Analysing cluster evolution using repeated cross-sectional ordinal data

Marta Disegna\textsuperscript{a,}\textsuperscript{*}, Pierpaolo D’Urso\textsuperscript{b}, Riccardo Massari\textsuperscript{b}

\textsuperscript{a}Accounting, Finance \& Economics Department, Faculty of Management, Bournemouth University, 89 Holdenhurst Road, Bournemouth, BH8 8EB, United Kingdom

\textsuperscript{b}Department of Social Sciences and Economics, Sapienza University of Roma, P.le Aldo Moro 5, 00185 Roma, Italy

Abstract

This study contributes to the existing literature on tourism market segmentation by providing a new matching-clustering procedure that allows patterns of behaviours to be identified using repeated cross-sectional surveys. By extracting equivalent samples over time, the matching method allows inter-temporal cluster analyses to be performed so that a deeper insight into a phenomenon can be obtained beyond the traditional aggregate level of understanding. The paper provides a step-by-step description of the matching-clustering procedure that can be easily replicated, both within and outside the tourism field, when repeated cross-sectional data are available. From a practical and managerial perspective, the proposed procedure helps destination managers and municipalities to describe and verify the efficacy of policy and strategies adopted over years without the necessity to rely on longitudinal surveys, which are often difficult to conduct.

Keywords: Repeated cross-sectional data; Matching; Evolution; Ordinal data; Fuzzy data; Fuzzy clustering

\textsuperscript{*}Corresponding author

Email addresses: disegna@bournemouth.ac.uk (Marta Disegna), pierpaolo.durso@uniroma1.it (Pierpaolo D’Urso), riccardo.massari@uniroma1.it (Riccardo Massari)
1. Introduction

In the recent tourism literature, the necessity to perform accurate analyses to obtain information on temporal, spatial or cultural changes regarding a particular phenomenon has been highlighted (Cang et al., 2017; Amaro et al., 2016; Song et al., 2012, 2011), especially in consideration of the increased multiculturalism that characterises the worldwide tourism markets (Jin et al., 2017). In this paper, a new matching-clustering procedure is suggested to obtain information on the evolution of tourists’ behaviour when repeated cross-sectional data are available. The increasing availability of large repeated cross-sectional and longitudinal surveys, at both national and international levels, has given researchers the possibility to study how different tourism-related phenomena evolve either within or between countries. While in longitudinal surveys the same sample of units (individuals, households, firms, cities, etc.) is tracked over time, in the repeated cross-sectional surveys different units are involved each time. Therefore, longitudinal surveys are designed to track individual behaviour gathering information on the evolution of a phenomenon while repeated cross-sectional surveys allows comparing years by using aggregate data (Wooldridge, 2012). When repeated cross-sectional surveys are used to obtain temporal comparisons at an individual level the results can be biased producing wrong policy and managerial conclusions. On the other hand, repeated cross-sectional surveys are frequently designed to capture feelings, emotions, motivation, and tourism behaviour (see for instance the Eurobarometer survey on “Preferences of Europeans towards tourism” or the World Tourism Barometer collected by the UNWTO), making it a necessity to develop an adequate procedure to compare this kind of individual information over time.

The idea behind the matching-clustering procedure suggested in this paper is that more detailed information on the evolution of a particular phenomenon can be gathered at cluster-level when units, observed by repeated cross-sectional surveys, are comparable across time. Figure 1 schematically describes the suggested 7-steps procedure and the “What-How-Why” questions (i.e. the aims,
the methods, and the reasons) behind the adoption of each particular step and

35 technique. The first and the last steps, i.e. the “Matching” and the “Compar-

ing partitions”, are the main innovations of this study. In fact, these two steps

are not usually included in any traditional cluster analysis since they serve the

particular aim of performing and describing comparisons when the same set of

segmentation variables is collected on different samples of units.

The core five steps, going from the “Data recoding” to the “Labelling & pro-

filing” phase, are the ones that generally characterise any clustering analysis

(Everitt et al., 2011). Consequently, the aims (“What”) of these five steps are

commonly known in the literature of market segmentation. However, the tech-

niques (“How”) that can be adopted to accomplish each of these steps are not

unique since the choice depends on both data and research aim (“Why”).

[Insert Figure 1]

The matching-clustering procedure illustrated in this paper combines (1)
the matching method with (2) fuzzy numbers and (3) fuzzy clustering algo-

rithm to overcome three problems, respectively: (1) the comparison between

cross-sectional samples, (2) the imprecision embedded in ordinal data, and (3) the

uncertainty in the assignment of each unit to clusters. Therefore, this pa-

per specifically focuses on the situation in which (1) repeated cross-sectional

surveys are available, (2) segmentation variables are measured by means of or-

dinal scale, such as Likert-type scale, and (3) the hidden market structure is

unclear, making it unreasonable to assign each unit solely to one cluster. Read-

ers should note that the focus of this paper is on the procedure itself, not on

the techniques (“How”) suggested. In fact, each technique can be replaced by

any other suitable technique better able to address the “What” questions ac-

cording to either the research aims or the nature of the segmentation variables;

however, the matching-clustering procedure should remain the same if repeated

cross-sectional surveys are used.

The next section provides a review of the relevant literature describing these

three problems while the third section provides a more technical discussion of

3
each step of the suggested procedure. The fourth section describes the empirical
data used to illustrate the matching-clustering procedure, while the findings are
discussed in the fifth section. The final section provides conclusions, highlighting
both limitations and further directions.

2. Literature review

2.1. Statistical matching

In order to carry out meaningful temporal comparisons, it is necessary to
perform quantitative analyses on harmonised samples, i.e. samples made up
from the same group of units (individuals, households, companies, cities, etc.)
observed over time, or paired samples, i.e. samples of units that are similar
according to a set of controlling factors defined \textit{a priori}. Harmonised samples
result from either longitudinal or panel surveys. Even if the same group of units
is studied across time, researchers have to interpret carefully changes observed at
the micro-level \cite{Brandt2018}. In fact, results can be compromised, for instance,
by a socio-demographic or economic change: an individual could change civil
status, becoming separated or getting married, or employment status, becoming
employee or retired; household composition can change on the arrival of new
babies or the death of a family member; a company can become bigger changing
its legal status. The main clustering approaches suggested in the literature to
aggregate units characterised by similar behaviour across time are the (1) model-
(2) feature- and (3) observation-based approaches \cite{Disegna2015, Caiado2015}

When repeated cross-sectional surveys are used to see how a key relation-
ship has changed over time, a common procedure is to create a pooled dataset
aggregating cross-sectional data from different years. Although extensive econo-
metric techniques have been developed for the analysis of pooled cross-sectional
data \cite{Wooldridge2012}, in tourism market segmentation the cluster analysis is
usually performed on the pooled dataset losing the information on the evolution
of the phenomenon \cite{Ferrer-RosellCoenders2018}. Another
common approach adopted in tourism literature, is to compare directly clusters over time, even if samples made by different units are used (see for instance Cang et al. [2017]). However, as stated at the beginning of this section, comparison should be made on either harmonised or paired samples. In fact, comparing clusters that have been obtained on different samples can lead to misleading and imprecise conclusions, since it is impossible to know whether the changes finally observed in the phenomenon are due to actual changes in the phenomenon or if they are due to structural differences in the samples.

In this study the adoption of a statistical matching procedure is suggested to overcome this limitation and identify paired samples in repeated cross-sectional surveys. As discussed in D’Orazio et al. [2006], statistical matching procedures aim to match two or more datasets in order to create a more informative integrated dataset. Statistical matching procedures can be adopted when (1) a set of variables is commonly observed, (2) different units, belonging to the same population, are observed in different datasets and (3) any couple of datasets are conditionally independent given a set of common variables used for the matching. Since datasets obtained through repeated cross-sectional surveys satisfy all these conditions, it is reasonable to adopt a statistical matching procedure. In this case, the final aim is not, as usual, the creation of an integrated dataset but the identification of a perfectly matched pair of units belonging to different datasets collected over years. When similar sub-samples are identified, a cluster analysis can be performed on each sub-sample and the results can be compared based on the variables used in the matching procedure and any other variables against which the sub-samples are statistically independent. In this way, the comparison will highlight changes in the phenomenon under observation and not changes in the samples used to obtain the clusters.

2.2. Data recoding

Feelings, emotions, motivation, and consumer behaviour are complex psychological processes vaguely captured through quantitative imprecise measurements (Hung & Yang 2005, D’Urso 2007). In both academia and industry, ordinal
scales, such as Likert-type scales, are commonly used to capture human feel-
ings since they are user-friendly, easy-to-develop and easy-to-administer (Coppi et al., 2012; Benítez et al., 2007; Li et al., 2013). However, the drawbacks of this kind of measurements are widely debated in the recent literature (Kampen & Swyngedouw, 2000; D'Urso et al., 2016). In particular, it is important to under-
line that (1) subjective feelings and judgements, vague by definition, are cap-
tured by means of linguistic expressions (Benítez et al., 2007; Coppi & D’Urso, 2002), that are (2) subjectively interpreted by respondents who attribute specific meanings to each expression that depend on their personal background (Davidov et al., 2014). Respondents are forced to convert their feelings and opinions into specific linguistic expressions, usually coded into natural numbers, and these conversions can be inaccurate, causing loss of information (Benítez et al., 2007; D’Urso 2007). Furthermore, these values seem inadequate to collect subjective evaluations since respondents might either attribute different meanings to the numbers or have difficulties in understanding the differences in the categories used in the Likert-type scale (see, for instance, D’Urso et al., 2013a; Benítez et al., 2007; Coppi & D’Urso, 2002). For instance, analysing European Social Survey (ESS) data, Piurko et al. (2011) demonstrated that the 11-point Likert-type scale (from 0 to 10) may be differently interpreted, assuming a variety of meaning according to the different political orientation of the European countries (i.e. liberal, traditional and post-communist). Respondents “have been socialized in diverse economic and cultural backgrounds, speak different languages, and might understand certain ideas or concepts in varying ways or in a culturally specific manner” (Davidov et al., 2014, p. 56) and “even if an item is similarly understood, the scale measuring the item may be used differently across populations”.

Partially to overcome these drawbacks and to be able to more precisely capture human feelings/thinking, the simple visual analogue scale or the fuzzy rating scale should be adopted instead of the traditional Likert-type scales (Gil & González-Rodríguez, 2012; De la Rosa de Sáa et al., 2015). The visual analogue scale requires respondents to express their opinion indicating their positions
along a continuous line between two end-points. On the other hand, the fuzzy rating scale is a modification of the traditional Likert-type scale that allows respondents to indicate both a preferred point on a scale and latitudes of acceptance on either side. Nowadays, these methods have not been extensively adopted in tourism and marketing research mainly because: (1) the mathematical and statistical methods/tools necessary to analyse the responses are quite complex; (2) adequate software is required to decode the visual analogue scale and fuzzy rating scale answers collected; and (3) many academic researches are based on secondary data in which the traditional ordinal scales are typically used [Dolnicar 2002; Grün & Dolnicar 2016].

When a priori correction mechanisms cannot be adopted, fuzzy sets theory, firstly introduced by Zadeh (1965), can be a posteriori embraced to limit the imprecision and vagueness inherent to both Likert-type variables and human thinking [Wang et al. 2014; Zhang & Lipkin 2013; Pérez-Gladish et al. 2010; Chang & Wang 2006; Hung & Yang 2005; Mikhailov 2003]. In particular, Likert-type variables can be recoded into fuzzy variables, before any further analysis, by associating a range of possible values to each individual score/expression. As demonstrated by Hu et al. (2010), this kind of recoding not only allows reduction of the imprecision entailed in ordinal scales, but it also allows improvement of the results in terms of reliance and effectiveness of further analysis.

2.3. Selection of the clustering algorithm

Among a posteriori segmentation approaches, cluster analysis remains the most popular method and the most frequently used in the literature [Dolnicar 2002; Jain 2010], even if it has been criticised for its overestimation of the validity of the segmentation results [Dolnicar 2002; Dolnicar & Lazarevski 2009]. Clustering algorithms are unsupervised methods, aiming to assign objects in clusters. There are different ways to fulfil this aim, but the most commonly used assignments are known as hard (or crisp) and fuzzy. The hard clustering algorithms are mainly used to classify objects into particular clusters while the
fuzzy clustering algorithms aim to extract and describe the hidden relationships amongst objects (Bai et al., 2014; Everitt et al., 2011). More specifically, hard clustering algorithms group objects into a specific number of mutually exclusive clusters, while fuzzy clustering algorithms (1) allow objects to belong simultaneously to several clusters, and (2) identify the strength of membership of each object to each cluster. As suggested by Steenkamp & Wedel (1991), and confirmed more recently by Zhang & Lipkin (2013), fuzzy clustering methods generally satisfy managerial needs of segmentation, providing a more realistic multidimensional description of the market place in which consumers are not forced to belong solely to one cluster.

Especially when personal judgements are used for market segmentation, forcing respondents to belong solely to one cluster (hard clustering) is not reasonable since people would not normally have such polar attitudes to products, experiences or destinations (Kotler, 1988; Chaturvedi et al., 1997; Chiang, 2011; Li et al., 2013). Consumers whose profiles mainly match with one cluster do not necessarily have to be assigned 100% to that cluster (Chaturvedi et al., 1997). Consumers who can be assigned to more than one cluster should not be solely assigned to one of them in an arbitrary way (Li et al., 2013). When consumers are arbitrarily allocated to one cluster only, the final market representation is not realistic but merely a selected picture of the reality (Chiang, 2011). The fuzzy clustering approach returns instead a measure of the uncertainty, i.e., the membership degree, that describes how close a unit is to each cluster. Therefore, the final partial classification defined by the membership degrees is more attractive than the hard classification since it allows us to identify unclear boundaries between clusters that more realistically represent the hidden markets' structures. Observing the individual vector of membership degrees, it is possible to rate the cluster(s) that better describes each unit, providing another place of information that traditional clustering methods cannot uncover.

Fuzzy clustering algorithms are also preferable to crisp clustering algorithms for philosophical, economic, managerial, and marketing reasons (see D’Urso et al., 2018; 2016; Wang et al., 2015; Goneos-Malka et al., 2014; Georgescu et al., 2013).
as well as for their technical and methodological advantages (Hwang et al., 2007; D’Urso, 2015). Moreover, fuzzy clustering methods are computationally more efficient since dramatic changes in the value of cluster membership are less likely to occur in estimation procedures (McBratney & Moore, 1985). Finally, it has been demonstrated that these methods are less affected by the local minimum problem (Heiser & Groenen, 1997; Klawonn et al., 2015). This final property is of particular interest since it implies that fuzzy clustering algorithms avoid the creation of undesired final clusters.

3. Methodology

3.1. Propensity score matching

The matching (or integration) of two or more datasets can be pursued either with a micro- or macro-approach. While the macro approach can be used when not all variables of interest are simultaneously observed in the datasets, the micro-approach is adopted when all the variables are available for every data source (D’Orazio et al., 2006). When repeated cross-sectional data are involved in the analysis, as in this study, the micro-approach is generally the most appropriate. Among the micro-approaches, the propensity score (PS) matching is the most popular method that allows the reduction of the multidimensionality of the matching problem to a single dimension, i.e., the propensity score (Rosenbaum & Rubin, 1983). In other words, individual units can be compared on the basis of their propensity scores, rather than matching on all values of the set of relevant variables. The propensity score is the estimated conditional probability of a unit belonging to one dataset given a set of relevant variables jointly observed. The set of joint variables is usually based on a priori knowledge on the relationships among covariates within and between data sources (Brookhart et al., 2006). When the aim is to compare clustering results across time, we suggest the consideration of all characteristics that can confound the description of the evolution of the phenomenon under observation and that, consequently, need to be fixed. When the set of relevant variables is defined,
an additional indicator variable, which identifies to which dataset each unit belongs, is included in the datasets. Afterwards, the datasets are combined and usually estimated using a binary or multinomial (either logit or probit) regression model, depending on the number of datasets involved in the matching. While in the binary case the logit and probit models allow the acquisition of similar results, in the multiple datasets case the multinomial probit model is preferable (Caliendo & Kopeinig, 2008). To compare propensity scores, finally matching pair of units, different matching algorithms can be implemented (for a review see Stuart, 2010; Caliendo & Kopeinig, 2008). The most common and simplest matching algorithm is the nearest neighbour (NN) matching (Caliendo & Kopeinig, 2008). The NN algorithm matches the units on the basis of their similarities, i.e. two units are paired if their propensity scores are the closest. Several variants have been suggested in the literature (Caliendo & Kopeinig, 2008). In this paper we suggest the use of the one-to-one NN matching algorithm without replacement in which each unit is paired only with another unit and no replacement is allowed.

As suggested by Rosenbaum & Rubin (1985), significant differences among samples are expected when the relevant PS variables are considered before the adoption of the matching procedure while, after matching, any significant difference between samples should be removed. The most common procedure to assess matching quality involves the computation of the well-known $t$-test, to assess difference in means (Rosenbaum & Rubin, 1985) and Chi-square test (or Fisher’s exact test for small samples), to assess differences in proportions when categorical variables are involved.

3.2. Fuzzy numbers

In order to convey mathematically linguistic and/or imprecise expressions, the adoption of fuzzy data is highly advised (Benítez et al., 2007; Coppi et al., 2012; D’Urso, 2007). A general class of fuzzy data is the LR (left and right) fuzzy data (Dubois & Prade, 1988; Zimmermann, 2011). Let us assume that $\tilde{x}_{ik}$ is the $k$-th LR fuzzy variable ($k = 1, \ldots, K$) observed on the $i$-th unit ($i = 1, \ldots, N$).
Usually $\tilde{x}_{ik}$ is denoted as $\tilde{x}_{ik} = (m_{1ik}, m_{2ik}, l_{ik}, r_{ik})_{LR}$, where $m_{1ik}$ and $m_{2ik}$ (with $m_{2ik} > m_{1ik}$) are respectively the left and right centres and the interval $[m_{1ik}, m_{2ik}]$ is usually called the “core” of the fuzzy number; $l_{ik}$ and $r_{ik}$ represent the left and right spreads, i.e. the vagueness of the observation. The membership function of $\tilde{x}_{ik}$ is represented in Appendix, Eq. (1). When the membership function is made by linear functions and only one centre is defined, the triangular fuzzy number is established. In this study, the triangular fuzzy number is used to recode Likert-type variables into fuzzy variables thanks to (1) its popularity in practical applications and (2) its effectiveness in expressing decision problems characterised by imprecise and subjective information (Dagdeviren & Yulsel, 2008).

3.3. Fuzzy distance

When fuzzy variables are used as segmentation variables, the traditional Euclidean distance between pairs of units cannot be used in the clustering algorithm. Instead, a suitable distance for fuzzy data has to be adopted. In this paper, the squared fuzzy distance suggested by Coppi et al. (2012) (see Appendix, Eq. (2)) has been used. This distance is a weighted sum of squared Euclidean distances between the centres and squared Euclidean distances between the left and right spread, respectively. The weights are suitable non-negative values that can be fixed either subjectively or objectively, under the normalisation (sum at one) and coherence conditions (the weight of the centres has to be no smaller than the weight of the spreads). In this study, the weights have been objectively fixed via optimisation procedure to extract this information directly from the data (Coppi et al., 2012).

3.4. Fuzzy clustering for fuzzy data

Within the fuzzy clustering literature, the fuzzy $C$-means clustering algorithm and the fuzzy $C$-medoids are the most popular methods (Bezdek, 1981; Hung & Yang, 2005; Klawonn et al., 2015). In general, the main difference between fuzzy $C$-means clustering algorithms and fuzzy $C$-medoids clustering
algorithms is that prototypes of each cluster are “virtual” and “real” units, respectively. In other words, the units that describe final clusters (prototypes) obtained using the fuzzy $C$-means clustering algorithm are weighted means computed over all units belonging to the same cluster, given for each segmentation variable, and using the level of membership as weight (see [Everitt et al.] 2011, [Coppi et al.] 2012). On the other hand, the units that describes final clusters obtained using the fuzzy $C$-medoids clustering algorithm are actually observed units.

The fuzzy clustering algorithm for fuzzy data suggested in the literature are: the fuzzy $C$-means for fuzzy data (FCM-FD) ([Coppi et al.] 2012), the fuzzy $C$-medoids for fuzzy data (FCMdC-FD) ([D’Urso et al.] 2018), the bagged fuzzy $C$-means for fuzzy data (BFCMC-FD) ([D’Urso et al.] 2015), and the bagged fuzzy $C$-medoids clustering methods for fuzzy data (BFCMdC-FD) ([D’Urso et al.] 2013b). Being a combination of fuzzy numbers and fuzzy clustering algorithm, the above-mentioned clustering algorithms inherit the benefits connected with both the fuzzy formalisation of imprecise information and the fuzzy assignment of units. In this study, the FCM-FD clustering algorithm has been adopted since, after the matching method, it is meaningless to compare final clusters using “real” units (i.e. using a fuzzy $C$-medoids clustering algorithm). Furthermore, the FCM-FD clustering algorithm has been adopted instead of the BFCMC-FD clustering algorithm for its simplicity. The FCM-FD objective function that has to be minimised is reported in Appendix, Eq. (3).

3.5. Fuzzy partition validity

To check replication and robustness of the final partition, the stability of the initial random values chosen to run the algorithm over several iterations should be checked ([D’Urso] 2015).

In general, internal validity measures provide useful guidelines in the identification of the best final partition ([Handl et al.] 2005, [D’Urso et al.] 2015). Suitable measures for the fuzzy clustering algorithm have been suggested by [Xie & Beni] 1991 and [Campello & Hruschka] 2006. Among them, the Xie and
Beni (henceforth XB) index (Xie & Beni, 1991) is a popular measure whose aim is to quantify the ratio between compactness (i.e. total variance within the clusters) and separation of clusters. See Appendix, Eq. (4) for the computation of the XB index when the FCM-FD clustering algorithm is adopted. A small XB value indicates that all clusters are overall compact and separate from each other. Thus the best fuzzy C-partition is the one with the smallest XB.

3.6. Labelling and profiling: rank and weighted data

The final partition obtained by means of the FCM-FD is characterised by C fuzzy prototypes. Each fuzzy prototype is a vector of fuzzy numbers, each of which represents the cluster-weighted means of one segmentation variable. Therefore, being able to rank the fuzzy numbers that make up each prototype allow both labelling and description of the main features of each cluster. Several methods for ordering fuzzy numbers have been suggested in the literature and, following Brunelli & Mezei (2013), they can be broadly divided into two kinds: (1) methods that transform each fuzzy number into a crisp real number; (2) methods that compare all pairs of fuzzy numbers generating pairwise orderings on the basis of which the final rank is determined. In this study, the method suggested by Abbasbandy & Hajjari (2009) is adopted for its simplicity (see Appendix, Eq. (5)). This method belongs to the first kind of ranking methods and it is based on the definition of the “magnitude” of fuzzy numbers: the larger the “magnitude” of a fuzzy number, the larger the fuzzy number.

Furthermore, the results of any fuzzy clustering algorithm can be summarised by a \((N \times C)\) matrix \(U\) containing the level of membership of each \(i\)-th unit to each \(c\)-th cluster. The higher the membership degree, the higher the strength of the association between the observed pair unit-cluster. Therefore, the \(U\) matrix can be used to weight the original crisp segmentation variables and further describe the final clusters.

The \(U\) matrix can also be used in the profiling phase (D’Urso et al., 2016, 2013a) where external variables (i.e. variables not involved in the cluster algorithm) are normally used to characterise the clusters (Everitt et al., 2011, Hair...
When fuzzy clustering algorithms are performed, a “defuzzification” procedure is usually adopted to obtain final crisp partitions that are easy to interpret. This procedure consists of assigning each unit to the cluster with which the unit is linked via the highest level of membership (or with a level of membership higher than a specified cut-off point). As discussed in D’Urso et al. (2016), the “defuzzification” is in contrast with the very essence of the fuzzy theory and fuzzy clustering and it implies a loss of information. In fact, the key idea of the fuzzy clustering algorithm is to allow units to belong to more than one cluster simultaneously describing individual membership distributions by means of a vector of degrees of membership. Using the “defuzzification”, this information is lost and the result of the fuzzy clustering algorithm give the same kind of information as any crisp clustering algorithm. To avoid this problem, the external variables are weighted using the U matrix and suitable tests for differences in means and proportions can be computed to describe the final cluster partition.

3.7. Comparing partitions through fuzzy distance

When the best partitions are obtained for each of the datasets, the computation of pairwise distances is necessary to assess differences between clusters, within and between years. Therefore, one can be interested in the computation of both intra- and inter-temporal distances between clusters. As an example, figure 2 schematically represents the idea of these distances by observing two periods \((t_1, t_2)\), two variables \((k_1, k_2)\), and three final clusters in each period. Any distance measure can be adopted for this purpose. Since in this research clusters are represented by fuzzy prototypes, a fuzzy distance has to be adopted Coppi et al. (2012).

4. Application: tourist satisfaction

In the following, each step of the matching-clustering procedure will be illustrated using empirical data. Data have been analysed from the Banca d’Italia,
“International Tourism in Italy” survey, in 2000-01 and 2010-11 among foreign tourists visiting any municipality located in South-Tyrol region (Northern Italy). The “International Tourism in Italy” survey is a repeated cross-sectional survey that collects socio-economic and demographic information on tourists, trip characteristics (including travel expenditure), and tourist satisfaction with different aspects of the destination. The inbound-outbound frontier survey is the technique adopted for the collection of the data. In this study, some individual characteristics (gender, age, employment status and country of origin) and travel purpose have been used as relevant variables for the calculation of the PS, while tourist satisfaction with 10 different aspects, which overall describe the destination, have been used as segmentation variables of the FCM-FD algorithm. Finally, PS variables and two more travel characteristics (i.e. travel expenditures for goods and services made in the country visited and duration of the trip), that we believe are important factors in explaining differences in tourism satisfaction levels, have been used to profile the final clusters.

4.1. Statistical matching

The original samples comprise 1,582 and 997 international tourists interviewed in 2000-01 and 2010-11, respectively. After the NN matching, two final samples of 806 international tourists, one in each period, were identified and further analysed. Table [1] in the Appendix allows the evaluation of the quality of the matching procedure: before the implementation of the NN matching, the samples collected in the two periods were significantly different given each relevant variable, while they were not significantly different after the implementation of the matching method. Furthermore, the majority (57.4%) of the paired tourists perfectly matched, while 41% of the paired tourists differ only for one variable, mainly for gender (98%), and the remaining 1.6% of the paired tourists differ for two characteristics (mainly the pair gender-age 54%).

4.2. Data recoding

Satisfaction has been measured by a 10-point Likert-type scale (1= Very unsatisfied and 10= Very satisfied). Figures [3(a)] and [3(b)] provide the list of rated
aspects, showing the percentage distributions per period. As we can observe, the proportion of tourists who overall were less satisfied with the destination, i.e. tourists who gave a rating of under 6 for any of the destination’s aspects considered, is low in both observed periods and it is mainly concentrated on one aspect, “Prices”. Over the years, the proportion of Very satisfied tourists (the dark red part of each bar) is significantly decreased with different aspects of the destination (in specifically “Friendliness”, “Art”, “Landscape”, “Accommodation” “Food and beverages”, and “Overall”) while tourists became significantly more satisfied only with “Safety” (two-sample tests for equality of proportions have been performed, level of significance 5%). The aspects that lost the most in terms of Very satisfied tourists, and that consequently deserve more managerial and strategic attention in the long-run, are “Landscape” (-19%) and “Food and beverages” (-13%).

[Insert Figure 3(a) and 3(b)].

As mentioned in the previous Section, the triangular fuzzy number has been used in this study to recode the linguistic expression of the Likert-type scale. The fuzzy recoding is displayed in Figure 4 showing that, for instance, the value 1 (Very unsatisfied) corresponds to a fuzzy number (1, 0, 1.5), while the value 5 (i.e. mild dissatisfaction) corresponds to the fuzzy number (5, 0.5, 0.5). In this study we chose to assign unequal degrees of vagueness, i.e. right and left spreads, to each linguistic expression of the Likert-type scale to capture partially the different degrees of uncertainty associated by respondents to each linguistic expression. In particular, we assigned the highest degree of vagueness to the extreme linguistic terms, i.e. 1 (Very unsatisfied) and 10 (Very satisfied), and a decreased degree of vagueness approaching the central values, i.e. 5 and 6. In fact, we believe respondents perceived quite clearly the difference between the values 5 and 6 since it is common practice to attribute a negative (positive) evaluation to each value below 5 (above 6). Therefore, these values are generally a little vague for respondents. On the other hand, it is generally more difficult to capture the difference between 1 and 2, or between 9 and 10, thus these values...
should incorporate a higher degree of uncertainty.

4.3. Clusters’ results and cross-sectional comparison

For this illustration, the weights of the dissimilarity measure in Eq. 2 have been objectively obtained via optimisation procedure and they are both equal to 0.5. The order of fuzziness \( p \) is 1.75. According to the \( XB \) index, in 2000-01 (2010-11) tourists are better classified into two (three) clusters. The initial random values after 50 iterations are stable in each analysis, indicating that the final cluster partitions are robust.

Figures 5(a)-6(c) represent (1) the weighted percentage distribution of the segmentation variables (graph on the left) and (2) the fuzzy prototypes (graph on the right) for each final cluster. These results clearly show that in 2000-01 tourists are simply split into “high” (cluster 1) and “low” (cluster 2) satisfied tourists. Therefore, the two clusters can be labelled for simplicity as “Enthusiasts” and “Unfulfilled”, respectively. In 2010-11 the differences between cluster 1 (the “Enthusiasts”) and 3 (the “Unfulfilled”) are stronger than in the previous period and a “Moderate satisfied” group of tourists (cluster 2) emerged in-between the “Enthusiasts” and the “Unfulfilled”. To appreciate and describe further the differences in satisfaction levels within clusters, the 10 aspects have been ranked in ascending order (from the most to the least satisfactory) according to their “magnitude” computed as in Eq. 5. As shown in Table 1, the most appreciable difference between the two final clusters obtained in 2000-01 regards the ranking of “Friendliness” and “Art” aspects. The “Enthusiasts” rated “Friendliness” (“Art”) in a higher (lower) position than the “Unfulfilled”, indicating a higher (lower) satisfaction with this aspect when compared to the other aspects characterising the destination. In 2010-11 the within cluster rankings show that the “Enthusiasts” are much more (less) satisfied with the “Accommodation” (“Friendliness”) than other tourists are. Furthermore, the “Unfulfilled” rated the “Overall” satisfaction destination in the 4th position while the other tourists ranked
this aspect lower in the ranking (6th and 7th position). Comparing the results obtained in the two periods we can state that “Landscape” is always the most preferable aspect of the South Tyrol region while “Prices”, “Products sold”, and “Information” are always the worst aspects on which the destination has to pay more attention. Over time, the “Enthusiasts” became more satisfied with the “Accommodation” but less satisfied with the “Friendliness” of inhabitants and “Overall” with the destination. Comparing with the past, in 2010-11 tourists feel generally safer in the destination but, unfortunately, “Food and beverages” satisfy them less.

To describe and investigate further the evolution of tourist satisfaction, the intra- and inter-temporal distances between fuzzy prototypes are computed and reported in Table 2. Within the period, we can observe that the two clusters identified in 2000-01 are quite distinguishable. In 2010-11 the clusters are more separated than in 2000-01 (they are characterised by higher distance values), meaning that the differences between clusters become more and more pronounced over time. Furthermore, in 2010-11 clusters 2 and 3 are the most similar followed by the pair of clusters 1-2. Focusing on the inter-temporal distances (i.e. distances between periods), we can observe that both clusters identified in 2000-01 are similar to the “Moderate satisfied” tourists identified in 2010-11. Moreover, the old “Enthusiasts” (“Unfulfilled”) are more similar to the new “Enthusiasts” (“Unfulfilled”) than to the new “Unfulfilled” (“Enthusiasts”).

In the profiling phase, travel expenditures and duration of the trip can be used, together with the PS variables, to characterise the final clusters. Since the samples are not significantly different against these variables (t-test for differences in means have been performed, level of significance 5%), any significant cross-sectional difference detected in one of these variables can be mainly ascribed to a change in tourism behaviour. The weighted percentage distributions and weighted means, respectively for categorical and quantitative variables, are reported in Table 3. As we can observe, in 2000-01 the two clusters signifi-
cantly differ only for the average amount of money spent (as a whole and per night) for the trip, revealing that the higher the satisfaction with the destination, the higher the expenditure. In 2010-11, the “Unfulfilled” are mainly Austrian tourists visiting South Tyrol for business or other personal reasons. Consequently, they stayed on average longer and spent less in the destination than other tourists. The “Enthusiasts” are mainly German tourists travelling for holiday purposes in an Italian region that offers them an astonishing landscape together with a perfect Italian-German mixture of cultures, traditions, foods, and languages. The result is a group of tourists very satisfied with the destination but which spends (in terms of both money and time) on average less than the “Moderate satisfied” who mainly come from other European and non-European countries.

5. Conclusions

The increased availability of data and the necessity to study dynamic changes, have led to the introduction of novel, more accurate and precise methods of analyses. The aim of this paper is to suggest a new matching-clustering procedure to study the change over time in tourists’ behaviour levels when repeated cross-sectional data are available. In tourism, cross-sectional surveys are widespread to collect information on individual behaviours, perceptions, feelings, and emotions. Since the sample of units interviewed each time is different, these kind of surveys would only allow us to obtain information on aggregate evolutions. In fact, any direct individual comparisons can be misleading, resulting in wrong policy and managerial conclusions. To overcome this limitation, the new matching-clustering procedure suggests the identification of paired samples by means of a matching method before conducting a clustering analysis and comparing the final results. The variables to be used in the NN matching have to be subjectively selected according to the availability of data (i.e. the set of equivalent variables collected over time) and the research questions (i.e. which variables should be used in the profiling phase).
The importance of both equivalent instrument measurements and accurate analyses of data collected by means of ordinal scales, such as Likert-type scales, have been discussed. Fuzzy numbers have been suggested to recode Likert-type scales and the main advantages have been highlighted. In this study we assumed that each respondent is characterised by the same uncertainty/vagueness function. To be more accurate, the membership function of the fuzzy number should be personalised at the respondent level to capture (1) the individual uncertainty and vagueness regarding the subjective opinions and judgements collected and (2) the uncertainty entitled in the subjective interpretation of the linguistic expression of the Likert-type scale. Unfortunately, with large anonymous datasets it is not possible to make this kind of personalisation, but further studies should be directed to overcome this aspect and produce more realistic and accurate results. Furthermore, it is important to underline that the form of the membership function of the fuzzy number, denoted as elicitation, is subjective and, in general, it is determined by experts in the problem area (Coppi et al., 2006). The membership function is context-sensitive meaning that it has to capture approximately the feelings/opinions of the person represented. In this respect, the elicitation of a membership function requires a deep psychological understanding (Coppi et al., 2006). For more details on the elicitation of the membership function, as well as on the analysis of robustness of the results obtained from a fuzzy data analysis, see De la Rosa de Sás et al. (2015).

Moreover, the fuzzy clustering algorithm has been suggested in this paper, even if any clustering algorithm can potentially be adopted. In general, this kind of algorithm is able to manage the uncertainty that arises in assigning the units to the clusters, allowing each unit to belong simultaneously to more than one cluster accordingly to a vector of membership degrees. As a result, the final partitions more realistically interpret the hidden relations that characterised a market allowing for more efficient understanding of managerial, marketing, and political strategies. The combination of fuzzy numbers and fuzzy clustering allow us to cope more precisely with different sources of uncertainty which arise in particular arising when human perceptions and judgments are investigated.
From a practical and managerial perspective, this 7-steps procedure will help
destination managers and municipalities to describe and verify more accurately
the efficacy of policies and strategies conducted over years. To illustrate the
suggested procedure, the evolution of satisfaction with different aspects that
classified the South-Tyrol region (Northern Italy) has been analysed. Data
collected by Banca d’Italia in 2000-01 and 2010-11 by means of a repeated cross-
sectional survey have been used. Being able to perform accurate analyses on

From the comparison of tourists’ satisfaction levels between the two periods, a general decrease in
the proportion of “Very satisfied” tourists emerged, in particular with regard to
key aspects of the destination like “Landscape” and “Food and beverages”. This
is an important and worrying result on which the government and the munici-
palities have to devote more resources and efforts. Furthermore, the findings
show that in 2000-01 the two identified clusters significantly differ only for the
average travel expenditure (as a whole and per night), revealing that the higher
the satisfaction with the destination, the higher the expenditure. This result is
only partially confirmed in 2010-11, since the most satisfied tourists are mainly
from Germany and they spend less than other European (excluding Austria)
and non-European tourists. The reason for this result can be ascribed to the
fact that German tourists are generally very satisfied with South-Tyrol. In this
destination, they feel like at home thanks to the similarity in language, tradi-
tions, cultures and foods. To obtain a more detailed profiling of the clusters,
and therefore obtain more practical and managerial information about the clus-
ters, the fractional multinomial logit model (Papke & Wooldridge, 1996), in
which the vector of membership degrees is set as dependent variable, can be
used (Khoo-Lattimore et al., 2018).

Since the majority of tourist satisfaction studies focus on one-off surveys,
preventing the possibility of monitoring the evolution of tourist satisfaction
(Song et al., 2011), this analysis is of particular interest in demonstrating how
evolutions in tourist satisfaction with a destination over time can be studied. Analysing tourist satisfaction with a destination in one year helps destination managers and municipalities to identify weak/strong aspects and define short-run strategies, while being able to extend the analyses to a period of time allows for the evaluation and definition of long-run policy and managerial strategies.
References


### Figures

<table>
<thead>
<tr>
<th>What</th>
<th>How</th>
<th>Why</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching</td>
<td>Propensity score matching</td>
<td>Cross-sectional samples not comparable</td>
</tr>
<tr>
<td>Data recoding</td>
<td>Fuzzy numbers</td>
<td>Ordinal data</td>
</tr>
<tr>
<td>Selection dissimilarity measure</td>
<td>Fuzzy distance</td>
<td>Fuzzy segmentation variables</td>
</tr>
<tr>
<td>Selection clustering algorithm</td>
<td>Fuzzy clustering for fuzzy data</td>
<td>Fuzzy assignment of units to cluster</td>
</tr>
<tr>
<td>Best partition &amp; cluster validation</td>
<td>Fuzzy validity index</td>
<td>Fuzzy prototypes</td>
</tr>
<tr>
<td>Labelling &amp; profiling</td>
<td>Rank of fuzzy numbers, Use of weighted data</td>
<td>Fuzzy prototypes, Membership degree</td>
</tr>
<tr>
<td>Comparing partitions</td>
<td>Fuzzy distance</td>
<td>Different partitions across time</td>
</tr>
</tbody>
</table>

Figure 1: Steps of the suggested matching-clustering procedure.
Figure 2: Inter- and intra-temporal distances between clusters.

Figure 3: % distribution for each aspect.

(a) 2000-01

(b) 2010-11
Figure 4: Linguistic satisfaction terms in the form of fuzzy numbers.
Figure 5: Cluster description 2000-01: weighted distribution of segmentation variables (on the left) and radar plots of fuzzy prototypes (on the right). In the radar plot, the black solid line represents the centres of the fuzzy prototypes, while the dashed lines represent the lower and upper values of the fuzzy prototypes.
Figure 6: Cluster description 2010-11: weighted distribution of the segmentation variables (on the left) and radar plot of the fuzzy prototypes (on the right). In the radar plot, the black solid line represents the centres of the fuzzy prototypes, while the dashed lines represent the lower and upper values of the fuzzy prototypes.
### Tables

**Table 1:** Magnitude of the fuzzy numbers that make up each prototype and ranking order (largest fuzzy number is ranked 1).

<table>
<thead>
<tr>
<th>Segmentation variables</th>
<th>2000-01</th>
<th></th>
<th>2010-11</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
<td></td>
<td>Cluster 2</td>
<td></td>
<td>Cluster 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Magnitude</td>
<td>Rank</td>
<td>Magnitude</td>
<td>Rank</td>
<td>Magnitude</td>
<td>Rank</td>
</tr>
<tr>
<td>Friendliness</td>
<td>9.096</td>
<td>3</td>
<td>8.339</td>
<td>5</td>
<td>9.221</td>
<td>8</td>
</tr>
<tr>
<td>Art</td>
<td>8.961</td>
<td>5</td>
<td>8.381</td>
<td>3</td>
<td>9.311</td>
<td>4</td>
</tr>
<tr>
<td>Landscape</td>
<td>9.531</td>
<td>1</td>
<td>8.972</td>
<td>1</td>
<td>9.596</td>
<td>1</td>
</tr>
<tr>
<td>Accommodation</td>
<td>8.821</td>
<td>6</td>
<td>7.768</td>
<td>7</td>
<td>9.345</td>
<td>3</td>
</tr>
<tr>
<td>Food and beverages</td>
<td>9.132</td>
<td>2</td>
<td>8.431</td>
<td>2</td>
<td>9.277</td>
<td>5</td>
</tr>
<tr>
<td>Prices</td>
<td>6.685</td>
<td>10</td>
<td>6.080</td>
<td>10</td>
<td>7.952</td>
<td>10</td>
</tr>
<tr>
<td>Products sold</td>
<td>8.273</td>
<td>9</td>
<td>7.627</td>
<td>9</td>
<td>8.806</td>
<td>9</td>
</tr>
<tr>
<td>Information</td>
<td>8.510</td>
<td>8</td>
<td>7.650</td>
<td>8</td>
<td>9.237</td>
<td>6</td>
</tr>
<tr>
<td>Overall</td>
<td>9.096</td>
<td>4</td>
<td>8.340</td>
<td>4</td>
<td>9.231</td>
<td>7</td>
</tr>
</tbody>
</table>

**Table 2:** Intra- and inter-temporal distances between fuzzy prototypes.

<table>
<thead>
<tr>
<th>Fuzzy prototypes</th>
<th>2000-01</th>
<th></th>
<th>2010-11</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cluster 1</td>
<td></td>
<td>Cluster 2</td>
<td></td>
<td>Cluster 1</td>
<td></td>
</tr>
<tr>
<td>2000-01 Cluster 1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000-01 Cluster 2</td>
<td>6.763</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuzzy prototypes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 1</td>
<td>4.038</td>
<td></td>
<td>18.858</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Cluster 2</td>
<td>3.198</td>
<td></td>
<td>2.277</td>
<td></td>
<td>9.865</td>
<td></td>
</tr>
<tr>
<td>Cluster 3</td>
<td>15.551</td>
<td></td>
<td>2.760</td>
<td></td>
<td>29.678</td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Cluster profile.

<table>
<thead>
<tr>
<th></th>
<th>2000-01</th>
<th></th>
<th></th>
<th>2010-11</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CL1</td>
<td>CL2</td>
<td>p-value</td>
<td>CL1</td>
<td>CL2</td>
<td>CL3</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>Cluster size (%)</td>
<td>52.86</td>
<td>47.14</td>
<td>0.005</td>
<td>30.04</td>
<td>35.01</td>
<td>34.95</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>PS variables (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>64.38</td>
<td>68.35</td>
<td>0.001</td>
<td>65.06</td>
<td>69.52</td>
<td>71.10</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 35 years old</td>
<td>20.59</td>
<td>22.97</td>
<td>0.050</td>
<td>19.65</td>
<td>21.70</td>
<td>22.78</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>35-44 years old</td>
<td>25.95</td>
<td>28.54</td>
<td>0.050</td>
<td>28.35</td>
<td>27.25</td>
<td>25.72</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>45-64 years old</td>
<td>41.45</td>
<td>37.74</td>
<td>0.050</td>
<td>41.59</td>
<td>39.09</td>
<td>37.98</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>More than 64 years old</td>
<td>12.01</td>
<td>10.75</td>
<td>0.050</td>
<td>10.40</td>
<td>11.96</td>
<td>13.51</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-employed</td>
<td>13.43</td>
<td>13.36</td>
<td>0.050</td>
<td>13.75</td>
<td>13.52</td>
<td>12.62</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Employee</td>
<td>68.03</td>
<td>68.48</td>
<td>0.050</td>
<td>68.04</td>
<td>69.31</td>
<td>66.98</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>13.48</td>
<td>12.52</td>
<td>0.050</td>
<td>11.95</td>
<td>11.84</td>
<td>14.43</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>5.06</td>
<td>5.64</td>
<td>0.050</td>
<td>6.25</td>
<td>5.33</td>
<td>5.97</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Country of origin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>Austria</td>
<td>13.05</td>
<td>13.53</td>
<td>0.050</td>
<td>8.95</td>
<td>9.10</td>
<td>21.88</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>61.98</td>
<td>59.47</td>
<td>0.050</td>
<td>67.25</td>
<td>63.74</td>
<td>52.29</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Other EU countries</td>
<td>20.09</td>
<td>21.42</td>
<td>0.050</td>
<td>20.49</td>
<td>21.07</td>
<td>20.92</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Outside EU</td>
<td>4.88</td>
<td>5.58</td>
<td>0.050</td>
<td>3.30</td>
<td>6.09</td>
<td>4.90</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Main purpose of travel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>Holiday</td>
<td>82.94</td>
<td>80.44</td>
<td>0.050</td>
<td>88.49</td>
<td>82.59</td>
<td>74.80</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>9.68</td>
<td>12.05</td>
<td>0.050</td>
<td>6.51</td>
<td>10.39</td>
<td>14.52</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Other personal motivations</td>
<td>7.39</td>
<td>7.51</td>
<td>0.050</td>
<td>5.00</td>
<td>7.02</td>
<td>10.68</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Other variables (average)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total expenditure</td>
<td>527.70</td>
<td>400.20</td>
<td>0.050</td>
<td>388.10</td>
<td>499.00</td>
<td>529.90</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Expenditure per night</td>
<td>77.20</td>
<td>63.88</td>
<td>0.050</td>
<td>49.23</td>
<td>54.51</td>
<td>44.72</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>Nr. nights</td>
<td>5.72</td>
<td>5.35</td>
<td>0.050</td>
<td>2.69</td>
<td>2.99</td>
<td>3.70</td>
<td>0.050</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Weighted percentage distributions and weighted means are reported. Significance of the χ² test of independence and z-test, respectively for qualitative and quantitative variables, is reported. All test results are not significant unless indicated otherwise: ***Significant at p ≤ 0.01, **Significant at p ≤ 0.05, *Significant at p ≤ 0.1.
Appendix

Fuzzy numbers

The membership function of the $k$–th LR fuzzy variable, $\mu_{\tilde{x}_ik}(a_{ik})$, is defined as:

$$
\mu_{\tilde{x}_ik}(a_{ik}) = \begin{cases} 
L \left( \frac{m_{1ik} - a_{ik}}{l_{ik}} \right) & a_{ik} \leq m_{1ik} (l_{ik} > 0) \\
1 & m_{1ik} \leq a_{ik} \leq m_{2ik} \\
R \left( \frac{a_{ik} - m_{2ik}}{r_{ik}} \right) & a_{ik} \geq m_{2ik} (r_{ik} > 0)
\end{cases}
$$

(1)

where both $L$ and $R$ are decreasing “shape” functions defined in $[0,1]$. When both $L$ and $R$ are linear functions, the trapezoidal fuzzy number is defined. A trapezoidal fuzzy number characterised by $m_{1ik} = m_{2ik}$, i.e. only one centre is defined, is called triangular fuzzy number.

Fuzzy distance

The squared fuzzy distance suggested by Coppi et al. (2012) between the $i$-th and $j$-th units, with $i \neq j$, is computed as follows:

$$
d_F^2(\tilde{x}_i, \tilde{x}_j) = w_M^2 (\|\mathbf{m}_i - \mathbf{m}_j\|^2) + w_S^2 (\|\mathbf{l}_i - \mathbf{l}_j\|^2 + \|\mathbf{r}_i - \mathbf{r}_j\|^2)
$$

(2)

where $\tilde{x}_i$ denotes the fuzzy vector of segmentation variables observed for the $i$-th unit; $\mathbf{m}_i$ is the vector of the centres; $\mathbf{l}_i$ and $\mathbf{r}_i$ are the vectors of the left and right spreads, respectively; $\|\mathbf{m}_i - \mathbf{m}_j\|^2$ is the squared Euclidean distances between the centres; $\|\mathbf{l}_i - \mathbf{l}_j\|^2$ and $\|\mathbf{r}_i - \mathbf{r}_j\|^2$ are the squared Euclidean distances between the left and right spread, respectively; $w_M$ and $w_S$ are suitable non-negative weights for the centre and the spread components.

Fuzzy clustering for fuzzy data

The FCM-FD objective function that has to be minimised is as follows:

$$
\begin{align*}
\min & : \sum_{i=1}^{N} \sum_{c=1}^{C} u_{ic}^p d_F^2(\tilde{x}_i, \tilde{h}_c) \\
\text{s.t.} & : \sum_{c=1}^{C} u_{ic} = 1, \quad u_{ic} \geq 0
\end{align*}
$$

(3)
where $p > 1$ is a weighting exponent that controls the fuzziness of the obtained partition; $u_{ic}$ indicates the level of membership, i.e. closeness, of the $i$-th fuzzy unit to the $c$-th ($c = 1, \ldots, C$) cluster; $d^2_F(\hat{x}_i, \hat{h}_c)$ is the fuzzy distance, computed as in Eq. (2), between the $i$-th fuzzy unit ($\hat{x}_i$) and the fuzzy prototype that describes the $c$-th cluster ($\hat{h}_c$).

The iterative optimal solutions of (3) are:

$$u_{ic} = \frac{\left[ w^2_M (\| m_i - h^c_M \|^2) + w^2_S (\| l_i - h^c_L \|^2 + \| r_i - h^c_R \|^2) \right]^{-\frac{1}{p-1}}}{\sum_{c'=1}^{C} \left[ w^2_M (\| m_i - h^{c'}_M \|^2) + w^2_S (\| l_i - h^{c'}_L \|^2 + \| r_i - h^{c'}_R \|^2) \right]^{-\frac{1}{p-1}}},$$

$$h^c_M = \frac{\sum_{i=1}^{N} u^p_{ic} m_i}{\sum_{i=1}^{N} u^p_{ic}}, \quad h^c_L = \frac{\sum_{i=1}^{N} u^p_{ic} l_i}{\sum_{i=1}^{N} u^p_{ic}}, \quad h^c_R = \frac{\sum_{i=1}^{N} u^p_{ic} r_i}{\sum_{i=1}^{N} u^p_{ic}}$$

$$w_M = \frac{\sum_{i=1}^{N} \sum_{c=1}^{C} u^p_{ic} (\| m_i - h^c_L \|^2 + \| r_i - h^c_R \|^2)}{\sum_{i=1}^{N} \sum_{c=1}^{C} u^p_{ic} (\| m_i - h^c_M \|^2 + \| l_i - h^c_L \|^2 + \| r_i - h^c_R \|^2)}$$

($w_S = 1 - w_M$)

where $h^c_M$, $h^c_L$ and $h^c_R$ represent respectively the centre and the left and right spreads of the $c$-th fuzzy prototype.

Fuzzy partition validity

In the case of FCM-FD, the $XB$ index is computed as follow:

$$\sum_{i=1}^{N} \sum_{c=1}^{C} u^p_{ic} d^2_F(\hat{x}_i, \hat{h}_c)$$

$$XB = \frac{\min_{c \neq c'} d^2_F(\hat{h}_c, \hat{h}_{c'})}{\sum_{i=1}^{N} \sum_{c=1}^{C} u^p_{ic} d^2_F(\hat{x}_i, \hat{h}_c)}$$

Labelling and profiling: rank and weighted data

The “magnitude” ($M(\hat{x})$) is a crisp real number computable using both centre ($m$) and spreads values ($l$ and $r$) of the triangular fuzzy number ($\hat{x}$) as follows:

$$M(\hat{x}) = \frac{1}{2} \int_{0}^{1} [2m + (l - r)z - l + r] f(z) dz$$

39
Table 1: Assessing matching quality.

<table>
<thead>
<tr>
<th></th>
<th>Before PSM</th>
<th></th>
<th>After PSM</th>
<th></th>
<th>p−value</th>
<th></th>
<th>p−value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2000-01</td>
<td>2010-11</td>
<td>2000-01</td>
<td>2010-11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>1582</td>
<td>997</td>
<td>806</td>
<td>806</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PS variables (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>64.79</td>
<td>68.91</td>
<td>**</td>
<td>66.25</td>
<td>68.73</td>
<td>0.288</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td>***</td>
<td></td>
<td></td>
<td>0.985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 35 years old</td>
<td>30.10</td>
<td>21.16</td>
<td>21.71</td>
<td>21.46</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44 years old</td>
<td>24.21</td>
<td>28.59</td>
<td>27.17</td>
<td>27.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-64 years old</td>
<td>38.59</td>
<td>36.41</td>
<td>39.70</td>
<td>39.45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More than 64 years old</td>
<td>7.10</td>
<td>13.84</td>
<td>11.71</td>
<td>12.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td>***</td>
<td></td>
<td></td>
<td>0.977</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-employed</td>
<td>21.18</td>
<td>11.57</td>
<td>13.40</td>
<td>13.28</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employee</td>
<td>53.92</td>
<td>69.92</td>
<td>68.24</td>
<td>68.11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>14.35</td>
<td>12.47</td>
<td>13.03</td>
<td>12.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>10.56</td>
<td>6.04</td>
<td>5.33</td>
<td>5.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country of origin</td>
<td></td>
<td>***</td>
<td></td>
<td></td>
<td>0.988</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Austria</td>
<td>9.42</td>
<td>21.06</td>
<td>13.28</td>
<td>13.52</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>68.77</td>
<td>50.85</td>
<td>60.79</td>
<td>60.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other EU countries</td>
<td>15.49</td>
<td>21.46</td>
<td>20.72</td>
<td>20.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outside EU</td>
<td>6.32</td>
<td>6.62</td>
<td>5.21</td>
<td>4.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main purpose of travel</td>
<td></td>
<td>***</td>
<td></td>
<td></td>
<td>0.981</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holiday</td>
<td>79.84</td>
<td>76.03</td>
<td>81.76</td>
<td>81.64</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td>12.07</td>
<td>10.53</td>
<td>10.79</td>
<td>10.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other personal motivations</td>
<td>8.09</td>
<td>13.44</td>
<td>7.44</td>
<td>7.69</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Significance of the $\chi^2$ test of independence and $z$-test, respectively for qualitative and quantitative variables, is reported. All test results are not significant unless indicated otherwise: ***Significant at $p \leq 0.01$, **Significant at $p \leq 0.05$, *Significant at $p \leq 0.1$. 

40