, W 2002, 'On Theories Explaining the Success of the Gravity Equation', *Journal of Political Economy*, vol. 110, no. 2, pp. 281–316.

Feenstra, R 2001, 'Using the gravity equation to differentiate among alternative theories of trade', *Canadian Journal of Economics*, vol. 34, no. 2, pp. 430–447.

Feenstra, RC, Hai, W, Woo, WT & Yao, S 1999, 'Discrepancies in International Data: An Application to China-Hong Kong Entrepot Trade', *American Economic Review*, vol. 89, no. 2, pp. 338–343.

Feenstra, RC, Inklaar, R & Timmer, MP 2015, 'The Next Generation of the Penn World Table', *forthcoming American Economic Review*.

Feenstra, RC, Lipsey, RE & Bowen, HP 1997, *World Trade Flows, 1970-1992, with Production and Tariff Data*, Working Paper Series 5910, National Bureau of Economic Research.

Feenstra, RC, Lipsey, RE, Deng, H, Ma, AC & Mo, H 2005, *World Trade Flows: 1962-2000*, Working Paper No 11040, National Bureau of Economic Research.

Feenstra, RC & Rose, AK 2000, 'Putting Things in Order: Patterns of Trade Dynamics and Growth', *Review of Economics and Statistics*, vol. 82, pp. 369–382.

Fisher, EO 2011, 'Heckscher–Ohlin theory when countries have different technologies', *International Review of Economics & Finance*, vol. 20, no. 2, pp. 202–210.

Goldberg, P, Khandelwal, A, Pavcnik, N & Topalova, P 2009, 'Trade liberalization and new imported inputs', *American Economic Review*, vol. 99, no. 2, pp. 494–500.

Greenaway, D, Morgan, W & Wright, P 1997, 'Trade liberalization and growth in developing countries: Some new evidence', *World Development*, vol. 25, no. 1, pp. 1885–1892.

Greenaway, D, Morgan, W & Wright, P 1998, 'Trade reform, adjustment and growth: what does the evidence tell us?', *The Economic Journal*, vol. 108, no. 1993, pp. 1547–1561.

Greenaway, D, Morgan, W & Wright, P 2002, 'Trade liberalisation and growth in developing countries', *Journal of development economics*, vol. 67, pp. 229–244.

Guillaume, G & Zignago, S 2010, *BACI: International Trade Database at the Product-Level. The 1994-2007 Version*, Working Paper No 2010-23, CEPII.

Gutin, G & Punnen, A 2002, *The Travelling Salesman Problem and Its Variations* Volume 12., Kluwer Dordrecht.

Haberler, G 1930, 'The Theory of Comparative Cost and Its Use in the Defense of Free Trade', *Weltwirtschaftliches Archiv*, vol. 32, pp. 349–370.

Hausmann, R & Hidalgo, C 2011, 'The network structure of economic output', *Journal of Economic Growth*, vol. 16, no. 4, pp. 309–342.

Hausmann, R & Hidalgo, C 2014, *The atlas of economic complexity: Mapping paths to prosperity*, Center for International Development (Harvard University).

Hausmann, R & Klinger, B 2006, 'Structural Transformation and Patterns of Comparative Advantage in the Product Space', *CID Working Paper*, vol. 128, pp. 1–38.

He, J & Deem, MW 2010, 'Structure and Response in the World Trade Network', *Physical Review Letters*, vol. 105, no. 19, p. 198701 (4 pages).

Head, K, Jing, R & Swenson, DL 2014, 'From Beijing to Bentonville: Do multinational retailers link markets?', *Journal of Development Economics*, vol. 110, pp. 79–92.

Helbing, D & Kirman, AP 2013, 'Rethinking economics using complexity theory', *Real-World Economics Review*, vol. 64, pp. 23–52.

Helpman, E & Krugman, P 1985, *Market Structure and Trade: Increasing Returns, Imperfect Competition, and the International Economy*, MIT Press.

Helpman, E, Melitz, M & Rubinstein, Y 2008, 'Estimating trade flows: Trading partners and trading volumes', *The Quarterly Journal of Economics*, vol. 123, no. 2, pp. 441–487.

Hidalgo, C 2009, *The Product Space and the Building Blocks of Economic Development*, Mimeo.

Hidalgo, C 2015, *Why Information Grows*, Basic Books.

Hidalgo, C & Hausmann, R 2008, 'A network view of economic development', *Developing alternatives*, vol. 12, no. 1, pp. 5–10.

Hidalgo, C, Klinger, B, Barabási, A & Hausmann, R 2007, 'The product space conditions the development of nations.', *Science*, vol. 317, no. 5837, pp. 482–7.

Hillman, A 1980, 'Observations on the relation between "revealed comparative advantage" and comparative advantage as indicated by pre-trade relative prices', *Review of World Economics*, vol. 116, no. 2, pp. 315–321.

Hinloopen, J & van Marrewijk, C 2008, 'Empirical relevance of the Hillman condition for revealed comparative advantage: 10 stylized facts', *Applied Economics*, vol. 40, no. 18, pp. 2313–2328.

Hubert, L, Arabie, P & Meulman, J 1987, *Combinatorial Data Analysis: Optimization by Dynamic Programming*, Society for Industrial Mathematics.

Hummels, D, Ishii, J & Kei-Mu, Y 2001, 'The nature and growth of vertical specialisation in world trade', *Journal of International Economics*, vol. 54, pp. 75–96.

Jackson, MO 2014, 'The Past and Future of Network Analysis in Economics', in *The Oxford Handbook on the Economics of Networks*, Oxford University Press, pp. 1–10.

Jackson, MO & Nei, SM 2015, *Networks of Military Alliances , Wars , and International Trade*, Working Paper.

Jackson, MO & Watts, A 2002, 'The Evolution of Social and Economic Networks', *Journal of Economic Theory*, vol. 106, no. 2, pp. 265–295.

Jacomy, M, Venturini, T, Heymann, S & Bastian, M 2014, 'ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software', *PLoS ONE*, vol. 9, no. 6, pp. 1–12.

Kali, R, Méndez, F & Reyes, J 2007, 'Trade structure and economic growth', *The Journal of International Trade & Economic Development: An International and Comparative Review*, vol. 16, no. 2, pp. 245–269.

Kali, R, Reyes, J, McGee, J & Shirrell, S 2013, 'Growth networks', *Journal of Development Economics*, vol. 101, pp. 216–227.

Kehoe, TJ & Ruhl, KJ 2013, 'How Important Is the New Goods Margin in International Trade?', *Journal of Political Economy*, vol. 121, no. 2, pp. 358–392.

Kornai, J 1992, *The Socialist System: The Political Economy of Communism*, Princeton University Press, Princeton, NJ.

Krishna, P & Levchenko, A 2012, *Comparative Advantage , Complexity , and Volatility*, Working Paper No 632 (Gerald R. Ford School of Public Policy).

Krueger, A 1997, 'Trade policy and economic development: how we learn', *The American Economic Review*, vol. 87, no. 1, pp. 1–22.

Krueger, A 1998, 'Why Trade Liberalisation is Good for Growth', *The Economic Journal*, vol. 108, no. 450, pp. 1513–1522.

Krugman, P 2008, 'Trade and Wages, Reconsidered', *Brookings Papers on Economic Activity*, vol. 2008, no. 1, pp. 103–154.

Kunimoto, K 1977, 'Typology of trade intensity indices', *Hitotsubashi Journal of Economics*, vol. 17, no. 2, pp. 15–32.

Lafay, G 1992, 'The Measurement of Revealed Comparative Advantages', in MG Dagenais & PA Muet (eds), *International Trade Modeling*, Chapman & Hill, London, pp. 209–220.

Lall, S 2000, 'The Technological Structure and Performance of Developing Country

Manufactured Exports, 1985-98', *Oxford Development Studies*, vol. 28, no. 3, pp. 337–369.

Laursen, K 2000, 'Do export and technological specialisation patterns co-evolve in terms of convergence or divergence? Evidence from 19 OECD countries, 1971–1991', *Journal of Evolutionary Economics*, vol. 10, pp. 415–436.

Leamer, E 1984, *Sources of International Comparative Advantage: Theory and Evidence*, MIT Press.

Lee, J 1997, 'The maturation and growth of infant industries: The case of Korea', *World Development*, vol. 25, no. 8, pp. 1271–1281.

Leromain, E & Orefice, G 2013, *New revealed comparative advantage index: dataset and empirical distribution*, CEPII (Working Paper).

Leromain, E & Orefice, G 2014, 'New revealed comparative advantage index: Dataset and empirical distribution', *International Economics*, vol. 139, pp. 48–70.

Liiv, I 2010, 'Seriation and Matrix Reordering Methods : An Historical Overview', *Statistical Analysis and Data Mining*, vol. 3, pp. 70–91.

Linden, G, Kraemer, KL & Dedrick, J 2009, 'Who captures value in a global innovation network?', *Communications of the ACM*, vol. 52, no. 3, p. 140.

Mahutga, M 2006, 'The persistence of structural inequality? A network analysis of international trade, 1965–2000', *Social Forces*, vol. 84, no. 4, p. 1863.

Marchese, S & Simone, F 1989, 'Monotonicity of indices of 'Revealed "comparative advantage: Empirical evidence on Hillman's condition"', *Review of World Economics: Weltwirtschaftliches Archiv*, vol. 125, no. 1, pp. 158–167.

Markusen, J 2013, 'Putting per-capita income back into trade theory', *Journal of International Economics*, vol. 90, no. 2, pp. 255–265.

McCormick, WT, Schweitzer, PJ & White, TW 1972, 'Problem Decomposition and Data Reorganization by a Clustering Technique', *Operations Research*, vol. 20, no. 5, pp. 993–1009.

Melitz, MJ 2005, 'When and how should infant industries be protected?', *Journal of International Economics*, vol. 66, no. 1, pp. 177–196.

Melitz, MJ & Ottaviano, GIP 2008, 'Market size, trade, and productivity', *The Review of*

Economic Studies, vol. 75, pp. 295–316.

Morrow, PM 2010, 'Ricardian–Heckscher–Ohlin comparative advantage: Theory and evidence', *Journal of International Economics*, vol. 82, no. 2, pp. 137–151.

Neary, JP 1978, 'Short-Run Capital Specificity and the Pure Theory of International Trade', *Economic Journal*, vol. 88, no. 351, pp. 488–510.

OECD 2014, *Agricultural Policy Monitoring and Evaluation 2014: OECD Countries*, OECD Publishing.

Ohlin, B 1933, *Interregional and International Trade*, Harvard University Press, Cambridge.

Papageorgiou, D, Choksi, AM & Michaely, M 1990, *Liberalizing foreign trade in developing countries: the lessons of experience*, The World Bank.

Price, VC 2001, 'Some Causes and Consequences of Fragmentation', in *Fragmentation: New Production Patterns in the World Economy*, pp. 88–107.

Proudman, J & Redding, S 2000, 'Evolving patterns of international trade', *Review of international economics*, vol. 8, no. 3, pp. 373–396.

Puga, D & Venables, T 1999, 'Agglomeration and economic development: import substitution versus trade liberalisation', *The Economic Journal*, vol. 109, no. 455, pp. 292–311.

Reyes, J, Schiavo, S & Fagiolo, G 2010, 'Using complex networks analysis to assess the evolution of international economic integration: The cases of East Asia and Latin America', *The Journal of International Trade & Economic Development*, vol. 19, no. 2, pp. 215–239.

Reyes, J, Wooster, R & Shirrell, S 2014, 'Regional Trade Agreements and the Pattern of Trade: A Networks Approach', *The World Economy*, vol. 37, no. 8, pp. 1128–1151.

Ricardo, D 1817, *On the Principles of Political Economic and Taxation*, London: John Murray, Albemarle-Street, London.

Rodriguez, F & Rodrik, D 2001, 'Trade policy and economic growth: a skeptic's guide to the cross-national evidence', *NBER Macroeconomics Annual 2000, Volume 15*, vol. 15, no. January, pp. 261–338.

Romalis, J 2004, 'Factor proportions and the structure of commodity trade', *American*

Economic Review, vol. 94, no. 1, pp. 67–97.

Sachs, J & Warner, A 1995, 'Economic-Reform and the Process of Global Integration', *Brookings Papers on Economic Activity*, no. 1, pp. 1–95.

Samuelson, PA 1949, 'International Factor-Price Equalisation Once Again', *The Economic Journal*, vol. 59, no. 234, pp. 181–197.

Santos-Paulino, A & Thirlwall, AP 2004, 'The Impact of Trade Liberalisation on Exports, Imports, and the Balance of Payments of Developing Countries', *The Economic Journal*, vol. 114, no. 493, pp. F50–F72.

Santos-Paulino, AU 2005, 'Trade Liberalisation and Economic Performance: Theory and Evidence for Developing Countries', *The World Economy*, vol. 28, no. 6, pp. 783–821.

Shi, P, Zhang, J, Yang, B & Luo, J 2014, 'Hierarchicality of trade flow networks reveals complexity of products.', *PloS one*, vol. 9, no. 6, p. e98247.

Shin, N, Kraemer, KL & Dedrick, J 2012, 'Value Capture in the Global Electronics Industry: Empirical Evidence for the "Smiling Curve" Concept', *Industry & Innovation*, vol. 19, no. 2, pp. 89–107.

Shirotori, M, Tumurchudur, B & Cadot, O 2010, *Revealed Factor Intensity Indices At the Product Level*, Working Paper No 44 (UNCTAD).

Silva, J & Tenreyro, S 2006, 'The log of gravity', *The Review of Economics and statistics*, vol. 88, no. 4, pp. 641–658.

Smith, D & White, D 1992, 'Structure and dynamics of the global economy: network analysis of international trade 1965–1980', *Social Forces*, vol. 70, no. 4, pp. 857–893.

Swenson, DL 2005, 'Overseas assembly and country sourcing choices', *Journal of International Economics*, vol. 66, no. 1, pp. 107–130.

Swenson, DL 2006, 'Country Competition and US Overseas Assembly', *The World Economy*, vol. 29, no. 7, pp. 917–937.

Swenson, DL 2008, 'Multinationals and the creation of Chinese trade linkages', *Canadian Journal of Economics*, vol. 41, no. 2, pp. 596–618.

Tacchella, A, Cristelli, M, Caldarelli, G, Gabrielli, A & Pietronero, L 2012, 'A new metrics for countries’ fitness and products' complexity.’, *Scientific reports*, vol. 2, p. 723.

Tacchella, A, Cristelli, M, Caldarelli, G, Gabrielli, A & Pietronero, L 2013, ‘Economic complexity: Conceptual grounding of a new metrics for global competitiveness’, *Journal of Economic Dynamics and Control*, vol. 37, no. 8, pp. 1683–1691.

Thompson, H 2001, ‘International Trade with Three Factors, Goods or Countries’, *Keio economic studies*, vol. 38, no. 1, pp. 43–52.

Topalova, P & Khandelwal, A 2011, ‘Trade Liberalization and Firm Productivity: The Case of India’, *Review of Economics and Statistics*, vol. 93, no. 3, pp. 995–1009.

Vanek, J 1963, ‘Variable Factor Proportions and Inter-Industry Flows in the theory of International Trade’, *The Quarterly Journal of Economics*, vol. 77, no. 1, pp. 129–142.

Vernon, R 1966, ‘International Investment and International Trade in the Product Cycle’, *The Quarterly Journal of Economics*, vol. 80, no. 2, pp. 190–207.

Vollrath, T 1991, ‘A theoretical evaluation of alternative trade intensity measures of revealed comparative advantage’, *Weltwirtschaftliches Archiv*, pp. 265–279.

Wacziarg, R & Welch, KH 2008, ‘Trade liberalization and growth: New evidence’, *World Bank Economic Review*, vol. 22, no. 2, pp. 187–231.

Wilson, CA 1980, ‘On the General Structure of Ricardian Models with a Continuum of Goods: Applications to Growth, Tariff Theory, and Technical Change’, *Econometrica*, vol. 48, no. 7, pp. 1675–1702.

Winters, LA 2004, ‘Trade Liberalisation and Economic Performance: An Overview’, *The Economic Journal*, vol. 114, no. 493, pp. F4–F21.

World Bank 2015, *World Development Indicators*, World Bank.

Yeats, AJ 1985, ‘On the Appropriate Interpretation of the Revealed Comparative Advantage Index: Implications of a Methodology Based on Industry Sector Analysis’, *Review of World Economics*, vol. 121, no. 1, pp. 61–73.

Yeats, AJ 2001, ‘Just How Big is Global Production Sharing?’, in AS W. & H Kierzkowski (eds), *Fragmentation: New Production Patterns in the World Economy*, pp. 108–143.

Yu, R, Cai, J & Leung, P 2008, ‘The normalized revealed comparative advantage index’, *The Annals of Regional Science*, vol. 43, no. 1, pp. 267–282.

Zaghini, A 2003, *Trade Advantages and Specialisation Dynamics in Acceding Countries*, European Central Bank (Working Paper 249).

Zhu, Z, Cerina, F, Chessa, A, Caldarelli, G & Riccaboni, M 2014, 'The Rise of China in the International Trade Network: A Community Core Detection Approach', *PLoS ONE*, vol. 9, no. 8, p. e105496.

APPENDICES

LIST OF APPENDIX FIGURES

LIST OF APPENDIX TABLES

APPENDIX A DATASET COMPILATION

This appendix describes the compilation process used in constructing datasets that are used throughout this thesis. The International Trade data are derived from two datasets that originally make use of UN Comtrade data. Other datasets used include World Development Indicators (WDI), Penn World Tables (Feenstra, Inklaar & Timmer 2015), and Trade Liberalisation indicators from Wacziarg and Welch (2008).

A.1. INTERNATIONAL TRADE DATASETS

Trade datasets are all effectively derived from one starting point – the UN Comtrade dataset. This thesis makes use of two datasets: NBER WTF and CEPII BACI, both which set out to clean the Comtrade data and exploit the maximum information available through the use of mirror flows. The use of mirror trade flows increases the number of available countries (in particular developing countries) as these countries export flows can be captured on the import side of more developed countries. However, this type of data recovery is not without complications. Export and Import data differ in how they capture “Free on Board” (FOB) and “Cost Insurance and Freight” (CIF), and each dataset handles these complexities in different ways. The main benefit to using both datasets is to obtain the maximum number of years available from 1962 to 2012.

The following discussion is split across the two different trade datasets NBER and CEPII BACI. This allows the discussion to address the specific needs of each dataset independently. This is followed by a discussion about joining the two datasets together. Both datasets are available independently for robustness checks.

A.2. NBER WORLD TRADE FLOW DATASET

The NBER dataset is compiled from United Nations merchandise trade data with extensive cleaning and checking for inconsistencies and corrections made for any country specific data reporting issues.[112] In general this dataset gives primacy to the import side (under the assumption that import reports are more accurate than export reports) or to high quality export reporters when reconciling trade flows. The dataset

[112] A detailed description of how the dataset was compiled can be found in the NBER documentation (Feenstra et al. 2005). The suggested China-Hong Kong adjustments are incorporated into these datasets (Feenstra et al. 1999), in addition to removing erroneous Zimbabwe and Malawi data points prior to 1965.

documents commodity trade data at the SITC 4-digit level for the years 1962 to 2000 as detailed in Table A.1.[113]

Due to some characteristics of this dataset additional cleaning measures were conducted to improve the dynamic consistency of product codes. First, the raw dataset is discussed and then all of the adjustments considered in this thesis to transform the data to be more inter-temporally consistent is detailed.

Table A.1: Description of RAW NBER Dataset

	RAW DATASET
Number of trade flows	27,573,764
Number of exporters	201
Number of importers	203
Number of product codes	1,690
Number of years	1962 to 2000

The RAW NBER World Trade Flow Dataset has a number of features that need to be considered when used in any analysis:

Country Codes

This dataset contains “World”, and “NES” categories. These need to be removed for Country Level Analysis. Aggregating the entire dataset provides twice the value of world trade.

Product Codes

This dataset contains non-standard product codes to ensure all levels of aggregation are equal and consistent. These non-standard product codes take the form of adjustment (A) values or residuals (X) using the identifier ‘A’ and ‘X’ integrated into the codes. For example, Spain (ESP) contains product codes 721A and 721X in 1993. These adjustments allow for aggregation between product levels. The majority of these codes exist at the most disaggregated level SITC4.

113 The NBER dataset is based on revision 2 of the SITC classification. However as this dataset was released in 1974 this doesn’t cover the full year range. While the dataset was constructed to maximize the data as much as possible, there are non-official SITC codes, which is why there are 1,690 product codes, contained in the raw dataset.

Not all of the product codes contained in the dataset are found in a single official SITC classification. SITC Revision 2 is used as the primary coding system. However, as there is data the pre-dates the SITC Rev. 2 classification system, the products are "back" coded. This is not always possible and as a result there are some product codes in the dataset from SITC Revision 0/1. (i.e. 0010).

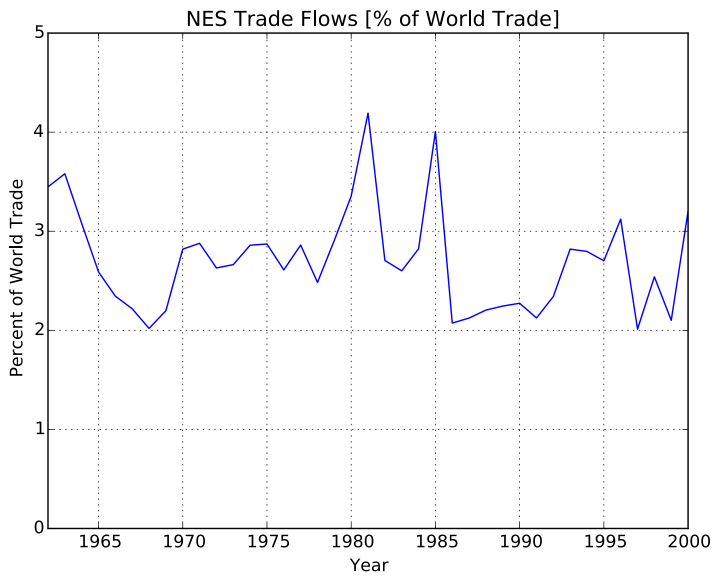
Due to changes in the product codes over time, some categories suffer from intertemporal compositional changes.[114] For example, Japan has exported Fish consistently from 1962 to 2000. However, when records began in 1962 the exports are recorded in '0340'. With SITC revision 2 this gets split into four additional codes '0341', '0342', '0343', and '0344', which correspond to export in "Live", "Frozen", "Whole Fish" and "Fish Fillet" categories. To harmonise these exports over time, a dataset is constructed that collapses these products for all countries into a single comparable code '034'.

A.2.1. TREATMENT OF COUNTRY CODES

Treatment of the additional country codes is relatively straightforward. Removing all "World" values removes trade value duplication in the dataset. However treatment of NES categories various depending on the type of data computed. For Trade data there is no way of reconciling those trade flows and are therefore dropped in trade analysis. These regions, or non-country areas, make up between 2 and 4 percent of the raw values contained in the NBER dataset. In Trade datasets these values are dropped.

[114] The NBER WTF dataset is itself an amalgam of datasets that has been harmonized. During the years 1962 to 1983 data was collected from other studies (Feenstra, Lipsey & Bowen 1997; Feenstra et al. 2005). The years 1984 to 2000 was procured from the UN Comtrade dataset and consisted of 72 reporting countries, which accounts for 98 percent of world trade.

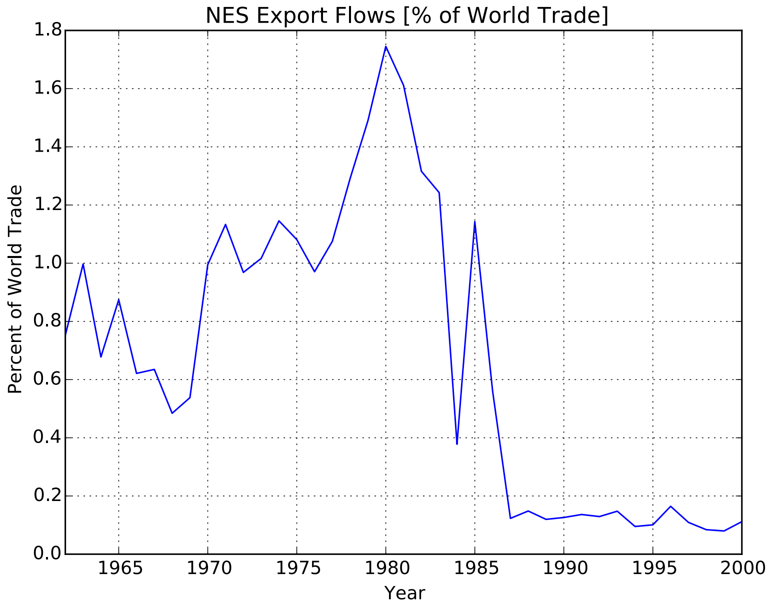
Figure A.1: Percent of World Trade Flows in NES (Non-Country) Country Codes



When considering export or import data, the NES categories can be summed across the partner dimension (i.e. exports – sum across all importers) and therefore many of the NES categories can be attributed to an exporter. By symmetry this also holds when computing import data. This thesis takes a maximal information approach and generates all datasets from the original raw data to capture as much information as possible.[115] This results in less data loss as exports to those regions are captured in the aggregation and assigned to the exporting country. NES Categories contain between 0.1% and 1.6% of world export values and these areas are dropped from the dataset.

[115] This is in contrast to generating a clean 'trade' dataset and then using this data to derive export and import datasets from this source dataset.

Figure A.2: Exporter NES (Non-Country) country codes as a percentage of world trade



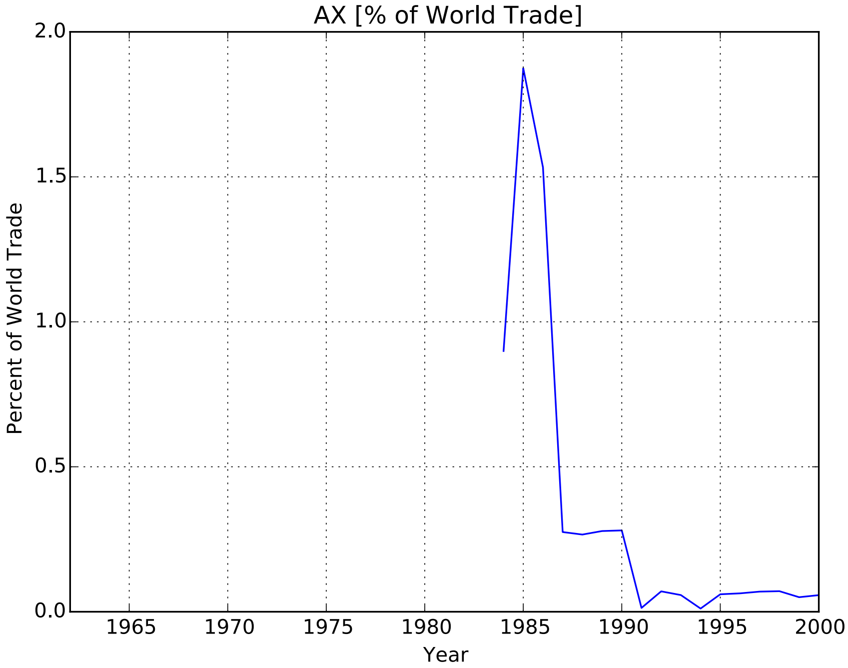
A.2.2. TREATMENT OF PRODUCT CODES

‘A’ and ‘X’ Codes

The value of data that is contained in product codes containing ‘A’ or ‘X’ codes is relatively small over time and in most cases represent supplemental information for countries that already export within that category. When considering datasets at a higher level of aggregation these codes are collapsed into the summation. At disaggregated levels of analysis these categories are dropped. These values as a percentage of world trade are shown in Figure A.1. The only year in which this data is relatively large is 1985 and this may warrant a year fixed effect in any future regression analysis.[116]

[116] It is feasible for lower levels of aggregation (1,2,3) to recover the some aspects of this data, when product codes are truncated. This is not done in the construction of datasets as these values represent residuals for the purpose of ensuring proper aggregation throughout the dataset levels. As a future robustness test this may be considered – but it is thought dropping these values will be of little consequence.

Figure A.3: 'A' and 'X' Codes as a % of World Trade



Non-Standard SITC Codes

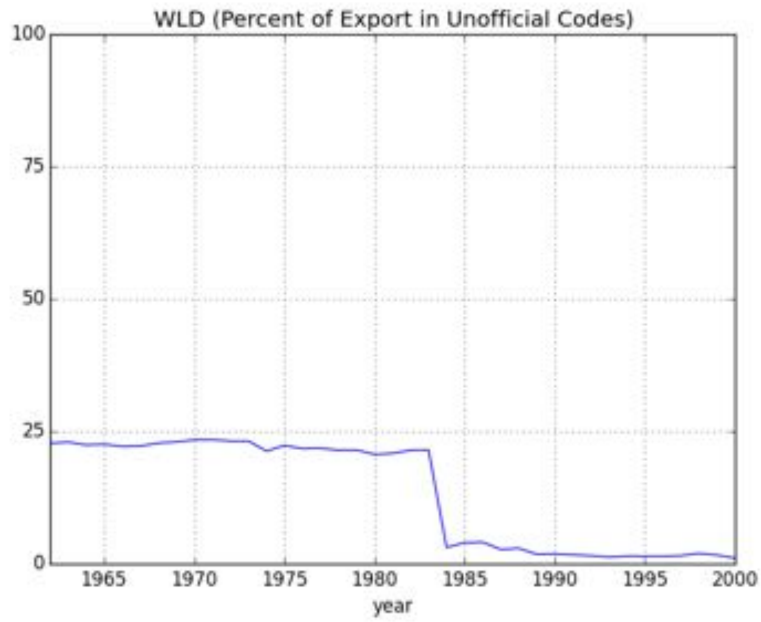
The SITC classification (Revision 2) consists of the following number of product codes. The raw NBER dataset contains 1,690 level 4 codes, which is in excess of the number of products available in the target classification, 786.

Table A.2: SITC Revision 2 Standard Classification Codes

SITC Level	Number of Product codes
1	10
2	69
3	239
4	786
5	1472

The value of data contained in the non-standard SITC revision 2 product codes is significant. These codes account for nearly 25 percent of world trade between the years 1962 and 1984. Adjustment is required to improve dynamic consistency. There is also a large degree of underlying heterogeneity at the country level. Therefore to be able to compare products more accurately, both across countries and across time, a modified product classification is considered (via aggregation) to ensure similar products are being compared. The downside to this aggregation approach is a loss of product heterogeneity (and information), which is important when considering the method of reflections used in the product space. In reality there is a trade-off between cross sectional product variation and a fully intertemporal product classification. The approach taken in this thesis is to analyse a fully consistent classification - recognising that more optimal comparisons may be fruitful by exploring this tradeoff in greater detail.

Figure A.4: Percentage of World Trade in Unofficial Export Codes



One typical approach that is often undertaken is to drop the problematic product codes. Dropping them would introduce significant bias to the data. This effectively removes products from country export baskets when in fact that country has in reality the capacity to export that product. Removing non-official codes also affects some countries more systematically than others. More export data for Australia, for example, is recorded in non-standard product codes prior to 1983 then was recorded by Fiji. The next section identifies ways to harmonise the data over time at the cost of cross-sectional product variety.

One version of the datasets contain an intertemporal product code classification which is what I consider a maximum full intertemporal type. In other words, it attempts to create a dataset that is completely intertemporally consistent while retaining the maximum amount of product variation. There is a tradeoff to be had however, where full intertemporal consistency isn't necessarily required as in the case of cross sectional analysis. From a dynamic perspective the difficulty with this tradeoff is that it isn't clear how to handle composition shifts as countries reclassify exports over time in new product categories, rather than have truly de novo product signals. The next section details the approach taken to improving data consistency throughout the full time period between 1962 and 2012.[117]

A.2.3. ADDITIONAL ADJUSTMENTS FOR DYNAMIC CONSISTENCY

When considering the consistency of trade over time there are two primary dimensions that need to be considered; Countries and Products. First the country dimension is explored and details definitions of special country codes that collapse countries into maximum geographic regions throughout 1962 to 2000. The most notable case is the Soviet Union, which is defined to consist of 15 countries. Second the product dimension is explored and a few options are presented.

Country Code Adjustments

There are a number of additional options explored when considering the intertemporal consistency of countries as a unit of analysis. During the period 1962 to 2000 there are a number of cases where countries either split into a number of countries or combine to form one country. These changes in country definitions create discontinuities in the time series data as exports are absorbed or split into other countries data lines. To improve consistency the following geographic units were recoded as the following special codes:

Table A.3: Country Splits during 1962 to 2000

Code	Country Code	ISO3C Membership
SP1	SUN	ARM, AZE, BLR, EST, GEO, KAZ, KGZ, LVA, LTU, MDA, RUS, TJK, TKM, UKR, and UZE
SP2	PAK	PAK, and BGD
SP3	YUG	BIH, HRV, MKD, MNE, SVN, and SRB
SP4	CSK	CZE, and SVK

[117] Other time periods are also considered such as 1974-2012 and 1984-2012 as each step in time reduces the number of problematic codes the require collapsing for comparability backwards in time.

Table A.4: Country Merges during 1962 to 2000

Code	**Country Code**	**ISO3C Membership**
SP5	DEU	DEU, and DDR
SP6	YEM	YEM, and YMD

Product Code Adjustments

Two primary approaches are adopted in addressing inconsistent product codes. First is to compute a product concordance that is intertemporally consistent through selective deletions and collapses of product codes. The second approach is to restrict the time period to remove problematic product codes. Let us consider Japanese fish exports in more detail as a motivating example.

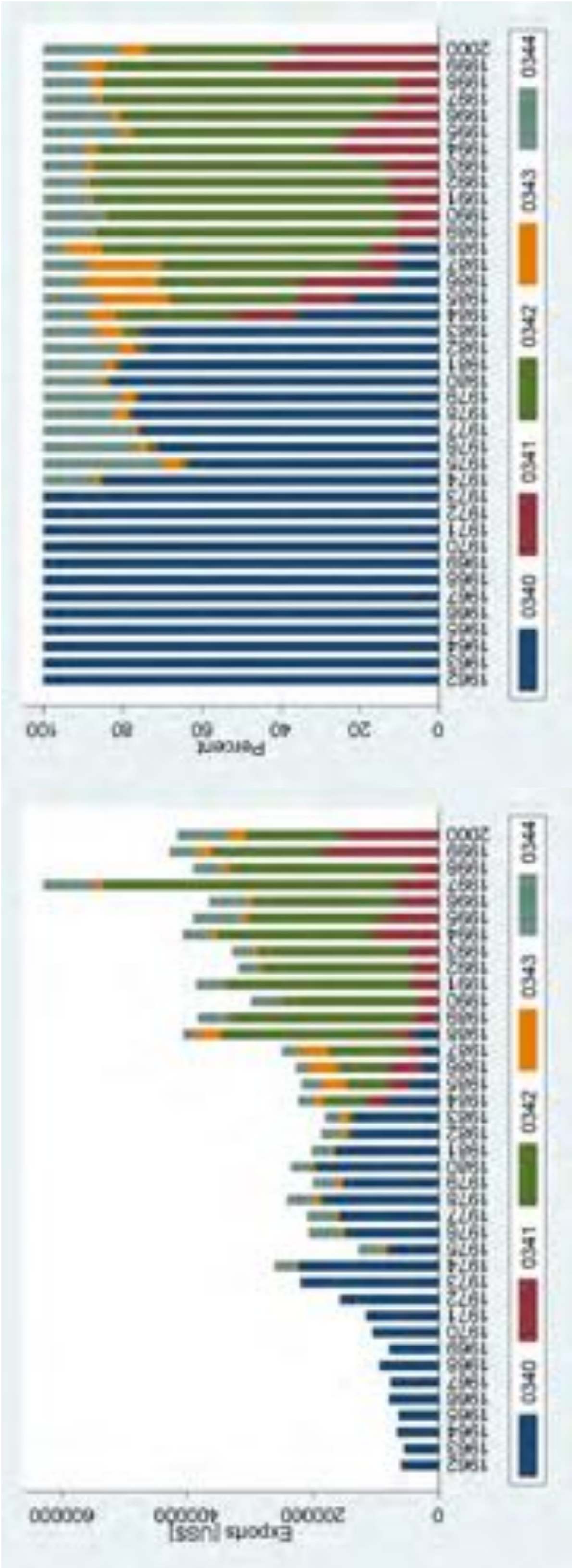
Example: Japanese Fish Exports ("034")

Table A.5: SITC "034" Fish Exports

SITC Codes		*Description*
L3	*L4*	
034	0	Fish, Fresh (live or dead), chilled or frozen
	1	Fish, Fresh (live or dead) or chilled
	2	Fish, Frozen (excl. Fillets)
	3	Fish Fillets, fresh or chilled
	4	Fish Fillets, Frozen

In 1974 the SITC international trade classification revision 2 was released. This transition is clearly observed in the NBER dataset, Japan moves between classifications such that "0340" (Fish Exports) is recorded in a disaggregated fashion (Live, Frozen, Fresh, Fillets) during the 1970s.

Figure A.5: Japanese Fish Exports (Values, Composition % of “034”)



Option #1: Restrict the time span of the dataset.

Time Span	Comments
1962 to 1973	Many trade flows use non-standard product codes in the dataset during this period as the original data is encoded SITC Revision 1.
1974 to 1983	A ***period of transition*** is observed in the data as items in countries are recoded to new SITC Revision 2 codes. This generates a lot of compositional change ***within*** SITC Level 3.
1984 to 2000	SITC Revision 2 Level 4 can be used

This is considered in robustness checks – but is not a preferable method as it greatly restricts the time span of data that is available for analysis. This is particularly important when considering periods when trade liberalisation is implemented in a number of countries.

Option #2: Compute an intertemporally consistent product classification through aggregation to the minimum level that allows for cross-country comparability and that are in addition consistent through time.

Using the Japanese Fish Export example, a product classification, which collapses the disaggregated flows to product code “034”, allows for comparison of Fish exports from 1962 to 2000 as demonstrated by Figure A.5.

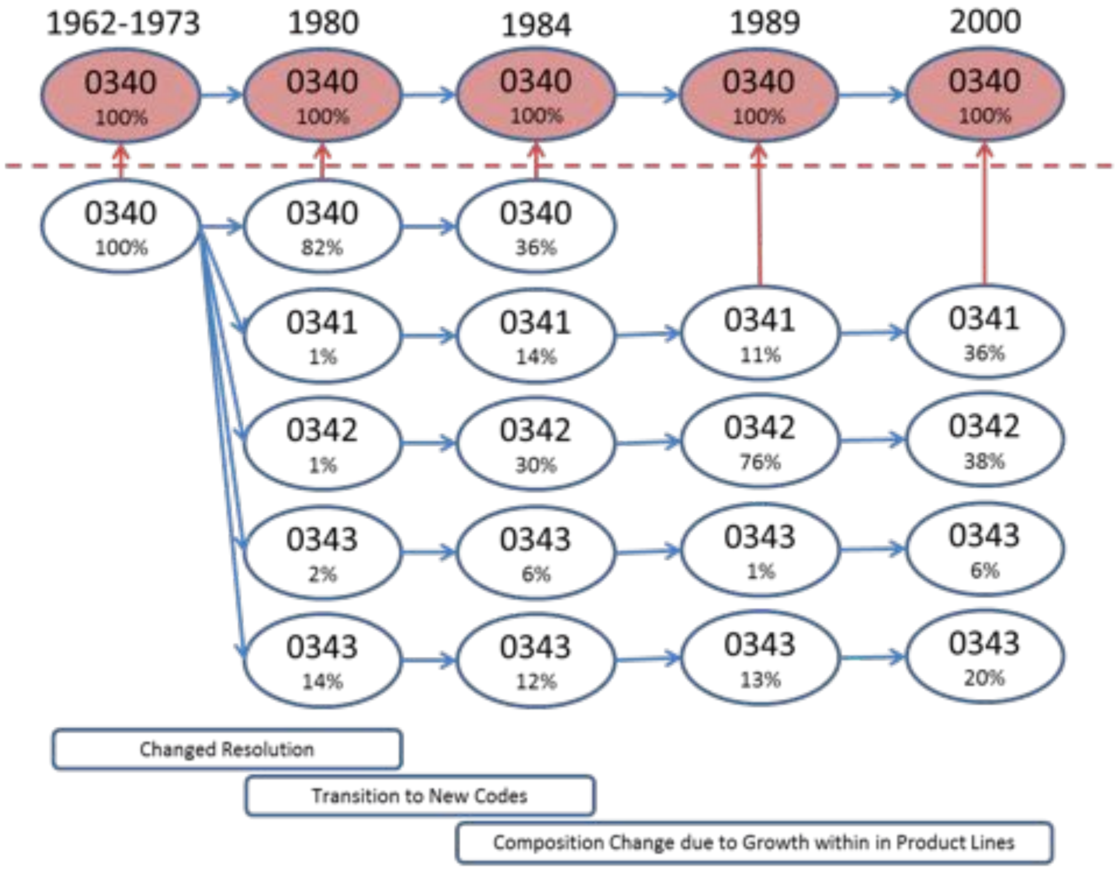


Figure A.6: Japanese Fish Exports Recoding Trade Data for Intertemporal Consistency

To assist with a classification that works across all products, while retaining as much product heterogeneity, an algorithmic approach was undertaken to minimise the number of product collapses. Six hierarchical rules were developed such that an algorithmic approach to developing the classification could be achieved.[118] This discussion is motivated using the most disaggregated form of the data: SITC4, but is equally relevant for the various higher levels of aggregation.

Rule #1: If an SITC4 product group, based on the SITC3 level, consists only of non-official SITC revision 2 codes then this group is not able to be collapsed and is marked for "Deletion".

[118] This approach also allows for context driven decisions based on the level of aggregation. When using the NBER dataset at SITC level 3 more codes are naturally rendered consistent and therefore a different set of product codes are specified to maximize the within chapter 2 level heterogeneity.

Example: SITC4 “0021” as “002” does not exist in SITC revision 2

Rule #2: If an SITC4 product group, based on the SITC3 level, contains codes that are non official SITCR2 (i.e. there are years across ALL countries when there exists no reported trade flows due to composition shift in reporting) this product group is marked for “Collapse”.

Rule #3: If an SITC4 product group, based on the SITC3 level, contains codes that are intertemporally inconsistent (i.e. there are years across ALL countries when there exists to reported trade flows) this product group is marked for “Collapse”.

Rule #4: If an SITC4 product group, based on the SITC3 level, contains codes that are not valuable based on the following criteria:

1. normalised average compositional value within the group that is less than 1 percent, and
2. a maximum compositional value of 4 percent across all years.

Then this product is deemed insignificant and is marked for “Deletion”. The remainder of the group is marked “Keep” which increases product heterogeneity by preventing unnecessary collapses of the product code data.

Example: “0110” represents 0.1% of “001” data normalised over the years 1962 to 2000, in addition to a maximum value of 1.97% in the year 1962. So that “0111” to “0118” isn’t collapsed this rule suggests deleting “0110” in preference for product heterogeneity.

Rule #5: If an SITC4 product group, based on the SITC3 level, contains official SITCR2 codes that are intertemporally inconsistent then for any “Keep” outcome as defined by *Rule #4* then these products are re-coded as “Collapse”.

Example: “0010” is initially marked for “Deletion” according to rule #4, however as “0019” is intertemporally inconsistent across all countries, these values are all marked “Collapse”.

Rule #6: Allows specifying certainty conditions for a number of special case products codes, and enforce “Deletion”, “Collapse”, and “Keep” rules to certain product codes.

Example: "9110" is marked for "Deletion" as chapter nine contains some products that have peculiar properties. In addition there are five product categories that are not available across all years (intertemporally consistent) at SITC level 3 and are therefore recoded to SITC level 2 codes (such as "84" - Apparel[119]).

As an example the net outcome of these rules for 1962 to 2000 at the fully disaggregated product level SITC4 is contained in Table A.6 below. In general this table varies based on sample time restrictions, in addition to levels of aggregation. The other definitions can be found in the online supplementary material:

https://github.com/mmcky/econ-phdthesis

It should also be noted that while this derivation pertains to the NBER dataset, the following adjustments are also made in the CEPII dataset to extend the dataset from the year 2000 to 2012 for full intertemporal consistency to 2012.

[119] In the case of '84' Apparel the breakdown of this chapter is available in other datasets 'A' through 'D' and is used when considering cross-sectional information.

Table A.6: Rules for NBER SITC Level 4 Dataset E (D=Delete, K=Keep, C=Collapse, R=Recode)

sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule
.	D	0371	C	0610	C	2114	K	2632	K	2814	C	3354	C	5154	C
0010	C	0372	C	0611	C	2116	K	2633	K	2815	C	3410	C	5155	C
0011	C	0410	C	0612	C	2117	K	2634	K	2816	C	3413	C	5156	C
0012	C	0411	C	0615	C	2119	K	2640	K	2820	K	3414	C	5157	C
0013	C	0412	C	0616	C	2120	K	2650	D	2829	D	3415	C	5160	C
0014	C	0420	C	0619	C	2200	D	2651	K	2860	K	3510	K	5161	C
0015	C	0421	C	0620	K	2220	C	2652	K	2870	C	4110	D	5162	C
0019	C	0422	C	0710	D	2221	C	2654	K	2871	C	4111	K	5163	C
0021	D	0430	K	0711	K	2222	C	2655	K	2872	C	4113	K	5169	C
0022	D	0440	K	0712	K	2223	C	2659	K	2873	C	4200	D	5220	C
0023	D	0450	C	0720	D	2224	C	2660	C	2874	C	4230	D	5221	C
0024	D	0451	C	0721	K	2225	C	2665	C	2875	C	4232	K	5222	C
0025	D	0452	C	0722	K	2226	C	2666	C	2876	C	4233	K	5223	C
0031	D	0459	C	0723	K	2230	C	2667	C	2877	C	4234	K	5224	C
0035	D	0460	K	0730	K	2231	C	2670	C	2879	C	4235	K	5225	C
0039	D	0470	K	0740	C	2232	C	2671	C	2880	C	4236	K	5230	C
0110	D	0480	C	0741	C	2234	C	2672	C	2881	C	4239	K	5231	C
0111	K	0481	C	0742	C	2235	C	2680	D	2882	C	4240	C	5232	C
0112	K	0482	C	0750	D	2238	C	2681	K	2890	K	4241	C	5233	C
0113	K	0483	C	0751	K	2239	C	2682	K	2910	C	4242	C	5239	C
0114	K	0484	C	0752	K	2320	K	2683	K	2911	C	4243	C	5240	C
0115	K	0488	C	0810	D	2321	D	2685	K	2919	C	4244	C	5241	C
0116	K	0540	C	0811	K	2330	D	2686	K	2920	D	4245	C	5249	C
0118	K	0541	C	0812	K	2331	K	2687	K	2922	K	4249	C	5310	C
0120	D	0542	C	0813	K	2332	K	2690	K	2923	K	4310	D	5311	C
0121	K	0544	C	0814	K	2400	D	2710	D	2924	K	4311	K	5312	C
0129	K	0545	C	0819	K	2440	K	2711	K	2925	K	4312	K	5320	C
0140	C	0546	C	0901	D	2450	K	2712	K	2926	K	4313	K	5322	C
0141	C	0548	C	0910	D	2460	K	2713	K	2927	K	4314	K	5323	C
0142	C	0560	C	0913	K	2470	C	2714	K	2929	K	5100	D	5330	C
0149	C	0561	C	0914	K	2471	C	2730	C	3200	D	5110	C	5331	C
0220	D	0564	C	0980	K	2472	C	2731	C	3220	C	5111	C	5332	C
0222	D	0565	C	1110	K	2479	C	2732	C	3221	C	5112	C	5334	C
0223	K	0570	C	1120	D	2480	D	2733	C	3222	C	5113	C	5335	C
0224	K	0571	C	1121	K	2481	K	2734	C	3223	C	5114	C	5410	C
0230	K	0572	C	1122	K	2482	K	2740	C	3224	C	5120	C	5411	C
0231	D	0573	C	1123	K	2483	K	2741	C	3231	K	5121	C	5413	C
0240	K	0574	C	1124	K	2510	C	2742	C	3232	K	5122	C	5414	C
0250	C	0575	C	1210	C	2511	C	2770	D	3330	K	5123	C	5415	C
0251	C	0576	C	1211	C	2512	C	2771	K	3340	C	5130	C	5416	C
0252	C	0577	C	1212	C	2516	C	2772	K	3341	C	5137	C	5417	C
0340	C	0579	C	1213	C	2517	C	2780	D	3342	C	5138	C	5419	C
0341	C	0580	C	1220	D	2518	C	2782	K	3343	C	5139	C	5500	D
0342	C	0581	D	1221	K	2519	C	2783	K	3344	C	5140	C	5510	C
0343	C	0582	C	1222	K	2610	D	2784	K	3345	C	5145	C	5513	C
0344	C	0583	C	1223	K	2613	K	2785	K	3350	C	5146	C	5514	C
0350	K	0585	C	2110	D	2614	K	2786	K	3351	C	5147	C	5530	K
0360	K	0586	C	2111	K	2630	D	2789	K	3352	C	5148	C	5540	D
0370	C	0589	C	2112	K	2631	K	2810	C	3353	C	5150	C	5541	K

sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule
5542	K	5981	C	6412	C	6561	D	6648	C	6785	C	6970	C	7212	C
5543	K	5982	C	6413	C	6570	C	6649	C	6790	C	6973	C	7213	C
5620	C	5983	C	6415	C	6571	C	6650	C	6791	C	6974	C	7219	C
5621	C	5986	C	6416	C	6572	C	6651	C	6793	C	6975	C	7220	C
5622	C	5989	C	6417	C	6573	C	6652	C	6794	C	6978	C	7223	C
5623	C	6000	D	6418	C	6574	C	6658	C	6810	R(68)	6990	C	7224	C
5629	C	6110	C	6419	C	6575	C	6659	D	6811	R(68)	6991	C	7230	C
5711	D	6112	C	6420	C	6576	C	6660	C	6812	R(68)	6992	C	7231	C
5720	C	6113	C	6421	C	6577	C	6661	D	6820	R(68)	6993	C	7232	C
5721	C	6114	C	6422	C	6579	C	6664	C	6821	R(68)	6994	C	7233	C
5722	C	6115	C	6423	C	6580	C	6665	C	6822	R(68)	6995	C	7234	C
5723	C	6116	C	6424	C	6581	C	6666	C	6830	R(68)	6996	C	7239	C
5800	D	6118	C	6428	C	6582	C	6670	C	6831	R(68)	6997	C	7240	C
5820	C	6120	C	6500	D	6583	C	6671	C	6832	R(68)	6998	C	7243	C
5821	C	6121	C	6510	C	6584	C	6672	C	6840	R(68)	6999	C	7244	C
5822	C	6122	C	6511	C	6585	C	6673	C	6841	R(68)	7000	D	7245	C
5823	C	6123	C	6512	C	6589	C	6674	C	6842	R(68)	7100	R(71)	7246	C
5824	C	6129	C	6513	C	6590	C	6710	D	6850	R(68)	7110	R(71)	7247	C
5825	C	6130	K	6514	C	6591	C	6712	K	6851	R(68)	7111	R(71)	7248	C
5826	C	6210	K	6515	C	6592	C	6713	K	6852	R(68)	7112	R(71)	7250	C
5827	C	6214	D	6516	C	6593	C	6716	K	6860	R(68)	7119	R(71)	7251	C
5828	C	6250	C	6517	C	6594	C	6720	C	6861	R(68)	7120	R(71)	7252	C
5829	C	6251	C	6518	C	6595	C	6724	C	6863	R(68)	7121	R(71)	7259	C
5830	C	6252	C	6519	C	6596	C	6725	C	6870	R(68)	7126	R(71)	7260	C
5831	C	6253	C	6520	C	6597	C	6726	D	6871	R(68)	7129	R(71)	7263	C
5832	C	6254	C	6521	C	6610	D	6727	C	6872	R(68)	7130	R(71)	7264	C
5833	C	6255	C	6522	C	6611	K	6730	C	6880	R(68)	7131	R(71)	7265	C
5834	C	6259	C	6529	D	6612	K	6731	C	6890	R(68)	7132	R(71)	7266	C
5835	C	6280	C	6530	C	6613	K	6732	C	6891	R(68)	7133	R(71)	7267	C
5836	C	6281	C	6531	C	6618	K	6733	C	6899	R(68)	7138	R(71)	7268	C
5837	C	6282	C	6532	C	6620	D	6735	D	6900	D	7139	R(71)	7269	C
5838	C	6289	C	6534	C	6623	K	6740	C	6910	C	7140	R(71)	7271	C
5839	C	6291	D	6535	C	6624	K	6741	C	6911	C	7144	R(71)	7272	C
5840	C	6292	D	6536	C	6630	C	6742	C	6912	C	7147	R(71)	7280	C
5841	C	6330	K	6538	C	6631	C	6744	C	6920	C	7148	R(71)	7281	C
5842	C	6340	C	6539	C	6632	C	6745	C	6921	C	7149	R(71)	7283	C
5843	C	6341	C	6540	C	6633	C	6746	C	6924	C	7160	R(71)	7284	C
5849	C	6342	C	6541	C	6635	C	6747	C	6930	D	7161	R(71)	7285	C
5850	C	6343	C	6542	C	6637	C	6749	C	6931	K	7162	R(71)	7300	D
5851	C	6344	C	6543	C	6638	C	6750	D	6932	K	7163	R(71)	7311	D
5852	C	6349	C	6544	C	6639	C	6760	C	6935	K	7164	R(71)	7313	D
5910	C	6350	D	6545	C	6640	C	6762	D	6940	K	7165	R(71)	7314	D
5911	C	6351	K	6546	C	6641	C	6768	C	6950	C	7169	R(71)	7315	D
5912	C	6352	K	6549	C	6642	C	6770	K	6951	C	7187	R(71)	7316	D
5913	C	6353	K	6550	C	6643	C	6780	C	6953	C	7188	R(71)	7317	D
5914	C	6354	K	6551	C	6644	C	6781	C	6954	C	7189	R(71)	7331	D
5921	C	6359	K	6552	C	6645	C	6782	C	6956	C	7200	D	7351	D
5922	C	6410	C	6553	C	6646	C	6783	C	6960	K	7210	C	7359	D
5980	C	6411	C	6560	K	6647	C	6784	C	6963	D	7211	C	7360	C

sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule	sitc4	rule
7361	C	7510	R(75)	7753	R(77)	7922	C	8463	R(84)	8924	R(89)
7362	C	7511	R(75)	7754	R(77)	7923	C	8464	R(84)	8928	R(89)
7367	C	7512	R(75)	7757	R(77)	7924	C	8465	R(84)	8930	R(89)
7368	C	7518	R(75)	7758	R(77)	7928	C	8470	R(84)	8931	R(89)
7369	C	7520	R(75)	7760	R(77)	7929	C	8471	R(84)	8932	R(89)
7370	C	7521	R(75)	7761	R(77)	7930	C	8472	R(84)	8933	R(89)
7371	C	7522	R(75)	7762	R(77)	7931	C	8480	R(84)	8935	R(89)
7372	C	7523	R(75)	7763	R(77)	7932	C	8481	R(84)	8939	R(89)
7373	C	7524	R(75)	7764	R(77)	7933	C	8482	R(84)	8940	R(89)
7400	R(74)	7525	R(75)	7768	R(77)	7935	D	8483	R(84)	8941	R(89)
7411	R(74)	7526	R(75)	7780	R(77)	7938	C	8484	R(84)	8942	R(89)
7412	R(74)	7528	R(75)	7781	R(77)	8000	D	8510	K	8946	R(89)
7413	R(74)	7590	R(75)	7782	R(77)	8120	C	8513	D	8947	R(89)
7414	R(74)	7591	R(75)	7783	R(77)	8121	C	8517	D	8950	R(89)
7415	R(74)	7599	R(75)	7784	R(77)	8122	C	8710	K	8951	R(89)
7416	R(74)	7610	C	7788	R(77)	8124	C	8720	K	8952	R(89)
7420	R(74)	7611	C	7800	D	8210	C	8721	D	8959	R(89)
7421	R(74)	7612	C	7810	K	8211	C	8722	D	8960	R(89)
7422	R(74)	7620	C	7820	D	8212	C	8730	C	8970	R(89)
7423	R(74)	7621	C	7821	K	8215	C	8731	C	8972	R(89)
7425	R(74)	7622	C	7822	K	8219	C	8732	C	8973	R(89)
7427	R(74)	7628	C	7830	D	8310	K	8740	C	8974	R(89)
7428	R(74)	7630	C	7831	K	8400	D	8741	C	8980	R(89)
7429	R(74)	7631	C	7832	K	8420	R(84)	8742	C	8981	R(89)
7430	R(74)	7638	C	7840	C	8421	R(84)	8743	C	8982	R(89)
7431	R(74)	7640	C	7841	C	8422	R(84)	8744	C	8983	R(89)
7432	R(74)	7641	C	7842	C	8423	R(84)	8745	C	8989	R(89)
7433	R(74)	7642	C	7843	C	8424	R(84)	8746	C	8990	R(89)
7434	R(74)	7643	C	7849	C	8429	R(84)	8747	C	8991	R(89)
7435	R(74)	7648	C	7850	C	8430	R(84)	8748	C	8993	R(89)
7436	R(74)	7649	C	7851	C	8431	R(84)	8749	C	8994	R(89)
7437	R(74)	7710	R(77)	7852	C	8432	R(84)	8810	D	8996	R(89)
7439	R(74)	7711	R(77)	7853	C	8433	R(84)	8811	K	8997	R(89)
7440	R(74)	7712	R(77)	7860	C	8434	R(84)	8812	K	8998	R(89)
7441	R(74)	7720	R(77)	7861	C	8435	R(84)	8813	K	8999	R(89)
7442	R(74)	7721	R(77)	7862	C	8438	R(84)	8820	D	9000	D
7447	R(74)	7722	R(77)	7868	C	8439	R(84)	8821	K	9110	D
7449	R(74)	7723	R(77)	7900	D	8440	R(84)	8822	K	9310	D
7451	R(74)	7724	R(77)	7910	C	8441	R(84)	8830	K	9410	D
7452	R(74)	7725	R(77)	7911	C	8442	R(84)	8841	K	9510	K
7473	R(74)	7730	R(77)	7912	C	8443	R(84)	8842	K	9610	D
7478	R(74)	7731	R(77)	7913	C	8450	R(84)	8850	D	9710	D
7482	R(74)	7732	R(77)	7914	C	8451	R(84)	8851	K		
7490	R(74)	7740	R(77)	7915	C	8452	R(84)	8852	K		
7491	R(74)	7741	R(77)	7916	D	8458	R(84)	8900	R(89)		
7492	R(74)	7742	R(77)	7918	C	8459	R(84)	8913	R(89)		
7493	R(74)	7750	R(77)	7919	C	8460	R(84)	8920	R(89)		
7499	R(74)	7751	R(77)	7920	C	8461	R(84)	8921	R(89)		
7500	R(75)	7752	R(77)	7921	C	8462	R(84)	8922	R(89)		

A.2.4. CONSTRUCTED DATASETS 1962 TO 2000

Bringing the above discussion together, the following datasets were constructed, at all levels of aggregation (SITC Level 1, 2, 3, and 4), and all types of data (Import, Export and Trade) in support of this research. Much of this thesis will focus on “export” type data, however “Import” data can be useful when considering other measures of comparative advantage such as the Lafay measure (Lafay 1992). This will also assist in further work using fully disaggregated “trade” flows. Below is a description of the various datasets that were computed.

A	Dataset which consists of RAW data from NBER World Trade Flows with a minimum of adjustments: Keeps country data only, removing “World” and relevant “NES” categories found in the dataset
B	Same as [A] with adjustments made to the data with the supplied Hong Kong and China Data Adjustment Tables
C	Same as [B] but removes products that contain ‘A’ or ‘X’ in the product code
D	Same as [C] but removes products that are not official SITCR2 codes
E	Same as [C] but recodes product codes to an intertemporally consistent classification across countries and products. This looks for the most disaggregated product classification while adjusting for composition changes within some product categories.
F	Same as [E] but additionally recodes countries to be dynamically consistent geographic units. As data cannot be split, maximum geographic membership approach is taken to smooth these changes in border changes across time. For example, countries that have belonged to the Soviet Union are aggregated across all years.

The primary focus will be on ‘E’ when considering dynamic export flows, and ‘C’ and ‘D’ when considering cross-sectional flows. Having a well defined intertemporally consistent product classification for dynamic analysis is import to avoid discontinuities across time. Without harmonisation, there are clear observable differences in the years 1974 and 1984 where we observe migration between classification systems for some product codes. Suppose one uses Dataset ‘C’, which keeps the NBER data in its basic form dropping only minor information through the removal of ‘A’ and ‘X’ codes.

In the context of the economic complexity literature, allowing the composition exports to change over these boundary years leads to two primary issues in dynamic analysis

1. Product ubiquity information changes over time. As some products are phased out, their ubiquity information will be incorrect and will be recorded as an increasingly rare product. This incorrect information is not because of product characteristics but as a feature of a dying product code.
2. Country diversity information will tend to increase within boundary years due to the tendency for a dying product code to be recorded as multiple product codes in subsequent years. A force countering this tendency is that all countries will tend to change at the same time, making relative changes less problematic.

Datasets taking into account country aggregations was computed 'F' and 'G' but are mainly used for robustness checks.[120]

It is evident that the number of products is reduced from ~786 official SITC codes to 283 intertemporally consistent product codes. In an effort to increase the lost product heterogeneity, the following section details a restricted dataset from 1974 to 2000 in addition to 1984 to 2000.

SITC Bilateral Trade Datasets (SITC Level 4, 3 and 2)

SITC L4	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Importers	179	179	179	179	179	155	155
Products	1687	1689	1070	786	283	283	786
Years	39	39	39	39	39	39	39
Trade Flows	21,675,345	21,726,374	21,646,226	19,515,216	12,204,084	11,835,995	18,976,240
% of WTF	~96-98%	~96-98%	~95-98%	~75-96%	~93-96%	~93-96%	~75-96%

SITC L3	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Importers	179	179	179	179	179	155	155
Products	419	420	273	239	180	180	239
Years	39	39	39	39	39	39	39
Trade Flows	13,414,271	13,435,254	13,409,290	13,117,346	10,492,293	10,151,728	12,693,665
% of WTF	~96-98%	~96-98%	~95-98%	~94-97%	~94-97%	~94-97%	~94-97%

[120] Datasets 'A' and 'B' are largely diagnostic datasets used to check for appropriate transformations and adjustments made to the data due to China-HK adjustments and comparing 'C' through 'H' with RAW data.

SITC L2	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Exporters	179	179	179	179	179	155	155
Importers	179	179	179	179	179	155	155
Products	93	93	73	69	64	64	69
Years	39	39	39	39	39	39	39
Trade Flows	6,472,016	6,434,904	6,429,724	6,257,092	6,141,735	5,926,663	6,016,245
% of WTF	~96-98%	~96-98%	~96-98%	~95-97%	~95-97%	~94-97%	~95-97%

SITC Export Datasets (SITC Level 4, 3, and 2)

SITC L4	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Products	1687	1689	1070	786	283	283	786
Years	39	39	39	39	39	39	39
Trade Flows	1,832,812	1,833,471	1,766,602	1,585,519	845,013	812,866	1,523,103
% of WTF	~98-99%	~98-99%	~98-99%	~75-98%	~95-98%	~95-98%	~75-98%

SITC L3	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Exporters	179	179	179	179	179	155	155
Products	419	420	273	239	180	180	239
Years	39	39	39	39	39	39	39
Trade Flows	847,310	847,486	826,383	804,931	631,630	605,542	771,181
% of WTF	~98-99%	~98-99%	~98-99%	**~96-98%**	~95-98%	~95-98%	~96-98%

SITC L2	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Products	93	93	73	69	64	64	69
Years	39	39	39	39	39	39	39
Trade Flows	318,137	318,167	314,133	304,316	289,600	276,583	290,943
% of WTF	~98-99%	~98-99%	~98-99%	~98-99%	~95-98%	~95-98%	~98-99%

Data tables for SITC levels 1, 2 and 3 in addition to NBER imports are available in the online appendix. They are not reported here as they are not used in much of the thesis except in Appendix B when constructing the Lafay RCA index.

A.2.5. CONSTRUCTED DATASETS 1974 TO 2000

Additional datasets were considered with the time restriction of removing data prior to 1974. This removes a large number of SITC revision 0/1 codes, but doesn't solve all of the compositional shifts that are observed from 1974 to 1984, and some product code aggregation is still required. However the same algorithmic approach was used on this subset of data in an effort to recover product heterogeneity and increase the number of product codes. This increases product variety to 499 products.

Bilateral Trade Datasets (SITC Level 4)

	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Importers	179	179	179	179	179	155	155
Products	1678	1680	1061	786	499	499	786
Years	27	27	27	27	27	27	27
Trade Flows	15,684,510	15,735,539	15,655,391	14,519,961	12,463,875	12,051,081	14,053,577
% of WTF	~96-98%	~96-98%	**~95-98%**	**~77-96%**	**~93-96%**	**~93-96%**	~77-96%

Export Datasets (SITC Level 4)

	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Products	1678	1680	1061	786	499	499	786
Years	27	27	27	27	27	27	27
Trade Flows	1,338,362	1,339,021	1,272,152	1,180,358	905,242	862,005	1,123,825
% of WTF	~98-100%	~98-100%	**~95-98%**	**~77-98%**	**~95-98%**	**~95-98%**	~77-98%

Data tables are also available for SITC Levels 1 through 3, in addition to import datasets through the online appendix. They are not reported here as they are not used in much of the thesis except in Appendix C when constructing the Lafay RCA index.

A.2.6. CONSTRUCTED DATASETS 1984 TO 2000

It is feasible to construct a dataset from 1984 to 2000 using only SITC revision 2 official codes. This provides a consistent classification over time and captures ~98% of the original NBER dataset (excluding codes containing 'A' and 'X').

Given this time restriction there is not much to be gained by adopting the more uniform algorithmic product classification (based on selective aggregation) in datasets E, F, and H. In this context dataset D becomes more attractive as it provides greater product heterogeneity with similar value coverage, which is uniform across years.

SITC Level 4 Bilateral Trade Datasets

	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Importers	179	179	179	179	179	155	155
Products	1617	1619	1000	786	678	678	786
Years	17	17	17	17	17	17	17
Trade Flows	8,988,742	9,039,771	8,959,623	8,920,155	8,141,154	7,801,093	8,556,152
% of WTF	~96-99%	~96-99%	~95-98%	**~93-97%**	~93-96%	~93-96%	~93-97%

SITC Level 4 Export Datasets

	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	156	156
Products	1617	1619	1000	786	678	678	786
Years	17	17	17	17	17	17	17
Trade Flows	816,518	817,177	750,308	740,875	661,775	619,973	?
% of WTF	~99%	~99%	~98-100%	**~96%-99%**	~96-98%	~96-98%	~96-99%

Data tables are available for SITC Levels 1 through 3, in addition to import datasets through the online appendix. They are not reported here, as they are not used in much of the thesis except in Appendix B when constructing the Lafay RCA index.

A.3. CEPII BACI WORLD TRADE FLOW DATASET

The CEPII group have produced a similarly improved dataset using UN Comtrade HS92, HS96, and HS02 trade data. BACI however uses a statistical technique to adjust for FOB and CIF differences between export and import trade flows, in addition to a gravity style estimation to adjust for transport costs. This statistical methodology produces a dataset representing a high number of countries at a large degree of product disaggregation through the harmonisation process of using mirror flows. The full details of this dataset are outlined in Table A.8 and a summary of the dataset in Table A.9.

Table A.7: BACI (HS) World Trade Dataset

TRADE	**HS96**
Number of exporters	248
Number of importers	250
Number of products	5111
Number of trade flows	97,495,894
Years	1998 to 2012 (15 Years)

The BACI dataset was concorded to SITC Revision 2 Level 4 codes using the UN statistical concordance table between HS96 and SITCR2.

Table A.8: BACI Concorded to SITC Revision 2

TRADE	**SITCR2L5**	**SITCR2L4**
# of Exporters	222	222
# of Importers	222	222
# of Products	1751	774
# of Trade Flows	51,084,449	32,994,513
% of Original Dataset	~99%	~99%
Years	1998 to 2012 (15 Years)	
Notes: 1. The dataset has been filtered for countries only excluding NES categories		

A.4. COMBINED (NBER/CEPII) TRADE DATASETS

When harmonising the CEPII dataset with the NBER datasets, the countries that are contained in the BACI dataset and are not found in the NBER dataset are dropped[121], in addition to collapsing the appropriate SITC codes through aggregation.[122] Due to the UN concordance between HS96 and SITC R2, "Petroleum and Oil" is coded at the 3rd digit level as "334".

A.4.1. BILATERAL TRADE DATASETS

The following tables provide summary statistics for SITC Revision 2 Level 4 Bilateral Trade datasets.

[121] These countries consist largely of small island countries and do not consist of large trade values.
[122] Matching the product code aggregation procedure is context specific when matching to the 1962 to 2000, 1974 to 2000, and 1984 to 2000 datasets.

Combined Bilateral Trade Datasets (1962 to 2012)

SITCL4	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Importers	179	179	179	179	179	155	155
Products	1681	1683	1066	783	283	283	783
Years	1962-2012	1962-2012	1962-2012	1962-2012	1962-2012	1962-2012	1962-2012
Trade Flows	50,434,530	50,473,159	50,393,385	48,276,758	24,933,933	22,724,377	43,822,111
Notes: 1. 783 Products in dataset 'D' is due to collapsing 3341, 3342, 3343, 3344 to 334 when merging with the NBER dataset							

Combined Bilateral Trade Datasets (1974 to 2012)

SITCL4	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Importers	179	179	179	179	179	155	155
Products	1672	1674	1057	783	499	499	783
Years	1974-2012						
Trade Flows	44,468,720	44,407,349	43,298,814	43,298,814	33,208,538	29,903,403	38,916,678
Notes: 1. 783 Products in dataset 'D' is due to collapsing 3341, 3342, 3343, 3344 to 334 when merging with the NBER dataset							

Combined Bilateral Trade Datasets (1984 to 2012)

SITCL4	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Importers	179	179	179	179	179	155	155
Products	1611	1613	996	783	675	675	783
Years	1984-2012						
Trade Flows	37,797,172	37,835,801	37,756,027	37,716,820	34,167,056	30,257,825	33,436,873
Notes: 1. 783 Products in dataset 'D' is due to collapsing 3341, 3342, 3343, 3344 to 334 when merging with the NBER dataset							

A.4.2. EXPORT DATASETS

The following tables provide summary statistics for the combined SITC Revision 2 Level 4 Export datasets. These datasets are the primary datasets used in Chapter 3 when computing export emergence.

Combined Export Datasets (1962 to 2012)

SITCL4	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Products	1681	1683	1066	783	283	283	783
Years	51	51	51	51	51	51	51
Trade Flows	3,019,233	3,019,839	2,953,496	2,774,537	1,312,310	1,220,617	2,555,962
Notes: 1. 783 Products in dataset 'D' is due to collapsing 3341, 3342, 3343, 3344 to 334 when merging with the NBER dataset							

Combined Export Datasets (1974 to 2012)

SITC L4	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	156	156
Products	1672	1674	1057	783	499	499	783
Years	39	39	39	39	39	39	39
Trade Flows	2,528,652	2,529,258	2,463,015	2,372,234	1,708,378	1,561,687	2,159,502
Notes: 1. 783 Products in dataset 'D' is due to collapsing 3341, 3342, 3343, 3344 to 334 when merging with the NBER dataset							

Combined Export Datasets (1984 to 2012)

SITC L4	A	B	C	D	E	F	G
Exporters	179	179	179	179	179	155	155
Products	1611	1613	996	783	675	675	783
Years	29	29	29	29	29	29	29
Trade Flows	2,010,462	2,011,068	1,944,825	1,935,565	1,714,077	1,533,443	1,731,484
Notes: 1. 783 Products in dataset 'D' is due to collapsing 3341, 3342, 3343, 3344 to 334 when merging with the NBER dataset							

A.5. ATLAS OF COMPLEXITY (REFERENCE DATASET)

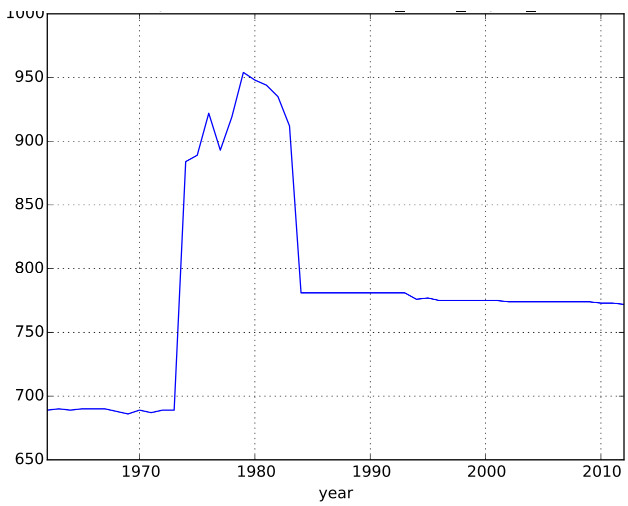
The dataset used in construction of the Atlas of Complexity has been obtained for comparison (Hausmann & Hidalgo 2014). Table A.9 provides basic statistics for the Atlas Export Datasets.

Table A.9: Atlas of Complexity SITC Raw Data

	SITCR2 L4	SITCR2 L3	SITCR2 L2	SITCR2 L1
Exporters	248	248	248	248
Products	986	260	72	10
Years	51	51	51	51
Trade Flows	4,958,468	1,849,841	594,833	94,680

The number of non-zero products over time found in the atlas dataset is shown in Figure A.7. As is evident from Table A.9, the number of products at the 4-digit level exceeds that of the official SITC classification. Therefore there exist a number of compositional changes in classification in the Atlas dataset. For conducing cross-sectional analysis this is of little consequence, but comparing the data over time would require some additional post processing.

Figure A.7: Number of non-zero products over time in atlas dataset



A.6. OTHER DATASETS (WDI, PENN, UNCTAD TRADE INTENSITIES)

Other datasets used in this thesis include

1. World Development Indicators (WDI) (World Bank 2015)
2. PENN World Tables (Feenstra, Inklaar & Timmer 2015)
3. UNCTAD Relative Factor Trade Intensities (Shirotori, Tumurchudur & Cadot 2010)
4. Trade Liberalization Indicators (Wacziarg & Welch 2008)

APPENDIX B REVEALED COMPARATIVE ADVANTAGE

> *"One wonders, therefore, whether more could not be gained if, instead of enunciating general principles and trying to apply these to explain actual trade flows, one took the observed pattern of trade as the point of departure, and subsequently attempted to find the main influences that have determined the pattern"* (Balassa 1965)

The theory of comparative advantage is the fundamental force driving cross-country patterns of trade. The fundamental lesson learnt from comparative advantage is to recognise that it is the differences in *relative opportunity costs* that are a driving force for determining the pattern of international trade.

However tests of comparative advantage have been notoriously difficult to undertake. One problem concerns the inability to observe the country autarkic state for static comparisons between autarky and free trade. Another issue concerns the lack of comparable cross-country data on productivity (Ricardian), and endowments such as capital, land, and labour (Hecksher-Ohlin). This has led to a considerable literature on considering "revealed" metrics of comparative advantage based on ex-post international trade flows.

Comparative advantage is measured indirectly in a variety of ways. The most popular method is the Balassa (1965) RCA measure, which compares product export shares within a country to the share of that product in world trade. If the product forms a larger share in the country export vector relative to that products share in total world trade it is said to have comparative advantage (RCA >= 1).

The influential Balassa (1965) RCA index, which is formally defined in the next section, is the primary starting point for this discussion. The literature is then surveyed to understand which metric should be used to support the research conducted in this thesis.

B.1. BALASSA RCA (BRCA)

Balassa (1965) is a metric that considers the product share in any given country's export basket relative to that products share in world trade. A measure that is greater than one suggests that country C exports product P with revealed comparative advantage as its product share is greater than the "average" or "expected" share in world trade. The measure is bounded by zero and unbounded above the cut-off point one which renders it asymmetric and not normally distributed.

$$BRCA_{cp} = \frac{\frac{X_{cp}}{X_c}}{\frac{X_p}{X_w}}$$

Benedictis & Tamberi (2001) provide a critique of the BRCA measure concluding that while other metrics of revealed comparative advantage may offer some degree of improved mathematical characteristics, in general they also introduce a new set of limitations. While the BRCA measure is not perfect, it can still be used with great effect through careful analysis, keeping in mind the limitations of the metric.

The BRCA metric is used widely in the literature including the cited literature that this thesis is largely draws on with regards to economic complexity and the product space. The primary disadvantage of this measure is that it is not firmly embedded in a theoretical framework. A probability framework is considered by Kunimoto (1977), which relates trade intensity indices to deviations of a countries export relative to a hypothetical world. When deviations are greater than one it is determined that a country is exporting a product with relative intensity. This is a concept strongly linked with BRCA. Another interpretation of BRCA is the case when exports exceed the "expected" level of exports relative to the hypothetical country. Bowen (1983) also makes a probabilistic argument computing RCA indices as a deviation from the expected level of trade, production, and consumption. This formulation however depends on having comparable production and consumption data. While consumption data can be derived from GDP statistics under the assumption of homothetic preferences, it is difficult to concord production data.

A link between BRCA and comparative advantage is also explored in the context of comparing changes in comparative advantage relative to changes in pre-trade relative prices (Bowen 1983).[123] By considering a static change in factor endowments, a

[123] It should be noted that the Hillman (1980) conditions are derived in a simple model where demand plays no role in shaping international trade. Demand is assumed to be identical and homothetic, and the economies are assumed to have identical

condition is derived to ensure that any relative price decline (due to the change in factor) is correspondent with an increase in comparative advantage.

$$1-\frac{X_i^c}{W_i}>\frac{X_i^c}{T^c}*(1-\frac{T^c}{Z})$$

where;

X_i^c= Export of Product 'i' in country 'c'

T^c = Total Export of country 'c'

Z = Total World Export

The Hillman condition could be considered as a filter for the BRCA measure to improve its correspondence with the theory of comparative advantage, particularly when considering changes in pre-trade relative prices due to differences in factor endowments. In practice this condition is violated infrequently (0 to 0.2 percent of all cases) as a share of the number of observations (Marchese & Simone 1989; Hinloopen & van Marrewijk 2008). However these violations correspond to relatively large individual export values (0 to 3.4 percent of world trade). Violation of the condition tends to occur when a country is "overly" specialised and therefore this mostly affects estimates of RCA for some developing countries, which are mainly concentrated in Africa, Middle East, Latin America, Caribbean and Eastern Europe.

Additional concerns with the BRCA index are that the values do not contain useful ordinal or cardinal information (Yeats 1985). BRCA values are not normally distributed (imposing difficulties for use in estimation), and it is not clear what the intertemporal properties are when comparing one time period with the next due to a time varying mean (Vollrath 1991; Proudman & Redding 2000; De Benedictis & Tamberi 2004). The latter is of most concern in parts of this thesis that consider product emergence and the dynamics of comparative advantage. It would be best to have a metric that is more intertemporally comparable, as is the case with many of the transformations and measures discussed below.

---

technologies and capital-labour endowment ratios. The variance in the static analysis comes from a difference in absolute endowments.

It is noted that Balassa (1965) also proposed an alternative specification that considered net trade flows (X – M). Balassa mainly focused on the relative export definition of revealed comparative advantage due to relatively higher distortions observed on import side data.

B.2. SYMMETRIC RCA (SRCA)

Symmetric RCA is a transformation of the BRCA measure (Laursen 2000), which compresses the information into an interval between -1 and +1. This moves the cut-off value to zero where country c with a positive value of SRCA export that product p with comparative advantage.

$$SRCA_{cp} = \frac{BRCA_{cp} - 1}{BRCA_{cp} + 1}$$

This transformation is essentially an approximation of the log transformation (Vollrath 1991) without suffering the mathematical limitations that the log function introduces with regards to zero values. The aim of this transformation is to improve the distributional characteristics of BRCA so that it more closely approximates a normal distribution, which can then be incorporated into regressions.

Given this measure is essentially a compression of information it is not considered further in this thesis as it incorporates the same information as that found in BRCA. Each measure will produce equivalent M_{cp} matrices, which defines the relationships between which products countries export with comparative advantage.

B.3. WEIGHTED RCA (WRCA)

The BRCA measure has two characteristics that impact on its quality when considering changes in RCA over time. The mean of the BRCA measure across sectors, and across time are variable, which makes cross-sectoral comparisons over time difficult, as they are not normalised values.

Proudman and Redding (2000) proposed an adjusted BRCA index through a cross sectional normalisation of the index to improve the comparability of the BRCA measure between sectors and over time.

$$WRCA_{cp} = \frac{BRCA_{cp}}{\frac{1}{n}\sum_{p=1}^{P} BRCA_{cp}}$$

B.4. LAFAY RCA (LRCA)

One suggested indicator (Lafay 1992) is notable in that it doesn't incorporate world trade conditions, but does incorporate a domestic macroeconomic factor in terms of national demand (Yi, Gross National Product). This measure is primarily useful for computing revealed comparative advantage in national level studies. The primary strength of this formulation is that it inherently accounts for intra-industry trade due to considering trade balance (X – M).

Due to the more favourable mathematical properties (invariant mean) analysis can be done effectively over time and is thus appropriately normalised for comparison.

$$LRCA_{ij} = \frac{1000}{Y_i} * \frac{2\left(X_{ij}M_i - X_iM_{ij}\right)}{X_i + M_i}$$

A related but modified version of this index was proposed by Bugamelli (2001), and used in various studies (Zaghini 2003; Alessandrini et al. 2011), which removes the proposed sector weights relative to national GDP.

$$LFI_{ij} = 100 * \left(\frac{x_{ij} - m_{ij}}{x_{ij} + m_{ij}} - \frac{\sum_{j=1}^{N}\left(x_{ij} - m_{ij}\right)}{\sum_{j=1}^{N}\left(x_{ij} + m_{ij}\right)}\right) * \left(\frac{x_{ij} + m_{ij}}{\sum_{p=1}^{N}\left(x_{ij} + m_{ij}\right)}\right)$$

This version of the index measures the revealed comparative advantage of country i in the export of product j as the deviation of the net trade flow of product j which has been normalised by the overall trade balance of country c which is then scaled by a products total trade share in total country trade. This measure is computed solely by each country's overall trade structure and doesn't incorporate any relative measures to the rest of the world. Therefore this measure is more suited for individual country analysis case studies, which accounts well for the increase of intra-industry trade from rising product fragmentation.

The disadvantage of this indicator is that it is not as commonly used in the literature, in preference for the intuitive Balassa (1965) measure. It makes use of import data, which is often criticized as having a larger degree of measurement error when compared with export data.[124] And finally while intra-industry trade is an important and growing phenomenon, this issue is partly alleviated by using highly disaggregated data with greater product heterogeneity.

Given this thesis uses disaggregated trade data covering the majority of world trade flows, it is suggested that the BRCA measure is more appropriate as it incorporates global world trade conditions, whereas the Lafay index is particularly well designed for country specific studies. It could also becomes increasingly useful at higher levels of product aggregation where intra-industry trade becomes a larger issue.

B.5. NORMALISED RCA (NRCA)

Normalised Revealed Comparative Advantage (Yu, Cai & Leung 2008) proposes an RCA index that is more comparable across commodity, country and time.

$$NRCA_{cp} = \frac{X_{cp}}{X_w} - \frac{X_{wp}X_c}{X_w X_w}$$

"The NRCA index measures the degree of deviation of a country's actual export from its comparative-advantage-neutral level in terms of its relative scale with respect to the world export market" (Yu, Cai & Leung 2008). This expression essentially reflects the idea that when exports are in excess of the expected export value based on the world composition of trade, a country has comparative advantage in exporting that product.[125]

A country c has a comparative advantage in product p when NRCA > 0. This metric has the added property such that the sum of NRCA over all products is equal to zero in addition to the sum of NRCA over a country's exports. Therefore NRCA is more conducive to dynamic analysis of RCA as the mean value of NRCA is invariant over time.

The case of zero exports is however a departure from other measures of revealed comparative advantage such as in the baseline BRCA. The BRCA measure would equate to a zero value meaning simply that country does not have a comparative advantage in

[124] This issue is tempered in this study as it focuses on two curated datasets that have been constructed in an effort to improve data quality.

[125] Due to the very small nature of NRCA a scale factor is typically applied with 10,000 as the recommended factor by the author.

that product. In this NRCA measure however, a country with zero exports will have a –ve value indicating a comparative disadvantage by design in zero export flows. The relative size of this –ve value is determined by both the country size operating through total exports (Xc), in addition to total product exports in world trade (X_{wp}) normalised by the square of total exports (X_w).

B.5.1. IMPACT ON PRODUCT RANKINGS

There are large departures between the BRCA and NRCA measures with respect to their identification when ranking the identified products with revealed comparative advantage. Using the Year 2000 as a sample year, Table B.1 provides the top five products were identified for China (CHN), Thailand (THA), and the United States of America (USA).

Table B.1: Balassa compared with Yu RCA Metrics for the Year 2000

CHN	Balassa (1965)		Yu (2009)	
	SITC4	**Description**	**SITC4**	**Description**
1	2613	Raw Silk	8942	Children's toys, indoor games
2	5723	Pyrotechnic articles	8510	Footwear
3	6597	Plaits, plaited products …	8310	Travel goods, handbags etc
4	8994	Umbrellas, canes …	8421	Outerwear knitted …
5	7622	Portable radio receivers	7525	Peripheral Units

THA	Balassa (1965)		Yu (2009)	
	SITC4	**Description**	**SITC4**	**Description**
1	2320	Natural rubber latex	7524	Digital central storage units …
2	9410	Animals, live nes (Zoo Animals)	7764	Electronic microcircuits …
3	0422	Rice	2320	Natural rubber latex
4	0548	Vegetable products roots …	7599	Parts, nes for machines …
5	0372	Crustaceans and molluscs	7525	Peripheral units …

USA	Balassa (1965)		Yu (2009)	
	SITC4	**Description**	**SITC4**	**Description**
1	6553	Knitted or crocheted fabrics …	7924	Aircraft (>15000 kg)
2	6344	Wood-based panels, nes	7764	Electronic microcircuits
3	0459	Buckweat	7849	Other parts for vehicles …
4	6880	Uranium depleted in U235 …	7149	Parts of engines and motors …
5	2222	Soya beans	7768	Crystals, and electronic parts .

As is evident from inspection of the top five products with RCA, the BRCA index is susceptible to some special products that produce large RCA numbers primarily due to small overall shares in world trade (i.e. Uranium and Live Animals, nes (Zoo Animals)) affecting the RCA measure through the denominator. The NRCA also produces more consistent rankings that fall in line with 'ad hoc' expectation, such as in the case of the USA.

Comparing all countries and computing the average within country rank correlation between BRCA and NRCA demonstrates the degree of the differences in how each indicator ranks products according to RCA. Given the results there is on average general broad agreement between the two measures, while NRCA is less susceptible to small denominators producing overly large RCA values

Table B.2: Balassa and Yu RCA metrics Rank Correlation

	Balassa (1965) compared with Yu (2009)
Spearman (Rho) Rank Measure	0.84 (std: 0.12)
Kendall (Tau) Rank Measure	0.68 (std: 0.13)

B.5.2. IMPACT ON PRODUCT IDENTIFICATION (M_{cp} MATRIX)

While the two measures do not provide rank equivalence, the set of products identified to have revealed comparative advantage are equivalent between the two measures. Therefore when computing the cross-country assignment matrix (M_{cp}) between countries and products, the two measures are equivalent. This can be shown algebraically.

For $BRCA_{cp} \geq 1$:

$$\frac{E_{cp}}{E_c} \geq \frac{E_p}{E}$$

Now considering the NRCA measure it can be rearranged to show the same relationship based on its cut-off value of 0.

For $NRCA_{cp} \geq 0$:

$$\frac{E_{cp}}{E} \geq \frac{E_c * E_p}{E * E}$$

Multiplying both sides by E/E_c provides the relation:

$$\frac{E_{cp}}{E_c} \geq \frac{E_p}{E}$$

Therefore these measures are perfectly equivalent when computing the M_{cp} matrix. When considering RCA rankings within country export vectors, in addition to changes in RCA over time, the NRCA approach appears to be a better method for normalising the data.

B.6. EK RCA (EKRCA)

A measure of Revealed Comparative Advantage from (Arkolakis, Costinot & Rodriguez-Clare 2012; Leromain & Orefice 2013) is more theoretically founded and based on the Eaton & Kortum (2002) Ricardian style model. Costinot (2012) adds an industry level and derives a measure of Revealed Comparative Advantage based on fixed effect regression techniques.[126]

Leromain & Orefice (2014) use this new approach and find that the resultant metric displays superior statistical properties to the Balassa RCA measure. The resultant metric exhibits a symmetric distribution, better time stability, and superior ordinal ranking properties. These are particularly important if one uses RCA within a regression framework.

While this metric is certainly attractive there are two reasons it is not used in this study. First, the dimensionality of my dataset posses a significant estimation issue due to the large number of fixed effects. If I estimate the proposed model for a single cross section of 179 countries and 283 products this amounts to 50,657 exporter-product fixed effects, 50,657 importer-product fixed effects, and 80,089 exporter-importer

[126] This measure is based on bilateral trade relationships. Aggregation would be required over partner countries to bring the results to the export level, but the advantage of this metric is that is more closely linked with the theoretical underpinnings of comparative advantage.

fixed effects (181,439 total fixed effects). Second, the estimate of θ, which is a parameter that measures intra-industry heterogeneity in the base model, has only been computed for a single cross section and at a relatively high levels of aggregation relative to the data used in this study. Therefore using the preferred estimate of $\theta = 6.53$ based on Costinot (2012) one would need to assume that this parameter is time invariant. Given this parameter is used to retrieve the measure of RCA: $z_{i,k} = e^{\frac{\delta_{i,k}}{\theta}}$, not allowing variation within industry heterogeneity over time may bias the estimates across the relatively long time space of data used in this thesis.

Given its firm roots embedded in the EK Ricardian model for many countries and many products this metric will be important in any further research. As has been documented in Leromain & Orefice (2014) a procedure such as that proposed in Abowd (2002), which estimates high dimensionality fixed effects through the use of grouping in sparse matrices, would be essential.

B.7. OTHER RCA MEASURES

While there is a range of other RCA measures proposed the above measures capture the main thread over time with regards to the primary contributions of revealed comparative advantage measures that deal primarily with export data. Additional threads of literature are more directly concerned with the relationship with trade theoretic approaches to measurement.

Bowen (1983) shows that a tight coupling between the theory of comparative advantage and the traditional measures of revealed comparative advantage (such as BRCA) is problematic without imposing the assumption that all countries export all commodities. These alternative theoretically consistent measures however typically rely on production and consumption data, which is difficult to source and compare for many countries.

B.8. EMPIRICAL CONSIDERATIONS – COMPUTATION OF BALASSA RCA

When considering the computation of the Balassa (1965) RCA indicator

$$BRCA_{cp} = \frac{\frac{X_{cp}}{X_c}}{\frac{X_p}{X_w}}$$

The denominator is influenced by the selective exclusion of some data that is contained in the NBER dataset but is unassigned to a specific exporter. These regions are typically NES areas. Therefore when computing this metric a comparison was made between defining X_p and X_w endogenously from the dataset through internal aggregation with values computed from the original raw NBER dataset.

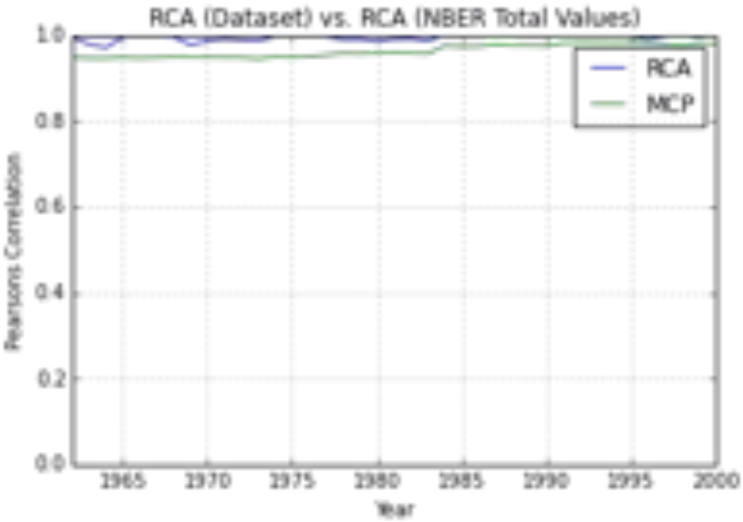
Using the most disaggregated SITC level 4 data for the year 2000 and conducting a pairwise comparison of RCA values using Pearsons, Kendall τ rank, and Spearman's rank correlations shows the values are all very similar.

Table B.3: RCA and M_{cp} correlation using country only or global aggregates

Year 2000	Pearson	Kendall Tau	Spearman Rank
RCA	0.998	0.987	0.998
M_{cp}	0.992	0.992	0.992

Comparing the results across years shows a largely stable relationship across time as shown in Figure B.1. Therefore exclusion of these other 'nes' categories does not greatly influence the outcome of the BRCA. Therefore computation of the BRCA index, in addition to other measures of revealed comparative advantage, use aggregate data computed from within the dataset.

Figure B.1: RCA values computed with dataset total values compared to total values in raw NBER dataset



B.9. HILLMAN CONDITIONS

A review of the Hillman (1980) conditions were applied to the construction of M_{cp} matrices for various datasets defined in Appendix A. In broad agreement with Hinloopen & van Marrewijk (2008) these conditions do not apply to a large proportion of revealed comparative advantage estimates. In the NBER dataset for the year 2000 there were only four significant cases where the condition was violated.

Table B.4: Hillman (1980) violations for NBER fro the Year 2000

Country	SITC4	Description
Iraq (IRQ)	3330	Petroleum Oils, Crude
Saudi Arabia (SAU)	3330	Petroleum Oils, Crude
United States (USA)	6344	Wood based panels, n.e.s
United States (USA)	6553	Knitted, crocheted fabrics, elastic and rubberised

Given this result, it does not appear that the Hillman conditions are violated for a significantly large number of BRCA cases identified in the context of data used in this thesis. In addition, while it may be true that Iraq and Saudi Arabia have an unduly large estimate of RCA, it is broadly considered that these countries poses a comparative advantage in oil production due to the relative size of their oil endowments. Therefore the Hillman conditions are not considered further as a filter on the M_{cp} matrix.

B.10. CONCLUSION AND RECOMMENDATION

The Balassa (1965) measure of RCA is used as a baseline metric for two primary reasons. This metric has properties that are well understood, and in addition, this measure is also used broadly in the empirical literature. This is not to dismiss recent literature such as Yu (2008) provide convincing alternatives for ranking countries products according to RCA. The normalised revealed comparative advantage appears to perform well when ordering products within country export vector according to their level of revealed comparative advantage. The BRCA metric suffers various weaknesses, such as the issue of small denominators for products that have small world export shares generating large RCA estimates. The NRCA performs a more globally appropriate normalisation strategy.

Most significantly the computation of the cross-country assignment of countries and products that have revealed comparative advantage (as is captured by the M_{cp} matrix) is indifferent to the use of BRCA or NRCA. Therefore many of the differences are in the internal consistency of the measure and affects how each measure ranks products with its associated degree of RCA.

APPENDIX C EXAMPLES OF COMPLEXITY RANKINGS

C.1. EXAMPLES OF COMPLEXITY INDICATORS

This appendix contains example data to demonstrate the results of the economic complexity indicator (ECI), the product complexity indicator (PCI) and proximity, computed for the year 2012 from the CEPII BACI world trade dataset. The year 2000 is primarily used in Chapter 2 as it is the year used most widely in the literature and therefore maintains comparability. The observed macro patterns discussed in Chapter 2 are largely similar to those presented here. There are however differences in the micro patterns over time in product and country rankings. These metrics have been computed using Dataset D (see Appendix A for the full definition). In 2012 this dataset consists of SITC revision 2 codes at the 4 digit level.

C.2. ECONOMIC COMPLEXITY INDICATOR (ECI)

See Table C.1 for full data table.

C.3. PRODUCT COMPLEXITY INDICATOR (PCI)

See Table C.2 for full data table.

C.4. PROXIMITY DATA

See Table C.3 for full data table.

C.5. THE ATLAS OF COMPLEXITY FORMULATION OF ECI AND PCI

This method of reflections produces very similar results to the latest method of computation used by the Atlas of Complexity, which uses the eigenvector associated with the second largest eigenvalue of the following recursive definition (Hausmann & Hidalgo 2014).

$$k_{c,N} = \sum_{c'} M_{cc'} * k_{c',N-2}$$

where,

$$M_{cc'} = \sum_{p} \frac{M_{cp} * M_{c'p}}{k_{c,0} * k_{p,0}}$$

ECI and PCI therefore are

$$ECI = \frac{\vec{K} - < \vec{K} >}{stdev(\vec{K})}$$

where,

$\vec{K}$ = Eigenvector of $\widetilde{M_{cc'}}$ associated with second largest eigenvalue

and

$$PCI = \frac{\vec{Q} - < \vec{Q} >}{stdev(\vec{Q})}$$

$\vec{Q}$ = Eigenvector of $\widetilde{M_{pp'}}$ associated with second largest eigenvalue

Table C.1: Economic Complexity Indicator (ECI) for the Year 2012

ISO3C	ECI	RANK	ISO3C	ECI	RANK	ISO3C	ECI	RANK	ISO3C	ECI	RANK
JPN	2.44	1	BIH	0.68	45	GTM	-0.14	89	CMR	-0.84	133
CHE	2.09	2	UKR	0.63	46	LKA	-0.17	90	ZMB	-0.84	134
DEU	2.00	3	TUR	0.63	47	KEN	-0.22	91	TCD	-0.84	135
KOR	1.99	4	BHS	0.60	48	SOM	-0.23	92	KWT	-0.86	136
TWN	1.98	5	ARM	0.54	49	OMN	-0.25	93	IRN	-0.89	137
SWE	1.81	6	GIB	0.52	50	MAR	-0.25	94	CIV	-0.94	138
SGP	1.79	7	PHL	0.52	51	AUS	-0.26	95	SUR	-0.94	139
AUT	1.79	8	TUN	0.48	52	HND	-0.27	96	BFA	-0.98	140
CZE	1.78	9	IND	0.44	53	TTO	-0.27	97	LBR	-0.99	141
FIN	1.76	10	PAN	0.37	54	RWA	-0.28	98	YEM	-1.00	142
HUN	1.61	11	JOR	0.34	55	SEN	-0.28	99	LAO	-1.02	143
USA	1.53	12	LBN	0.34	56	UZB	-0.29	100	MMR	-1.02	144
SVN	1.52	13	WSM	0.33	57	KIR	-0.31	101	PNG	-1.04	145
GBR	1.50	14	CRI	0.33	58	ALB	-0.33	102	TJK	-1.07	146
SHN	1.49	15	ISL	0.32	59	UGA	-0.33	103	VEN	-1.07	147
FRA	1.47	16	NZL	0.32	60	KAZ	-0.33	104	MWI	-1.07	148
ISR	1.42	17	GRC	0.29	61	BHR	-0.35	105	TKM	-1.09	149
SVK	1.42	18	BRA	0.28	62	SPM	-0.36	106	FLK	-1.09	150
ITA	1.40	19	RUS	0.27	63	ARE	-0.37	107	AZE	-1.14	151
KNA	1.33	20	MDA	0.27	64	PER	-0.40	108	GHA	-1.18	152
BEL	1.29	21	BMU	0.26	65	SYC	-0.42	109	DZA	-1.19	153
DNK	1.27	22	MUS	0.19	66	TZA	-0.45	110	MNG	-1.19	154
IRL	1.27	23	SLV	0.16	67	SLE	-0.49	111	NER	-1.19	155
MEX	1.20	24	ARG	0.15	68	GRL	-0.51	112	MOZ	-1.21	156
CHN	1.19	25	KGZ	0.14	69	PAK	-0.51	113	BEN	-1.34	157
MLT	1.18	26	SYR	0.12	70	PRY	-0.51	114	MRT	-1.40	158
NLD	1.14	27	COL	0.10	71	TGO	-0.52	115	NGA	-1.45	159
POL	1.13	28	IDN	0.09	72	QAT	-0.57	116	SDN	-1.54	160
MYS	1.07	29	NPL	0.07	73	ECU	-0.57	117	GNQ	-1.55	161
ESP	1.05	30	ZAF	0.07	74	BOL	-0.58	118	COG	-1.74	162
HKG	1.00	31	CHL	0.06	75	AFG	-0.60	119	GNB	-1.83	163
BLR	0.98	32	URY	0.04	76	ZWE	-0.62	120	GAB	-1.93	164
THA	0.98	33	MKD	0.03	77	CUB	-0.63	121	GIN	-2.07	165
EST	0.93	34	VNM	0.00	78	KHM	-0.65	122	LBY	-2.10	166
HRV	0.93	35	DOM	-0.02	79	BDI	-0.65	123	IRQ	-2.39	167
BRB	0.92	36	SAU	-0.04	80	BLZ	-0.66	124	AGO	-2.68	168
MAC	0.89	37	PRK	-0.04	81	MLI	-0.68	125			
LTU	0.88	38	DJI	-0.05	82	NIC	-0.70	126			
CAN	0.83	39	JAM	-0.07	83	MDG	-0.72	127			
PRT	0.80	40	NCL	-0.10	84	GUY	-0.73	128			
CYP	0.79	41	EGY	-0.10	85	BGD	-0.76	129			
LVA	0.79	42	CAF	-0.10	86	ETH	-0.78	130			
BGR	0.76	43	FJI	-0.10	87	GMB	-0.82	131			
NOR	0.70	44	GEO	-0.11	88	HTI	-0.82	132			

Table C.2: Product Complexity Indicator (PCI) for the Year 2012

SITC4	PCI	RANK	Description
7133	2.39	1	Internal combustion piston engines, marine propulsion
7284	2.13	2	Machinery for specialized industries and parts thereof, nes
6643	2.07	3	Drawn or blown glass (flashed glass), unworked, in rectangles
8813	2.06	4	Photographic and cinematographic apparatus and equipment, nes
5843	2.04	5	Cellulose acetates
7367	2.03	6	Other machines-tools for working metal or metal carbides, nes
7244	2.00	7	Machines for extruding man-made textile; other textile machinery
8710	1.97	8	Optical instruments and apparatus
7252	1.86	9	Machinery for making paper pulp, paper, paperboard; cutting machines
8744	1.82	10	Nonmechanical or electrical instruments for physical, etc, analysis
5826	1.80	11	Epoxide resins
6641	1.80	12	Glass in the mass, in balls, rods or tubes (nonoptical); waste
7369	1.79	13	Parts, nes of and accessories for machine-tools of heading 736
5827	1.78	14	Silicones
8821	1.73	15	Chemical products and flashlight materials for use in photografy
7742	1.73	16	X-ray apparatus and equipment; accessories; and parts, nes
8748	1.73	17	Electrical measuring, controlling, etc, instruments, apparatus, nes
6745	1.70	18	Sheet, plates, rolled of thickness 3mm to 4,75mm, of iron or steel
7246	1.70	19	Auxiliary machinery for use with those of headings 72451 to 72453
8983	1.63	20	Sound recording tape, discs
3345	1.61	21	Lubricating petroleum oils, and preparations, nes
7499	1.61	22	Other non-electric parts and accessories of machinery, nes
5223	1.60	23	Halogen and sulphur compounds of non-metals
7361	1.60	24	Metal cutting machine-tools
7741	1.60	25	Electro-medical equipment
7451	1.59	26	Power hand tools, pneumatic or non-electric, and parts thereof, nes
7373	1.57	27	Welding, brazing, cutting, etc machines and appliances, parts, nes
7492	1.57	28	Cocks, valves and similar appliances, for pipes boiler shells, etc
7368	1.57	29	Work holders, dividing heads for machine-tools, etc; tool holders
6954	1.54	30	Interchangeable tools for hand or machine tools (tips, blades, etc)
8822	1.53	31	Photographic film, plates and paper (other than cinematograph film)
7431	1.52	32	Air pumps, vacuum pumps and air or gas compressors
5154	1.51	33	Organo-sulphur compounds
7259	1.50	34	Parts, nes of the machines falling within heading 725
5838	1.49	35	Ion exchangers of the polymerization or copolymerization type
7416	1.49	36	Machinery, plant, laboratory equipment for heating and cooling, nes
6940	1.49	37	Nails, screws, nuts, bolts, rivets, etc, of iron, steel or copper
7439	1.49	38	Parts, nes of the machines falling within headings 7435 and 7436
7712	1.48	39	Other electric power machinery, parts, nes
5148	1.47	40	Other nitrogen-function compounds
5839	1.46	41	Other polymerization and copolymarization products
7493	1.45	42	Shaft, crank, bearing housing, pulley and pulley blocks, etc
8743	1.44	43	Gas, liquid control instruments and apparatus, non-electrical
8742	1.43	44	Drawing, marking-out and mathematical calculating instruments, etc
7432	1.42	45	Parts, nes of the pumps and compressor falling within heading 7431
7269	1.42	46	Parts, nes of machines falling within headings 72631, 7264, 7267
7452	1.41	47	Other non-electrical machines and parts thereof, nes

Examples of Complexity Rankings

SITC4	PCI	RANK	Description
7132	1.40	48	Motor vehicles piston engines, headings: 722; 78; 74411 and 95101
6648	1.39	49	Glass mirror, unframed, framed or backed
7732	1.37	50	Electrical insulating equipment
6992	1.37	51	Chain and parts thereof, of iron or steel
7423	1.36	52	Rotary pumps (other than those of heading 74281)
6632	1.35	53	Abrasive power or grain, on a base of woven fabrics
5836	1.35	54	Acrylic and methaacrylic polymers; acrylo-methacrylic copolymers
7281	1.31	55	Machine-tools for specialized industries; parts or accessories, nes
6573	1.31	56	Coated or impregnated textile fabrics and products, nes
7783	1.31	57	Automotive electrical equipment; and parts thereof, nes
7441	1.31	58	Work trucks, of the type use in factories, dock areas, etc
7491	1.31	59	Ball, roller or needle roller bearings
7784	1.30	60	Electro-mechanical hand tools, and parts thereof, nes
5332	1.30	61	Printing inks
7842	1.30	62	Bodies, for vehicles of headings 722, 781-783
8935	1.30	63	Articles of electric lighting of plastic
7161	1.29	64	Motors and generators, direct current
5829	1.29	65	Other condensation, polycodensation or polyaddition products
7129	1.28	66	Parts, nes of steam power units
5825	1.28	67	Polyurethanes
5113	1.28	68	Halogenated derivatives of hydrocarbons
7263	1.28	69	Machinery, accessories for type-setting, for printing blocks, etc
7648	1.27	70	Telecommunications equipment, nes
5415	1.27	71	Hormones, natural, or reproduce by synthesis, in bulk
7251	1.27	72	Machinery for making, finishing cellulose pulp, paper or paperboard
6994	1.27	73	Springs and leaves for springs, of iron, steel or copper
7421	1.26	74	Reciprocating pumps (other than those of heading 74281)
8749	1.26	75	Parts, nes, and accessories of headings 873, 8743, 87454 or 8748
7264	1.25	76	Printing presses
7422	1.25	77	Centrifugal pumps (other than those of heading 74281)
5163	1.25	78	Inorganic esters, their salts and derivatives
7429	1.25	79	Parts, nes of pumps and liquids elevators falling in heading 742
8732	1.24	80	Counting devices non-electrical; stroboscopes
7849	1.24	81	Other parts and accessories, for vehicles of headings 722, 781-783
7362	1.24	82	Metal forming machine-tool
0913	1.23	83	Lard, pig and poultry fat, rendered or solvent-extracted
7247	1.22	84	Textile machinery, nes for cleaning, cutting, etc, and parts nes
7754	1.22	85	Electric shavers and hair clippers, parts thereof, nes
7442	1.21	86	Lifting, handling, loading machinery, telphers and conveyors
8981	1.20	87	Pianos, other string musical instruments
8959	1.20	88	Other office and stationary supplies
5419	1.20	89	Pharmaceutical goods, other than medicaments
6953	1.20	90	Other hand tools
5416	1.19	91	Glycosides, glands, antisera, vaccines and similar products
7188	1.19	92	Engines and motors, nes (wind, hot air engines, water wheel, etc)
7413	1.18	93	Industrial and laboratory furnaces and ovens, etc, parts, nes
5157	1.18	94	Sulphonamides, sultones and sultams
5331	1.18	95	Other colouring matter; inorganic products use as luminophores

SITC4	PCI	RANK	Description
7763	1.18	96	Diodes, transistors, photocells, etc
7518	1.18	97	Office machines, nes
5156	1.18	98	Heterocyclic compound; nucleic acids
7753	1.17	99	Domestic dishwashing machines
8996	1.17	100	Orthopaedic appliances, hearing aids, artificial parts of the body
7412	1.17	101	Furnace burners; mechanical stokers, etc, and parts thereof, nes
7149	1.17	102	Parts, nes of the engines and motors of group 714 and item 71888
5849	1.16	103	Other chemical derivatives of cellulose; vulcanized fibre
6546	1.15	104	Fabrics of glass fibre (including narrow, pile fabrics, lace, etc)
7649	1.15	105	Parts, nes of and accessories for apparatus falling in heading 76
7523	1.15	106	Complete digital central processing units; digital processors
7758	1.14	107	Electro-thermic appliances, nes
7434	1.14	108	Fans, blowers and the like, and parts thereof, nes
6631	1.14	109	Hand polishing stone, grindstones, grinding wheels, etc
8745	1.13	110	Measuring, controlling and scientific instruments, nes
5123	1.11	111	Phenols and phenol-alcohols, and their derivatives
7764	1.11	112	Electronic microcircuits
7112	1.11	113	Auxiliary plant for boilers of heading 7111; condensers
5162	1.11	114	Aldehyde, ketone and quinone-function compounds
7139	1.10	115	Piston engines parts, nes, falling in headings: 7132, 7133 and 7138
7599	1.10	116	Parts, nes of and accessories for machines of headings 7512 and 752
5833	1.10	117	Polystyrene and its copolymers
7528	1.10	118	Off-line data processing equipment, nes
7525	1.10	119	Peripheral units, including control and adapting units
6832	1.10	120	Nickel and nickel alloys, worked
5835	1.09	121	Copolymers of vinyl chloride and vinyl acetate
7914	1.09	122	Railway, tramway passenger coaches, etc, not mechanically propelled
6997	1.09	123	Articles of iron or steel, nes
8989	1.09	124	Parts, nes of and accessories for musical instruments; metronomes
7788	1.08	125	Other electrical machinery and equipment, nes
5983	1.08	126	Organic chemical products, nes
7768	1.08	127	Crystals, and parts, nes of electronic components of heading 776
5824	1.07	128	Polyamides
7371	1.07	129	Metallurgy and metal foundry equipment, and parts thereof, nes
7435	1.06	130	Centrifuges
6642	1.06	131	Optical glass and elements of optical glass (unworked)
7212	1.03	132	Harvesting and threshing machines; fodder presses, etc; parts nes
7641	1.02	133	Electrical line telephonic and telegraphic apparatus
6635	1.02	134	Wool; expanding or insulating mineral materials, nes
6289	1.01	135	Other articles of rubber, nes
6412	1.01	136	Printing paper and writing paper, in rolls or sheets
5413	1.01	137	Antibiotics, not put up as medicaments
5145	1.01	138	Amine-function compounds
7148	1.00	139	Gas turbines, nes
7224	1.00	140	Wheeled tractors (other than those falling in heading 74411, 7832)
5822	0.99	141	Aminoplasts
8841	0.98	142	Lenses and other optical elements of any material
7436	0.98	143	Filtering and purifying machinery, apparatus for liquids and gases

SITC4	PCI	RANK	Description
7722	0.98	144	Printed circuits, and parts thereof, nes
5138	0.98	145	Polycarboxylic acids and their derivatives
0451	0.97	146	Rye, unmilled
6744	0.97	147	Sheet, plates, rolled of thickness 4,75mm plus, of iron or steel
8951	0.97	148	Office and stationary supplies, of base metal
7721	0.96	149	Switches, relays, fuses, etc; switchboards and control panels, nes
6647	0.96	150	Safety glass consisting of toughened or laminated glass, cut or not
8852	0.96	151	Clocks, clock movements and parts
7138	0.96	152	Internal combustion piston engines, nes
7187	0.95	153	Nuclear reactors, and parts thereof, nes
7913	0.94	154	Mechanically propelled railway, tramway, trolleys, etc
8121	0.94	155	Central heating equipment, not electrically heated, parts, nes
5841	0.94	156	Regenerated cellulose
6637	0.94	157	Refractory goods, nes
8974	0.94	158	Other articles of precious metals or rolled precious metals, nes
7449	0.93	159	Parts, nes of the machinery falling within heading 7442
8942	0.93	160	Children's toys, indoor games, etc
7757	0.92	161	Domestic electro-mechanical appliances; and parts thereof, nes
5311	0.92	162	Synthetic organic dyestuffs, etc, natural indigo and colour lakes
6760	0.91	163	Rails and railway track construction materials, of iron or steel
6649	0.91	164	Glass, nes
5335	0.90	165	Glazes, driers, putty etc
7126	0.90	166	Steam power units (mobile engines but not steam tractors, etc)
0113	0.89	167	Pig meat fresh, chilled or frozen
6998	0.88	168	Articles, nes, of copper, nickel, aluminium, lead, zinc and tin
6785	0.88	169	Tube and pipes fittings, of iron or steel
7268	0.88	170	Bookbinding machinery; parts thereof, nes
5137	0.88	171	Monocarboxylic acids and their derivatives
6411	0.87	172	Newsprint
6658	0.87	173	Articles made of glass, nes
7621	0.87	174	Radio receivers for motor-vehicles
2671	0.87	175	Regenerated fibre suitable for spinning
6418	0.86	176	Paper and paperboard, coated, impregnated, etc, in rolls or sheets
7169	0.86	177	Parts, nes, of rotating electric plant
7841	0.86	178	Chassis fitted with engines, for vehicles of headings 722, 781-783
7415	0.86	179	Air conditioning machines and parts thereof, nes
7245	0.86	180	Weaving, knitting, etc, machines, machines for preparing yarns, etc
7810	0.86	181	Passenger motor vehicles (excluding buses)
8946	0.85	182	Non-military arms and ammunition therefor
7643	0.85	183	Television, radio-broadcasting; transmitters, etc
8842	0.85	184	Spectacles and spectacle frames
7213	0.85	185	Dairy machinery, nes (including milking machines), and parts nes
6727	0.84	186	Iron or steel coils for re-rolling
6674	0.83	187	Synthetic or reconstructed precious or semi-precious stones
7723	0.83	188	Fixed, variable resistors, other than heating resistors, parts, nes
7411	0.82	189	Gas generators, and parts, nes of gas generators
7219	0.81	190	Agricultural machinery and appliances, nes, and parts thereof, nes
7832	0.80	191	Road tractors for semi-trailers

SITC4	PCI	RANK	Description
6782	0.79	192	Seamless tubes, pipes; blanks for tubes and pipes, of iron or steel
6991	0.79	193	Locksmiths wares, safes, etc, and hardware, nes, of base metal
7919	0.79	194	Railway track fixtures, and fittings, etc, parts nes of heading 791
5169	0.78	195	Organic chemicals, nes
7929	0.78	196	Parts, nes of the aircraft of heading 792
5851	0.78	197	Modified natural resins etc; derivatives of natural rubber
5417	0.77	198	Medicaments (including veterinary medicaments)
6639	0.77	199	Articles of ceramic materials, nes
7267	0.77	200	Other printing machinery; machines for uses ancilliary to printing
7243	0.77	201	Sewing machines, furniture, needles etc, and parts thereof, nes
7591	0.76	202	Parts, nes of and accessories for machines of headings 7511 or 7518
8720	0.76	203	Medical instruments and appliances, nes
8939	0.76	204	Miscellaneous articles of plastic
2518	0.75	205	Chemical wood pulp, sulphite
0013	0.75	206	Swine, live
7611	0.75	207	Television receivers, colour
5155	0.74	208	Other organo-inorganic compounds
7762	0.74	209	Other electronic valves and tubes
6415	0.73	210	Paper and paperboard, in rolls or sheets, nes
5821	0.73	211	Phenoplasts
5122	0.73	212	Cyclic alcohols, and their derivatives
6579	0.73	213	Special products of textile materials
7631	0.72	214	Gramophones and record players, electric
6210	0.71	215	Materials of rubber
7928	0.71	216	Aircraft, nes and associated equipment
7522	0.70	217	Complete digital data processing machines
6975	0.70	218	Base metal indoors sanitary ware, and parts thereof, nes
8941	0.70	219	Baby carriages and parts thereof, nes
6251	0.70	220	Tires, pneumatic, new, for motor cars
6812	0.69	221	Metals of platinum group, unwrought, unworked, or semi-manufactured
7162	0.69	222	Electric motors, generators (not direct current); generating sets
6793	0.68	223	Steel and iron forging and stampings, in the rough state
6572	0.68	224	Bonded fibre fabrics, etc, whether or not impregnated or coated
7372	0.67	225	Rolling mills, rolls therefor, and parts, nes of rolling mills
6891	0.67	226	Tungsten, molybdenum, tantalum, magnesium, unwrought; waste, scrap
5842	0.66	227	Cellulose nitrates
6282	0.66	228	Transmission, conveyor or elevator belts, of vulcanized rubber
6419	0.65	229	Converted paper and paperboard, nes
5922	0.64	230	Albuminoid substances; glues
6623	0.64	231	Refractory bricks and other refractory construction materials
0252	0.63	232	Eggs, birds', egg yolks, fresh, dried or preserved, not in shell
8982	0.63	233	Musical instruments, nes
7782	0.62	234	Electric filament lamps and discharge lamps; arc-lamps
6872	0.62	235	Tin and tin alloys worked
6252	0.61	236	Tires, pneumatic, new, for buses and lorries
7414	0.61	237	Non-domestic refrigerators and refrigerating equipment, parts, nes
6633	0.61	238	Manufactures of mineral materials, nes (other than ceramic)
2516	0.61	239	Chemical wood pulp, dissolving grades

SITC4	PCI	RANK	Description
8922	0.60	240	Newspapers, journals and periodicals
7751	0.59	241	Household laundry equipment, nes
5239	0.59	242	Inorganic chemical products, nes
6259	0.58	243	Other tires, tire cases, tire flaps and inner tubes, etc
8932	0.58	244	Plastic sanitary and toilet articles
5161	0.58	245	Ethers, epoxides, acetals
6652	0.58	246	Glassware (other than heading 66582), for indoor decoration
4311	0.57	247	Processed animal and vegetable oils
5823	0.57	248	Alkyds and other polyesters
6577	0.57	249	Wadding, wicks and textiles fabrics for use in machinery or plant
5411	0.56	250	Provitamins and vitamins
5989	0.56	251	Chemical products and preparations, nes
6551	0.56	252	Knitted etc, not elastic nor rubberized, of synthetic fibres
8952	0.56	253	Pens, pencils and, fountain pens
6999	0.56	254	Other base metal manufactures, nes; and of cermets
0121	0.56	255	Bacon, ham, other dried, salted or smoked meat of domestic swine
7868	0.55	256	Other not mechanically propelled vehicles; and parts, nes
5233	0.55	257	Salts of metallic acids; compounds of precious metals
7111	0.55	258	Steam and other vapour-generated boilers; super-heated water boiler
8928	0.54	259	Printed matter, nes
6544	0.54	260	Fabrics, woven, of flax or of ramie
5982	0.53	261	Anti-knock preparation, anti-corrosive; viscosity improvers; etc
1122	0.53	262	Other fermented beverages, nes (cider, perry, mead, etc)
8731	0.52	263	Gas, liquid and electricity supply or production meters; etc
8811	0.51	264	Photographic cameras, flashlight apparatus, parts, accessories, nes
6542	0.51	265	Fabrics, woven, 85% plus of sheep's or lambs' wool or of fine hair
7638	0.51	266	Other sound recording and reproducer, nes; video recorders
7622	0.50	267	Portable radio receivers
6880	0.50	268	Uranium depleted in U235, thorium, and alloys, nes; waste and scrap
6646	0.50	269	Bricks, tiles, etc of pressed or moulded glass, used in building
8124	0.48	270	Lighting fixture and fittings, lamps, lanterns, and parts, nes
7239	0.47	271	Parts, nes of machinery and equipment of headings 72341 to 72346
5147	0.47	272	Amide-function compounds; excluding urea
7915	0.47	273	Railway and tramway freight, etc, not mechanically propelled
7512	0.46	274	Calculating, accounting, cash registers, ticketing, etc, machines
6591	0.46	275	Linoleum and similar floor covering
6645	0.45	276	Cast, rolled glass (flashed or wired), unworked, in rectangles
6747	0.45	277	Tinned sheets, plates of steel (not of high carbon or alloy steel)
8998	0.45	278	Small-wares and toilet articles, nes; sieves; tailors' dummies, etc
7781	0.45	279	Batteries and electric accumulators, and parts thereof, nes
5334	0.45	280	Varnishes and lacquers; distempers etc
0619	0.45	281	Sugars and syrups nes; artificial honey; caramel
5834	0.45	282	Polyvinyl chloride
2331	0.45	283	Synthetic rubber, latex; factice derived from oils
6644	0.44	284	Glass, cast, rolled, etc, surface-ground, but no further worked
0730	0.44	285	Chocolate and other preparations containing cocoa, nes
7211	0.44	286	Agricultural and horticultural machinery for soil preparation, etc
6352	0.44	287	Casks, barrels; other coopers products and parts, including staves

SITC4	PCI	RANK	Description
6911	0.43	288	Structures and parts of, of iron, steel; plates, rods, and the like
7223	0.43	289	Track-laying tractors
6912	0.42	290	Structures and parts of, of aluminium; plates, rods, and the like
9510	0.42	291	Armoured fighting vehicles, war firearms, ammunition, parts, nes
0142	0.42	292	Sausages and the like, of meat, meat offal or animal blood
2512	0.42	293	Mechanical wood pulp
4113	0.42	294	Animals oils, fats and greases, nes
6770	0.42	295	Iron or steel wire (excluding wire rod), not insulated
8851	0.41	296	Watches, watch movements and case
6746	0.41	297	Sheet, plates, rolled of thickness less 3mm, of iron or steel
5312	0.41	298	Synthetic organic luminophores, indigo, lakes
7428	0.40	299	Other pumps for liquids and liquid elevators
6731	0.40	300	Wire rod of iron or steel
7119	0.40	301	Parts, nes of boilers and auxiliary plant of headings 7111 and 7112
6112	0.39	302	Composition leather, in slabs, sheets or rolls
3353	0.39	303	Mineral tar pitch, pitch coke
5139	0.39	304	Oxygen-function acids, and their derivatives
7931	0.39	305	Warships
5146	0.38	306	Oxygen-function amino-compounds
0482	0.38	307	Malt, roasted or not, including flour
6993	0.38	308	Pins, needles, etc, of iron, steel; metal fittings for clothing
6543	0.38	309	Fabrics, woven, of sheep's or lambs' wool or of fine hair, nes
2226	0.37	310	Rape and colza seeds
6353	0.34	311	Builders` carpentry and joinery (including prefabricated)
0114	0.33	312	Poultry, dead and edible offal, fresh, chilled or frozen
7163	0.33	313	Rotary converters
5514	0.32	314	Mixtures of odoriferous substances, used in perfumery, food etc
8830	0.32	315	Cinematograph film, exposed and developed
6541	0.32	316	Fabrics, woven, of silk, of noil or other waste silk
6422	0.32	317	Correspondence stationary
6281	0.32	318	Hygienic, pharmaceutical articles of unhardened vulcanized rubber
5249	0.32	319	Other radio-active and associated materials
2471	0.31	320	Sawlogs and veneer logs, of coniferous species
6960	0.31	321	Cutlery
0014	0.31	322	Poultry, live
6121	0.31	323	Articles of leather use in machinery or mechanical appliances, etc
5543	0.30	324	Polishes and creams, for furniture, floors, footwear, metals etc
0452	0.30	325	Oats, unmilled
8211	0.30	326	Chairs and other seats; and parts thereof, nes
8924	0.30	327	Picture postcards, decalcomanias, etc, printed
7711	0.29	328	Transformers, electrical
7761	0.29	329	Television picture tubes, cathode ray
8219	0.29	330	Other furniture and parts thereof, nes
7272	0.29	331	Other food-processing machinery and parts thereof, nes
2666	0.29	332	Continuous filament tow for synthetic (discontinuous) fibres
5837	0.28	333	Polyvinyl acetate
0488	0.28	334	Malt extract; cereals preparations with less 50% of cocoa
7853	0.28	335	Invalid carriages; parts, nes of articles of heading 785

SITC4	PCI	RANK	Description
8947	0.27	336	Other sporting goods and fairground amusements, etc
7628	0.27	337	Other radio receivers
6571	0.27	338	Felt, articles of felt, nes, whether or not impregnated or coated
6254	0.27	339	Tires, pneumatic, new, for motorcycles and bicycles
7752	0.26	340	Domestic refrigerators and freezers
5831	0.26	341	Polyethylene
6996	0.25	342	Miscellaneous articles of base metal
6921	0.25	343	Iron, steel, aluminium reservoirs, tanks, etc, capacity 300 lt plus
7911	0.25	344	Rail locomotives, electric
2665	0.25	345	Discontinuous synthetic fibres, not carded or combed
7511	0.24	346	Typewriters; cheque-writing machines
6539	0.23	347	Pile and chenille fabrics, woven, of man-made fibres
6519	0.23	348	Yarn of textile fibres, nes
5852	0.23	349	Other artificial plastic materials, nes
9610	0.22	350	Coin (other than gold coin), not being legal tender
7923	0.22	351	Aircraft of an unladen weight from 2000 kg to 15000 kg
3224	0.21	352	Peat, not agglomerated
6973	0.21	353	Domestic, non-electric, heating, cooking apparatus, and parts, nes
6514	0.20	354	Yarn 85% of synthetic fibres, not for retail; monofil, strip, etc
7642	0.19	355	Microphones; loud-speakers; audio-frequency electric amplifiers
6749	0.19	356	Other sheet and plates, of iron or steel, worked
6794	0.19	357	Castings of iron or steel, in rough state
5542	0.19	358	Organic surface-active agents, nes
6428	0.19	359	Articles of paper pulp, paper, paperboard or cellulose wadding, nes
6535	0.19	360	Fabric, woven of continuous regenerated textile materials
6822	0.18	361	Copper and copper alloys, worked
4239	0.16	362	Other fixed vegetable oils, soft
8994	0.16	363	Umbrellas, canes and similar articles and parts thereof
2120	0.15	364	Furskins, raw
6351	0.15	365	Wood packing cases, boxes, cases, crates, etc, complete
6842	0.15	366	Aluminium and aluminium alloys, worked
5912	0.14	367	Fungicides, for sale by retail or as preparation
6595	0.14	368	Carpets, rugs, mats, of man-made textile materials, nes
5913	0.14	369	Herbicides, for sale by retail or as preparation
6931	0.14	370	Wire, cables, cordage, ropes, plaited bans, sling and the like
8921	0.14	371	Printed books, pamphlets, maps and globes
6416	0.14	372	Fibre building board of wood or other vegetable material
8960	0.12	373	Works of art, collectors' pieces and antiques
6343	0.12	374	Improved wood and reconstituted wood
6538	0.12	375	Fabrics, woven, less 85% of discontinuous regenerated fibres
5914	0.11	376	Disinfectants, etc, for sale by retail or as preparation
8972	0.10	377	Imitation jewellery
7283	0.10	378	Other mineral working machinery; and parts thereof, nes
6330	0.10	379	Cork manufactures
0118	0.09	380	Other fresh, chilled or frozen meat or edible meat offal
5111	0.09	381	Acyclic hydrocarbons
6118	0.08	382	Leather, specially dressed or finished, nes
7821	0.08	383	Motor vehicles for the transport of goods or materials

SITC4	PCI	RANK	Description
2224	0.07	384	Sunflower seeds
7144	0.07	385	Reaction engines
7612	0.06	386	Television receivers, monochrome
0149	0.06	387	Other prepared or preserved meat or meat offal
6417	0.06	388	Paper and paperboard, creped, crinkled, etc, in rolls or sheets
0240	0.05	389	Cheese and curd
6253	0.05	390	Tires, pneumatic, new, for aircraft
0230	0.05	391	Butter
2772	0.05	392	Other natural abrasives
7861	0.04	393	Trailers and transports containers
5981	0.04	394	Woods and resin-based chemical products
6424	0.04	395	Paper and paperboard cut to size or shape, nes
6618	0.04	396	Construction materials, of asbestos-cement or fibre-cements, etc
6129	0.04	397	Other articles of leather or of composition leather
5114	0.03	398	Hydrocarbons derivatives, nonhaloganeted
6413	0.02	399	Kraft paper and paperboard, in rolls or sheets
2672	0.02	400	Waste of man-made fibres, not carded or combed
4241	0.01	401	Linseed oil
5723	0.01	402	Pyrotechnic articles
2482	0.00	403	Wood of coniferous species, sawn, planed, tongued, grooved, etc
6638	0.00	404	Manufactures of asbestos; friction materials
7851	0.00	405	Motorcycles, auto-cycles; side-cars of all kind, etc
7912	0.00	406	Other rail locomotives; tenders
3352	-0.01	407	Mineral tars and products
7922	-0.01	408	Aircraft of an unladen weight not exceeding 2000 kg
6924	-0.02	409	Cask, drums, etc, of iron, steel, aluminium, for packing goods
7248	-0.02	410	Machinery for preparing, tanning, working leather, etc; parts nes
6664	-0.02	411	Porcelain or china house ware
6811	-0.02	412	Silver, unwrought, unworked, or semi-manufactured
2667	-0.02	413	Discontinuous synthetic fibres, carded or combed
6781	-0.02	414	Tubes and pipes, of cast iron
8741	-0.03	415	Surveying, navigational, compasses, etc, instruments, nonelectrical
4313	-0.04	416	Fatty acids, acid oils, and residues; degras
5112	-0.04	417	Cyclic hydrocarbons
2517	-0.06	418	Chemical wood pulp, soda or sulphate
0223	-0.07	419	Milk and cream fresh, not concentrated or sweetened
3231	-0.07	420	Briquettes, ovoids, from coal, lignite or peat
6359	-0.07	421	Manufactured articles of wood, nes
2652	-0.08	422	True hemp, raw or processed but not spun, its tow and waste
2712	-0.09	423	Natural sodium nitrate
0430	-0.09	424	Barley, unmilled
6935	-0.09	425	Gauze, cloth, grill, netting, reinforced fabric and the like
8212	-0.09	426	Furniture for medical, surgical, dental or veterinary practice
5530	-0.10	427	Perfumery, cosmetics, toilet preparations, etc
6594	-0.10	428	Carpets, rugs, mats, of wool or fine animal hair
3415	-0.10	429	Coal gas, water gas and similar gases
6978	-0.10	430	Household appliances, decorative article, etc, of base metal, nes
6951	-0.11	431	Hand tools, used in agriculture, horticulture or forestry

SITC4	PCI	RANK	Description
6733	-0.11	432	Angles, shapes, sections and sheet piling, of iron or steel
5221	-0.12	433	Chemical elements
5323	-0.12	434	Synthetic tanning substances; tanning preparations
7831	-0.12	435	Public service type passenger motor vehicles
7924	-0.12	436	Aircraft of an unladen weight exceeding 15000 kg
6517	-0.12	437	Yarn of regenerated fibres, not for retail, monofil, strip, etc
5414	-0.12	438	Vegetable alkaloids and derivatives, not put up as medicaments
5231	-0.12	439	Metallic salts and peroxysalts of inorganic acids
6863	-0.13	440	Zinc and zinc alloys worked
5222	-0.14	441	Inorganic acids and oxygen compounds of non-metals
2519	-0.14	442	Other cellulosic pulps
7271	-0.15	443	Machinery for the grain milling industry; working cereals, parts
6531	-0.16	444	Fabrics, woven, of continuous synthetic textile materials
8122	-0.17	445	Ceramic plumbing fixtures
6349	-0.17	446	Wood, simply shaped, nes
0141	-0.17	447	Meat extracts and juices; fish extracts
6724	-0.17	448	Puddled bars, pilings; ingots, blocks, lumps, etc, of iron or steel
0481	-0.18	449	Cereal grains, worked or prepared, not elsewhere specified
2881	-0.18	450	Ash and residues, nes
0115	-0.19	451	Meat of horses, asses, mules and hinnies, fresh, chilled or frozen
6783	-0.19	452	Other tubes and pipes, of iron or steel
0980	-0.19	453	Edible products and preparations, nes
7234	-0.19	454	Construction and mining machinery, nes
0251	-0.20	455	Eggs, birds', and egg yolks, fresh, dried or preserved, in shell
6560	-0.20	456	Tulle, lace, embroidery, ribbons, trimmings and other small wares
6532	-0.20	457	Fabrics, woven, 85% plus of discontinuous synthetic fibres
8931	-0.20	458	Plastic packing containers, lids, stoppers and other closures
6574	-0.20	459	Elastic fabrics and trimming (not knitted or crocheted)
0819	-0.21	460	Food waste and prepared animal feed, nes
2686	-0.21	461	Waste of sheep's or lambs' wool, or of other animal hair, nes
7731	-0.21	462	Insulated electric wire, cable, bars, etc
0484	-0.22	463	Bakery products
1124	-0.22	464	Distilled alcoholic beverages, nes
1223	-0.23	465	Tobacco, manufactured; tobacco extract and essences
3232	-0.23	466	Coke and semi-coke of coal, of lignite or peat; retort carbon
6421	-0.24	467	Packing containers, box files, etc, of paper, used in offices
0712	-0.24	468	Coffee extracts, essences or concentrates
6974	-0.24	469	Base metal domestic articles, nes, and parts thereof, nes
1221	-0.25	470	Cigars, cheroots: cigarillos
8310	-0.25	471	Travel goods, handbags etc, of leather, plastics, textile, others
1123	-0.26	472	Beer made from malt (including ale, stout and porter)
7852	-0.26	473	Cycles, not motorized
0111	-0.26	474	Bovine meat, fresh, chilled or frozen
8993	-0.28	475	Candles, matches, combustible products, etc
2511	-0.28	476	Waste paper and paperboard, etc
2734	-0.28	477	Pebbles, gravel, crushed or broken stone, etc
5832	-0.28	478	Polypropylene
6861	-0.28	479	Zinc and zinc alloys, unwrought

SITC4	PCI	RANK	Description
0574	-0.29	480	Apples, fresh
6666	-0.30	481	Ornaments, personal articles of porcelain, china, or ceramic, nes
6651	-0.31	482	Bottles etc of glass
0583	-0.33	483	Jams, jellies, marmalades, etc, as cooked preparations
6516	-0.33	484	Yarn containing less than 85% of discontinuous synthetic fibres
8997	-0.33	485	Basketwork, wickerwork; brooms, paint rollers, etc
8973	-0.35	486	Precious jewellery, goldsmiths' or silversmiths' wares
6130	-0.35	487	Furskins, tanned or dressed; pieces of furskin, tanned or dressed
8812	-0.35	488	Cinematographic cameras, projectors, etc, parts, accessories, nes
2926	-0.36	489	Live plants, bulbs, etc
0586	-0.36	490	Fruit, temporarily preserved
2614	-0.36	491	Silk worm cocoons and silk waste
7822	-0.36	492	Special purpose motor lorries and vans
5623	-0.36	493	Mineral or chemical fertilizer, potassic
5322	-0.37	494	Dyeing, tanning extracts, tannins and their derivatives
5911	-0.37	495	Insecticides, for sale by retail or as preparations
0576	-0.38	496	Figs, fresh or dried
6423	-0.38	497	Registers, exercise books, file and book covers, etc, of paper
5232	-0.39	498	Metallic salts and peroxysalts of inorganic acids
8991	-0.39	499	Articles and manufacture of carving, moulding materials, nes
6831	-0.39	500	Nickel and nickel alloys, unwrought
2711	-0.39	501	Animal or vegetable fertilizer, crude
2332	-0.40	502	Reclaimed rubber, waste, scrap of unhardened rubber
6518	-0.40	503	Yarn of regenerated fibres, put up for retail sale
0620	-0.41	504	Sugar confectionery and preparations, non-chocolate
8483	-0.41	505	Fur clothing (not headgear) and other articles made of furskins
0015	-0.41	506	Equine species, live
3351	-0.41	507	Petroleum jelly and mineral waxes
6611	-0.41	508	Lime, quick, slaked and hydraulic (no calcium oxide or hydroxide)
8481	-0.42	509	Articles of apparel, clothing accessories of leather
1121	-0.42	510	Wine of fresh grapes etc
0129	-0.42	511	Meat and edible meat offal, nes, in brine, dried, salted or smoked
6536	-0.42	512	Fabrics, woven, 85% plus of discontinuous regenerated fibres
6665	-0.43	513	Articles of domestic or toilet purposes, of other kind of pottery
5921	-0.43	514	Starches, insulin and wheat gluten
1110	-0.43	515	Non-alcoholic beverages, nes
4235	-0.43	516	Olive oil
0582	-0.43	517	Fruit, fruit-peel and parts of plants, preserved by sugar
6552	-0.44	518	Knitted, not elastic nor rubberized, of fibres other than synthetic
6512	-0.44	519	Yarn of wool or animal hair (including wool tops)
5722	-0.44	520	Fuses, caps, igniters, detonators
6549	-0.45	521	Fabrics, woven, nes
6852	-0.45	522	Lead and lead alloys, worked
4245	-0.46	523	Castor oil
0224	-0.46	524	Milk and cream, preserved, concentrated or sweetened
6553	-0.46	525	Knitted or crocheted fabrics, elastic or rubberized
6123	-0.48	526	Parts of footwear of any material except metal and asbestos
5224	-0.48	527	Metallic oxides of zinc, iron, lead, chromium etc

SITC4	PCI	RANK	Description
6122	-0.48	528	Saddlery and harness, of any material, for any kind of animal
0722	-0.49	529	Cocoa powder, unsweetened
7521	-0.50	530	Analogue and hybrid data processing machines
7131	-0.50	531	Internal combustion piston engines, for aircraft, and parts, nes
6672	-0.51	532	Diamonds (non-industrial), not mounted or set
6732	-0.51	533	Bars, rods (not wire rod), from iron or steel; hollow mining drill
2742	-0.52	534	Iron pyrites, unroasted
6624	-0.52	535	Non-refractory ceramic bricks, tiles, pipes and similar products
2786	-0.53	536	Slag, scalings, dross and similar waste, nes
6534	-0.53	537	Fabrics, woven, less 85% of discontinuous synthetic fibres
0412	-0.54	538	Other wheat and meslin, unmilled
6671	-0.54	539	Pearls, not mounted, set or strung
2460	-0.54	540	Pulpwood (including chips and wood waste)
6725	-0.54	541	Blooms, billets, slabs and sheet bars, of iron or steel
2222	-0.55	542	Soya beans
8433	-0.55	543	Womens, girls, infants outerwear, textile, not knitted or crocheted; dresses
6575	-0.55	544	Twine, cordage, ropes and cables and manufactures thereof
0541	-0.56	545	Potatoes, fresh or chilled, excluding sweet potatoes
7932	-0.57	546	Ships, boats and other vessels
7933	-0.57	547	Ships, boats and other vessels for breaking up
2234	-0.57	548	Linseed
0011	-0.58	549	Animals of the bovine species (including buffaloes), live
6596	-0.58	550	Carpets, rugs, mats, of other textile materials, nes
2239	-0.59	551	Flour or meals of oil seeds or oleaginous fruit, non-defatted
5622	-0.59	552	Mineral or chemical fertilizers, phosphatic
3354	-0.59	553	Petroleum bitumen, petroleum coke and bituminous mixtures, nes
6613	-0.59	554	Building and monumental stone, worked, and articles thereof
0116	-0.61	555	Edible offal of headings 0011-5 and 0015, fresh, chilled or frozen
6515	-0.61	556	Yarn containing 85% or more of synthetic fibres, put up for retail
2925	-0.62	557	Seeds, fruits and spores, nes, for planting
8933	-0.62	558	Personal adornments and ornaments articles of plastic
2782	-0.63	559	Clay and other refractory minerals, nes
5721	-0.63	560	Propellent powders and other prepared explosives
5629	-0.64	561	Fertilizers, nes
8465	-0.66	562	Corsets, garters, etc, not knitted or crocheted, elastic or not
6354	-0.66	563	Manufactures of wood for domestic or decorative use
0546	-0.66	564	Vegetables, frozen or in temporary preservative
8461	-0.67	565	Under-garments, knitted or crocheted; of wool or fine animal hair, not elastic nor rubberized
8421	-0.67	566	Men's and boys' outerwear, textile fabrics not knitted or crocheted; overcoats and other coats
6821	-0.68	567	Copper and copper alloys, refined or not, unwrought
1222	-0.69	568	Cigarettes
2481	-0.70	569	Railway or tramway sleepers (ties) of wood
6871	-0.70	570	Tin and tin alloys, unwrought
8424	-0.71	571	Men's and boys' outerwear, textile fabrics not knitted or crocheted; jackets, blazers and the like
7233	-0.72	572	Road rollers, mechanically propelled

SITC4	PCI	RANK	Description
4249	-0.73	573	Fixed vegetable oils, nes
8431	-0.73	574	Womens, girls, infants outerwear, textile, not knitted or crocheted; coats and jackets
0411	-0.74	575	Durum wheat, unmilled
5241	-0.75	576	Radio-active chemical elements, isotopes etc
2919	-0.75	577	Other materials of animal origin, nes
6341	-0.76	578	Wood sawn lengthwise, veneer sheets etc, up to 5 mm in thickness
0470	-0.76	579	Other cereal meals and flour
8484	-0.77	580	Headgear and fitting thereof, nes
0565	-0.77	581	Vegetables, prepared or preserved, nes
6932	-0.77	582	Barbed iron or steel wire: fencing wire
2771	-0.77	583	Industrial diamonds
2732	-0.78	584	Gypsum, plasters, limestone flux and calcareous stone
0585	-0.79	585	Fruit or vegetable juices
6716	-0.79	586	Ferro-alloys
6576	-0.80	587	Hat shapes, hat-forms, hat bodies and hoods
2890	-0.81	588	Ores and concentrates of precious metals, waste, scrap
2783	-0.82	589	Common salt; pure sodium chloride; salt liquors; sea water
0575	-0.82	590	Grapes, fresh or dried
0914	-0.82	591	Margarine, imitation lard and other prepared edible fats, nes
6899	-0.84	592	Base metals, nes and cermets, unwrought (including waste and scrap)
2440	-0.84	593	Cork, natural, raw and waste
4312	-0.84	594	Hydrogenated animal or vegetable oils and fats
2882	-0.85	595	Other non-ferrous base metal waste and scrap, nes
6582	-0.85	596	Tarpaulins, sails, tents, camping goods, etc, of textile fabrics
6583	-0.85	597	Travelling rugs, blankets (non electric), not knitted or crocheted
0561	-0.85	598	Vegetables (excluding leguminous), dried, evaporated, etc
2733	-0.85	599	Sands, excluding metal-bearing sands
6511	-0.86	600	Silk yarn and spun from noil or waste; silkworm gut
8471	-0.86	601	Clothing accessories, of textile fabrics, not knitted or crocheted
8434	-0.87	602	Womens, girls, infants outerwear, textile, not knitted or crocheted; skirts
8510	-0.87	603	Footwear
8464	-0.87	604	Under-garments, knitted or crocheted; of other fibres, not elastic nor rubberized
6612	-0.87	605	Cement
0589	-0.88	606	Fruit prepared or preserved, nes
2923	-0.88	607	Vegetable plaiting materials
8432	-0.88	608	Womens, girls, infants outerwear, textile, not knitted or crocheted; suits and costumes
0343	-0.88	609	Fish fillets, fresh or chilled
8422	-0.89	610	Men's and boys' outerwear, textile fabrics not knitted or crocheted; suits
2111	-0.89	611	Bovine and equine hides, raw, whether or not split
2814	-0.89	612	Roasted iron pyrites
8482	-0.89	613	Articles of apparel, clothing accessories of plastic or rubber
6712	-0.90	614	Pig iron, cast iron, spiegeleisen, in pigs, blocks, lumps, etc
0544	-0.90	615	Tomatoes, fresh or chilled
4232	-0.90	616	Soya bean oil
0564	-0.90	617	Flour, meals and flakes of potatoes, fruit and vegetables, nes
5121	-0.91	618	Acyclic alcohols, and their derivatives
8472	-0.92	619	Clothing accessories, knitted or crocheted, nes

SITC4	PCI	RANK	Description
6344	-0.93	620	Wood-based panels, nes
0811	-0.94	621	Hay and fodder, green or dry
0571	-0.94	622	Oranges, mandarins, etc, fresh or dried
4111	-0.95	623	Fat and oils of fish and marine mammals
0612	-0.95	624	Refined sugar etc
5621	-0.95	625	Mineral or chemical fertilizers, nitrogenous
6584	-0.97	626	Linens and furnishing articles of textile, not knitted or crocheted
0572	-0.97	627	Other citrus fruits, fresh or dried
7938	-0.99	628	Tugs, special purpose vessels and floating structures
2479	-0.99	629	Pitprops, poles, piling, post and other wood in the rough, nes
2875	-1.00	630	Zinc ores and concentrates
4236	-1.01	631	Sunflower seed oil
2659	-1.03	632	Vegetable textile fibres, nes, and waste
2820	-1.04	633	Waste and scrap metal of iron or steel
2685	-1.04	634	Horsehair and other coarse animal hair, not carded or combed
8463	-1.04	635	Under-garments, knitted or crocheted; of synthetic fibres not elastic nor rubberized
6851	-1.05	636	Lead, and lead alloys, unwrought
8435	-1.06	637	Womens, girls, infants outerwear, textile, not knitted or crocheted; blouses
8452	-1.07	638	Outerwear knitted or crocheted, not elastic nor rubberized; womens, girls, infants, suits, dresses, etc, knitted, crocheted
2683	-1.07	639	Fine animal hair, not carded or combed
8429	-1.09	640	Men's and boys' outerwear, textile fabrics not knitted or crocheted; other outer garments
4234	-1.10	641	Groundnut (peanut) oil
0460	-1.10	642	Meal and flour of wheat and flour of meslin
8443	-1.10	643	Under garments of textile fabrics, not knitted or crocheted; womens, girls, infants under garments, textile, not knitted, etc
0813	-1.11	644	Oilcake and other residues (except dregs)
2785	-1.11	645	Quartz, mica, felspar, fluorspar, cryolite and chiolite
334	-1.12	646	Petroleum products, refined
5541	-1.12	647	Soaps, organic products and preparations for use as soap
0341	-1.12	648	Fish, fresh or chilled, excluding fillet
2927	-1.12	649	Cut flowers and foliage
6593	-1.13	650	Kelem, Schumacks and Karamanie rugs and the like
2789	-1.13	651	Minerals, crude, nes
6521	-1.14	652	Cotton fabrics, woven, unbleached, not mercerized
0616	-1.14	653	Natural honey
0344	-1.14	654	Fish fillets, frozen
3221	-1.15	655	Anthracite, not agglomerated
7921	-1.16	656	Helicopters
0483	-1.17	657	Macaroni, spaghetti and similar products
6513	-1.17	658	Cotton yarn
2450	-1.17	659	Fuel wood and wood charcoal
2651	-1.18	660	Flax and ramie, flax tow, ramie noils, and waste
6673	-1.18	661	Precious and semi-precious stones, not mounted, set or strung
8441	-1.18	662	Under garments of textile fabrics, not knitted or crocheted; mens and boys shirts
5513	-1.18	663	Essential oil, resinoid, etc
2690	-1.18	664	Old clothing and other old textile articles; rags

SITC4	PCI	RANK	Description
2911	-1.19	665	Bones, ivory, horns, coral, shells and similar products
6589	-1.19	666	Other made-up articles of textile materials, nes
8999	-1.20	667	Manufactured goods, nes
2816	-1.21	668	Iron ore agglomerates
2116	-1.21	669	Sheep and lamb skin with the wool on, raw, whether or not split
3223	-1.21	670	Lignite, not agglomerated
2874	-1.22	671	Lead ores and concentrates
8423	-1.22	672	Men's and boys' outerwear, textile fabrics not knitted or crocheted; trousers, breeches and the like
2872	-1.22	673	Nickel ores and concentrates; nickel mattes, etc
6114	-1.22	674	Leather of other bovine cattle and equine leather
6522	-1.22	675	Cotton fabrics, woven, bleached, dyed, etc, or otherwise finished
2654	-1.23	676	Sisal, agave fibres, raw or processed but not spun, and waste
2682	-1.23	677	Wool degreased, uncombed of sheep or lambs
6713	-1.23	678	Iron and steel powders, shot or sponge
2112	-1.24	679	Calf skins, raw, whether or not split
0371	-1.24	680	Fish, prepared or preserved, nes
0112	-1.24	681	Meat of sheep and goats, fresh, chilled or frozen
6342	-1.24	682	Plywood consisting solely of sheets of wood
0723	-1.25	683	Cocoa butter and paste
3414	-1.27	684	Petroleum gases, nes, in gaseous state
0814	-1.27	685	Flours and meals, of meat, fish,etc, unfit for human; greaves
8442	-1.27	686	Under garments of textile fabrics, not knitted or crocheted; mens, boys under garments; other than shirts
8439	-1.28	687	Womens, girls, infants outerwear, textile, not knitted or crocheted; other outer garments of textile fabrics, not knitted, crocheted
6597	-1.29	688	Plaits, plaited products for all uses; straw envelopes for bottles
2613	-1.29	689	Raw silk (not thrown)
5225	-1.30	690	Inorganic bases and metallic oxides, hydroxides and peroxides
2238	-1.31	691	Oil seeds and oleaginous fruits, nes
6841	-1.31	692	Aluminium and aluminium alloys, unwrought
0440	-1.33	693	Maize, unmilled
2119	-1.33	694	Hides and skins, nes; waste and used leather
0012	-1.34	695	Sheep and goats, live
2731	-1.35	696	Building and monumental (dimension) stone, roughly squared, split
8462	-1.36	697	Under-garments, knitted or crocheted; of cotton, not elastic nor rubberized
0342	-1.36	698	Fish, frozen, excluding fillets
2687	-1.37	699	Sheep's or lambs' wool, or of other animal hair, carded or combed
8459	-1.38	700	Outerwear knitted or crocheted, not elastic nor rubberized; other, clothing accessories, non-elastic, knitted or crocheted
2713	-1.38	701	Natural calcium phosphates, natural aluminium, etc
8451	-1.39	702	Outerwear knitted or crocheted, not elastic nor rubberized; jerseys, pullovers, slip-overs, cardigans, etc
0579	-1.42	703	Fruit, fresh or dried, nes
2929	-1.42	704	Other materials of vegetable origin, nes
6581	-1.42	705	Bags, sacks of textile materials, for the packing of goods
0752	-1.43	706	Spices, except pepper and pimento
2655	-1.43	707	Manila hemp, raw or processed but not spun, its tow and waste

SITC4	PCI	RANK	Description
9410	-1.44	708	Animals, live, nes, (including zoo animals, pets, insects, etc)
3222	-1.44	709	Other coal, not agglomerated
6592	-1.44	710	Carpets, carpeting and rugs, knotted
4244	-1.45	711	Palm kernel oil
0545	-1.47	712	Other fresh or chilled vegetables
0372	-1.47	713	Crustaceans and molluscs, prepared or prepared, nes
0615	-1.49	714	Molasses
0350	-1.51	715	Fish, dried, salted or in brine; smoked fish
0459	-1.54	716	Buckwheat, millet, etc, and other cereals, unmilled, nes
1213	-1.55	717	Tobacco refuse
4233	-1.56	718	Cotton seed oil
2871	-1.56	719	Copper ore and concentrates; copper matte; cement copper
0421	-1.57	720	Rice in the husk or husked, but not farther prepared
2232	-1.57	721	Palm nuts and kernels
0751	-1.58	722	Pepper of "piper"; pimento of "capsicum or pimenta"
1212	-1.58	723	Tobacco, wholly or partly stripped
0360	-1.59	724	Crustaceans and molluscs, fresh, chilled, frozen, salted, etc
2924	-1.61	725	Plants and parts of trees used in perfumery; in pharmacy; etc
2231	-1.63	726	Copra
4243	-1.64	727	Coconut (copra) oil
0742	-1.64	728	Mate
2483	-1.68	729	Wood, non-coniferous species, sawn, planed, tongued, grooved, etc
2873	-1.68	730	Aluminium ores and concentrates (including alumina)
1211	-1.68	731	Tobacco, not stripped
2235	-1.69	732	Castor oil seeds
2879	-1.70	733	Ores and concentrates of other non-ferrous base metals
0711	-1.71	734	Coffee green, roasted; coffee substitutes containing coffee
2681	-1.73	735	Wool greasy or fleece-washed of sheep or lambs
6116	-1.74	736	Leather of other hides or skins
2632	-1.74	737	Cotton linters
0548	-1.75	738	Vegetable products roots and tubers, nes, fresh, dried
4242	-1.75	739	Palm oil
0812	-1.76	740	Bran, sharps and other residues derives of cereals
9710	-1.78	741	Gold, non-monetary (excluding gold ores and concentrates)
2634	-1.79	742	Cotton, carded or combed
2633	-1.79	743	Cotton waste, not carded or combed
6545	-1.80	744	Fabrics, woven of jute or other textile bast fibres of heading 2640
0542	-1.82	745	Beans, peas, other leguminous vegetables, dried, shelled
4314	-1.82	746	Waxes of animal or vegetable origin
2117	-1.83	747	Sheep and lamb skin without the wool, raw, whether or not split
2472	-1.85	748	Sawlogs and veneer logs, of non-coniferous species
2784	-1.87	749	Asbestos
2741	-1.90	750	Sulphur (other than sublimed, precipitated or colloidal)
0741	-1.90	751	Tea
0611	-1.90	752	Sugars, beet and cane, raw, solid
6115	-1.95	753	Sheep and lamb skin leather
0422	-1.98	754	Rice, semi-milled or wholly milled
2922	-2.03	755	Natural gums, resins, lacs and balsams

Examples of Complexity Rankings

SITC4	PCI	RANK	Description
2815	-2.04	756	Iron ore and concentrates, not agglomerated
0577	-2.05	757	Nuts edible, fresh or dried
2223	-2.08	758	Cotton seeds
2640	-2.09	759	Jute, other textile bast fibres, nes, raw, processed but not spun
2221	-2.14	760	Groundnuts, green
0573	-2.21	761	Banana, plantain, fresh or dried
2860	-2.22	762	Ores and concentrates of uranium and thorium
2877	-2.35	763	Manganese ore and concentrates
3413	-2.37	764	Petroleum gases and other gaseous hydrocarbons, nes, liquefied
2631	-2.38	765	Raw cotton, excluding linters, not carded or combed
2876	-2.69	766	Tin ores and concentrates
2225	-2.73	767	Sesame seeds
0721	-2.86	768	Cocoa beans, raw, roasted
2320	-2.96	769	Natural rubber latex; natural rubber and gums
2114	-3.01	770	Goat and kid skins, raw, whether or not split
3330	-3.02	771	Crude petroleum and oils obtained from bituminous materials

Table C.3: Sample of proximity values for the Years 2000 and 2012

Product 1	**Product 2**	**2000**	**2012**
Men's & Boys Trousers (8423)	Men's and Boys Shirts (8441)	0.81	0.75
	Jersey's / Cardigans (8451)	0.77	0.69
	Linens / Textile Coverings (6584)	0.36	0.48
	Motor Vehicles (7810)	0.09	0.06
	Commercial Aircraft (7924)	0.06	0.04
Coffee (Green, Roasted) (0711)	Sugars, Raw (0611)	0.5	0.37
	Cut Flowers (2927)	0.45	0.23
	Jersey's / Cardigans (8451)	0.33	0.32
	Motor Vehicles (7810)	0.03	0.05

APPENDIX D A COUNTRY SPACE

Given the M_{cp} matrix is the starting point of the product space, an alternative proximity matrix ($\phi_{pp'}$) can be compiled that represents country similarity ($\phi_{cc'}$). By employing the same proximity metric, similarity scores can be compiled regarding the overlap of country export vectors conditioned on the maximum overall level of diversification. Rather than considering all column permutations, the row permutations are now considered to see what information this network approach reveals at the country level. Considering that the fundamental starting point is the same, and therefore contains the same initial information, it is reasonable to expect a similar story, with a densely connected core structure represented by relatively developed countries, and developing countries to occupy the periphery. Applying the same technique over countries, rather than products:

$$\phi_{c_1c_2} = \frac{\sum_p M_{c_1p} \bullet M_{c_2p}}{max(k_{c_1,0}, k_{c_2,0})}$$

where,

k_{c_1} = Diversity of Country 1

k_{c_2} = Diversity of Country 2

A sparse square country proximity matrix is computed that contains comparisons between all country pairs. Using the same data as was used in Chapter 3 (Section 3.2.2), this analysis was undertaken to demonstrate which countries are most similar in terms of overall exports, in addition to how countries cluster together using this analytical technique. For consistency the year 2000 is considered.[127]

The country proximity matrix is notably sparser with lower average similarity scores, when compared to the product proximity matrix. There is less similarity in overlaps, at the country level, when considering export vectors. For the year 2000, the largest similarity score obtained is between Austria (AUT) and the Czech Republic (CZE). Austria has a diversity score of 296 exports, and the Czech Republic 298 exports. They also jointly export 173 of the same products. This evidence suggests Austria and the

[127] Similar network diagrams are computed when using other year cross-sections.

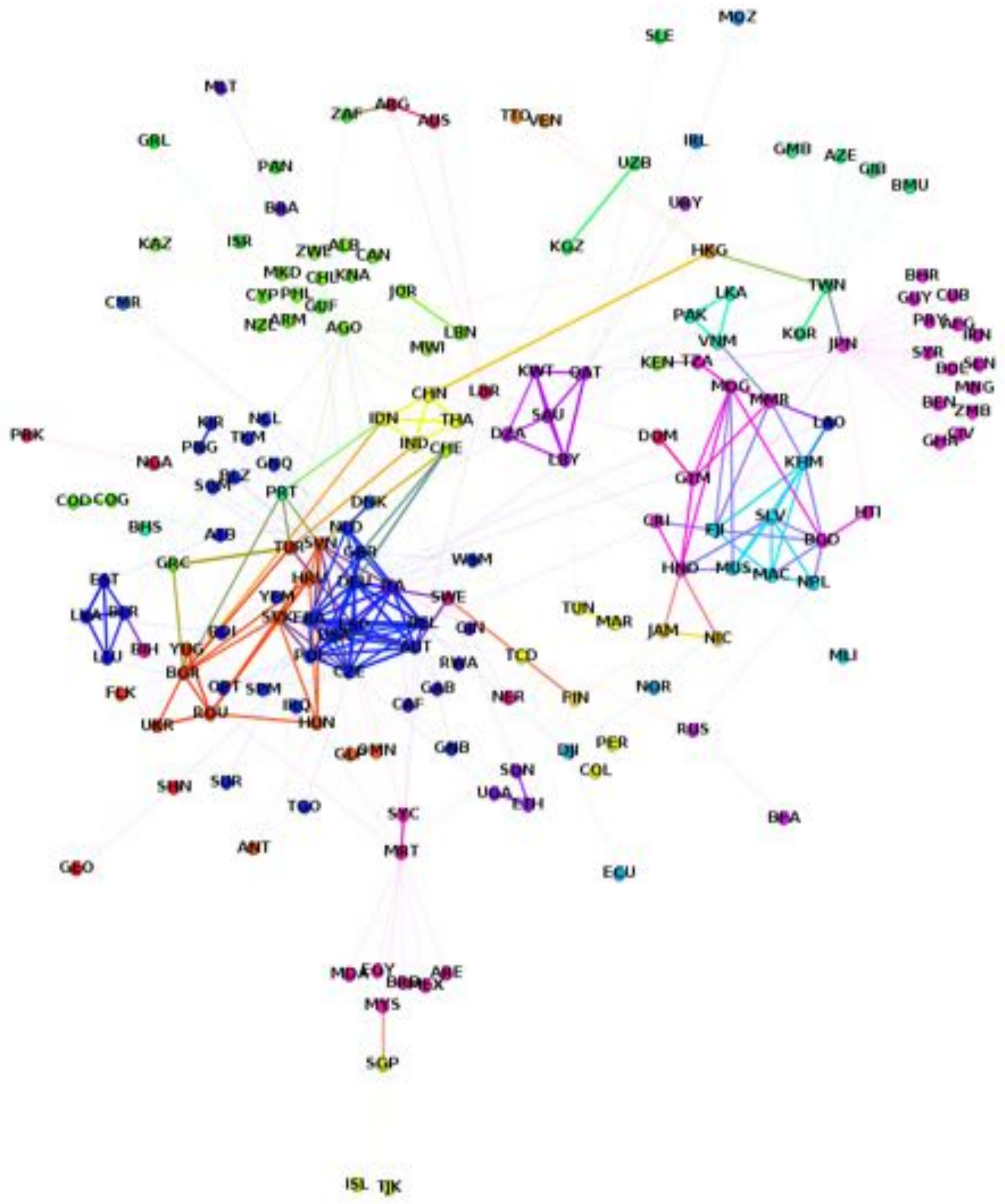
Czech Republic have similar production capabilities given the large overlap in exported products, in addition to exhibiting similar levels of diversification.

To conduct a network analysis of the country relationships, a similar methodology was applied by computing a minimum spanning tree. To ensure the network remains fully connected, in addition to adding high value proximity relationships that were greater than 0.4. The lower cut-off value, compared to that used in the product proximity found in the literature, is a direct result of lower average proximity scores over the country dimension. A force-directed network layout algorithm, ForceAtlas2, was used to produce a network diagram (Jacomy et al. 2014). The way these algorithms fundamentally work is where nodes repulse and edges attract in a gravitational sense, which leads to strongly connected groups of nodes to form clusters while weakly connected nodes move away from the centre of the graph. A Markov clustering algorithm (MCL) was then used to identify clusters of countries within the graph (van Dongen 2000). This algorithm was selected because it is similar in spirit to that used in the literature for product space analysis.[128] The product space computes random walks through the graph and statistically translates these into neighbourhoods.

The results of this analysis are shown in Figure D.1. As expected a number of industrialised developed economies cluster together to form a tightly held group as identified in blue in the core of the network. This cluster, defined solely by the network structure, consists of the Czech Republic, France, Germany, Great Britain, Italy, Poland, Spain, and the United States. Other interesting clusters are also identified that paint a picture of the importance of geography. Some less developed economies are grouped together (Ukraine, Romania, Hungary, and Slovakia) that are geographically located in close proximity to Europe are identified as a separate cluster that contains some links into the industrial core through countries such as Turkey and Slovenia. Within Asia, China (CHN), India (IND), Indonesia (IDN), and Thailand (THA) are identified as similar and are all successful exporters.

[128] This algorithm is also available in the GEPHII marketplace as a plugin. This greatly simplifies exploration of clusters as the code has already been implemented.

Figure D.1: Country Proximity Network Diagram



Notes: 1. A country space network diagram, which uses the force atlas to reveal clustering structures observed in the data. **2.** Data used are world export data for the year 2000. **3.** The network layout is achieved using the ForceAtlas2 algorithm (Jacomy et al. 2014) with minor manual adjustments to prevent node overlap. **4.** The observed clusters have been generated using a Markov Clustering (MCL) with default parameters (power parameter = 2, inflation parameter = 2, pruning parameter = 0.02) (van Dongen 2000) using Gephi software.

Notably China exhibits strong connections with Hong Kong (HKG) but they are not classified in the same group, instead Taiwan (TWN), Korea (KOR), Hong Kong (HKG), and Japan (JPN) exhibit relatively closer ties as their export compositions are more similar to one another than that of China.

While many clusters are aligned geographically, this is not the always the case. The resource-rich oil exporting countries are clearly identified as a cluster as they have similar export vectors despite some of the members being located in different regions. These countries include Saudi Arabia, Kuwait, Qatar, Algeria, and Libya. Another interesting case is that of Australia. Its closest similarity score is to Argentina, and indirectly South Africa. These countries are similar with respect to resources, agriculture, and light manufactures; despite Argentina's GDP per capita ranking 64th compared to Australia's 10th in PPP-adjusted terms. Export profiles do not solely determine income levels and other economic considerations remain important such as institutional quality and other non-physical sectors such as services. However, Figure D.1 does a good job of characterising countries' underlying productive similarities at the country level.[129]

These results are broadly similar to those found by Tacchella et al. (2012), who characterise country similarities using a different similarity matrix. However the underlying similarity measure (analogues to the product space approach) and the clustering algorithm used here identifies sensible within-region clusters, such as JPN-HKG-TWN compared to CHN-THA-IND-IDN, and exhibits less emphasis on geography, pairing Australia with Argentina rather than New Zealand (cf. Tacchella et al., 2012). The geographic component to both studies remains interesting, as it appears the similar geographic neighbourhoods tend to produce similar exports. This could be a result of shared history, similar cultural ties and values, transportation costs, spillovers, and product fragmentation.

[129] Peripheral countries are not necessarily similar to the nodes that they connect to. They are mainly countries that are dissimilar to all other countries in the world but remain connected due to the use of a minimum spanning tree. For example, Syria, Zimbabwe, and Iran are not similar to Japan's export profile. These are countries that are weakly connected to the entire graph and therefore end up in the periphery, and do not share strong links with Japan. The minimum spanning tree algorithm as just severed all of the other weak links to other countries.

The network approach, introduced by the product space, is more widely applicable then just the application to products. Equally interesting are similarity measures considered over the country dimensions. The main benefit of considering the product dimension is that it is more disaggregated and intuitive when understanding capabilities that are required in the production of any two physical products. Country information can also be overlayed with the product space representation to understand country positions in the network.

APPENDIX E A SERIATION APPROACH

As highlighted in Chapter 2, the *method of reflections* can be viewed as a sorting heuristic aligning countries and products along the diagonal line of specialisation. Another technique, used in a variety of fields such as computational archaeology, biology and manufacturing is that of seriation. Its primary object is to reorder matrices by rows and columns to maximize in a way that infers some constrained structure based on the relational data contained in the matrix (Liiv 2010). Anthropology uses it for sequence dating and biology uses it to cluster gene expression. The connection with this thesis is that it offers an analytical technique that fits with the aim to reorder the M_{cp} matrix in line with specialisation. This is in effect a Robinson matrix, a matrix that has the highest values in the matrix along the matrix with monotonically decreasing values as one moves away from the diagonal.

This section demonstrates how well the ECI and PCI approach is at uncovering this underlying structure, within each matrix, as a sorting heuristic. It is faster and produces better results than the simple matrix reordering (seriation) approach considered here: travelling salesman algorithm (TSP) and evaluated using the Bond Energy Algorithm (BEA).

In the context of this thesis, seriation is used as a comparative tool with the results obtained from the method of reflections. This technique was used to reorder M_{cp} {0,1} binary matrices and M_{cp} {RCA} RCA intensity matrices in an effort to maximize their density along the quasi-diagonal of the rectangular country-product matrix. There are two approaches usually considered in the literature. First is two-way two-mode data in which patterns in the data is jointly determined by row and column positions. Second is two-way one-mode data, which is used to organise similarity matrices such as the Mpp and Mcc proximity matrices, discussed in Chapter 2. The focus here is on an algorithm that is used for two-way two-mode data.

The main barrier to seriation is that they are computationally difficult problems to solve.[130] This is due to the nature of the problem in solving the combinatorial

[130] In the computer science literature this type of problem is know as NP-hard problems. This refers to the curse of dimensionality and brute-force solvers are

optimization problem. Suppose you have a small relational matrix that contains 5 rows and 10 columns. To compare all possible combinations in matrices that can be defined by permuting rows and columns amounts to 5! = 120 row combinations and 10! =3,628,800 column combinations. Therefore there are a total of 5! x 10! = 399,168,000 combinations jointly determined by all possible row and column positions. While computational feasibility is determined by available computing capacity it is generally advised that brute force combinatorial methods cannot be applied except in very small dimensional cases to find an optimal solution. There is however a rich literature discussing alternative approaches such as partial enumeration methods (Hubert, Arabie & Meulman 1987; Brusco & Stahl 2001), traveling salesperson solvers (Gutin & Punnen 2002), various clustering simplifications and heuristic approaches. The differences between these approaches are largely between the trade-off between optimality of solution and computational requirements.

Heuristics on the other hand make simplifications, typically transforming discrete combinatorial optimization methods into continuous counterparts and using these relationships to compute, in some cases, a 'close-approximate' to an optimal solution. The downside of heuristics is that you are never fully confident that you will have a global optimal – so typically these methods are iterated many times in search of a global optimum.

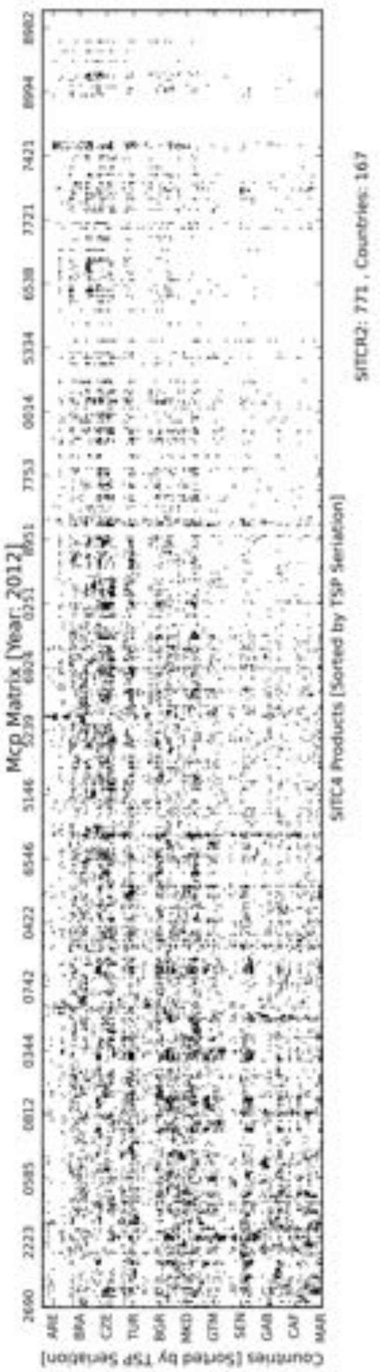
Two-Way Two-Mode Analysis

First the full two-way two-mode matrices are exploring using the often observed rectangular matrices mapping countries and products. The binary matrix M_{cp} will be considered using the bond energy algorithm (BE), which is a simple heuristic approach that each cell in the matrix is surrounded by neighbours that are as closely related to it as possible (McCormick, Schweitzer & White 1972). Given the dependence only on local neighbourhoods, this method was iterated for a million trials and the ordering with the largest clustering of relationships around the diagonal was recorded. The results are shown in Figure D.1 and Figure D.2.

---

typically not available due to the high dimensional nature of combinatorial optimization (in this case).

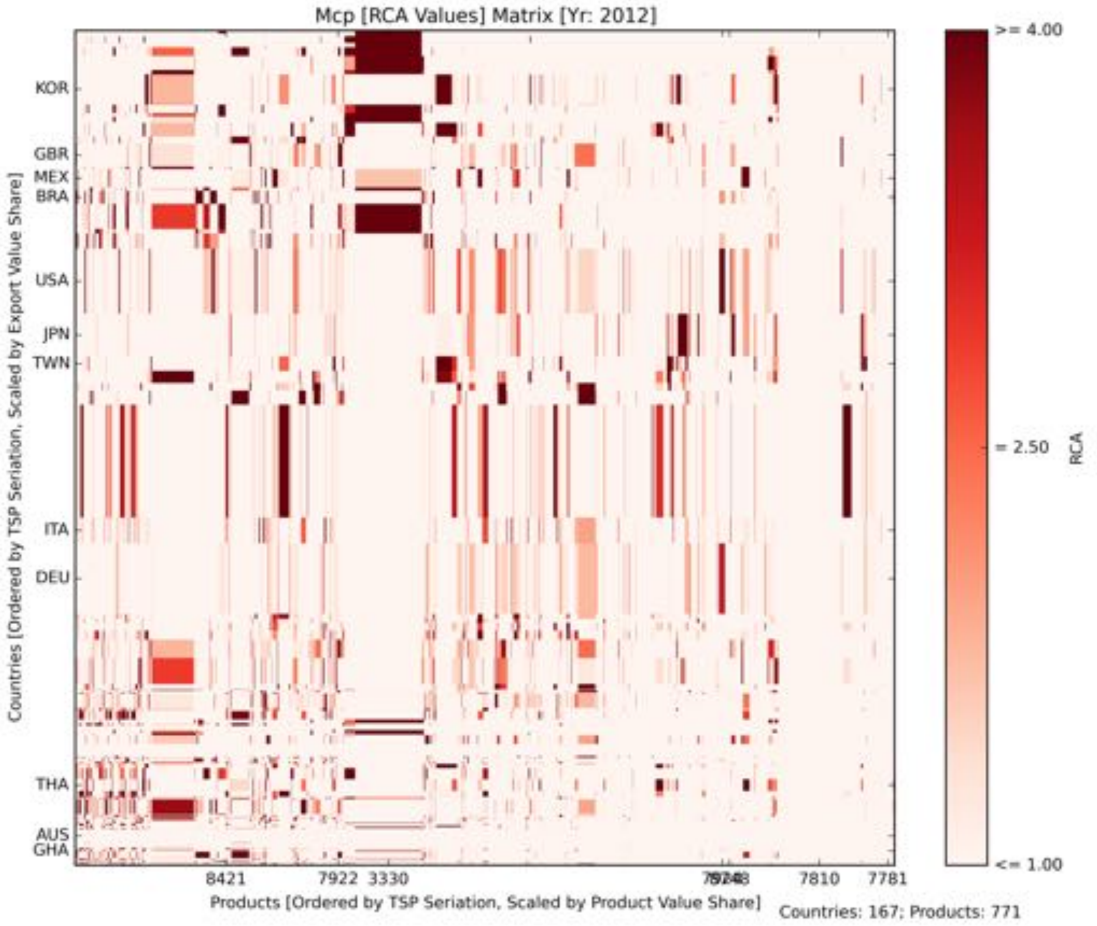
Figure E.1: Seriation TSP Algorithm [Year: 2012]



Notes: 1. This ordering of countries and products was determined using a seriation algorithm TSP using the software package R. This algorithm particularly focuses on local neighbourhoods when solving the dynamic programming problem and therefore products clusters of relationships but performs badly at producing an overall macro sorting pattern. **2.** This algorithm was run for 100000 iterations looking for a maximal density around the diagonal line of specialisation. **3.** This seriation was performed on the binary M_{cp} matrix.

This exercise demonstrates how good the ECI and PCI method is as a sorting heuristic.

Figure E.2: Seriation TSP Algorithm applied to Scaled RCA Heatmap for the Year 2012



Notes: 1. This diagram shows how ineffective the simple BEA TSP algorithm is at maximising information along the diagonal of the M_{cp} matrix. The intensity of RCA is superimposed on the matrix. **2.** This diagram is constructed using data generated by the same BEA TSP algorithm in Figure E.1

E.1. CONCLUSIONS

While the results presented here are used to show that ECI and PCI are relatively good sorting heuristics in which products and countries are sorted in line with a macro pattern of specialisation. This is compared with the BEA algorithm, which considers local patterns of neighbourhood similarities in the constrained optimisation problem as an alternative sorting heuristic. Given sparse data this heuristic works relatively well as local neighbourhoods chain together to form diagonalised patterns in the matrix. In this context, the matrix has a triangular density to the country-product assignments. This generates blocks of clustered countries and products, but fails to identify a solid macro-pattern around the diagonal.

E.1.1. FUTURE WORK

This exercise made use of the bond energy algorithm (BEA) as a sorting heuristic. Given the dimensionality of the problem – heuristic approaches are the only analytically tractable way forward. However more complex strategies could be investigated that incorporates higher orders of dependence that is observed in the matrix patterns to produce alternative rankings of country and product complexity indicators by finding the ordering that maximises export values around the diagonal line of specialisation.

Another estimation strategy that could be employed is to decouple the problem and perform seriation on the country and product dimensions separately through the use of a similarity matrix (such as proximity in the case of products).

APPENDIX F ROBUSTNESS RESULTS FOR PRODUCT EMERGENCE AND TRADE LIBERALISATION REGRESSIONS

This appendix includes the country definition tables found in the regression datasets used in Chapter 4, in addition to alternative regression results used for robustness checks. The robustness checks are mainly composed of various restricted datasets by considering the NBER dataset without the extension from 2000 to 2012, in addition to a subset of the NBER dataset from 1984 to 2000, which is a period containing the most cleanly defined product classification in Feenstra et al. (2005).

F.1. COUNTRY TABLES

Countries that are included in the ALL COUNTRY regression datasets are defined in Table F.1, and countries included in the DEVELOPING ONLY regression datasets are defined in Table F.2

Table F.1: Countries included in All Country Regression Dataset

All Country Datasets [Countries: 129]												
AGO	BFA	CHL	DNK	GAB	HKG	ISL	LKA	MOZ	NPL	RUS	TCD	UKR
ALB	BGD	CHN	DOM	GBR	HND	ISR	LTU	MRT	NZL	RWA	TGO	URY
ARG	BGR	CIV	DZA	GEO	HRV	ITA	LVA	MUS	PAK	SEN	THA	USA
ARM	BLR	CMR	ECU	GHA	HTI	JOR	MAR	MWI	PAN	SGP	TJK	UZB
AUS	BOL	COG	EGY	GIN	HUN	JPN	MDA	MYS	PER	SLE	TKM	VEN
AUT	BRA	COL	ESP	GMB	IDN	KAZ	MDG	NER	PHL	SLV	TTO	YEM
AZE	BRB	CRI	EST	GNB	IND	KEN	MEX	NGA	PNG	SVK	TUN	ZAF
BDI	CAF	CYP	ETH	GRC	IRL	KGZ	MKD	NIC	POL	SVN	TUR	ZMB
BEL	CAN	CZE	FIN	GTM	IRN	KOR	MLI	NLD	PRT	SWE	TZA	ZWE
BEN	CHE	DEU	FRA	GUY	IRQ	LBR	MLT	NOR	PRY	SYR	UGA	

Table F.2: Countries included in Developing Country Regression Datasets

Developing Country Datasets [Countries: 107]										
AGO	BLR	COL	GAB	HRV	KAZ	MDG	NER	RUS	TGO	URY
ALB	BOL	CRI	GEO	HTI	KEN	MEX	NGA	RWA	THA	UZB
ARG	BRA	CYP	GHA	HUN	KGZ	MKD	NIC	SEN	TJK	VEN
ARM	BRB	CZE	GIN	IDN	KOR	MLI	NPL	SGP	TKM	YEM
AZE	CAF	DOM	GMB	IND	LBR	MLT	PAK	SLE	TTO	ZAF
BDI	CHL	DZA	GNB	IRN	LKA	MOZ	PAN	SLV	TUN	ZMB
BEN	CHN	ECU	GTM	IRQ	LTU	MRT	PER	SVK	TUR	ZWE
BFA	CIV	EGY	GUY	ISR	LVA	MUS	PHL	SVN	TZA	
BGD	CMR	EST	HKG	JOR	MAR	MWI	PNG	SYR	UGA	
BGR	COG	ETH	HND	JPN	MDA	MYS	PRY	TCD	UKR	

Notes: 1. The set of developing countries include some current developed countries in the Asia region (i.e. Japan – JPN). This is because the list was constructed based on 1962 income levels. **2.** An alternative specification was considered by removing JPN, SGP, and KOR. Using dataset 'D' the results remain similar to those reported – except that the standard error is reduced to the 5% level for probable product emergence. Less reduction in the standard error is found for probable and persistent products.

F.2. ALTERNATIVE DATASETS

Given the dataset used in Chapter 4 is a composite dataset that has been harmonized to extend the NBER world trade flow datasets, an alternative specification has been considered on the reduced NBER dataset only. This robustness checks the implications of averaging over the years 1998 to 2000, in addition to checking the status quo assumption that trade liberalisation based on the pattern set in 2000 when extending the dataset from 2000 to 2012.[131]

The NBER only all country results are found in Table F.3. While strong evidence persists for datasets C and D, weaker and insignificant (in the case of all countries) results are found for dataset E in this reduced dataset. This is of concern given dataset E is computed to consider a more consistent intertemporal product definition. This estimate remains weakly significant in the sample that considers developing countries only (as shown in Table F.4). This could be indicative of the possibility that product re-classifications in dataset 'C' and 'D' may be causing an increase in the number of overall product emergence events which, by chance, happen to coincide with typical trade liberalization episodes. Alternatively the reduction in product codes greatly

[131] This status quo assumption extends the trade liberalization indicator from 2000 to 2012 based on pattern observed in 2000. It is suggested that country trade liberalisation states have not changed significantly in 2000 to 2012.

reduces the opportunity for product emergence, which captures less emergent product variation.

The first concern should be addressed with the inclusion of year fixed effects in all regression specifications. Inspecting the estimated coefficient on the year fixed effects for 1974 and 1984 (the product classification boundary years) found that they are larger and more significant than in any other year periods.[132] Therefore any year specific shocks in product emergence are controlled for. As an additional robustness check to ensure these boundary affects were not driving the results a restricted sample between 1984 and 2000, where no product reclassification boundaries are observed, was examined. The results are contained in Table F.5 and discussed in the next section. Despite the restriction in years, the results remain significant for probable and persistent emergent exports.

The second possibility is that the underlying number of reduced products, in the product level export dataset, does not provide sufficient product variation to compute as meaningful emergence statistics. The algorithmic approach to aggregation may aggregate away much of the within cluster product emergence dynamic. To check for potential aggregation issues introduced in the construction of dataset E, an additional robustness check was undertaken using SITC L3 product codes. This dataset aggregates in line with the hierarchy of the SITC chapter structure. This series of regressions produces similar results with slightly lower estimated standard errors than those presented in Chapter 3. Using SITC level 3 product codes, an estimated coefficient for trade liberalisation of 0.34 (0.18) for probable product emergence, and 0.33 (0.18) for probable and persistent product emergence. Both estimates are associated with a p-value of 0.06. The lower magnitude of the estimate is again unsurprising given the greatly reduce number of opportunities for product emergence given the smaller total number of products available at the 3 digit level (239).

[132] These boundary years 1974 and 1984 are discussed in more detail in Appendix A.

Table F.3: Fixed Effects Panel Regression over All Countries [NBER Dataset Only]

SITC Products (#)	**All Countries**					
	C	**D**	**E**	**C**	**D**	**E**
	1070	786	283	1070	786	283
	Probable			Probable and Persistent		
Trade Liberalisation (0=NotLib,1=Lib)	2.98*** (0.78)	2.97*** (0.70)	0.65** (0.33)	1.94*** (0.49)	2.06*** (0.44)	0.30 (0.22)
Observations	3935	3935	3935	3807	3807	3807
Country Clusters	129	129	129	129	129	129
R-Squared						
Within	0.19	0.19	0.07	0.18	0.19	0.07
Between	0.12	0.10	0.00	0.12	0.06	0.02
Overall	0.04	0.15	0.03	0.07	0.15	0.02
Other Controls	GDPPC (Constant US$ 2005)***					
Fixed Effects	Country Year					
Years	1962-2000					

Notes: 1. This model has been estimated using stata and xtreg [options: fe and vce(cluster cid)] with the addition of year fixed effects. **2.** The reported standard errors in this table have been adjusted for country clusters. **3.** Standard errors are reported in parenthesis. **4.** Levels of significant are indicated by pvalues ***= <0.01, **=<0.05, *=<0.1. **5.** GDPPC in constant US$ 2005 is sourced from WDI and is a control for the level of development. This is significant at the 1 percent level across all regressions. The estimated coefficients are not reported. **6.** Additional controls such as XR/PPP were considered and discussed but did not change the estimate significantly. **7.** The country list can be found in Appendix ##.

A final avenue explored was to consider additional explanatory controls of product emergence to improve the estimate. The exchange rate is likely to contain explanatory power of the emergence of new products as changes in relative costs change demand patterns. Two additional variables, the exchange rate (XR) variable and the price level of GDP (XR/PPP) from the PENN dataset were added to the regression model[133]. In the NBER only 1962 to 2000, the estimate for trade liberalization changed slightly 0.45 with a p-value of 0.051. This provides some evidence that the current specification may be subject to some degree of omitted variable bias – although the actual estimated coefficient does not change significantly in the case of adding these additional controls.

[133] The indicators used from PWT 8.1 were 'xr', and 'pl_gdpo'

Table F.4: Fixed Effects Panel Regression over Developing Countries
[NBER Dataset Only]

SITC Products (#)	Developing Countries Only					
	C	**D**	**E**	**C**	**D**	**E**
	1070	786	283	1070	786	283
	Probable			Probable and Persistent		
Trade Liberalisation (0=NotLib, 1=Lib)	2.24*** (0.80)	2.43*** (0.72)	0.76** (0.34)	1.32*** (0.48)	1.53** (0.45)	0.39* (0.23)
Observations	3171	3171	3171	3065	3065	3065
Country Clusters	107	107	107	107	107	107
R-Squared						
Within	0.14	0.14	0.07	0.12	0.12	0.06
Between	0.17	0.04	0.04	0.20	0.08	0.08
Overall	0.01	0.03	0.01	0.01	0.03	0.01
Other Controls	GDPPC (Constant US$ 2005)					
Fixed Effects	Country Year					
Years	1962-2012					

Notes: 1. This model has been estimated using stata and xtreg [options: fe and vce(cluster cid)] with the addition of year fixed effects, 2. The reported standard errors in this table have been adjusted for country clusters, 3. Standard errors are reported in parenthesis, 4. Levels of significant are indicated by pvalues ***= <0.01, **=<0.05, *=<0.1, 5. GDPPC in constant US$ 2005 is sourced from WDI and is a control for the level of development. This is significant at the 1 percent level across all regressions. The estimated coefficients are not reported.

F.3. NBER 1984 TO 2000 – RESTRICTED SAMPLE #2

As briefly discussed in the preceding section an alternate time period restriction was analysed. The benefit of using this time period is that it enables the full use of the disaggregated product classification system with 786 different products. This timeframe does not suffer from any peculiarities in relation to the dynamic characteristics of the dataset, therefore only dataset D needs to be considered. The disadvantage is that the time frame is greatly reduced from 50 to just 16 years. Despite this the results are reasonably robust, albeit with weaker standard errors than those observed in Chapter 3. In this case, the probable and persistent product indicator appears to provide a cleaner signal with a stronger estimate. Results in Table E.5 are for all countries, while Table E.6 contains results for the restricted sample of developing countries.

Table F.5: Restricted NBER Dataset 1984 to 2000 [All Countries]

	All Countries	
SITC Products (#)	**D**	**D**
	786	786
	Probable	Probable and Persistent
Trade Liberalisation (0=NotLib, 1=Lib)	2.203* (1.14)	1.56** (0.66)
Observations	1912	1784
Country Clusters	129	129
R-Squared		
Within	0.06	0.05
Between	0.05	0.05
Overall	0.01	0.01
Other Controls	GDPPC (Constant US$)	
Fixed Effects	Country Year	
Years	1984 to 2000	

Table F.6: Restricted NBER Dataset 1984 to 2000 [Developing Countries in 1962]

	Developing Countries	
SITC Products (#)	**D**	**D**
	786	786
	Probable	Probable and Persistent
Trade Liberalisation (0=NotLib, 1=Lib)	2.208* (1.14)	1.62** (0.66)
Observations	1565	1459
Country Clusters	106	106
R-Squared		
Within	0.07	0.05
Between	0.09	0.09
Overall	0.01	0.01
Other Controls	GDPPC (Constant US$)	
Fixed Effects	Country Year	
Years	1984 to 2000	

F.4. ESTIMATES OF WIDTH OF DIFFUSION IN NBER 1962 TO 2000

Estimates for the width of diffusion are also computed for the restricted NBER only dataset. The results are presented in Table E.7 and are broadly consistent with those discussed in Chapter 3.

Table F.7: Robustness test for the Estimates of the Width of Diffusion

	All Countries			Developing Countries		
	C	D	E	C	D	E
SITC Products (#)	1070	786	283	1070	786	283
	Width of Diffusion			Width of Diffusion		
Trade Liberalisation (0=NotLib, 1=Lib)	0.041** (0.019)	0.047*** (0.017)	0.035** (0.015)	0.041** (0.019)	0.046** (0.018)	0.035** (0.016)
Observations	3099	3624	3935	2712	2998	3171
Country Clusters	129	129	129	107	107	107
R-Squared						
Within	0.12	0.13	0.05	0.12	0.14	0.06
Between	0.12	0.12	0.13	0.08	0.06	0.08
Overall	0.01	0.00	0.00	0.03	0.03	0.00
Other Controls	GDPPC (Constant US$ 2005)					
Fixed Effects	Country Year					
Years	1962-2000					
Notes: 1. This model has been estimated using stata and xtreg [options: fe and vce(cluster cid)] with the addition of year fixed effects						

Robustness Results for Product Emergence and Trade Liberalisation Regressions

APPENDIX G EXAMPLE PROXIMITY VALUES AND SYMMETRY

This appendix contains some examples of computed proximity values between for range of different types of products. This data is computed for the year 2000 from the NBER world trade flows dataset.

Table G.1: Example Proximity Values NBER

P1	P1 Description	P2	P2 Description	Proximity
8423	Men's and boys' outerwear, textile fabrics not knitted or crocheted; trousers, breeches and the like	8441	Under garments of textile fabrics, not knitted or crocheted; mens and boys shirts	0.74
		8451	Outerwear knitted or crocheted, not elastic nor rubberized; jerseys, pullovers, slip-overs, cardigans, etc	0.71
		6584	Linens and furnishing articles of textile, not knitted or crocheted	0.49
		0611	Sugars, beet and cane, raw, solid	0.29
		2927	Cut flowers and foliage	0.20
		7810	Passenger motor vehicles (excluding buses)	0.09
		7924	Aircraft of an unladen weight exceeding 15000 kg	0.03
0711	Coffee green, roasted; coffee substitutes containing coffee	0611	Sugars, beet and cane, raw, solid	0.55
		2927	Cut flowers and foliage	0.43
		8441	Under garments of textile fabrics, not knitted or crocheted; mens and boys shirts	0.30
		8451	Outerwear knitted or crocheted, not elastic nor rubberized; jerseys, pullovers, slip-overs, cardigans, etc	0.29
		6584	Linens and furnishing articles of textile, not knitted or crocheted	0.24
		7810	Passenger motor vehicles (excluding buses)	0.05
		7924	Aircraft of an unladen weight exceeding 15000 kg	0.02

G.1. SYMMETRIC VERSUS ASYMMETRIC PROXIMITY MATRICES

The existing literature focuses on a symmetric definition of the proximity matrix. This has a number of advantages in that product relationships are symmetric and therefore is representative of an undirected network. This appendix briefly explores the implications of an asymmetric definition of proximity.

A small sample of computed proximity values are used to motivate this discussion. The symmetric (or minimum proximity values) are presented in Table G.2. This small sample consists of one textile and three apparel products. As is evident from the table the textile product (6584) is less related to the three apparel products they the three apparel products are to each other. This forms a natural clustering which suggest the various apparel products share more similarities or underlying factors of production. Similar sets of countries export each product, and each product has similar ubiquity.

Table G.2: Example Symmetric Relationships

P1	P1 Description	Proximity			
		6584	**8423**	**8441**	**8451**
6584	Linens and furnishing articles of textile, not knitted or crocheted	1.00	0.49	0.53	0.46
8423	Men's and boys' outerwear, textile fabrics not knitted or crocheted; trousers, breeches and the like	0.49	1.00	0.74	0.71
8441	Under garments of textile fabrics, not knitted or crocheted; mens and boys shirts	0.53	0.74	1.00	0.79
8451	Outerwear knitted or crocheted, not elastic nor rubberized; jerseys, pullovers, slip-overs, cardigans, etc	0.46	0.71	0.79	1.00

Notes: 1. The shaded cells contain the same information as the top left hand quadrant of the comparison matrix.

Considering the asymmetric relationships, presented in table G.4, conveys different information. The relationship is now directed and can be directly interpreted as the conditional probability of exporting product 1 given product 2. Understanding the relative outcomes in this matrix, it is useful to see the ubiquity characteristics of each product contained in Table G.3.[134]

[134] In each relative comparison the numerator remains constant.

Table G.3: Number of countries that export sample products (Ubiquity)

SITC4	SITC 4 Description	Ubiquity
6584	Linens and furnishing articles of textile, not knitted or crocheted	38
8423	Men's and boys' outerwear, textile fabrics not knitted or crocheted; trousers, breeches and the like	70
8441	Under garments of textile fabrics, not knitted or crocheted; mens and boys shirts	53
8451	Outerwear knitted or crocheted, not elastic nor rubberized; jerseys, pullovers, slip-overs, cardigans, etc	56

First consider the relationship between textile product (6584) and apparel product (8423).[135] The textile product has a lower ubiquity score relative to the apparel product. Exporting linens and textile products given one exports trousers produces a proximity score of 0.49. However exporting trousers given one exports linens and textile products produces a much higher proximity score of 0.89. As linens are relatively less ubiquitous it is likely to require more sophisticated capabilities. This suggests that if a country has acquired the capabilities to export linens and textiles then it is relatively simple to export all of the apparel products given then are less sophisticated. Using asymmetric values may be useful when considering the diffusion through the product space as diffusion may occur more readily having acquired the capabilities to export trousers, a country may be closer to exporting linens along a chain of required improvements of sophistication.

Comparing the overall distribution of symmetric and asymmetric proximity values is shown in Figure G.1. The asymmetric proximity matrix consists of a larger frequency of higher proximity value.

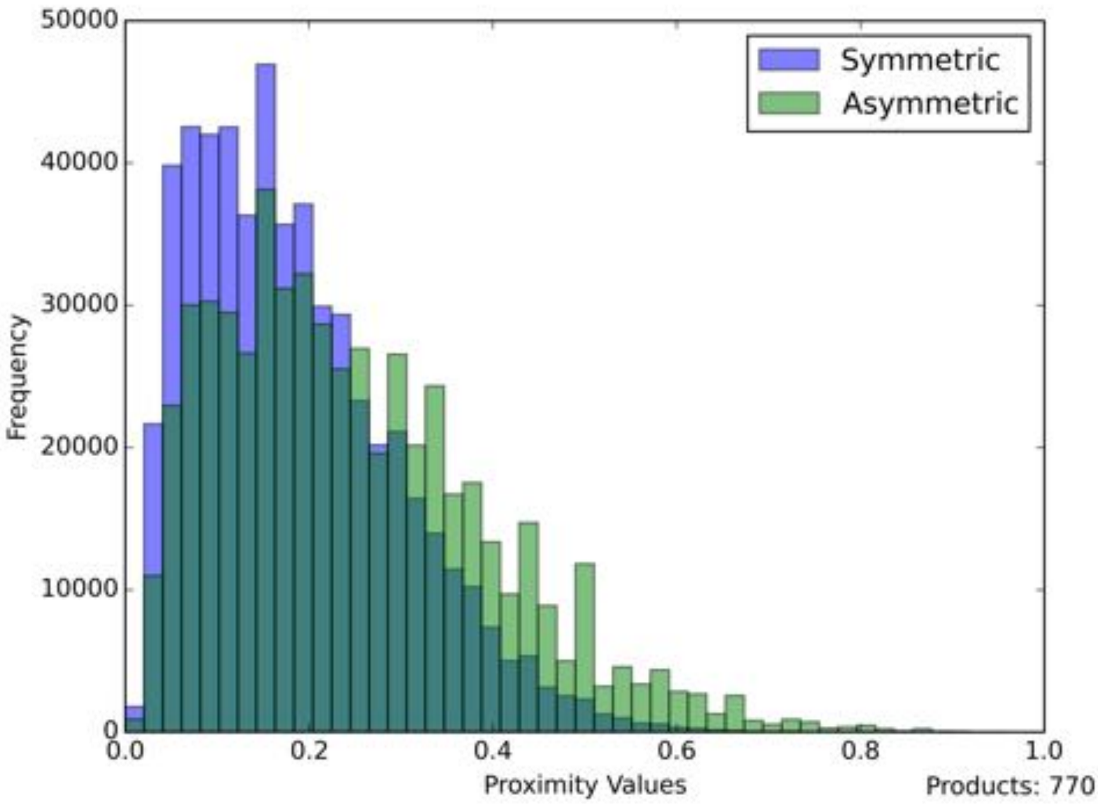
[135] The convention used in computing these relationships is that P1 is in the denominator when computing these relationships in the lower left off-diagonal, and P2 is in the denominator in the upper right off-diagonal.

Table G.4: Example Asymmetric Relationships

P1	P1 Description	P2			
		6584	8423	8441	8451
6584	Linens and furnishing articles of textile, not knitted or crocheted	1.00	0.49	0.53	0.46
8423	Men's and boys' outerwear, textile fabrics not knitted or crocheted; trousers, breeches and the like	0.89	1.00	0.98	0.89
8441	Under garments of textile fabrics, not knitted or crocheted; mens and boys shirts	0.74	0.74	1.00	0.79
8451	Outerwear knitted or crocheted, not elastic nor rubberized; jerseys, pullovers, slip-overs, cardigans, etc	0.68	0.71	0.83	1.00

Notes: 1. The shaded squares highlight the maximum proximity value in each comparison between the four products selected.

Figure G.1: Histogram of Symmetric and Asymmetric Proximity Values



Notes: 1. This figure overlays the histogram of symmetric with non-symmetric proximity. There are a large number of higher proximity values when considering asymmetric proximity, as the minimum conditional probability is no longer a restriction. A number of these relationships exist between items such as textile and apparel products. The data suggests that textile exporters more often co-export apparel products but that apparel exporters do not necessarily export textiles.

APPENDIX H PRODUCT FRAGMENTATION AND PARTS AND COMPONENTS

This Appendix provides the reference material for Chapter 5. It contains the Parts and Components list used in the analysis, and a number of graphs that provide different views of the data by product type that not included in the chapter. These include the Lall (2000), Leamer (1984), and SITC chapter level product classifications.

H.1. PARTS AND COMPONENTS DEFINITION

The parts and components list is defined in Table H.1.

Table H.1: Parts and Components (SITC rev 3)

SITC	Description	SITC	Description	SITC	Description
58291	Cellular plastic sheet	71321	Recip piston engs<1000cc	72689	Pts nes of bookbind mchn
58299	Non-cellular plast sheet	71322	Recip piston engs>1000cc	72691	Type-setting machn parts
59850	Doped chemicals (electr)	71323	Diesel etc engines	72699	Printing press parts
61290	Leather manufactures nes	71332	Marine spark-ign eng nes	72719	Cereal/dry legm mach pts
62141	Uh rubber tube no fittng	71333	Marine diesel engines	72729	Indus food proc mach pts
62142	Uh metal-reinf rubr tube	71381	Spark-ign piston eng nes	72839	Pts nes of machy of 7283
62143	Uh text-reinf rubbr tube	71382	Diesel engines nes	72847	Isotopic separators
62144	Uh nes-reinf rubber tube	71391	Parts nes spark-ign engs	72851	Glass-working machy part
62145	Uh rubber tube + fitting	71392	Parts nes diesel engines	72852	Plastic/rubber mach part
62921	Conveyor/etc belts v	71441	Turbo-jets	72853	Tobacco machinery parts
62999	Uh non-cell rub articles	71449	Reaction engines nes	72855	Parts nes, machines 7284
65621	Woven textile labels etc	71481	Turbo-propellers	73511	Tool holder/slf-open die
65629	Non-woven text label etc	71489	Other gas turbines nes	73513	Metal mch-tl work holder
65720	Non-woven fabrics nes	71491	Parts nes turbo-jet/prop	73515	Dividing head/spec atach
65751	Twine/cordage/rope/cable	71499	Parts nes gas turbines	73591	Pts nes metal rmvl tools
65752	Knotted rope/twine nets	71610	Electric motors <37.5w	73595	Pts nes mtl nonrmvl tool
65771	Textile wadding nes etc	71620	Dc motor(>37w)/generator	73719	Foundry machine parts
65773	Industrial textiles nes	71631	Ac,ac/dc motors >37.5w	73729	Roll-mill pts nes, rolls
65791	Textile hosepiping etc	71632	Ac generators	73739	Mtl weld/solder eq parts
65792	Machinery belts etc,text	71651	Gen sets with pistn engs	73749	Parts gas welders etc.
66382	Asbestos manuf-friction	71690	Pts nes motors/generator	74128	Furnace burner parts
66471	Tempered safety glass	71819	Parts nes hydraul turbin	74135	Elect furnace/oven parts
66472	Laminated safety glass	71878	Nuclear reactor parts	74139	Parts ind non-el furn/ov
66481	Vehicle rear-view mirror	71899	Parts nes of engines nes	74149	Pts nes indus refrig equ
66591	Laboratory etc glass	72119	Agric machine(7211)parts	74155	Air-conditioners nes
66599	Other glass articles nes	72129	Pts nes of machy of 7212	74159	Air-conditioner parts
69551	Band saw blades	72139	Pts nes dairy machinery	74172	Water proc gas gen parts
69552	Steel circular saw blade	72198	Parts wine/etc machines	74190	Parts indus heat/cool eq
69553	Circular saw blades nes	72199	Pts nes agric machines	74220	Piston eng fuel/wtr pump
69554	Chain saw blades	72391	E-m bucket/grab/shovels	74291	Pump parts

SITC	Description	SITC	Description	SITC	Description
69555	Straight saw bl for metl	72392	Bulldozer etc blades	74295	Liquid elevator parts
69559	Saw blades nes	72393	Boring/sink machry parts	74363	Engine oil/petrol filter
69561	Cutting blades for machn	72399	Pts nes earth-movg mach	74364	Engine air filters
69562	Carbide tool tips etc	72439	Sew mch needles/furn/pts	74391	Parts for centrifuges
69563	Rock etc drilling tools	72449	Pts nes textile machines	74395	Parts filters/purifiers
69564	Parts to insert in tools	72461	Auxil weave/knit machine	74419	Pts nes of work trucks
69680	Knives and blades nes	72467	Weaving loom parts/acces	74443	Jacks/hoists nes hydraul
69915	Base mtl vehicle fitment	72468	Loom/knitter etc pts/acc	74491	Parts for winches/hoists
69933	Base metal buckles etc	72488	Parts for leather machns	74492	Lift truck parts
71191	Pts nes of boilers 711.1	72491	Washing machine parts	74493	Lift/skip h/escalat part
71192	Pts nes boiler equ 711.2	72492	Textile machinry pts nes	74494	Lifting equip parts nes
71280	Stm turbine(712.1)parts	72591	Paper manuf machine pts	74519	Pts nes of tool of 7451
71311	Aircraft piston engines	72599	Paper product mach parts	74529	Packing etc mchy pts nes
71319	Pts nes a/c piston engs	72635	Printing type,plates,etc	74539	Weighng mach wts,pts nes

H.2. PARTS AND COMPONENTS COMPOSITION OF DIFFERENT PRODUCT CLASSIFICATIONS

The focus in Chapter 5 is on country shares in world trade, in addition to the within-country composition of PSB trade in total manufacturing exports. At the level of individual products some *types* of goods are more influenced by the process of product fragmentation than others. This is captured well by grouping trade data into different categories, such as those defined by the technology intensity concordance defined by Lall (2000). The share of parts and components in different Lall categories is presented in Figure H.1. The 10 Lall categories are based on SITC revision 2 products at the 3-digit level, which in this study have been matched to the SITC revision 3 data used here. Primary Products (PP) consists of raw agricultural products such as fresh fruit and rice; while resource based (RB) manufactures capture prepared meats and fruits that require some processing prior to export.[136] The manufacturing chapters are split into

[136] The original Lall classification is provided in SITC rev 2 codes. Therefore, SITC rev 2 codes where added to the dataset using the UN conversion tables. These codes were then used to match the Lall classification.

low technology manufactures (LT), such as textile fabrics and clothing, in a spectrum across 7 categories, to high technology (HT) manufactures such as aircraft and pharmaceuticals.

Acknowledging that visibility is restricted due to the nature of the SITC classification, Lall categories vary in their respective abilities to be fragmented.[137] This is a process that occurs in partnership with industry-specific technological changes in manufacturing. A high percentage (by value) of parts and components is exported in categories High Technology 1 (HT1) of around 55% and Medium Technology 3 (MT3) of approximately 49%.[138] Digging into the concordance shows that much of the electronics industry is contained within the HT1 category, which consists of office and data processing equipment (i.e. computers), telecommunications equipment, and televisions. MT3 is also characterised by a relatively high percentage of parts and components within the category (by value) and is composed of engines, motors, industrial machinery, ships, and watches.

The next largest category, MT1, consists largely of automotive products, for which parts and components make up around 25 percent of global exports by value. Toyota has in recent decades set up final assembly of automobiles in Thailand, which has led to a flourishing network of parts and components suppliers in that region.[139] When disaggregating the data at the country level, this trend can be observed as Chapter 78 captures an increasing share of Thailand’s total parts and components exports, as shown in Figure H.3.

Parts and components do not make up a significant segment of the MT2 category. MT2 consists largely of chemicals, paints, and fertilizers. Chemicals and industrial manufacturing are production processes that are more difficult to fragment into

[137] One could also argue that given the lack of detail in the SITC product classification, this is indicative of less trade in parts and component for categories to be set up in the first place.

[138] These percentage values are mean values over 1995 to 2013.

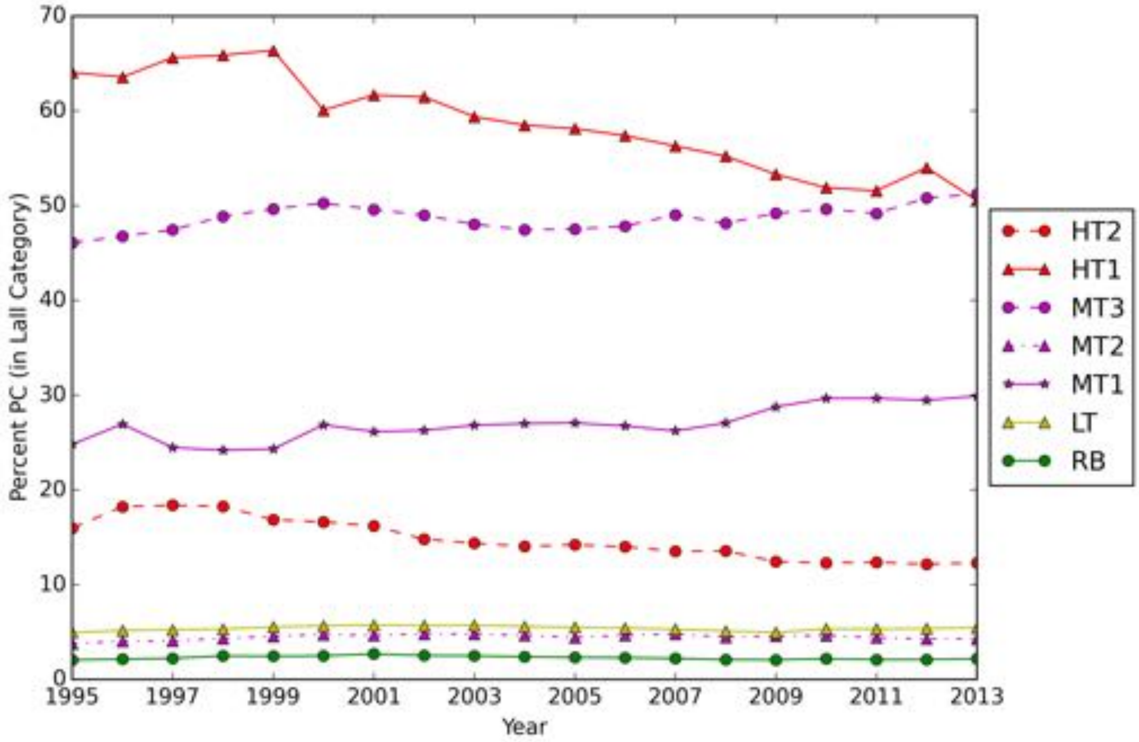
[139] As can be observed in Figure 5.6, SITC Chapter 78 “Road Vehicles” is capturing an increasing share in Thailand’s export composition. This suggests that final assembly in Thailand has potential spillovers into parts and components suppliers, which subsequently export, to other countries in the region.

discrete segments.[140] LT products such as textile and apparel, pottery and jewellery do not contain high parts and component values, as these products contain much less detailed parts and components definitions in the SITC rev 3, making it impossible to estimate for these industries. Yeats (2001) showed that these industries do however participate strongly in product fragmentation. Therefore values discussed here are likely to be underestimated.

The equivalent Leamer (1984) categories are shown in Figure H.2. This grouping of products aggregates over much of the variation that has been observed in the above discussion of the Lall (2000) classification.

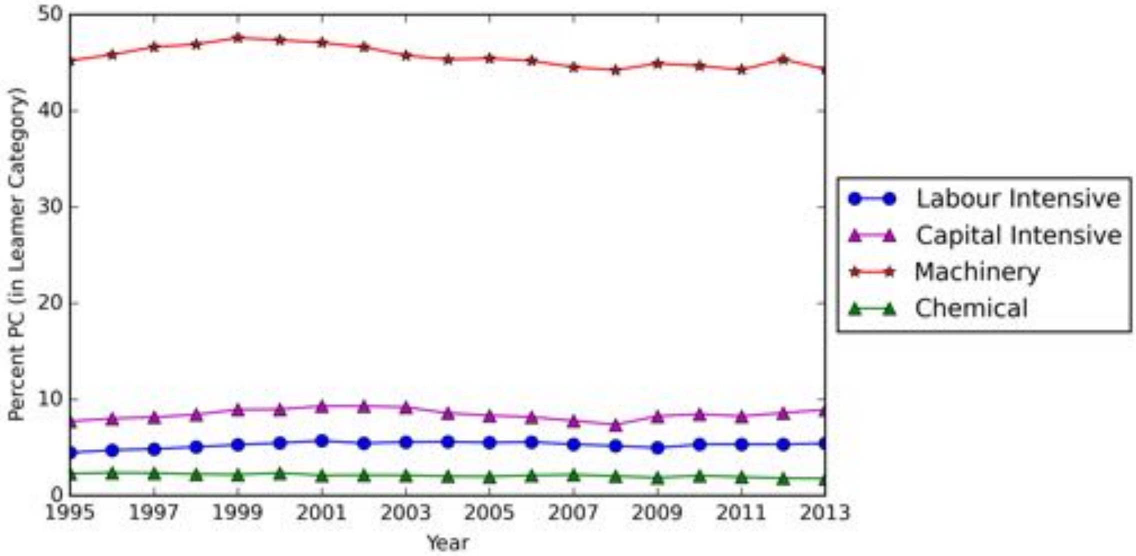
[140] There are many reasons why this may be so. Chemical processes require large investments in capital, which produces *stickiness* in geography. Low value-weight ratio makes some products uneconomic to transport.

Figure H.1: Percent Parts and Components within Lall (2000) Technology Intensity Categories



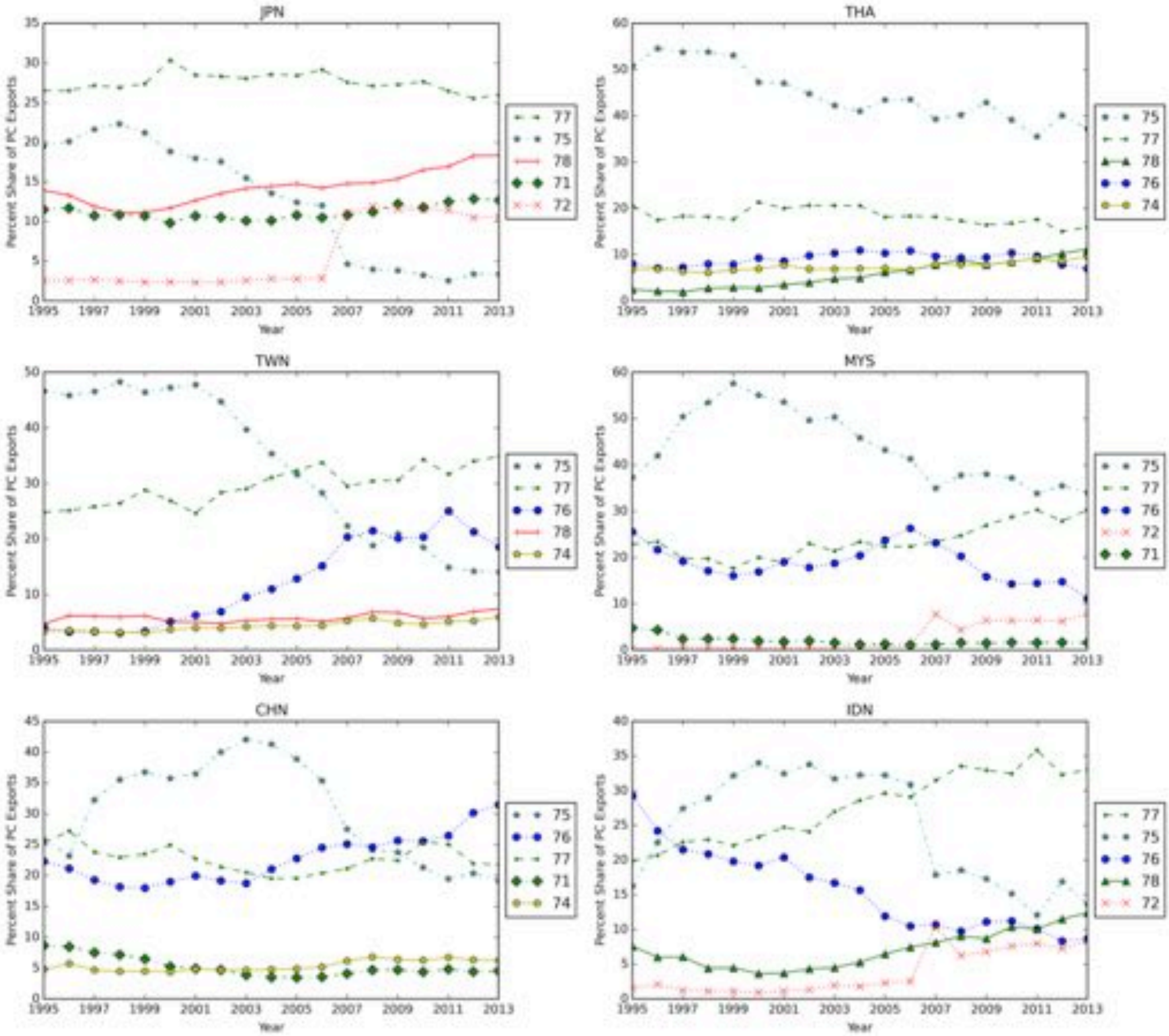
Notes: 1. HT = High technology manufactures, MT = Medium technology manufactures, LT = Low technology manufactures, RB = Resource based manufactures. **2.** In the original Lall (2000) classification there are two resource-based categories (RB1 and RB2). These categories are similar in composition so the mean is presented here. The same applies for light technological intensity (LT1 and LT2). **3.** The Lall technological intensity classification includes a Primary Product (PP) category. This category does not contain any defined parts and components. **4.** The Lall (2000) classification is defined in SITC rev. 2 at the 3-digit level. A UN conversion table between SITC rev. 3 and SITC rev.2 was used to match Lall categories with the data used in this study.

Figure H.2: Percent Parts and Components within Leamer (1984) product type categories



Notes: 1. There are only four (of 10) Leamer categories that contain parts and components as presented in the figure. **2.** The UN concordance between SITC revision 3 and SITC revision 2 was used to match each Leamer (1984) category as it is based on SITC revision 2.

Figure H.3: Percent Share of Parts and Components within major SITC2 categories (Manufacturing Exports)



Notes: 1. 71 = Power generating equipment, 72 = Machinery specialised for particular industries, 74 = General industrial machinery, 75 = Data processing machines, 76 = Telecommunication equipment, 77 = Electrical machinery nes, 78 = Road Vehicles.

H.3. PERCENTAGE PARTS AND COMPONENTS OF MANUFACTURING EXPORTS IN DIFFERENT REGIONS

Figure H.4: Percent parts and components in manufacturing exports of UN defined areas.

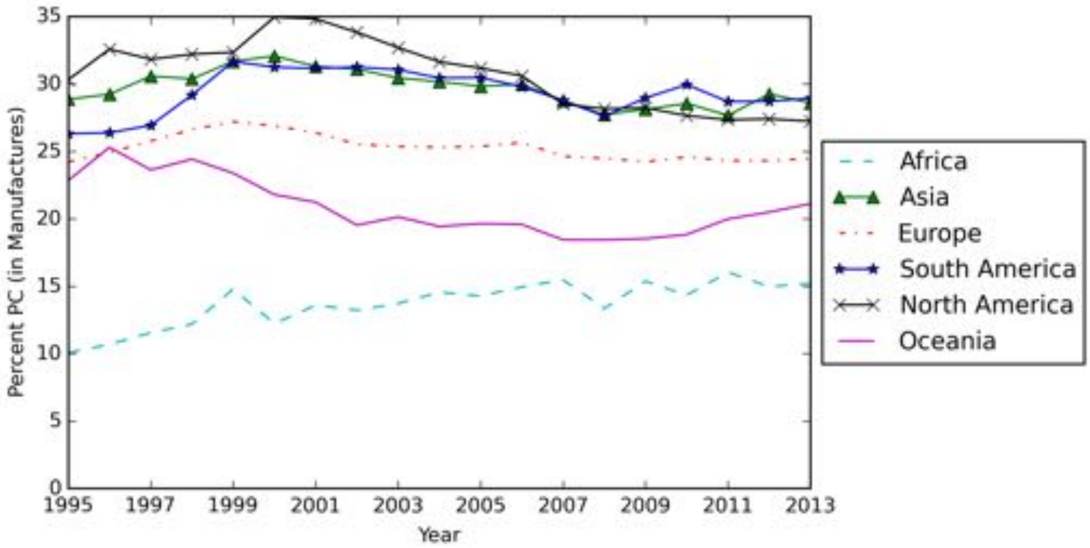
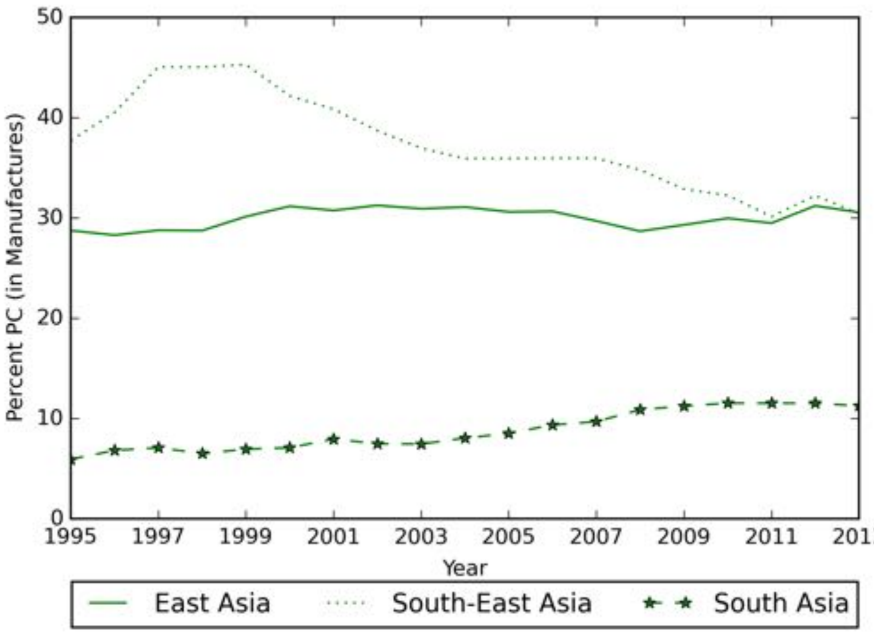


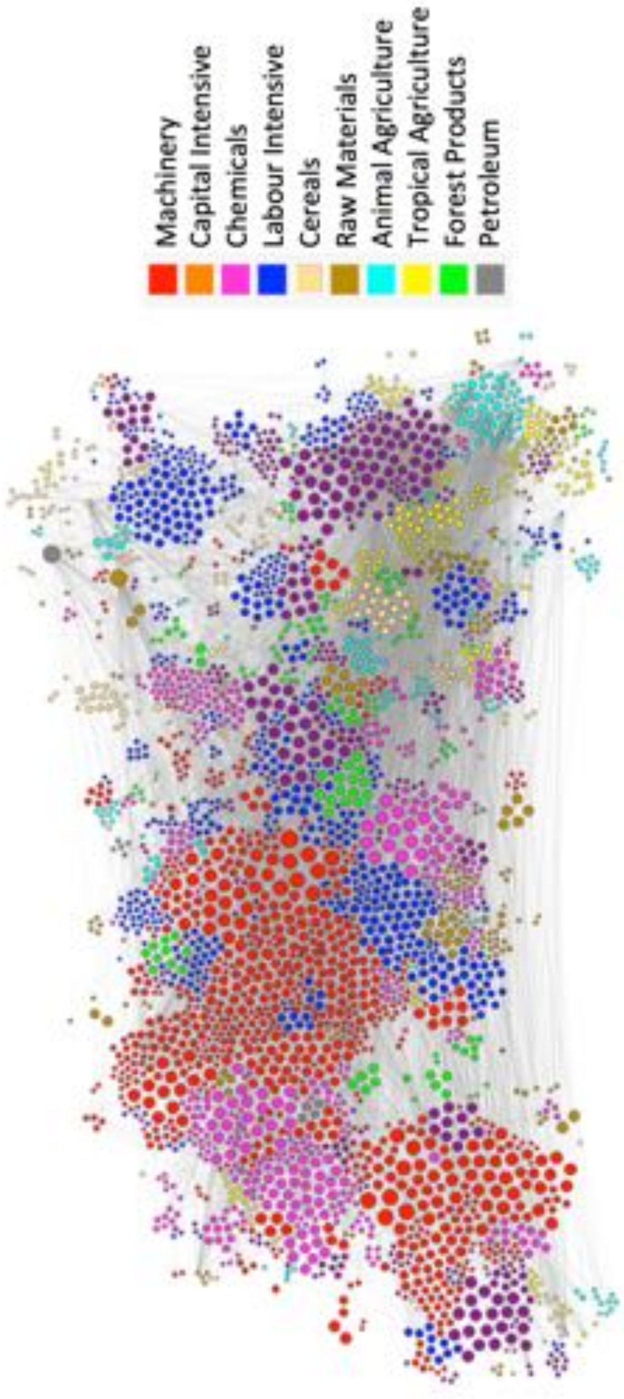
Figure H.5: Percent Parts and Components in manufacturing exports in three Asian regions



Notes: 1. East Asia contains CHN, HKG, JPN, KOR, MAC, MNG, PRK, and TWN. **2.** South-East Asian contains BRN, IDN, KHM, LAO, MMR, MYS, PHL, SGP, THA, and VNM. **3.** South Asia contains AFG, BGD, BTN, IND, IRN, LKA, MDV, NPL, and PAK.

H.4. PRODUCT SPACE IN LEAMER PRODUCT CATEGORIES

Figure H.6: Product Space (SITC Rev 3 L5) Atlas Layout in Leamer (1984) Categories



Notes: 1. Legend is organised by decreasing cluster size based on the number of within cluster nodes. Capital Intensive category features in a number of different atlas clusters. **2.** This diagram contains 3111 products.

APPENDIX I ‘PYECONLAB’ PYTHON PACKAGE

This appendix briefly outlines the software written in support of the analytics in this thesis. The package largely serves two purposes. First it provides a number of data structures that are useful when working with disaggregated international trade data. For example, the **ProductLevelExportSystem** and **DynamicProductLevelExportSystem** provide python objects that incorporate the computation of useful trade statistics. This makes it easier to work with different datasets and ensure computations remain consistent. Once a **ProductLevelExportSystem** is populated with international trade data one can easily transform the data and produce RCA or M_{cp} matrices, compute ECI and PCI vectors and obtain trade statistics and various other representations of the data.

The second major component of the library is that it turns a number of commonly used datasets into python objects that makes working with and accessing the data much simpler. These datasets include the PENN world tables and the World Development Indicators (WDI). Interfaces and constructors are also defined for parsing raw international trade data from specific sources. The NBER world trade flow dataset, for example, has particular characteristics such as the presence of non-official SITC revision 2 product codes. These constructors build methods that account for this dataset specific context.

I.1. SUB PACKAGES

The **pyeconlab** library is composed of subpackages, that provides different analytical functions. These subpackages include **country**, **penn**, **trade**, **util**, and **wdi**.

I.1.1. COUNTRY

The **COUNTRY** includes meta data and functions specific for working with country data. This includes concordances for converting between ISO3N, ISO3C and Country Names using ISO3166 definitions. This is useful as one can work with ISO3C country codes and easily convert them to country names for publication.

Examples

In:
```
import pyeconlab as econ

cntrys = econ.country.ISO3166(verbose=True)

cntrys.iso3c_to_name['ZMB']
```

Out: Zambia

In:
```
for cntry in ["AFG","USA","ZWE"]:
    print cntrys.iso3c_to_name[cntry]
```

Out:
```
Afghanistan
United States
Zimbabwe
```

I.1.2. PENN

The **PENN** subpackage is a wrapper around the PENN world tables providing a data object that can be used in analytical projects or dataset construction.

Examples

In:
```
import pyeconlab as econ

penn = econ.penn.PENN(source_dir="/Users/matthewmckay/work-data/datasets/2c2e8d593f39ee74aeb2c7c17047ea3f/")

print penn.series_descriptions['avh']

print penn.get(cntry="AUS",series_code="avh", year=2000)
```

Out:
```
Loading data for PENN world tables from: /Users/matthewmckay/work-data/datasets/2c2e8d593f39ee74aeb2c7c17047ea3f/pwt81.dta

Average annual hours worked by persons engaged

1794.6400000000001
```

I.1.3. TRADE

The **TRADE** subpackage is a package for working with disaggregated trade data. Currently it provides python objects for working with **ProductLevelExportSystem**, which describes a cross-section of world trade. This class includes common analytics that can be conducted on this type of trade data, such as computing matrices of revealed comparative advantage, producing networkx data structures for analysing network features, and computing economic and product complexity vectors. These objects are general and can be used with any underlying trade data and is built around pandas data tables.

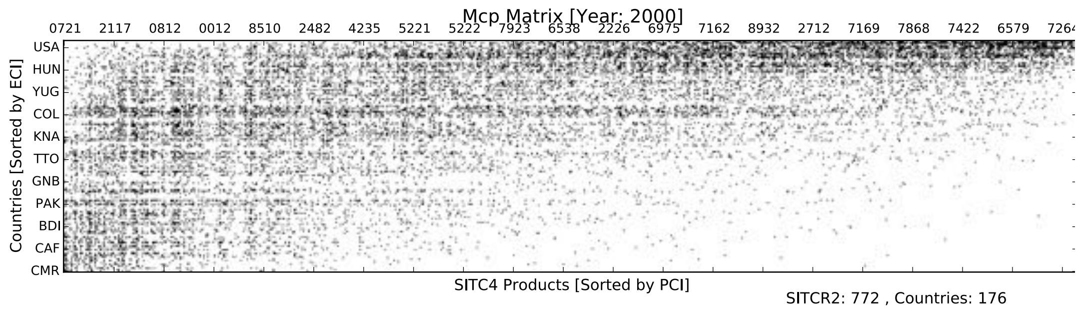
Example

In:

```
#-Example for Computing the Mcp Diagram for the Year 2000-#

import pandas as pd
from pyeconlab import CPExportData, ProductLevelExportSystem
from dataset_info import TARGET_DATASET_DIR, CHAPTER_RESULTS, SOURCE_DIR

#-Year:2000-#
DATASET_DIR = TARGET_DATASET_DIR['nber']
RESULTS_DIR = CHAPTER_RESULTS[2]
DATASET = 'D'
YEAR = 2000

#-Retrieve Export Data-#
data = pd.read_hdf(DATASET_DIR+"nber-export-sitcr2l4-1962to2000.h5", DATASET)
data = data.loc[data.year == YEAR]
data = data.rename(columns={'eiso3c':'country','sitc4':'productcode', 'value':'export'})
data = data.reset_index()
del data["index"]
del data["year"]
data = data.set_index(["country", "productcode"])

#-Product Level Export System-#
s = ProductLevelExportSystem()
s.from_df(data, country_classification="ISO3c", product_classification="SITCR2",
compile_dtypes=["DataFrame"], year=YEAR)
s.rca_matrix(complete_data=True)
```

```
s.mcp_matrix()

eci = s.compute_eci(auto_adjust_sign=True)

pci = s.compute_pci(auto_adjust_sign=True)

s.mcp = s.sorted_matrix(s.mcp, row_sortby=eci.copy(), column_sortby=pci.copy())

fig = s.plot_mcp(row_sortby_label="ECI", column_sortby_label="PCI")

plt.show()
```

Out:



The **DynamicProductLevelExportSystem** provides a data object that contains any number of yearly cross-section objects. This allows for easy computation across many different cross-sections, in addition implements the algorithms that require dynamic data in addition to smoothing algorithms such as moving averages or 3-year averages.

I.1.4. UTIL

The **UTIL** subpackage provides a number of developer tools and functions, such as regular expression based search functions for matching country names, and functions that switch between different representations of the same data.

I.1.5. WDI

The **WDI** subpackage is a wrapper around the World Development Indicators (WDI) and provides easy routines for accessing different data series such as GDP per Capita. It also provides lookup function for identifying WDI series codes.

Example

In:

```
import pyeconlab as econ

wdi = econ.wdi.WDI(source_dir="/Users/matthewmckay/work-
data/datasets/d1352f394ef8e7519797214f52ccd7cc/")
```

```
print wdi.codes.GDPPCGrowth
```

Out: 'NY.GDP.PCAP.KD.ZG'

In:

```
data = wdi.cntry_series(series_code=wdi.codes.GDPPCGrowth, cntry="AUS")
data
```

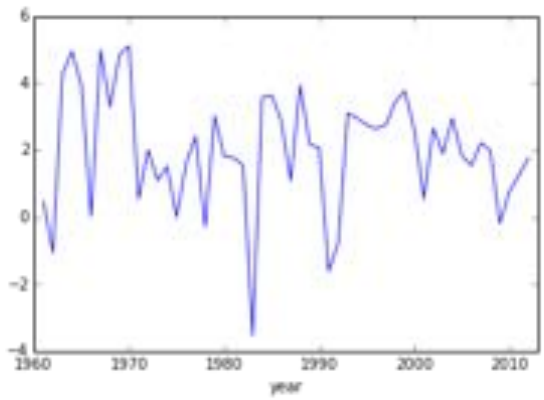
Out:

```
year
1960        NaN
1961   0.465882
1962  -1.063828
1963   4.257714
1964   4.944559
1965   3.928663
1966   0.024405
1967   4.983139
1968   3.290107
… (continued)
```

In:

```
data.plot()
```

Out:



I.2.

I.3. GITHUB

This python library is available on http://github.com/mmcky/econ-phdthesis

Key algorithms used for computing product diffusion in Chapter 3, and state transitions for Chapter 4 are available through the open source github repository.