From Quantity to Quality: Capturing Higher Spending Markets through a Segmenta- tion of Travellers' Expenditure

From Quantity to Quality: Capturing Higher Spending Markets through a Segmentation of Travellers' Expenditure

Consuelo R. Nava[1*], Linda Osti[2] and Maria Grazia Zoia[3]

[*] *Corresponding author*

[1] *Department of Economics and Statistics "Cognetti de Martiis", University of Turin, Lungo Dora Siena 110A, Turin, Italy. Corresponding Author. Tel:+390116704406. Fax:+390116703895*

consuelorubina.nava@unito.it

[2] *Faculty of Economics and Management, Free University of Bozen-Bolzano, Brunico, Italy,*

linda.osti@unibz.it

[3] *Department of Economic Policy, Università Cattolica del Sacro Cuore, Largo Gemelli 1, 20123, Milano, Italy. maria.zoia@unicatt.it*

From Quantity to Quality: Capturing Higher Spending Markets through a Segmentation of Travellers' Expenditure

Abstract

Many tourism destinations aim at expanding their market share of high spending visitors by shifting from quantity to quality. The COVID-19 pandemic has forced the introduction of social distancing requiring hotspots and mass destinations to reduce their capacity. This paper proposes a two-step approach for identifying top spending European countries over time, distinguishing between leisure and business travellers. The methodology employs the Country Product Dummy index with a hierarchical clusterization, enriched by a convergence analysis. This approach overcomes general shortcomings of descriptive statistics and cluster analyses directly applied to raw expenditure data. The outcomes of this analysis provide a detailed picture of the European travellers' expenditure across time and geographical area. The identified top spending countries of leisure and business travellers can be targeted through ad-hoc marketing campaigns and specific packages for privileging quality tourism and planning economic recovery in the post-COVID-19 reopening phase, while shifting away from mass tourism.

Keywords: travellers' expenditure; Country Product Dummy index; quality tourism; clustering; destination profitability.

1 Introduction

Over past decades, an in depth understanding of the potentially negative impacts of mass tourism and the awareness of including the economic pillar in any sustainable tourism development plan have induced many destinations to shift their focus from quantity to quality, in an attempt to gain higher yields through a lower number of visitors. Indeed, in many destinations, the increase in the number of visitors has generated excessive pressure on natural assets and local infrastructures. If this increase if further coupled with a simultaneous decrease in travellers' expenditure, tourism turns into a nonviable road to community development and wellbeing. For this reason, many destinations are aiming at expanding their market share with high spending visitors. As argued by **?**, a smaller number of visitors with a higher level of expenditure are to be preferred over a higher number of visitors with lower expenditures as, at an aggregate level, the former would generate lower negative externalities than the latter ones. In fact, if on the one hand income generated from leisure and business travellers has been recognized as a key factor for the economic growth and the development of regions and countries (**???**), on the other hand, the important issue of intensity and extensity of profits must be considered.

In fact, **?** stressed the importance of distinguishing if the same profit in a given destination is generated either through a high level of expenditure per visitor with a low volume of visitors (intensive), or a low level of expenditure per visitor and a high number of visitors (extensive). Extensive profitability inevitably leads to an increase of negative externalities due to the high number of visitors.

Thus, destinations' competitiveness has also moved from quantity to quality, that is from "the ability of a destination to maintain its market position and share and/or to improve upon them through time" (**?**, p.23) to "the ability to increase tourism expenditure" (**?**, p.2) while providing visitors "with satisfying, memorable experiences, and to do so in a profitable way, while enhancing the well-being of destination residents and preserving the natural capital of the destination for future generations" (**?**, p.2).

Therefore, attracting groups of travellers generating the highest possible yield has become paramount for many destinations. In fact, understanding visitors' expenditure patterns is es-

sential for both destination marketing and policy makers, to develop successful strategies to increase travellers' receipts (**?**).

This task was initially challenged by the 2008 financial crisis, one of the most severe since the Great Depression. This crisis inevitably affected European (EU) countries and travellers' purchasing power (**?**), as the Asian crisis did in the late 90s (**?**). Developing successful strategies to increase travellers' receipts in an effective way has become even more crucial in 2021, in order to respond to the challenges raised by COVID-19.

In this regard, it is worth noting that the tourism industry is currently facing a global crisis stretching from a health emergency to a social and economic collapse. The COVID-19 pandemic is having an unprecedented impact on travel, hospitality and tourism. In this respect, it represents an economic super-shock (**?**) which requires that destinations and operators undertake strategic analyses and decisions in order to overcome the financial losses faced in 2020, and to pursue a sustainable recovery of the sector, as indicated by the World Economic Forum.

Recently, the UNWTO has affirmed that "international tourism declined over 70% in 2020, back to levels of 30 years ago" (**?**). In particular, international tourist arrivals reduced by 72% from January to October 2020 compared to the same period in 2019. A slight improvement occurred during the 2020 summer months. The **?** has talked about "positive signs of a gradual but still cautious change in trend for the upcoming Northern Hemisphere peak summer season". Unfortunately, the second COVID-19 wave has once again dramatically affected the world tourism sector. Thus, with the subsequent reintroduction of travel restrictions, further losses have been registered: - 900 million international tourist arrivals (compared to 2019), and of - 935 billion US$ in export revenues from international visitors during the first ten months of 2020. This loss has an unprecedented magnitude: it is 10 times higher than the loss registered under the global economic crisis started in 2008 **?**.

Not only has the COVID-19 pandemic currently reduced, and in some case cancelled, the number of source markets, but it has also introduced the concept of social distancing, whereby hotspots and mass destinations are required to reduce their capacity. With tourism accounting for 29% of the global exports in the service industries (**?**), the economic recovery of many

EU countries turns out to crucially depend upon their ability to generate higher yields from a lower number of visitors.

While acknowledging the human, social and financial devastation of COVID-19, scholars (**??????**) have also seen this crisis as a chance to reset and reopen tourism in the post-pandemic phase in a more sustainable way (**?**). There is the call for a more equitable form of development for this sector, which should encompass the three main pillars of sustainability (environmental, social, and economic) with the scope to protect natural as well as cultural resources while fostering the wellbeing of the local community. Local wellbeing can be enhanced through different ways including, among the others, employment opportunities, local economic growth, local business development and increments of local employees' wages. The growth in travellers' expenditure can be an underlying promoter of such outputs. To increase travellers' expenditure, tourism destinations and operators have two possibilities: either increase their capacity, or improve the tourism quality by increasing their spending per travelling day (**?**), that is shifting from quantity to quality, and from extensive to intensive profits, as suggested by **?**.

2 Purpose of the paper

InIn light of this premise, the purpose of this paper is to apply an innovative two-step procedure to analyse the variation of travellers' expenditure over time. This methodology, when applied to an EU dataset of leisure and business travellers, is capable of identifying groups of EU countries from which the top-spending travellers come from. The methodology here proposed consists of two steps. In the first step, variations of travellers' expenditures aggregated by countries of origin are computed by using the Country Product Dummy (**?**) (CPD hereafter) index. In the second step, clusters of EU countries are formed, according to the values of the CPD index. This approach enables the overcoming of the general shortcomings of descriptive statistics which are not capable of capturing the dynamics of a given phenomenon over time and geographical area simultaneously, such as expenditures on different types of commodities incurred by travellers from different places of residence over time.

In the first step, the CPD index which is the main index used by the International Comparison Program at the World Bank, is employed to compare the expenditure incurred by leisure and business travellers, aggregated by EU countries of origin. To provide detailed results on how travellers belonging to different groups of EU countries spend on different types of commodities, a further analysis is carried out.

The latter consists of a hierarchical cluster analysis of travellers' countries of origin, based on the Ward's method and the CPD values estimated in the first step.

The travellers' expenditure convergence, between and within clusters, is also investigated with the Theil index, by using both the purchasing power parities (PPP) and the original series of travellers' expenditure (for each commodity). This analysis is relevant because it highlights expenditure disparities, within and between clusters. In fact, it aims at investigating if travellers' expenditure inequalities tend to decrease (increase) over time either between groups of countries (which are internally homogeneous for overall expenditure levels) or between single countries within the same group. This analysis proves useful to detect trends in the travellers' behaviour with respect to global and commodity-specific consumption.

The rest of the paper is organized as follows. Section **??** gives a brief overview of the literature on travellers' expenditure and price indexes. Section **??** explains the methodology behind the two-step procedure proposed in the paper. Section **??** presents the data used for the empirical analysis undertaken in Section **??**. The latter provides the main results of the empirical application of the proposed methodology together with a convergence analysis. Finally, concluding remarks and further hints are provided in Section **??**. For the sake of better readability, further graphs have been placed in the Appendix.

3 Travellers' Expenditure and Price Indexes

Travellers' expenditure can be studied either at an aggregate level, to identify the economic impacts of tourism on a given destination/nation, or at a disaggregate level (as in **?**), to identify the profitability of specific market segments (**?**). Therefore, understanding visitors' expenditure is of vital importance as it allows to identify valuable market segments, to formulate

market segmentation strategies (**?**) and, "consequently, to develop powerful analytical models for tracing tourists' spending patterns" (**?**, p.101).

Moreover, the scholar's interest on travellers' expenditure is high, both in terms of methods and empirical findings, and the literature on the topic is remarkably vast (see, among the others, **?????**). Reviews on methods to assess visitors' expenditure have been proposed, for instance in **?**. An expenditure-based segmentation is proposed in **?** while **?** evaluate international tourists' expenditure.

As anticipated in the previous section, the first step of the methodology proposed here is based on the estimation of the CPD index to analyse the variations of travellers' expenditure aggregate by countries of origin.

The use of price indexes, both in terms of time and geographical area, is still a challenging topic when it comes to tourism research. Price indexes have proven useful as starting points for the creation of composite indicators to describe, for instance, destination competitiveness (**?**).

Furthermore, it should be considered that price indexes are not only related to the supply, but they are the principal tool for measuring inflation. The same US Bureau of Economic Analysis analyzes personal consumption expenditures with a price index, detecting changes in the expenditure of goods and services purchased by consumers in the United States. The Bureau of Labor Statistics' consumer price index for urban consumers measures the inflation (or deflation) across a wide range of consumers' expenditure by considering their variation over time (www.bea.gov).

It is also worth considering that price indexes provide dimensionless and comparable values over time or geographical area. The information that can be inferred from price index variations – in terms of sign or order of magnitude – is analogous to what can be inferred from the data, even if a price index working on the aggregation of different goods and services, better synthesizes price/expenditure information. Thus, a positive variation of the index over time (or geographical area) indicates an increase in the level of expenditure: the greater the variation, the greater the expenditure. Therefore, even if price indexes and expenditures are different concepts, they are deeply related: the values of a price index can be seen as di-

mensionless indicators of the ability to spend by a certain community compared to others. In this case, the CPD index — being computed on travellers' expenditure from different EU countries — measures the (variations of) travellers' expenditure (by country of origin) in the commodities included in the reference basket.

Here, the CPD index – commonly used to measure changes in commodity prices sold in different countries, i.e. a multilateral price index – is estimated for goods/services purchased by travellers. The prices, which are used to determine the CPD index, are obtained as the ratio of the overall commodity expenditure and the number of travellers' overnight stays for each country of origin.

4 Methodology

Figure **??** shows the methodological proposal is finalized to cluster travellers' place of residency according to their expenditures over different commodities, discussed in what follows.

[Figure 1 about here.]

4.1 CPD index

In the first step, the CPD index (see e.g. **?**) is computed for each year. The CPD is an index derived within the framework of the stochastic approach (see, among others, **?**). In order to introduce the said index, let p_{ij} be the real price of the i^{th} commodity ($i = 1,\ldots,N$) for the j^{th} country ($j = 1,\ldots,M$) expressed in Euros, and let PPP_j be the purchasing power parity (PPP) of the currency of country j with respect to a reference country currency, M hereafter. Accordingly, $PPP_{jk} = PPP_k/PPP_j$ is the PPP between currencies of any two countries j and k.

The additive form of the model, based on the Jevons geometric specification (**?**), yielding

the CPD index is

$$\begin{aligned}\ln p_{ij} &= \ln p_i + \ln PPP_j + \ln u^*_{ij} = \eta_i + \pi_j + \varepsilon_{ij} \\ &= \sum_{i=1}^{N} \eta_i D_i^P + \sum_{j=1}^{M-1} \pi_j D_j^C + \varepsilon_{ij} \end{aligned} \tag{1}$$

where D_i^P, D_j^C are binary variables and ε_{ij} is the random error term. The former dummy takes values 1 for commodity i and 0 otherwise, while the latter dummy is equal to 1 for country j and 0 otherwise. In matrix form, Eq. (**??**) can be written as

$$\underset{(MN,1)}{\boldsymbol{y}} = \underset{(MN,M+N-1)}{\boldsymbol{X}} \underset{(M+N-1,1)}{\boldsymbol{\beta}} + \underset{(MN,1)}{\boldsymbol{\varepsilon}}$$

where

$$\boldsymbol{y}' = \ln p_{11}, \ldots, \ln p_{1M}, \ldots, \ln p_{N1}, \ldots, \ln p_{NM},$$

$$\boldsymbol{X} = [D_1^P, \ldots, D_N^P, D_1^C, \ldots, D_{M-1}^C],$$

$$\boldsymbol{\beta}' = [\eta_1, \ldots \eta_N, \pi_1, \ldots, \pi_{M-1}]'.$$

The model in Eq. (**??**) has been estimated by ordinary least squares:

$$\hat{\pi}_j = \frac{1}{N} \sum_{i=1}^{N} \left(\ln p_{ij} - \ln p_{iM} \right)$$

and, accordingly $\hat{PPP}_j = e^{\hat{\pi}_j}$.

In the empirical application conducted for this study, the entries of $\boldsymbol{y}$ denote the log average annual real price per overnight stays aggregated by travellers' country of origin. In particular, they represent the prices paid by either leisure or business travellers (i.e. travellers' expenditure) living in one of the considered EU countries and travelling abroad. More precisely, p_{ij} is obtained as the ratio between the overall turnover value of the i^{th} commodity consumed by travellers from country j, say v_{ij}, and the (observed) number q_j of these travellers' overnight stays, that is

$$p_{ij} = \frac{v_{ij}}{q_j}. \tag{2}$$

Next, expressing v_{ij} as the product of the (non-observed) unitary price of commodity i by travellers from country j, say p^u_{ij}, and the (non-observed) number of purchased units of commodity i by the same travellers, say q_{ij}, allows to express p_{ij} as follows

$$p_{ij} = \frac{p^u_{ij} q_{ij}}{q_j} = p^u_{ij} \frac{q_{ij}}{q_j} = p^u_{ij} \omega_{ij}. \tag{3}$$

According to Eq. (**??**), p_{ij} can be read as the product of the (non-observed) unitary price of commodity i weighted by a coefficient ω_{ij}, which is the ratio between the total number of units of commodity i and the total number of overnight stays purchased by travellers from country j. Interestingly, the weight ω_{ij} depends on both the type of commodity i, which may be a lump purchase (like transport) rather than a commodity more frequently consumed and the behaviour of travellers from country j.

Thus, the use of overnight stays by travellers from a given country j, in computing prices, introduces a system of weights which, depending on the purchased quantities by the same travellers, reflects both the commodity specificity consumption and heterogeneity. In this way, expenditure items which are lump sums turn out to be weighed differently from commodities which are consumed more frequently.

Finally, it is worth noting that, overnight stays (also referred to as "nights"), being the sum of the number of people that are present at a given destination each night in a set time-frame, reflect both the number of visitors and the length of stay. Moreover, overnight stays represent a variable often employed in the tourism literature to compute price indexes/expenditure of travellers (see e.g., **???**).

For the scope of this analysis, the reference basket of the CPD index includes commodities typically consumed by travellers, such as transport, restaurants/cafés, accommodation, durable and valuable goods, as well as the other types of goods (not previously cited). Given this basket, the multilateral CPD index has been computed to assess travellers' expenditure variations of the said commodities across EU countries and over time.

Positive variation of the index implies an increase in the level of travellers' expenditures. To see why, let us consider the Laspeyres price index which, being expressed like the Paasche one in terms of prices and quantities in two periods, is of easy and immediate understanding.

To this end, bearing in mind that v_{ij} denotes the value (or aggregated expenditure) of the i-th commodity consumed by travellers from country j and q_j the number of travellers' overnight stays, the price of the i-th commodity consumed by travellers from country j can be expressed as $p_{ij} = \frac{v_{ij}}{q_j}$. Given this premise, the Laspeyresprice index for time/country j is given by

$$L_j = \frac{\sum_i p_{ij} q_0}{\sum_i p_{i0} q_0} = \frac{\sum_i v_{ij} \overline{q}_j}{\sum_i p_{i0} q_0}$$

where p_{i0} is the (average) price of the i^{th} commodity in the base country, q_0 the (average) number of travellers' overnight stays in EU and $\overline{q}_j = \frac{q_j}{q_0}$ is the percentage of overnight stays of travellers from country j compared to the EU average. In a similar manner, the Laspeyresprice index for time/country k would be

$$L_k = \frac{\sum_i p_{ik} q_0}{\sum_i p_{i0} q_0} = \frac{\sum_i v_{ik} \overline{q}_k}{\sum_i p_{i0} q_0}.$$

Looking at L_j and L_k, we see that an increase of the price of the i-th commodity implies an increase of both the expenditure v_{ij} and v_{ikl} as well as the price indexes L_j and L_k. The higher the price increase, the higher the increase in the expenditures and, thus, in the price indexes. Furthermore, the increase in the price index L is also directly proportional to the travellers' overnight stays (namely, $\overline{q}_j$ and $\overline{q}_k$). The greater is q_j with respect to q_k, the greater the expenditure v_{ij} and the variation of L_j with respect to v_{ik} and L_k, respectively. What proved here for the Laspeyere index also holds for multilateral/multi-period indexes, like the CPD one.

4.2 Cluster analysis

In the second step, the clusterization of travellers' countries of origin has been carried out according to the CPD index values. A hierarchical cluster analysis has been chosen due to the reduced number of observations (the EU countries) and the interest in studying the agglomeration process. The chosen procedure does not assume an ex-ante defined number of clusters. Among all the linkage criteria for the hierarchical clustering, the popular Ward's method (**?**) has been selected for this scope. This choice is mainly due to the fact that it is the only method based on a sum-of-squares criterion, and forms clusters by minimizing the within-

group variance at each step of the algorithm. It forms clusters by maximizing the inter-cluster variance and minimizing the intra-cluster variance, resulting in a sort of weighted squared distance between clusters (**?**). Thus, the Ward's method is particularly suitable to obtain the most accurate classification when there is the suspect of several clusters, all of somewhat substantial size (**?**).[1]

Using the CPD index implies a price level comparison across different EU countries at a specific time. Thus, the use of this multilateral price index leads to a cross-section. Alternatively, a time series would result if a multiperiod price index was constructed, based on the study of the prices referring to a unique country over time.

In acluster analysis, which is essentially an exploratory approach, the crucial issue is the identification of the optimal partition. The analysis may not lead to a unique solution, given that the interpretation of the hierarchical structure is context-dependent. To this aim, a visual approach can be undertaken by studying a dendrogram (which represents the sequence of nested partitions) and silhouette plots, together with several numerical criteria such as Dunn's validity index, Hubert's gamma, G2/G3 coefficient, or the corrected Rand index (just to cite a few; see for further details **??**). Unfortunately, flattenning dendrograms in k clusters to cut them off at constant height $k-1$ leads in some cases to poor clusters (**?**). Thus, the latter approach should be mixed with the researcher's knowledge of the phenomenon under study. To accommodate the latter, **?** have proposed a step-wise methodology to select a cluster analysis with representative clusters, based on the definition of a minimum threshold of observation belonging to each cluster.

Finally, considering available data, it is worth noting that in the presence of space data, such as the distance between travellers' destinations and their country of origin, additional approaches could be employed to the one proposed here. In the presence of units described by both multivariate time series and spatial information, the Dynamic Time Warping FuzzyC-Medoids for Spatial–Temporal Trajectories clustering algorithm with penalty terms can be implemented with several advantages (**?**). Similarly, the COFUST algorithm performs a fuzzy Partitioning Around Medoids clustering using copula-based approach which directly includes spatial information (**?**). This is however out of the scope of our research aims.

4.3 Theil index

The analysis is completed by taking advantage of the Theil index, which proves to be more suitable to investigate expenditure inequalities within and between clusters. The Theil index represents one of the possible measures to study the sigma-convergence which aims at investigating the reduction of the expenditure disparities across the EU countries (or groups of countries) over time (**?**). There is evidence of sigma-convergence when the dispersion/variation of the travellers' expenditure between different countries decreases over time. The closer to zero the Theil index of a set of countries, the more uniform the expenditure by travellers from these countries.

For the purpose of this paper, the Theil index is particularly useful to study the dispersion of the travellers' expenditure within and between the identified clusters. To this end, the index has been computed by using both the purchasing power parities (PPP) obtained from the CPD index values, and the original expenditure series when the aim was the measure of the dispersion of travellers' expenditure on a specific commodity between and within identified clusters.

In details, the Theil index for the k^{th} cluster, including N_k EU countries, is defined as follows

$$T_k = \sum_{j=1}^{N_k} \frac{P\hat{P}P_j}{\mu_k} \log\left(\frac{P\hat{P}P_j}{\mu_k}\right)$$

where $P\hat{P}P_j$ is the estimated purchasing power parity of the j^{th} country and

$$\mu_k = \frac{1}{N_k} \sum_{j=1}^{N_k} P\hat{P}P_j.$$

Alternatively, considering only the original expenditure series of the i^{th} commodity, the Theil index can be computed as

$$T_{ik} = \sum_{j=1}^{N_k} \frac{p_{ij}}{\mu_{ik}} \log\left(\frac{p_{ij}}{\mu_{ik}}\right)$$

where

$$\mu_{ik} = \frac{1}{N_k} \sum_{j=1}^{N_k} p_{ij}.$$

The Theil index for all clusters and commodities (T) as well as for all clusters and a specific

commodity i (T_i) can be decomposed as follows (**?**):

$$T = \sum_{k=1}^{K} s_k T_k + T_c = T_n + T_c$$

$$T_i = \sum_{k=1}^{K} s_k T_{ik} + T_{ic} = T_{in} + T_{ic}$$

where K is the number of clusters and s_k is the share of the population of cluster k, i.e

$$s_j = \frac{\sum_{j=1}^{N_k} Pop_j}{\sum_{s=1}^{N} Pop_s}$$

with Pop_j and $\sum_{s=1}^{N} Pop_s$ denoting the population of country j belonging to the cluster k and the EU population, respectively. The Theil index T is computed by using the estimated PPP of the EU countries, while the Theil index T_i uses the raw expenditure series of the commodity i. The terms T_n and T_{in} represent the weighted Theil indexes within clusters and capture the extent of the disparities of travellers' expenditure (overall and with reference to the ith commodity) within the EU country clusters. The terms T_c and T_{ic} are the Theil indexes between clusters and measure the disparities of travellers' expenditure between these clusters – overall and with reference to the ith commodity.

5 Data

The yearly data employed in the analysis comes from the official Eurostat statistics, refers to the period 2012-2018, and consist of macro data of the travel sector. The decision to take European data rests on the following three reasons: a) they are detailed in terms of expenditure categories (i.e., transport, restaurants/cafés, accommodation, durable and valuable goods, and other types of goods); b) they provide historic yearly information up to the year 2018 for almost all the EU countries; c) they meet the aim of this paper which is the individuation of countries of origin of higher spending travellers by means of an original procedure based on a travellers' expenditure analysis. This application offers interesting results for (European) destinations, in a time when the COVID-19 pandemic has stopped intercontinental leisure travel-

ling and only domestic and intra-European travelling (mainly by car) is partially allowed.

It is worth noting that, at the moment, only the data in the time span 2012-2018 are available to undertake the analysis. In fact, before 2012, the data have not been collected for the majority of the EU countries in a systematic way, and post-2018 travellers' expenditure by countries have not yet been published by Eurostat. Unfortunately, a measure of the COVID-19 impact on travellers' expenditure requires data not available yet.

Data are aggregated by travellers' country of origin and by type of expenditure for the following commodities: transport, restaurants/cafés, accommodation, durable and valuable goods, and other types of goods, incurred by travellers. Thus, the reference basket turns out to be composed of $N = 5$ commodities. The prices of these commodities have been computed as in Eq. (**??**). Since the data partially cover the stretch of time affected by the global economic and financial crisis before COVID-19, this analysis also highlights the potential short- and medium- term effects of the 2008 crisis on the travellers' behaviour and it can help in the post-COVID-19 recovery phase. Moreover, available data do not refer to travellers' destinations, but only to travellers' origin countries and are in aggregated form.

Due to an excessive lack of information on the mentioned expenditure categories for Norway, Poland, Romania, Sweden, and the UK, these countries have been excluded from the analysis to avoid distortions. Switzerland, for its geographic position, has been included in the analysis; thus 24 countries have been considered. The CPD index has been computed separately for leisure and business travels, assuming the euro-zone as the base country. The latter is a common practice followed, for instance, by Eurostat for the estimation of other multilateral price indexes with the euro-zone representing the average expenditure per overnight stays for all EU countries.

It is worth noting that this approach is easily reproducible and not computation intensive. The analysis has been conducted using the software R with own written functions for both the CPD and Theil index, while the cluster analysis takes advantage of functions in the STAT package, such as HCLUST.

6 Results

6.1 Price indexes and cluster analysis

Figure **??** shows the results obtained from the segmentation of EU countries of both leisure (left panels) and business (right panels) travellers, performed through a hierarchical cluster analysis based on the Ward's method and the Euclidean distances of the standardized values of the CPD index. Note that leisure and business partitions can be obtained using other hierarchical clustering methods, such as the complete, average, median or McQuitty method instead of the Ward's. Both leisure and business travellers have been grouped into three clusters: cluster C1 composed of countries with a constantly above average CPD index; cluster C3 including countries with a constantly below average CPD index; and cluster C2 embodying countries with an average CPD index.[2]

Looking at the dendrograms in Figure **??**, two clusters emerge for both leisure and business travellers. However, these clusters still exhibit a too high heterogeneity. Indeed, for leisure travellers, one of the two clusters collects low and average spenders, while for business travellers high and average spenders are assigned to the same group. Given our research aim, these results are scarcely informative and the partitions are too poor (**?**). Thus, following **?** and **?**, we have repeated and compared the cluster extraction process for different numbers of clusters (from five to two, given the sample size of 24 EU countries). The related percentage of representativeness in each cluster is shown in Table **??**. Then, in order to obtain a cluster analysis with representative clusters, we select a 15% threshold. From Table **??**, if four or more groups are selected, minority groups accounting for less than 15% of the sample are obtained (thus cluster with no more than three countries). Therefore, considering the complexity of the results' interpretation and the number of variables involved, a partition composed by three groups was examined for both leisure and business travellers. To confirm this choice, we have compared 26 different performance indexes taking advantage of the NBCLUST package in R (**?**). The latter confirms that for leisure travellers, the resulting optimal partition is the one with three clusters (according to 12 indexes, including the Silhouette one). Differently, concerning business travellers, there is an equal distribution of the number of indexes suggest-

ing a partition with two and three clusters. Thus, for a better comparability with the leisure partition and according to Table **??**, we have confirmed also for this case the three clusters solution as the optimal partition.

[Table 1 about here.]

[Figure 2 about here.]

[Table 2 about here.]

Differences emerge between clusters and market segments. As indicated in Table **??**, for both leisure and business segments, C1 can be labelled as "top-spenders", C2 as "average-spenders", C3 as "low-spenders".

In particular, Figure **??** shows that among leisure travellers, eight countries are by far the top spenders (Austria, Belgium, Denmark, Finland, Germany, Ireland, Luxembourg, and Switzerland). For business travellers, besides the aforementioned countries, top spenders also includes the Netherlands. On a temporal perspective, travellers from Denmark and Austria outperform in the leisure category, while Portugal under performs in the business one. Interestingly, travellers from Luxembourg experienced an expenditure boom in the last few years in both travel segments; Swiss business travellers have an oscillating expenditure, while Swiss leisure travellers have a steady declining expenditure over years. This analysis confirms that business travellers are higher spenders, and that the expenditure in restaurant, transport and accommodation by the average business cluster (C2B) exceeds that of the top leisure cluster (C1L) (Table **??**).

In this regard, it is also worth investigating the expenditure patterns across different categories. When targeting market segments with the highest yields, it should be remembered that travellers' expenditures on different type of commodities are dependent upon each other.

Indeed, travellers first distribute their budget between travel activities, and other goods and services; second, they allocate their travel budget to a specific, or multiple destinations, including their home country; third, they choose how to split their budget among various goods and services offered by the selected destination(s) (**?**). This empirical evidence is coherent with results shown in Table **??**.

Figure **??** describes the correlation between the expenditure categories considered in this analysis. As far as leisure travellers are concerned, a medium correlation emerges between restaurant, accommodation, and transport expenditure for the top-spenders (C1L), together with an inverse relation between durable goods and spending at restaurants. For the average-spenders (C2L), there is a positive correlation between restaurant and transport expenditure, together with an inverse relationship between durable and other types of goods. Finally, there is a strong negative correlation between restaurants and other types of goods, and a mild one between expenditure on accommodation and transport for the low-spenders (C3L).

Among business travellers, there is a strong and constant correlation between accommodation and transportation expenditure. Moreover, in general, a positive correlation between restaurant, accommodation, and transport, and a negative correlation between restaurant and other types of goods emerge for both top-spenders (C1B) and low-spenders (C3B).

[Figure 3 about here.]

6.2 Expenditure convergence

The convergence of travellers' expenditure, between and within clusters, concludes this empirical analysis. It has been carried out at two levels: the first one, more general, is based on the use of the PPPs obtained from the CPD index[3], while the second one is commodity specific as it focuses on travellers' expenditure for each good in the reference basket.

Figure **??** provides the Theil index, computed by using the PPPs, and its decomposition highlighting the disparity of travellers' expenditure, within and between clusters, for both leisure and business travellers (**?**).

Considering leisure travellers, the PPPs disparities increase in 2013, decrease in 2014-2015, and increase again in the last three years. For business travellers, the PPPs disparities increase in 2015, to decrease again in the following two years, to finally show a progressive divergence in 2018. In both cases, these trends are mainly due to the T_n component while the between cluster component, T_c, progressively decreases across business travellers clusters and slightly increases for leisure ones. Thus, business travellers' expenditure shows a meaningful sigma-convergence, reflecting a reduction of dispersion/variability across clusters and over

time. In other words, the average PPPs of low, average and high spending travellers converge over the considered timespan.

[Figure 4 about here.]

Figure **??** reports the ratios T_n/T and T_c/T corresponding to the share of expenditure disparities explained by country disparities within clusters (intra-cluster inequalities), and cluster disparities (inter-cluster inequalities), respectively. Figure **??** shows the different expenditure disparity (ED, hereafter) patterns across travellers' types. On the one side, the intra-cluster ED of leisure travellers reaches the minimum value of 34% in 2015 and its maximum value of 62% in 2018, while the inter-cluster ED reaches its maximum value of 66% in 2015 and its minimum value of 36% in 2018. On the other side, the intra-cluster ED of business travellers has the minimum value (48%) in 2012 and the maximum one (78%) in 2018, while the inter-cluster ED reaches the maximum (52%) in 2012 and the minimum (22%) values in 2012 and 2018, respectively.

Therefore, there are high intra-cluster EDs for business travellers and high inter-cluster ED for leisure travellers. As a result, the disparities of the travellers' expenditure in the last year, based on the PPPs, may be largely imputed to the intra-cluster inequalities (62% and 78% for leisure and business travellers respectively).

[Figure 5 about here.]

Finally, Figure **??** shows the Theil index of each cluster (without population weights). For both traveller types, the average spender clusters are the ones whose expenditure remains more stable over years, with a slight increase for business travellers' expenditure over time. Among leisure travellers, ED increase since 2016 for both top and low spenders with the ED of the latter reaching the level of top spenders in 2018. Business travellers' expenditure shows higher divergence within top and low spending countries when compared to the same segments of the leisure travellers.

[Figure 6 about here.]

The evaluation of the commodity-specific sigma-convergence of the expenditure concludes the analysis, using raw expenditure series. From Figure **??**, it can be noticed that the ED shows the highest divergence for durable goods. This can be justified by the fact that these commodities are of a heterogeneous type. Moreover, while leisure travellers' EDs are quite stable for the remaining commodities, business travellers' EDs show a meaningful reduction for durable goods and a slight convergence for accommodation and restaurant expenditure. Differently, the EDs of other types of goods have an opposite trend.

[Figure 7 about here.]

Moreover, Figure **??** shows the commodity-specific average ED for each cluster. Some interesting stylized facts emerge. Considering business travellers, a great difference between CB1 and the other clusters in terms of ED is revealed. The increasing trend of ED over time for business clusters seems to be confirmed in almost all categories while divergence across clusters seems to be almost constant. Exceptions refers to: restaurant (CB2 approaches CB1 levels), durable goods (CB1 reduces, reaching CB2 levels) and other type of goods (CB32 shows a reduction over time). Similar observations also hold for leisure travellers. In particular, they show in general a lower ED than business travellers, which is cognisant of the fact the latter do not pay by themselves for restaurant, transport, and accommodation costs.

Figure **??** further confirms the effectiveness of this methodological proposal. Even if the segmentation has been done using only the multilateral CPD index over time, also cluster expenditure levels are coherent with the two obtained partitions.

Finally, Figure **??** compares travellers' ED across different commodities, travellers' types and segments evaluated by the Theil index. In general, the ED is low in all clusters for restaurant, transport and accommodation. Main differences appear, as expected, for durable goods and other type of goods (only for business travellers). The expenditure convergence observed here further enriches the profile of each cluster and suggests possible long-term trend.

The COVID-19 pandemic might change these patterns. However, it must be remarked that EU countries, which have been exposed to both European and nation-specific pandemic measures, have been considered. Thus, it might be expected that the empirical evidence may be partially confirmed also in 2019 and 2020, given the use of dimensionless multilateral PPPs.

[Figure 8 about here.]

[Figure 9 about here.]

7 Discussion and conclusions

In summary, the main aim of this research is to apply an innovative two-step procedure to analyse travellers' aggregate expenditure, with the objective of identifying top spending countries for intra-European travellers. This information can be very useful in targeting the reopening of the tourism sector post-COVID-19. According to this approach, first, the expenditure variations are computed by means of the CPD index, then the values of the same are used to determine clusters of countries.

The paper does not aim at developing a new statistical methodology to analyze travellers' expenditure over time but uses, in a proper way, some statistical tools so as to capture the dynamics of the phenomenon under study. The idea of clustering countries, on the basis of the values of a price index expressing travellers' expenditure changes over time, is definitely new, relevant and leads to interesting empirical results. To gain a deeper insight of travellers' expenditure, we have also developed a convergence analysis between and within clusters by resorting to the Theil index with either the purchasing power parities or the original series of travelers' expenditure. This innovative analysis of expenditure has brought to the fore the existence and dynamics over time of the disparities within and between clusters. The use of both the convergence analysis and the Theil index, not commonly and jointly used in tourism analyses, has allowed the analysis of traveller's trend expenditures under several profiles, both globally and with reference to a commodity-specific consumption.

Moreover, the correlation analysis highlights spending patterns and allows the understanding of how expenditure is allocated. This analysis is particularly useful to destination managers as they can package different services together in such a way as to maximise travellers' expenditure. Further insight on the importance of focusing on expenditure categories can be found in **?**. By identifying top spending countries, destinations can set suitable marketing strategies to target clusters offering higher yields through a lower number of travellers and therefore exerting less pressure on the local environment and community. Thanks to this information, destination planners can make informed decisions in targeting business and leisure travellers. For example, according to the results obtained from the empirical analysis, business travellers in the top and medium segments spend more than leisure travellers in the same

segments. However, travellers in the low business segment spend less than leisure travellers in the top segment. Therefore, despite business travellers being normally considered as a high spending segment, this analysis has revealed that it is more profitable to target top leisure spenders rather than low business spenders.

The results of this study also show a difference in the variability of leisure and business expenditure. The convergence analysis highlights the dynamics, stable or variable over time, of travellers' expenditure disparities. This allows a check if expenditure disparities of different types of travellers (from different countries) tend to exhibit a common trend or not. This analysis has been also carried out for each specific commodity. According to the results obtained in this paper, the higher expenditure disparities are registered for business travellers in general. This means that there is a high variability between the expenditure of the business segment. As for the single commodities, the highest variability has been registered in the purchases of durable and other types of goods by top business spenders. When marketing top-spenders, tourism destinations and operators have to design suitable packages for this segment, accounting for variations across countries. Moreover, the approach introduced here can be usefully employed, also in the post-COVID-19 phase, to reduce the effects of mass tourism which, in some cases, has led to residents being discontent and resistant to further tourism expansion in their area (**?**). The proposed methodological approach overcomes the general shortcomings of classical descriptive statistics and clustering techniques. The latter are not capable of capturing simultaneously time and geographical area dynamics of a phenomenon, such as commodity specific information (expenditures) across countries and times.

By identifying top-spenders, destinations can set suitable marketing strategies to attract visitors who have the means to spend the most. This could lead to higher yields with a lower number of travellers, thus reducing the probability of overcrowding hotspots and destinations and thereby supporting the current, and likely ongoing, measures of social distancing.

On the one hand, the results obtained with this methodology can be considered preliminary to further analyses aiming at the construction of competitiveness and tourist attractiveness indexes and, on the other hand, they can provide support for market research at a macro level.

Furthermore, the aggregation of countries which share common features of interest is an

approach which provides useful insights for redefining marketing strategies. For instance, it can be used to unify communication channels and marketing campaigns (e.g. German-speaking countries).

Finally, it is worth noting that the groups identified by the approach implemented here reveal an interesting territorial contiguity, for both the business and the leisure sectors. This feature can be justified by resorting to a multidisciplinary approach (sociology, economy, culture, etc.).

The scope of this study could be broadened under several aspects: by considering the expenditure patterns of leisure and business over a larger time frame, incorporating non-EU countries, and adding the total number of residents per country so as to identify the potential size of the market to attract. When space information is also collected, specific time-space clustering algorithm, such as the one proposed in **?**, can be implemented.

Moreover, a similar analysis could be conducted to describe the new post COVID-19 pandemic scenario, given that current travel flows have been completely disrupted around the world. Finally, the dependence between the different expenditure categories, preliminarily studied here, could be further investigated by using models such as copulas. Indeed, copulas could be integrated in a fuzzy clustering algorithm to perform a time-space cluster analysis (**?**).

Notes

[1]Alternatively, and especially in the presence of many observations to be clustered, a *k*-means algorithm could be easily implemented.

[2]It is worth noting that the same composition of clusters emerges if a *k*-means algorithm is used. Indeed, its application, based on the identification of three clusters, results in the same EU countries partition. This is just a robustness check which validates the use of the Ward's linkage method for the scope of the present analysis.

[3]This is motivated by the fact that the nature of the Theil index is suitable only for non-negative quantities. The PPPs are just the exponential of the CPD indexes as illustrated in Section **??**.

A Supplementary materials

[Table 3 about here.]

[Figure 10 about here.]

Tables

Table 1: Percentage of sample within each cluster

Clusters Leisure	5 (%)	4 (%)	3 (%)	2 (%)
A	33	33	33	33
B	25	33	33	67
C	8	12	33	-
D	12	21	-	-
E	21	-	-	-
Clusters Business	5 (%)	4 (%)	3 (%)	2 (%)
A	25	38	38	58
B	12	12	42	42
C	29	29	21	-
D	12	21	-	-
E	21	-	-	-

Table 2: Average expenditure per night by clusters. Standard deviations in parentheses

Travellers	Cluster	Restaurant	Transport	Accommodation	Durable goods	Other goods
Leisure	C1L (top-spenders)	20,76 (7,1)	33,61 (6,62)	41,00 (11,77)	3,85 (3,77)	29,89 (13,92)
	C2L (average-spenders)	11,83 (5,28)	17,87 (4,75)	18,4 (5,49)	1,34 (1,49)	16,06 (6,42)
	C3L (low-spenders)	8,24 (2,96)	9,94 (2,79)	9,58 (3,82)	0,53 (1,11)	8,93 (3,24)
Business	C1B (top-spenders)	37,93 (20,38)	133,06 (55,25)	115,15 (31,65)	6,09 (5,63)	41,36 (30,63)
	C2B (average-spenders)	29,05 (14,42)	64,00 (25,44)	61,09 (19,19)	2,87 (2,60)	27,64 (8,73)
	C3B (low-spenders)	16,31 (6,97)	30,65 (15,95)	33,38 (15,44)	1,86 (1,88)	11,77 (6,51)

Table 3: Average expenditure per night given each commodity by years and clusters

Commodity	Year	CL2	CL3	CL1	CB1	CB2	CB3
Total	2012	57.21	32.76	100.45	311.92	145.68	72.89
	2013	57.40	32.39	105.22	344.85	180.47	70.29
	2014	57.58	35.91	104.82	287.50	159.06	76.62
	2015	61.54	39.14	105.87	336.57	182.32	100.59
	2016	63.40	37.17	111.12	326.08	188.80	99.07
	2017	69.27	41.59	118.73	353.65	220.26	113.59
	2018	78.18	44.18	120.48	374.55	215.94	124.74
Restaurant	2012	7.75	5.92	16.72	35.47	18.66	10.13
	2013	10.33	5.81	16.00	35.68	28.67	11.68
	2014	10.28	7.44	16.37	32.82	25.93	12.11
	2015	12.14	9.09	19.16	37.47	30.79	18.17
	2016	11.74	8.95	20.39	43.41	30.30	18.15
	2017	12.68	10.47	17.05	34.12	33.81	21.15
	2018	14.94	10.43	22.06	46.57	35.19	22.77
Transport	2012	15.85	10.47	25.51	125.39	51.68	22.24
	2013	15.51	9.33	27.59	141.36	61.14	21.22
	2014	15.87	10.63	27.56	113.71	55.25	24.59
	2015	17.06	10.49	26.15	136.32	63.47	35.61
	2016	16.91	10.13	28.61	126.15	65.76	33.70
	2017	18.42	10.78	31.34	141.40	76.53	35.83
	2018	21.07	12.08	31.63	147.08	74.19	41.36
Accommodation	2012	16.55	7.63	32.61	103.36	50.06	23.88
	2013	16.32	7.60	32.84	117.49	59.38	24.19
	2014	16.52	8.82	32.67	99.22	51.49	27.13
	2015	17.69	9.39	32.12	115.70	59.77	34.95
	2016	18.40	9.11	34.35	115.81	60.97	35.68
	2017	20.45	10.42	37.15	125.88	73.78	41.42
	2018	22.49	11.13	36.96	128.60	72.19	46.42
Durable goods	2012	2.11	0.77	4.19	11.20	1.49	2.78
	2013	1.72	0.55	3.85	7.66	2.84	0.73
	2014	1.34	0.56	3.39	5.05	4.07	1.33
	2015	1.01	0.66	2.90	4.99	2.55	1.51
	2016	0.99	0.51	3.29	3.67	3.08	1.42
	2017	0.70	0.44	2.46	6.21	3.75	3.37
	2018	0.57	0.41	2.43	3.84	2.31	1.90
Other goods	2012	14.94	7.98	21.41	36.51	23.79	13.87
	2013	13.53	9.09	24.95	42.67	28.46	12.48
	2014	13.58	8.45	24.83	36.70	22.33	11.47
	2015	13.63	9.51	25.54	42.09	25.74	10.35
	2016	15.36	8.47	24.48	37.04	28.69	10.11
	2017	17.02	9.49	30.73	46.04	32.40	11.83
	2018	19.11	10.14	27.39	48.46	32.06	12.30

Figures

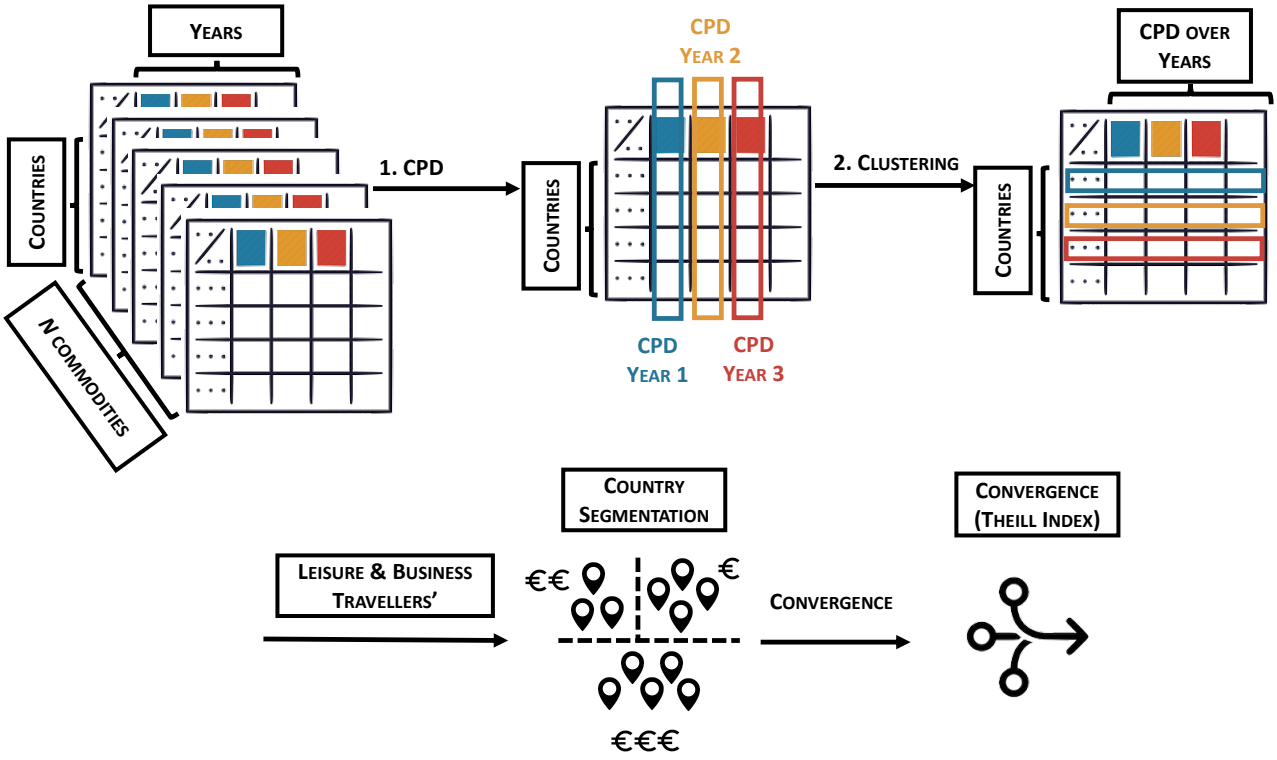

Figure 1: The methodological proposal. *Source*: own elaboration

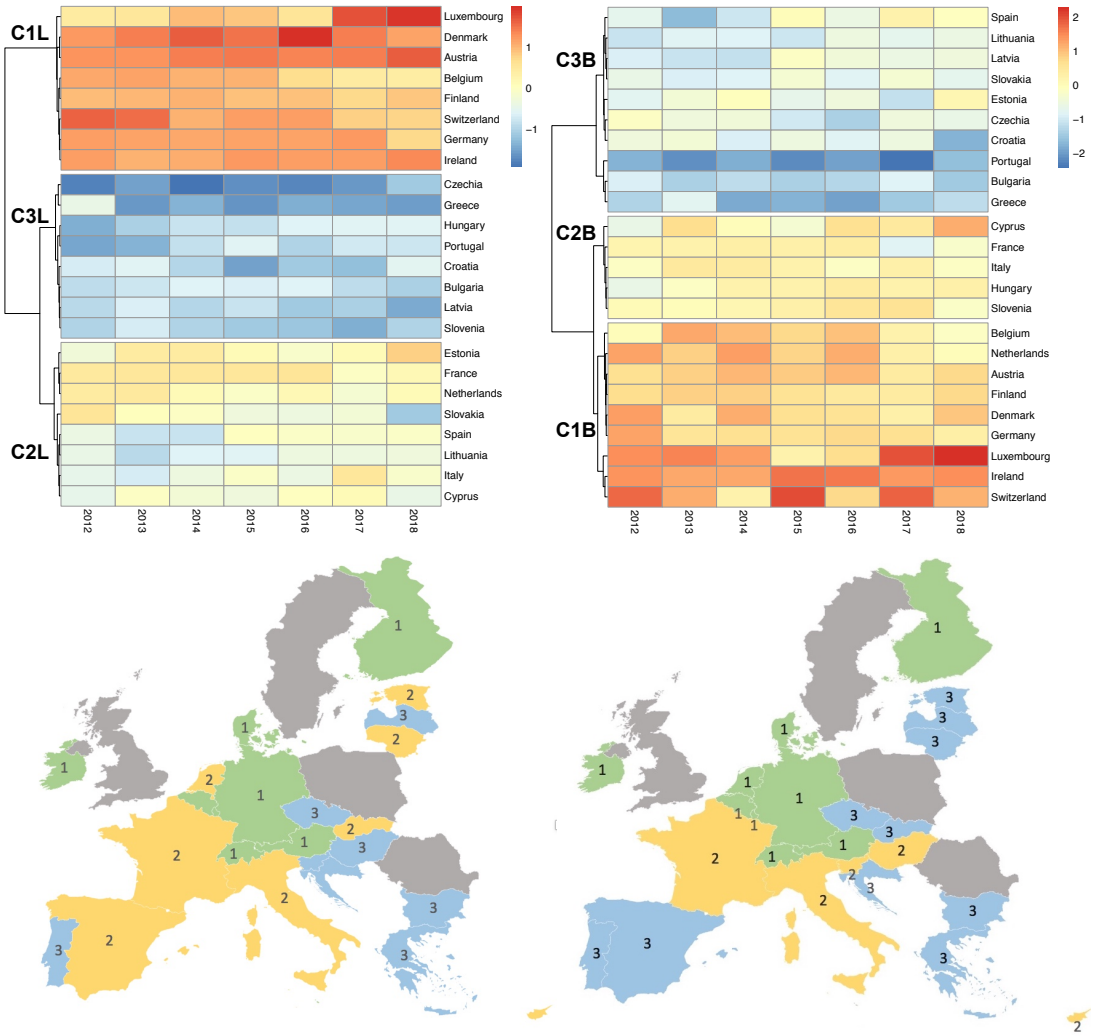

Figure 2: Clusters (top) and maps (bottom) of leisure (left) and business (right) travellers
Source: own elaboration

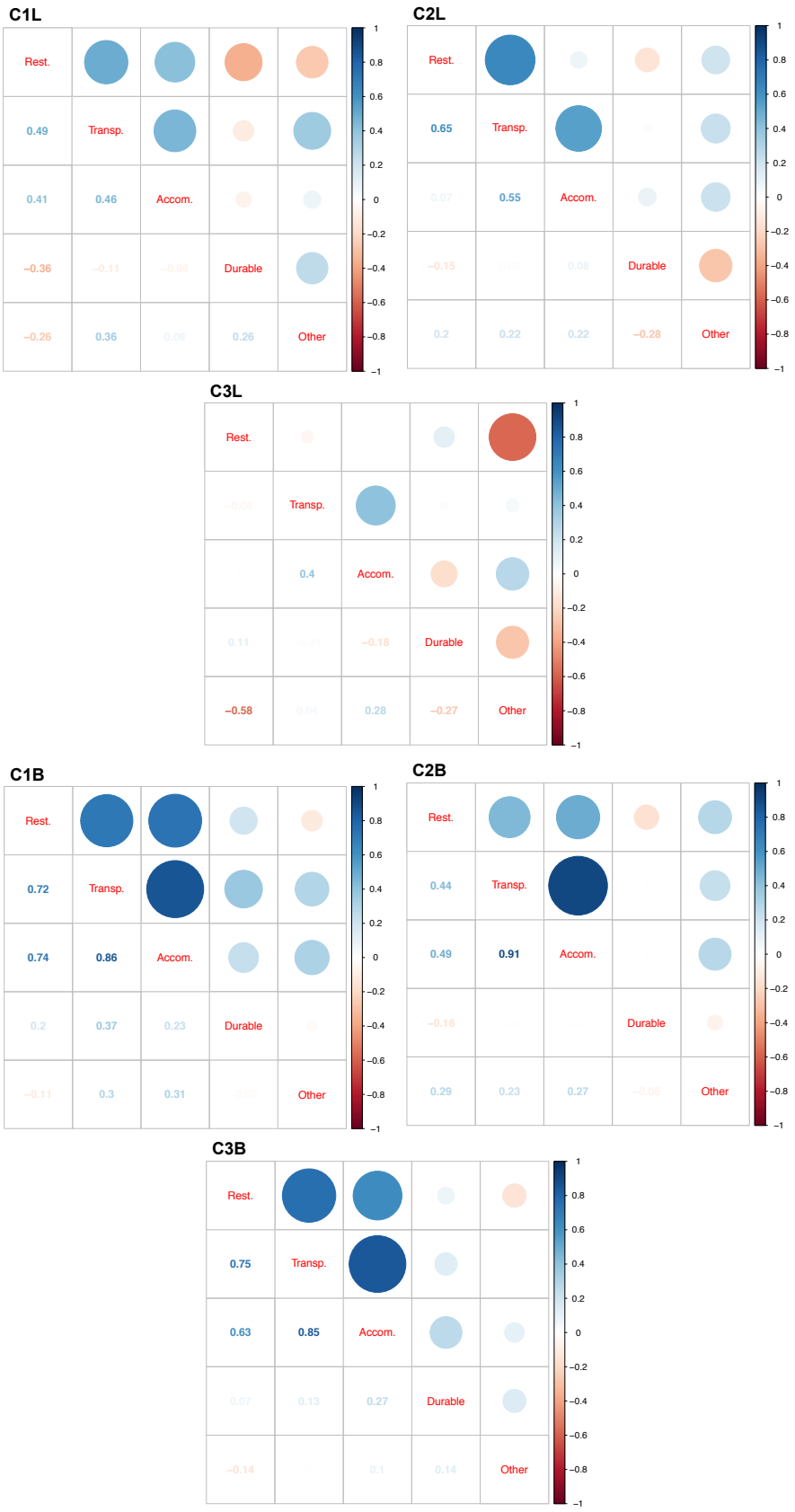

Figure 3: Correlation plot among expenditure categories for travellers: leisure (top), business (bottom)
Source: own elaboration

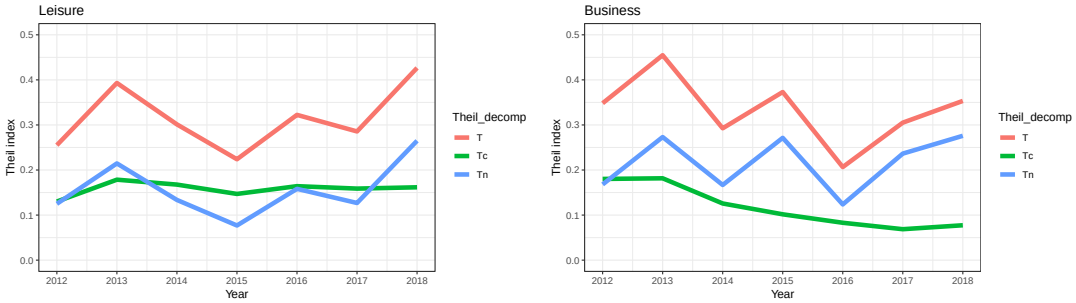

Figure 4: Decomposition of the Theil index for leisure (left) and business (right) travellers. *Source*: own elaboration

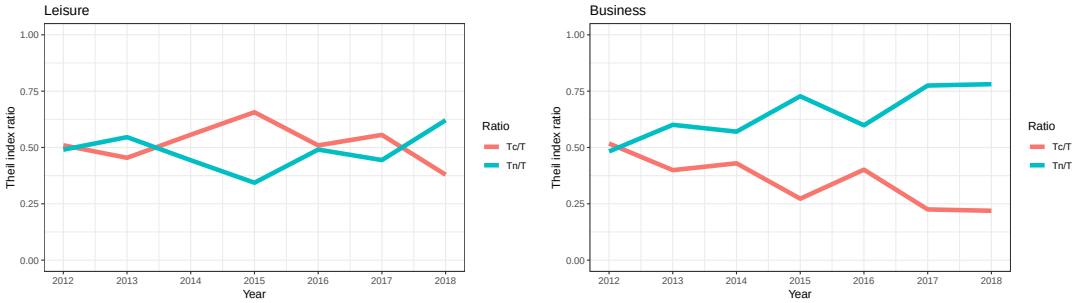

Figure 5: Ratio of the Theil index component for leisure (left) and business (right) travellers. *Source*: own elaboration

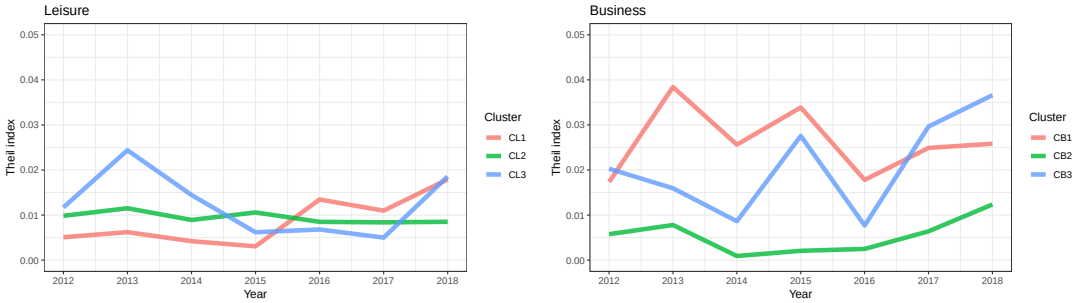

Figure 6: Theil index by clusters of leisure (left) and business (right) travellers. *Source*: own elaboration

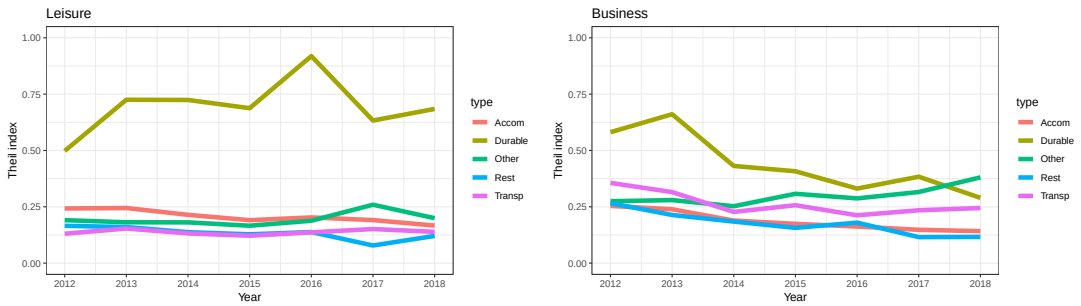

Figure 7: Theil index of the five commodity expenditure by leisure (left) and business (right) travellers. *Source*: own elaboration

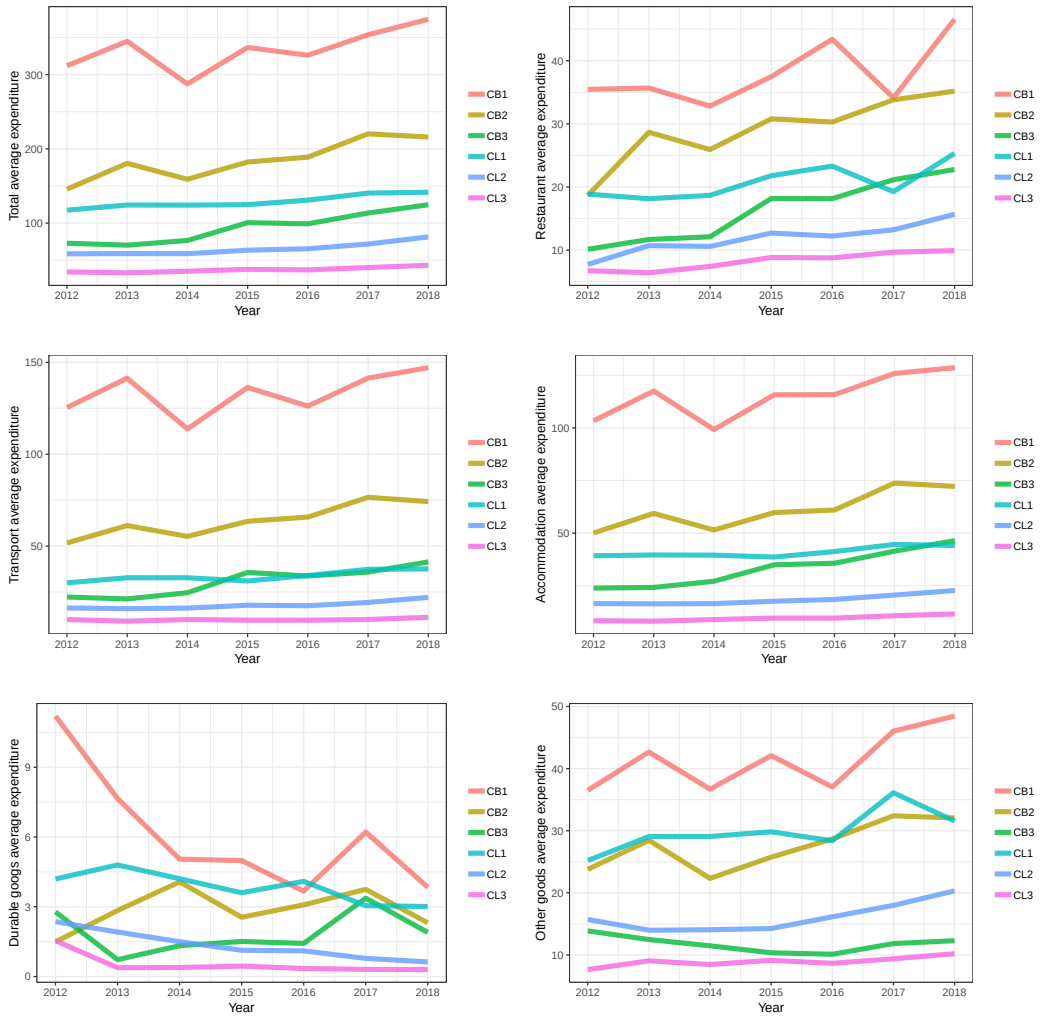

Figure 8: Average expenditure per night given cluster of the different available commodities. *Source*: own elaboration

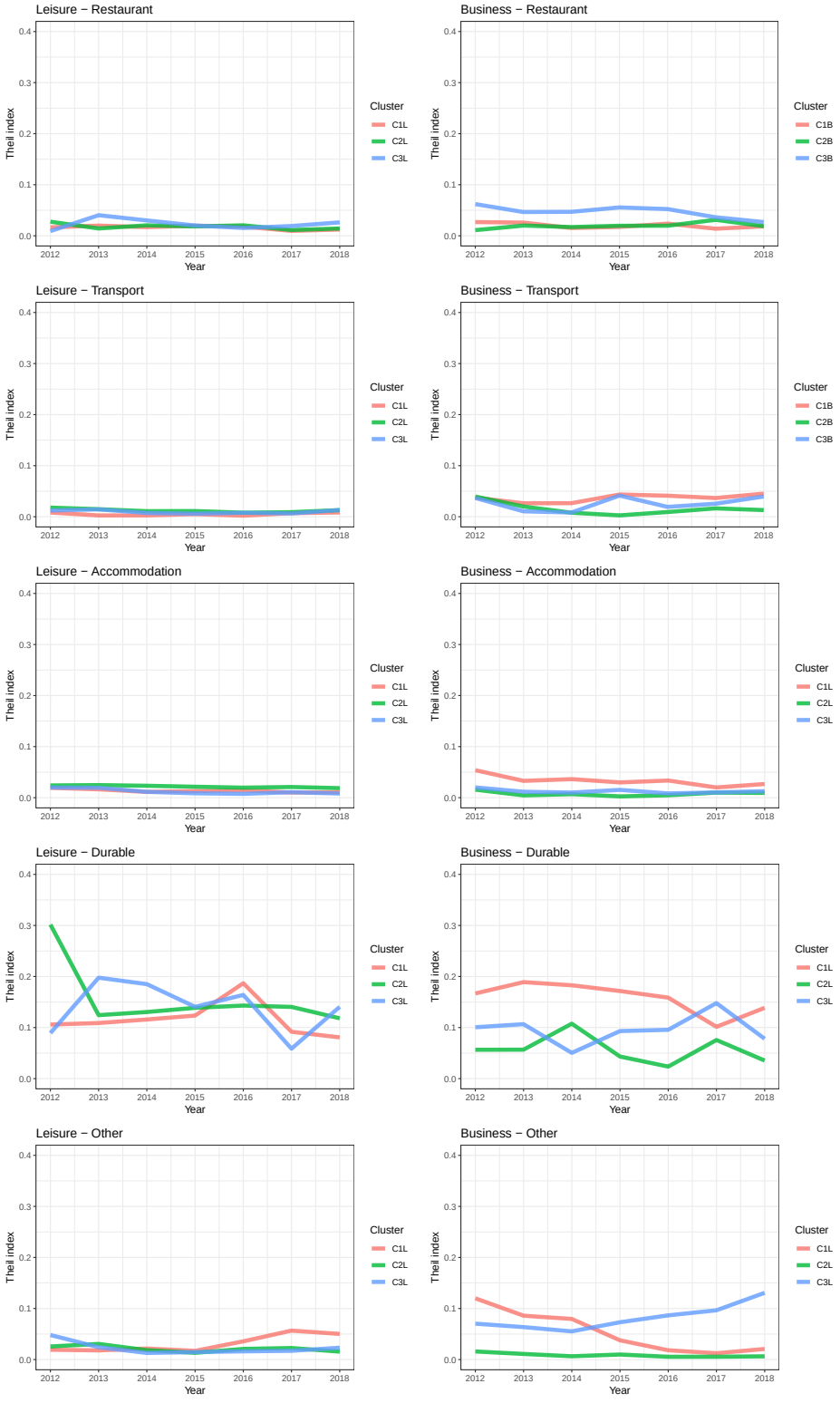

Figure 9: Theil indexes of commodity expenditure given clusters of leisure (left) and business (right) travellers. *Source*: own elaboration

Figure 10: Expenditure series of each country given clusters of leisure (left) and business (right) travellers. *Source*: own elaboration