

Analysing cluster evolution using repeated cross-sectional ordinal data

Marta Disegna^{a,*}, Pierpaolo D'Urso^b, Riccardo Massari^b

^a*Accounting, Finance & Economics Department, Faculty of Management, Bournemouth University, 89 Holdenhurst Road, Bournemouth, BH8 8EB, United Kingdom*

^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

*Corresponding author

Email addresses: disegnam@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 technique. The first and the last steps, i.e. the “Matching” and the “Comparing 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 & profiling” 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 techniques (“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 algorithm 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 paper specifically focuses on the situation in which (1) repeated cross-sectional surveys are available, (2) segmentation variables are measured by means of ordinal 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. Readers 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 according 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

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.

65 **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.)
70 observed over time, or paired samples, i.e. samples of units that are similar according to a set of controlling factors defined *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 (Brandt, 2018). In fact, results can be compromised, for instance,
75 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
80 aggregate units characterised by similar behaviour across time are the (1) model- (2) feature- and (3) observation-based approaches (for more details, see Disegna et al., 2017; Caiado et al., 2015)

When repeated cross-sectional surveys are used to see how a key relationship has changed over time, a common procedure is to create a pooled dataset
85 aggregating cross-sectional data from different years. Although extensive econometric techniques have been developed for the analysis of pooled cross sectional data (Wooldridge, 2012), in tourism market segmentation the cluster analysis is usually performed on the pooled dataset losing the information on the evolution of the phenomenon (see for instance Ferrer-Rosell & Coenders, 2018). Another

90 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, compari-
son 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
95 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
100 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 inte-
grated 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
105 conditionally independent given a set of common variables used for the match-
ing. 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
110 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
115 not changes in the samples used to obtain the clusters.

2.2. Data recoding

Feelings, emotions, motivation, and consumer behaviour are complex psycho-
logical processes vaguely captured through quantitative imprecise measurements
(Hung & Yang, 2005; D’Urso, 2007). In both academia and industry, ordinal

120 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-
125 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
130 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
135 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 coun-
140 tries (i.e. liberal, traditional and post-communist). Respondents “have been
socialized in diverse economic and cultural backgrounds, speak different lan-
guages, 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
145 across populations”.

Partially to overcome these drawbacks and to be able to more precisely cap-
ture human feelings/thinking, the simple visual analogue scale or the fuzzy rat-
ing scale should be adopted instead of the traditional Likert-type scales (Gil &
González-Rodríguez, 2012; De la Rosa de Sàa et al., 2015). The visual analogue
150 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
155 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
160 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.,
165 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
170 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,
175 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
180 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,

2010, 2007; De Wilde, 2004) 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
215 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.

220 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
225 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 multidimension-
230 ality 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
235 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,
240 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.

245 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

250 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 algo-

255 rithm 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 dif-

260 ference 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.

265 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

270 k -th LR fuzzy variable ($k = 1, \dots, K$) observed on the i -th unit ($i = 1, \dots, N$).

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 (Dağdeviren & Yüksel, 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

300 algorithms is that prototypes of each cluster are “virtual” and “real” units, re-
spectively. In other words, the units that describe final clusters (prototypes)
obtained using the fuzzy C -means clustering algorithm are weighted means com-
puted 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;
305 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
310 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 con-
315 nected 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
320 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
325 be checked (D’Urso, 2015).

In general, internal validity measures provide useful guidelines in the iden-
tification 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

330 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
335 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.
340 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)
345 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
350 the “magnitude” of a fuzzy number, the larger the fuzzy number.

Furthermore, the results of any fuzzy clustering algorithm can be sum-
marised by a $(N \times C)$ matrix \mathbf{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,
355 the \mathbf{U} matrix can be used to weight the original crisp segmentation variables
and further describe the final clusters.

The \mathbf{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 algo-
rithm) are normally used to characterise the clusters (Everitt et al., 2011; Hair

360 et al., 2010). 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.
365 (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
370 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 \mathbf{U} matrix and suitable tests for differences in means and proportions can be computed to describe the final cluster partition.

375 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,
380 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).

385 [Insert Figure 2.]

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
390 tourists visiting any municipality located in South-Tyrol region (Northern Italy).
The “International Tourism in Italy” survey is a repeated cross-sectional sur-
vey that collects socio-economic and demographic information on tourists, trip
characteristics (including travel expenditure), and tourist satisfaction with dif-
ferent aspects of the destination. The inbound-outbound frontier survey is the
395 technique adopted for the collection of the data. In this study, some individ-
ual 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 al-
400 gorithm. 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

405 The original samples comprise 1,582 and 997 international tourists inter-
viewed 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,
410 the samples collected in the two periods were significantly different given each
relevant variable, while they were not significantly different after the implemen-
tation 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
415 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 un-*
satisfied 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,
420 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
425 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
430 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.
435 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
440 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
445 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.

450

[Insert Figure 4.]

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
455 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
460 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 pe-
465 riod 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
470 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
475 "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.

[Insert Figure 5(a)-6(c).]

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
540 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
545 (2) the uncertainty entailed 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
550 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 understand-
555 ing (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 Saa 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
560 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
565 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 characterised 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 tourist satisfaction with a destination is of particular relevance since more satisfied tourists are more willing to revisit and to recommend the destination to relatives and friends the destination (Prayag et al., 2017). 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 municipalities 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, traditions, cultures and foods. To obtain a more detailed profiling of the clusters, and therefore obtain more practical and managerial information about the clusters, 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

- 605 Abbasbandy, S., & Hajjari, T. (2009). A new approach for ranking of trapezoidal fuzzy numbers. *Computers & Mathematics with Applications*, *57*, 413 – 419. doi:<http://dx.doi.org/10.1016/j.camwa.2008.10.090>.
- Amaro, S., Duarte, P., & Henriques, C. (2016). Travelers' use of social media: A clustering approach. *Annals of Tourism Research*, *59*, 1–15. doi:<http://dx.doi.org/10.1016/j.annals.2016.03.007>.
- 610 //dx.doi.org/10.1016/j.annals.2016.03.007.
- Bai, C., Dhavale, D., & Sarkis, J. (2014). Integrating Fuzzy C-Means and TOPSIS for performance evaluation: An application and comparative analysis. *Expert Systems with Applications*, *41*, 4186–4196. doi:<http://dx.doi.org/10.1016/j.eswa.2013.12.037>.
- 615 Benítez, J. M., Martín, J. C., & Román, C. (2007). Using fuzzy number for measuring quality of service in the hotel industry. *Tourism Management*, *28*, 544–555.
- Bezdek, J. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York.
- 620 Brandt, D. (2018). Wage determinants in the Swedish tourism sector. *Scandinavian Journal of Hospitality and Tourism*, *18*, 1–21. doi:10.1080/15022250.2016.1206832.
- Brookhart, M., Schneeweiss, S., Rothman, K., Glynn, R., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, *163*, 1149–1156.
- 625 //dx.doi.org/10.1016/j.ijar.2013.01.009.
- Brunelli, M., & Mezei, J. (2013). How different are ranking methods for fuzzy numbers? A numerical study. *International Journal of Approximate Reasoning*, *54*, 627–639. doi:<http://dx.doi.org/10.1016/j.ijar.2013.01.009>.
- Caiado, J., Maharaj, E. A., & D'Urso, P. (2015). Time series clustering. In C. Hennig, M. Meila, F. Murtagh, & R. Rocci (Eds.), *Handbook of Cluster Analysis* (pp. 241–264). Chapman and Hall/CRC.
- 630 //dx.doi.org/10.1016/j.ijar.2013.01.009.

- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*, 31–72. doi:10.1111/j.1467-6419.2007.00527.x.
- 635 Campello, R. J., & Hruschka, E. R. (2006). A fuzzy extension of the silhouette width criterion for cluster analysis. *Fuzzy Sets and Systems*, *157*, 2858–2875.
- Cang, S., Sun, Y.-Y., & Li, S. (2017). Impact of Taiwan open policy on Chinese tourist satisfaction. *Current Issues in Tourism*, *20*, 1140–1161. doi:10.1080/13683500.2014.964190.
- 640 Chang, C.-H., P.-C. Liu, & Wang, Y.-W. (2006). A hybrid model by clustering and evolving fuzzy rules for sales decision supports in printed circuit board industry. *Decision Support Systems*, *42*, 1254–1269. doi:10.1016/j.dss.2005.10.013.
- Chaturvedi, A., Carroll, J. D., Green, P. E., & Rotondo, J. A. (1997). A
645 feature-based approach to market segmentation via overlapping k -centroids clustering. *Journal of Marketing Research*, *34*, 370–377.
- Chiang, W.-Y. (2011). Establishment and application of fuzzy decision rules: an empirical case of the air passenger market in Taiwan. *International Journal of Tourism Research*, *13*, 447–456.
- 650 Coppi, R., & D’Urso, P. (2002). Fuzzy K–mean clustering models for triangular fuzzy time trajectories. *Statistical Methods and Applications*, *11*, 21–24.
- Coppi, R., D’Urso, P., & Giordani, P. (2012). Fuzzy and possibilistic clustering for fuzzy data. *Computational Statistics & Data Analysis*, *56*, 915–927.
- Coppi, R., Giordani, P., & D’Urso, P. (2006). Component models for fuzzy
655 data. *Psychometrika*, *71*, 733–761.
- Dağdeviren, M., & Yüksel, I. (2008). Developing a fuzzy analytic hierarchy process (AHP) model for behavior-based safety management. *Information Sciences*, *178*, 1717–1733. doi:http://dx.doi.org/10.1016/j.ins.2007.10.016.

- 660 Davidov, E., Meuleman, B., Ciecuch, J., Schmidt, P., & Billiet, J. (2014). Measurement Equivalence in Cross-National Research. *Annual Review of Sociology*, *40*, 55–75. doi:10.1146/annurev-soc-071913-043137.
- De Wilde, P. (2004). Fuzzy utility and equilibria. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, *34*, 1774–1785.
- 665 Disegna, M., D’Urso, P., & Durante, F. (2017). Copula-based fuzzy clustering of spatial time series. *Spatial Statistics*, *21*, 209–225.
- Dolnicar, S. (2002). A review of data-driven market segmentation in tourism. *Journal of Travel & Tourism Marketing*, *12*, 1–22.
- Dolnicar, S., & Lazarevski, K. (2009). Methodological reasons for the theory/practice divide in market segmentation. *Journal of Marketing Management*, *25*, 357–373.
- 670 D’Orazio, M., Di Zio, M., & Scanu, M. (2006). *Statistical Matching: Theory and Practice*. Wiley Series in Survey Methodology (1st ed.). Wiley.
- Dubois, D., & Prade, H. (1988). *Possibility theory*. Plenum press, New York.
- 675 D’Urso, P. (2007). Clustering of fuzzy data. In J. V. De Oliveira, & W. Pedrycz (Eds.), *Advances in fuzzy clustering and its applications* (pp. 155–192). J. Wiley and Sons.
- D’Urso, P. (2015). Fuzzy clustering. In C. Hennig, M. Meila, F. Murtagh, & R. Rocci (Eds.), *Handbook of Cluster Analysis* (pp. 545–573). Chapman and
680 Hall.
- D’Urso, P., De Giovanni, L., Disegna, M., & Massari, R. (2013a). Bagged Clustering and its application to tourism market segmentation. *Expert Systems with Applications*, *40*, 4944–4956. doi:10.1016/j.eswa.2013.03.005.
- D’Urso, P., De Giovanni, L., Disegna, M., & Massari, R. (2013b). Bagged
685 clustering and its application to tourism market segmentation. *Expert Systems with Applications*, *40*, 4944–4956.

- D'Urso, P., Disegna, M., & Massari, R. (2018). Fuzzy clustering in travel and tourism analytics. In P. Moscato, & N. De Vries (Eds.), *New ideas in Business and Consumer Analytics*. Springer. In press.
- 690 D'Urso, P., Disegna, M., Massari, R., & Osti, L. (2016). Fuzzy segmentation of postmodern tourists. *Tourism Management*, *55*, 297–308. doi:<http://dx.doi.org/10.1016/j.tourman.2016.03.018>.
- D'Urso, P., Disegna, M., Massari, R., & Prayag, G. (2015). Bagged fuzzy clustering for fuzzy data: An application to a tourism market. *Knowledge-Based Systems*, *73*, 335–346.
- 695 *Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster Analysis. (Fifth ed. ed.). John Wiley & Sons, Ltd. Wiley series in probability and statistics.*
- Ferrer-Rosell, B., & Coenders, G. (2018). Destinations and crisis. Profiling tourists' budget share from 2006 to 2012. *Journal of Destination Marketing & Management*, *7*, 26–35. doi:<http://dx.doi.org/10.1016/j.jdmm.2016.07.002>. In press.
- 700 *Georgescu, I. (2007). Similarity of fuzzy choice functions. Fuzzy Sets and Systems, 158, 1314–1326.*
- Georgescu, I. (2010). Arrow index of a fuzzy choice function. *Fundamenta Informaticae*, *99*, 245–261.
- 705 *Gil, M. A., & González-Rodríguez, G. (2012). Fuzzy vs. Likert Scale in Statistics. In E. Trillas, P. P. Bonissone, L. Magdalena, & J. Kacprzyk (Eds.), Combining Experimentation and Theory (pp. 407–420). Springer volume 271 of Studies in Fuzziness and Soft Computing.*
- 710 *Goneos-Malka, A., Strasheim, A., & Grobler, A. (2014). Conventionalists, Connectors, Technoisseurs and Mobilarti: Differential profiles of mobile marketing segments based on phone features and postmodern characteristics of consumers. Journal of Retailing and Consumer Services, 21, 905–916.*

- 715 Grün, B., & Dolnicar, S. (2016). Response style corrected market segmentation for ordinal data. *Marketing Letters*, *27*, 729–741. doi:10.1007/s11002-015-9375-9.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. (7th ed.). Prentice Hall.
- 720 Handl, J., Knowles, J., & Kell, D. B. (2005). Computational Cluster Validation in Post-Genomic Data Analysis. *Bioinformatics*, *21*, 3201–3212.
- Heiser, W. J., & Groenen, P. J. F. (1997). Cluster differences scaling with a within-clusters loss component and a fuzzy successive approximation strategy to avoid local minima. *Psychometrika*, *62*, 63–83.
- 725 Hu, H.-Y., Lee, Y.-C., & Yen, T.-M. (2010). Service quality gaps analysis based on Fuzzy linguistic SERVQUAL with a case study in hospital out-patient services. *The TQM Journal*, *22*, 499–515. doi:10.1108/17542731011072847.
- Hung, W. L., & Yang, M. S. (2005). Fuzzy clustering on LR-type fuzzy numbers with an application in Taiwanese tea evaluation. *Fuzzy sets and systems*, *150*,
730 561–577.
- Hwang, H., Desarbo, W. S., & Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika*, *72*, 181–198.
- Jain, A. K. (2010). Data clustering: 50 years beyond k -means. *Pattern Recognition Letters*, *31*, 651–666.
- 735 Jin, H., Moscardo, G., & Murphy, L. (2017). Making sense of tourist shopping research: A critical review. *Tourism Management*, *62*, 120–134. doi:http://dx.doi.org/10.1016/j.tourman.2017.03.027.
- Kampen, J., & Swyngedouw, M. (2000). The ordinal controversy revisited. *Quality and Quantity*, *34*, 87–102. doi:10.1023/A:1004785723554.
- 740 Khoo-Lattimore, C., Prayag, G., & Disegna, M. (2018). Me, My Girls and the Ideal Hotel: Segmenting Motivations of the Girlfriend Getaway Market using

- Fuzzy *C*-Medoids for Fuzzy Data. *Journal of Travel Research*, (pp. 1–19).
doi:<https://doi.org/10.1177/0047287518778154>.
- Klawonn, F., Kruse, R., & Winkler, R. (2015). Fuzzy clustering: More than
745 just fuzzification. *Fuzzy Sets and Systems*, *281*, 272–279. doi:<http://dx.doi.org/10.1016/j.fss.2015.06.024>.
- Kotler, P. (1988). *Marketing Management*. (6th ed.). Prentice-Hall: Englewood Cliffs, NJ.
- Li, X., Meng, F., Uysal, M., & Mihalik, B. (2013). Understanding China’s long-
750 haul outbound travel market: An overlapped segmentation approach. *Journal of Business Research*, *66*, 786–793.
- McBratney, A. B., & Moore, A. W. (1985). Application of fuzzy sets to climatic classification. *Agricultural and Forest Meteorology*, *35*, 165–185.
- Mikhailov, L. (2003). Deriving priorities from fuzzy pairwise comparison judgements.
755 *Fuzzy Sets and Systems*, *134*, 365–385. doi:[10.1016/S0165-0114\(02\)00383-4](https://doi.org/10.1016/S0165-0114(02)00383-4).
- Papke, L., & Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. *Journal of Applied Econometrics*, *11*, 619–632.
- 760 Pérez-Gladish, B., Gonzalez, I., Bilbao-Terol, A., & Arenas-Parra, M. (2010). Planning a TV advertising campaign: A crisp multiobjective programming model from fuzzy basic data. *Omega*, *38*, 84–94.
- Piurko, Y., Schwartz, S., & Davidov, E. (2011). Basic personal values and the meaning of left-right political orientations in 20 countries. *Political Psychology*,
765 *32*, 537–561. doi:[10.1111/j.1467-9221.2011.00828.x](https://doi.org/10.1111/j.1467-9221.2011.00828.x).
- Prayag, G., Hosany, S., Muskat, B., & Del Chiappa, G. (2017). Understanding the Relationships between Tourists’ Emotional Experiences, Perceived

- Overall Image, Satisfaction, and Intention to Recommend. *Journal of Travel Research*, 56, 41–54. doi:10.1177/0047287515620567.
- 770 Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41–55. doi:10.1093/biomet/70.1.41.
- Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician*, 39, 33–38.
- 775 De la Rosa de Sàa, S., Gil, M., Gonzalez-Rodriguez, G., López, M. T., & Lubiano, M. (2015). Fuzzy rating scale-based questionnaires and their statistical analysis. *IEEE Transactions on Fuzzy Systems*, 23, 111–126. doi:10.1109/TFUZZ.2014.2307895.
- 780 Song, H., Li, G., van der Veen, R., & Chen, J. L. (2011). Assessing mainland Chinese tourists' satisfaction with Hong Kong using tourist satisfaction index. *International Journal of Tourism Research*, 13, 82–96. doi:10.1002/jtr.801.
- Song, H., van der Veen, R., Li, G., & Chen, J. (2012). The Hong Kong tourist satisfaction index. *Annals of Tourism Research*, 39, 459–479. doi:10.1016/j.annals.2011.06.001.
- 785 Steenkamp, J.-B., & Wedel, M. (1991). Segmenting retail markets on store image using a consumer-based methodology. *Journal of Retailing*, 67, 300–320.
- Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25, 1–21. doi:10.1214/09-STS313.
- 790 Wang, D., Niu, Y., Lu, L., & Qian, J. (2015). Tourism spatial organization of historical streets – A postmodern perspective: The examples of Pingjiang Road and Shantang Street, Suzhou, China. *Tourism Management*, 48, 370–385.

- Wang, Y., Ma, X., Lao, Y., & Wang, Y. (2014). A fuzzy-based customer clustering approach with hierarchical structure for logistics network optimization. *Expert Systems with Applications*, *41*, 521–534.
- Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach*. (5th ed.). South-Western, Cengage Learning, Nelson Education, Ltd.
- Xie, X., & Beni, G. (1991). A Validity Measure for Fuzzy Clustering. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, *13*, 841–847.
- Zadeh, L. (1965). Fuzzy sets. *Information and control*, *8*, 338–353.
- Zhang, E., J. Prater, & Lipkin, I. (2013). Feedback reviews and bidding in online auctions: An integrated hedonic regression and fuzzy logic expert system approach. *Decision Support Systems*, *55*, 894–902. doi:10.1016/j.dss.2012.12.025.
- Zimmermann, H. J. (2011). *Fuzzy set theory and its applications*. Kluwer Academic Press, Norwell.

Figures

	What	How	Why
0.	Matching	Propensity score matching	Cross-sectional samples not comparable
1.	Data recoding	Fuzzy numbers	Ordinal data
2.	Selection dissimilarity measure	Fuzzy distance	Fuzzy segmentation variables
3.	Selection clustering algorithm	Fuzzy clustering for fuzzy data	Fuzzy assignment of units to cluster
4.	Best partition & cluster validation	Fuzzy validity index	Fuzzy prototypes
5.	Labelling & profiling	Rank of fuzzy numbers Use of weighted data	Fuzzy prototypes, Membership degree
6.	Comparing partitions	Fuzzy distance	Different partitions across time

Figure 1: Steps of the suggested matching-clustering procedure.

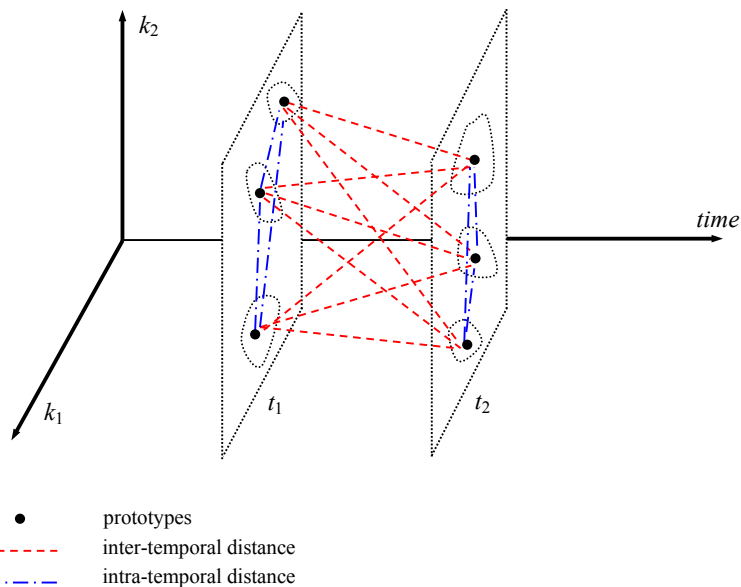
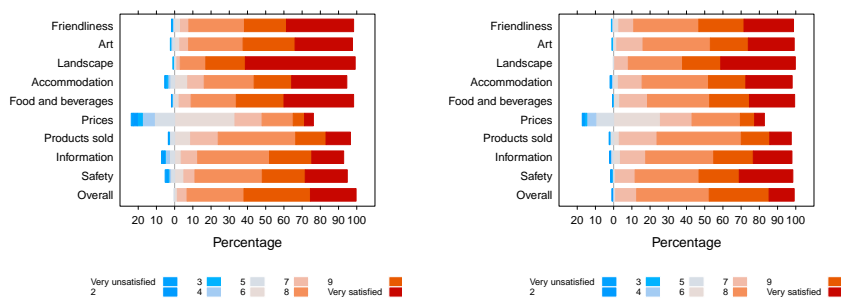


Figure 2: Inter- and intra-temporal distances between clusters.



(a) 2000-01

(b) 2010-11

Figure 3: % distribution for each aspect.

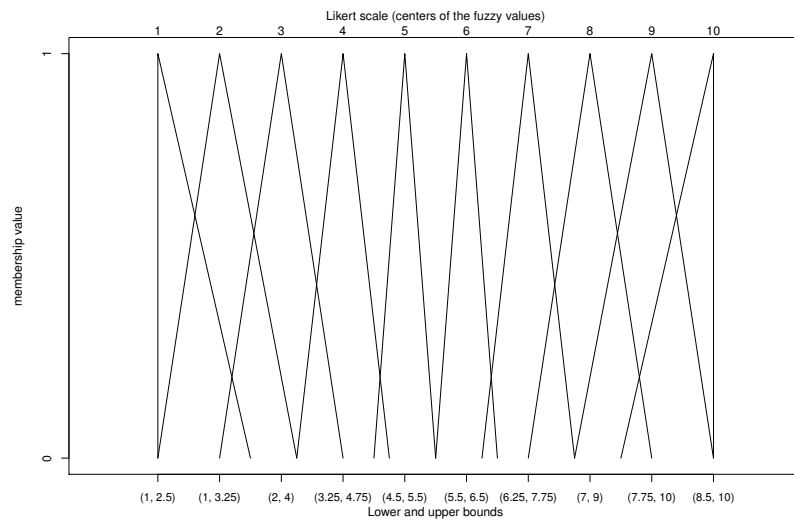
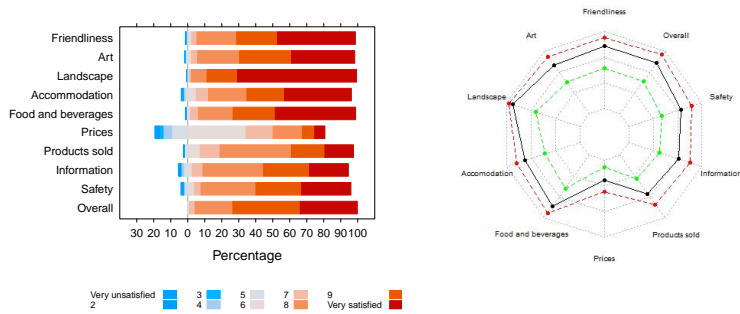
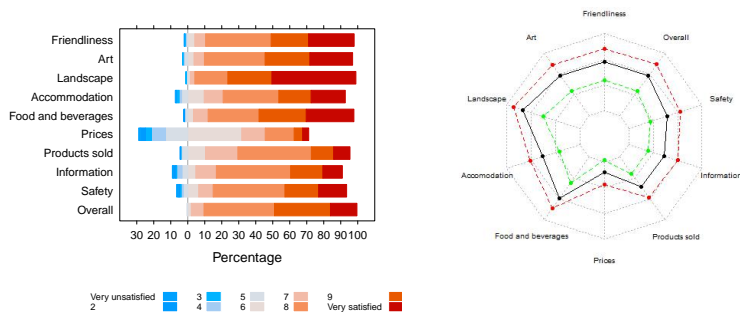


Figure 4: Linguistic satisfaction terms in the form of fuzzy numbers.

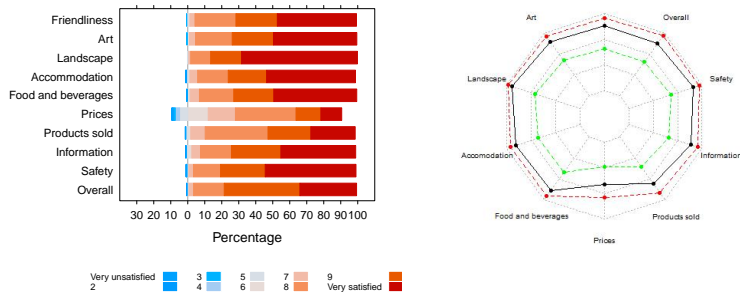


(a) 2000-01: Cluster 1

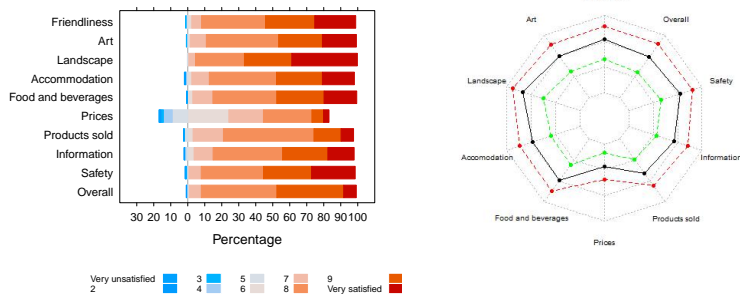


(b) 2000-01: Cluster 2

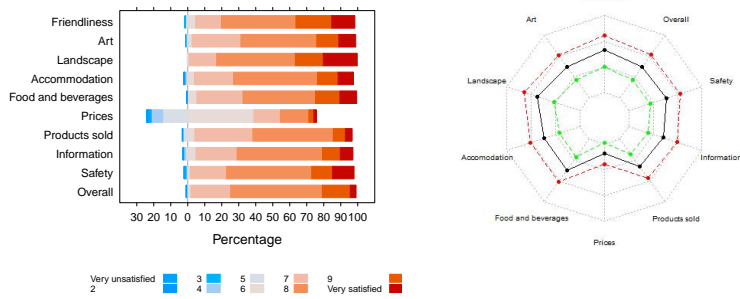
Figure 5: Cluster description 2000-01: weighted distribution of segmentation variables (on the left) and radar plots of fuzzy prototypes (on the right). In the radar plot, the black solid line represents the centres of the fuzzy prototypes, while the dashed lines represent the lower and upper values of the fuzzy prototypes.



(a) 2010-11: Cluster 1



(b) 2010-11: Cluster 2



(c) 2010-11: Cluster 3

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).

Segmentation variables	2000-01				2010-11					
	Cluster 1		Cluster 2		Cluster 1		Cluster 2		Cluster 3	
	Magnitude	Rank	Magnitude	Rank	Magnitude	Rank	Magnitude	Rank	Magnitude	Rank
Friendliness	9.096	3	8.339	5	9.221	8	8.607	3	7.995	2
Art	8.961	5	8.381	3	9.311	4	8.482	4	7.711	5
Landscape	9.531	1	8.972	1	9.596	1	8.969	1	8.099	1
Accommodation	8.821	6	7.768	7	9.345	3	8.378	7	7.685	7
Food and beverages	9.132	2	8.433	2	9.277	5	8.436	5	7.711	6
Prices	6.685	10	6.080	10	7.952	10	6.811	10	6.045	10
Products sold	8.273	9	7.627	9	8.806	9	7.951	9	7.470	9
Information	8.510	8	7.650	8	9.237	6	8.286	8	7.623	8
Safety	8.664	7	7.846	6	9.391	2	8.637	2	7.795	3
Overall	9.090	4	8.340	4	9.231	7	8.408	6	7.725	4

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

			Fuzzy prototypes					
			2000-01			2010-11		
			Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 3	
Fuzzy prototypes	2000-01	Cluster 1	0					
		Cluster 2	6.763	0				
	2010-11	Cluster 1	4.038	18.858	0			
		Cluster 2	3.198	2.277	9.865	0		
		Cluster 3	15.551	2.760	29.678	5.830	0	

Table 3: Cluster profile.

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

Notes: Weighted percentage distributions and weighted means are reported. Significance of the χ^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 $p \leq 0.1$.

Appendix

Fuzzy numbers

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

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

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:

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

where $\tilde{\mathbf{x}}_i$ 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:

$$\begin{cases} \min : & \sum_{i=1}^N \sum_{c=1}^C u_{ic}^p d_F^2(\tilde{\mathbf{x}}_i, \tilde{\mathbf{h}}_c) \\ s.t. & \sum_{c=1}^C u_{ic} = 1, \quad u_{ic} \geq 0 \end{cases} \quad (3)$$

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 ($c = 1, \dots, C$) cluster; $d_F^2(\tilde{\mathbf{x}}_i, \tilde{\mathbf{h}}_c)$ is the fuzzy distance, computed as in Eq. (2), between the i -th fuzzy unit ($\tilde{\mathbf{x}}_i$) and the fuzzy prototype that describes the c -th cluster ($\tilde{\mathbf{h}}_c$).

The iterative optimal solutions of (3) are:

$$u_{ic} = \frac{[w_M^2 (\|\mathbf{m}_i - \mathbf{h}_c^M\|^2) + w_S^2 (\|\mathbf{l}_i - \mathbf{h}_c^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_c^R\|^2)]^{-\frac{1}{p-1}}}{\sum_{c'=1}^C [w_M^2 (\|\mathbf{m}_i - \mathbf{h}_{c'}^M\|^2) + w_S^2 (\|\mathbf{l}_i - \mathbf{h}_{c'}^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_{c'}^R\|^2)]^{-\frac{1}{p-1}}},$$

$$\mathbf{h}_c^M = \frac{\sum_{i=1}^N u_{ic}^p \mathbf{m}_i}{\sum_{i=1}^N u_{ic}^p}, \quad \mathbf{h}_c^L = \frac{\sum_{i=1}^N u_{ic}^p \mathbf{l}_i}{\sum_{i=1}^N u_{ic}^p}, \quad \mathbf{h}_c^R = \frac{\sum_{i=1}^N u_{ic}^p \mathbf{r}_i}{\sum_{i=1}^N u_{ic}^p},$$

$$w_M = \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^p (\|\mathbf{l}_i - \mathbf{h}_c^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_c^R\|^2)}{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^p (\|\mathbf{m}_i - \mathbf{h}_c^M\|^2 + \|\mathbf{l}_i - \mathbf{h}_c^L\|^2 + \|\mathbf{r}_i - \mathbf{h}_c^R\|^2)}$$

$(w_S = 1 - w_M)$

where \mathbf{h}_c^M , \mathbf{h}_c^L and \mathbf{h}_c^R represent respectively the centre and the left and right spreads of the c -th fuzzy prototype.

835 *Fuzzy partition validity*

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

$$XB = \frac{\sum_{i=1}^N \sum_{c=1}^C u_{ic}^p d_F^2(\tilde{\mathbf{x}}_i, \tilde{\mathbf{h}}_c)}{\min_{c \neq c'} d_F^2(\tilde{\mathbf{h}}_c, \tilde{\mathbf{h}}_{c'})} \quad (4)$$

Labelling and profiling: rank and weighted data

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

$$M(\tilde{x}) = \frac{1}{2} \int_0^1 [2m + (l - r)z - l + r] f(z) dz \quad (5)$$

Table 1: Assessing matching quality.

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

Notes: Significance of the χ^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 $p \leq 0.1$.