Analysing Imprecise and Dependent Information

Elena Barzizza[1], Nicolò Biasetton[1], Marta Disegna[1], Alberto Molena[1]
[1]University of Padova, Department of Management and Engineering, Stradella San Nicola, 3, 36100 Vicenza, Italy
elena.barzizza@phd.unipd.it, nicolo.biasetton@phd.unipd.it, marta.disegna@unipd.it,
alberto.molena.1@phd.unipd.it

Abstract – *This paper presents a discussion on how to analyze imprecise and dependent information using traditional econometric models, supervised and unsupervised Machine Learning techniques. The discussion includes the presentation and analysis of real example data from the tourism field to familiarize the readers with imprecise and dependent information. Further developments in the treatment of such specific data are discussed in the conclusion.*

Keywords: imprecise data, dependent data, supervised Machine Learning, unsupervised Machine Learning

1. Introduction

Imprecise and dependent information can be found in various real case studies, with one of the most common applications being related to expenditures. In this paper, we will present a case study on tourism expenditure, providing a real data example. Dealing with dependent and imprecise information raises several issues that must be considered in both prediction and clustering frameworks. One of the main issues raising while analysing these kinds of data is the possibility that many individuals may spend nothing or very little during certain parts of their journey, leading to a high incidence of zeros in the data.

This paper presents an overview of the issues and solutions found in the existing literature and it is structured as follows. In Section 2 will be discussed and presented the kind of imprecise and dependent information with a focus of the tourism field given a real example data in sub-section 2.1. In Section 3 will be presented the economic theories at the basis of dependent expenditure, the main supervised and unsupervised machine learning algorithms used so far to analyse imprecise and dependent information. Conclusions and future directions are presented in the final section.

2. Imprecise and dependent information: the case of the tourism sector

The economic impact of tourism flows is often essential for those regions/local communities in which tourism is considered the major source of income [1]. To improve the economic effects of tourism visits, appropriate data and tools are needed to study the determinants of tourism expenditure and to analyse the tourists' spending behaviour in depth.

Studies on tourism demand have generally focused on the macroeconomic dynamics. Nevertheless, as stated for instance by Alegre and Pou [2], even though the micro-level perspective has been seldom analysed (e.g., see [3]–[5]), it provides several advantages over the macro-level studies. Among others, the micro-level approach makes it possible to observe individual choices regarding the consumption of a tourism commodity or service, and to analyse the heterogeneity and diversity that characterize individual tourism consumption behaviour. See [6] for an extensive review of the most common microeconomic models adopted in recent microdata studies.

The main advantage in adopting a micro-level approach is the possibility to simultaneously consider both the consumer behaviour theory on the decision-making process to purchase, and the neoclassical economic theory of budget constraint.

Specifically, the consumer behaviour theory assumes that the individual purchase process for a

tourism good or service is a two-decision process [7], i.e. the decision to purchase something followed by the decision on how much to spend on it. The economic theory of budget constraint is based on the assumption of weak separability between goods and services that leads tourists to allocate their budgets in accordance with a three-stage tourist spending process [8]: firstly, tourists decides how much of their budget to allocate for travel; secondly, they decide where to go on vacation; thirdly, they choose how to allocate their tourist budget among various goods and services offered by the selected destination. Obviously, these two economic theories are not disjoint but overlap; this means that an individual must make a two-stage decision process in each stage comprised in the three-stage tourist spending process. Since an in-depth knowledge of the determinants that affect tourist consumption behaviour is of primary importance for a destination [9], it is fundamental to accurately identify the set of factors that affect each stage of the decision-making process. In fact, the factors that determine tourism participation and tourism expenditure can be different and/or cannot have the same impact ([2], [10], [11]). At the same time, it is also fundamental to study the ways by which tourists choose a bundle of goods and services they consume/purchase at the destination to maximize their utility within certain budget constraints ([9], [12]). However, only few studies have analysed tourism expenditure behaviour by simultaneously considering both the consumer behaviour theory and the budget constraint theory ([13] [14]). Furthermore, many surveys conducted at either the national or international level, which investigate tourism expenditures, do not require proof (i.e., receipts) of expenditures, resulting in imprecise and vague information.

2.1 Real example data

The "International Tourism in Italy" survey, conducted annually by the Bank of Italy, has the aim of monitoring both travel expenditures and length of stay of inbound and outbound visitors from/to Italy determining the tourism balance of payments. The stratified sampling method is applied (using different types of stratified variables per each type of frontier) and face–to–face interviews are carried out at national borders (including highways, railway, airports, and harbours). Sampling is done independently at each type of frontier. Tourists are interviewed at the end of the trip when they are returning to their place of habitual residence and no proof of actual expenditure is required.

Therefore, tourism expenditures provide imprecise and vague information since they reflect what tourists remember about their trip, which can be different from what was spent. Interviews are conducted at different times of the day, during both working days and holidays, and month by month, with a fixed number of interviews per each period of survey. The questionnaires are anonymous and are offered in 14 languages. Socio–demographic characteristics of the interviewee (such as age, occupation, and country of origin), information on the trip (such as travel group size) and information on travel expenditures are collected.

3. Literature review
3.1 The economic theories of tourists' expenditure behaviour

Following the neoclassical economic theory of budget constraint, consumers are supposed to be rational and able to maximize their utility function by choosing among a set of available alternatives. Consumers are hence able to rank goods and services, so that they select the combinations from which their utility function gains the largest possible value, given budget and time constraints, relative prices, and preferences. Furthermore, the consumer's utility function is assumed to be "separable". The separability and the assumption of weak separability, only implies independence among broad aggregates of commodities, and not independence among individual commodities belonging to the same aggregate. It implies that the budgeting procedure by which individuals allocate their incomes among different goods and services is split in two stages [15]. Firstly, the individual decides in which broad commodity groups (i.e. food, tourism, housing, clothes, etc.) to allocate her/his income. Secondly, the individual decides which goods and services she/he wants to buy within each group, with no reference to expenditure in the other groups. Syriopoulos and Sinclair [8] applied this approach to the field of tourism suggesting a three-stage budgeting process. In the first stage, visitors allocate budget between the total tourism expenditure and the consumption of other goods and services. In the second stage, visitors allocate their tourist budget among different destinations, including the home country. Thirdly, visitors choose how to allocate their tourist budget among various goods and services offered by the selected destination. This means that the expenditure on different goods and services at the destination are dependent among each other but are independent of

tourism expenditure for goods and services made in another destination.

In each stage of this tourism consumption budgeting process, the consumer firstly divides the set of goods and services into two categories: the goods and services that he/she wants to obtain and the group of goods and services in which he/she is not interested. Obviously, this decision depends on both economic factors, and non-economic factors (ethnicity, gender, psychological, etc.). Subsequently, the consumer will decide the amount of money that he/she is willing to pay for each good and service belonging to the desired group of products. Therefore, following the consumer choice theory, each decision-making process to purchase can be split in two stages, or decisions [7]: the decision to spend or not (the selection stage) and, if the consumer decides to spend, how much money to spend (the outcome stage).

3.2 Supervised Machine Learning algorithms
3.2.1 Modelling dependent information

As introduced in the previous paragraph, following the consumer choice theory, the decision-making process to purchase a tourism good or service can be schematically represented by a two-stage decision process. For this reason, the tourism literature has been devoted along the years to the study of each of these stages, both separately and jointly. The binary regression models, such as Logit and Probit models ([2], [10], [16], [17]) but also the more recent Scobit model [18], have been intensively used in order to find those variables that determine the probability, or propensity, to consume a specific tourism good or service (i.e. the selection stage). As regards the second stage, i.e. the outcome stage, the OLS method has been extensively used to estimate multiple regression models in which the dependent variable was the amount of money spent for a particular tourism expenditure category or for the whole trip (among others, see [5], [17]–[19]). More recently, the quantile regression model has been introduced in the tourism literature [1], [14], [20], [21]. The main advantage of the quantile regression model compared to the traditional OLS regression model is that it can identify the determinants of the whole distribution of the dependent variable, i.e., the tourism expenditure, instead of only the determinants of the average expenditure. The Tobit model, introduced by Tobin in the late 50s, has been the first model created with the aim to study simultaneously the two-stages of the decision-making process [22]. Along the years this

model became quite common in the tourism literature [11], [23]–[27] but probably due to its lack in identifying separately the determinants of the two stages, the double-hurdle models, e.g. Cragg [28] and Heckman [29] models, have grown in popularity [30]–[36].

With the aim to incorporate the neoclassical economic theory of budget constraint in the study of the decision-making process, i.e. considering the dependence that might exist amongst alternative tourism expenditures on goods and services, two different system-of-equation approaches have been adopted in the tourism literature: the system of tobit equations [13], [37], [38]; and the almost ideal demand system (AIDS) model. The AIDS model, firstly introduced by Deaton and Muellbauer [15], has been commonly used to estimate the effect of relative prices and real expenditure on aggregate travel expenditures, but few studies have adopted this model also to estimate individual budget allocations of tourist expenditure for specific trips [14], [39] and to examine the third stage of the tourist budget allocation process [39]–[42]. Unfortunately, none of these systems of equations account for the imprecision of tourists' expenditure.

3.2.2 Modelling imprecise information

The standard linear regression model is the most employed approach for assessing the cause-and-effect connection between a response variable and a set of predictor variables or covariates. However, when one or more of the involved variables (whether they are the response or predictors) are not precisely defined or accurately measured, the standard linear regression model proves inadequate in capturing the relationship between the response and predictor variables. In such cases, it becomes necessary to turn to the fuzzy regression model [43]–[46] and hence work in the framework of fuzzy set theory [47] and fuzzy numbers (see Remark 1).

Remark 1 Fuzzy numbers

The LR-type (Left and Right) fuzzy data is a general class of fuzzy data that can be defined in a matrix form as follows [48]:

$$\tilde{Y} \equiv \{\tilde{y}_i = (m_{1i}, m_{2i}, l_i, r_i)_{LR} : i = 1, \dots, n\}$$

where $\tilde{y}_i$ is the LR fuzzy data observed on the ith unit; m_{1i} and m_{2i} (with $m_{1i} < m_{2i}$) represent the left and right centers of the fuzzy number, l_i and r_i represent the left and right spreads, i.e., the vagueness of the data. Once defined the general form of the fuzzy number, the membership function (see Remark 2) must be chosen [49].

Remark 2 Elicitation problem

The definition of the membership function (elicitation) and its specification are two important issues widely discussed in the literature [71]. The membership function is commonly developed based on expert's capabilities and knowledge of the topic under investigation [72], [73]. The membership function is normally defined at macro level, i.e., the sample, rather than at individual level. In other words, the MF is defined equal and constant for a variable, regardless of individual characteristics, both physical, cultural and psychological, and question's characteristics.

As per the comprehensive review conducted by Chukhrova and Johannssen [50], fuzzy regression analysis has seen substantial expansion and has gained increased importance in the field of fuzzy statistics, as confirmed by Li et al. [49]. In the literature, two main approaches to estimate a fuzzy regression model have been developed: the possibilistic and the fuzzy least squares approaches (refer to D'Urso [45], and Li et al., [49]). More recently, the Machine Learning approach has been added to these methods (as discussed in [50]). The possibilistic approach is based on the pioneering works of Tanaka et al [51]. The idea behind this method is to minimize the entire fuzziness (i.e., the vagueness of a phenomenon that cannot be expressed by randomness) of the predicted dependent variable by minimizing the total spread of the fuzzy parameters. This approach is known as possibilistic since the membership functions of fuzzy sets can be seen as possibility distributions. Since its introduction, this approach has been extensively used in applications, and several theoretical advancements have been suggested so far ([43], [52]–[55]). On the other hand, the least squares approach (firstly introduced by Celmiņš [56] and Diamond [57]) is an extension of the well-known least squares criterion. Therefore, its aim is to identify the linear model that best approximates the observed data in a metric space, in other words, the aim is to minimize the distance between the estimated fuzzy outputs and the observed fuzzy outputs using a suitable distance measure between fuzzy numbers. This approach has been extensively developed in the literature, as demonstrated by the large number of research papers published so far [43], [45], [49], [57]–[62]. Finally, the performance of the fuzzy regression has been enhanced by incorporating machine learning techniques, such as evolutionary algorithms, neural networks or support vector machines. See the extensive review of Chukhrova and Johannssen [50] for applications and developments of this approach.

3.4 Unsupervised Machine Learning algorithms

Cluster analysis is an unsupervised Machine Learning (ML) method used for conducting post hoc market segmentation. The primary objective of cluster analysis is to uncover concealed relationships among data points while grouping items to maximize similarity within groups and minimize dissimilarity between them. Given its exploratory nature, each clustering algorithm may yield a distinct partition for the same dataset. Consequently, there are not inherently right or wrong results; the utility of the outcomes depends on the compatibility of the clustering algorithm with the data's characteristics and requirements.

3.4.1 Clustering dependent information

According to Lee and Beeler [63] as well as Koc and Altinay [64], the development and maintenance of a competitive advantage in a highly competitive tourism markets hinge significantly on the degree to which visitors are well known and understood. Kau and Lim [65] emphasize that market segmentation enables destination planners to allocate resources more efficiently to attract distinct and unique groups of travellers. So, the process of market segmentation consists in the identification of groups of consumers that are similar for one or more behaviour, and then devises marketing strategies that appeal to one or more groups. Consequently, this kind of analysis can shed light on the existence of groups of tourists whose shared attitudes may lead to similar levels of spending and, consequently, a comparable economic impact on the region. In literature, research aiming to identify clusters of tourists who share the same expenditure behaviour can be found in [58]–[62], [66]–[68]. In all these articles, clustering analysis is performed considering different kind of expenditures as independent variables, i.e., without considering the dependence relationship that intrinsically characterise these kinds of data. To the best of our knowledge, no clustering algorithms for dependent variables, such as probabilistic clustering or copula-based clustering algorithms, have been suggested in the literature so far.

3.4.2 Clustering imprecise information

When imprecise information is used as segmentation variables in a clustering algorithm, a suitable distance measure must be adopted. Typically, imprecise information is converted into fuzzy numbers before being used in the clustering algorithm, and a

distance measure for fuzzy data must be selected. For a discussion on fuzzy distance, you can refer to Coppi et al. [69]. Recent applications of clustering algorithms for fuzzy data can be found in [70] and [6].

4. Conclusions and future directions

Tourism expenditures are vague and imprecise information usually collected through national repeated-cross sectional surveys or dedicated surveys. To model such imprecise information, Fuzzy set theory can be adopted. Practically, tourism expenditures can be converted into fuzzy numbers before the implementation of either a fuzzy regression model or a clustering algorithm using fuzzy distance. Since the imprecision by which tourists give information about their expenditures can be assumed to be a function of both length of stay (the longer the holiday, the higher the imprecision) and amount of money spent at the destination (the higher the expenditure, the higher the imprecision), a new method to compute the latent individual imprecision of the fuzzy number must be investigated.

According to the economic theory of budget constraint, tourists generally allocate their budget following to a three-stage process [8]. This theory assumes the principle of weak separability among goods and services: expenditures on different goods and services belonging to the same category (i.e. travel) are dependent on each other while broad aggregates of commodities are independent among them. Moreover, expenditures are generally characterised by a high level of zero values. The dependence among censored variables (hence tourism expenditures) should be further investigated and the applicability of either suitably Copula functions should be analysed.

Finally, to the best of our knowledge, no clustering algorithm has been developed so far to identify group of visitors with similar expenditure behaviour at the destination accounting for both the imprecision of data collected and the dependence among different expenditure categories. Therefore, there is a need in developing adequate clustering algorithms for dependent and imprecise information.

Concluding, we emphasize that imprecise information and dependent information have been extensively analysed in the literature, bridging from econometric modelling to machine learning algorithms. However, to the best of our knowledge, these kinds of information have been studied mostly separately, and no methods have been implemented so far to analyse information that is both dependent and imprecise.

References

[1] W. Hung, J. Shang, and F. Wang, "Another look at the determinants of tourism expenditure.," *Annals of Tourism Research*, vol. 39, no. 1, pp. 495–498, 2012.

[2] J. Alegre and L. Pou, "Micro-economic determinants of the probability of tourism consumption," *Tourism Economics*, vol. 10, no. 2, pp. 125–144, 2004.

[3] G. I. Crouch, "The study of international tourism demand: A review of findings," *Journal of Travel research*, vol. 33, no. 1, pp. 12–23, 1994.

[4] P. Fredman, "Determinants of visitor expenditures in mountain tourism," *Tourism Economics*, vol. 14, no. 2, pp. 297–311, 2008.

[5] Y. Wang and M. C. Davidson, "A review of micro-analyses of tourist expenditure," *Current issues in Tourism*, vol. 13, no. 6, pp. 507–524, 2010.

[6] P. D'Urso, L. De Giovanni, M. Disegna, R. Massari, and V. Vitale, "A tourist segmentation based on motivation, satisfaction and prior knowledge with a socio-economic profiling: A clustering approach with mixed information," *Social Indicators Research*, vol. 154, pp. 335–360, 2021.

[7] S. Pudney, *Modelling individual choice: the econometrics of corners, kinks and holes*. Basil Blackwell, 1991.

[8] T. C. Syriopoulos and M. Thea Sinclair, "An econometric study of tourism demand: the AIDS model of US and European tourism in

Mediterranean countries," *Applied economics*, vol. 25, no. 12, pp. 1541–1552, 1993.

[9] S. Divisekera, "Economics of tourist's consumption behaviour: Some evidence from Australia," *Tourism Management*, vol. 31, no. 5, pp. 629–636, 2010.

[10] J. Alegre, S. Mateo, and L. Pou, "An analysis of households' appraisal of their budget constraints for potential participation in tourism," *Tourism Management*, vol. 31, no. 1, pp. 45–56, 2010.

[11] J. G. Brida, M. Disegna, and L. Osti, "Visitors' expenditure behaviour at cultural events: the case of Christmas markets," *Tourism Economics*, vol. 19, no. 5, pp. 1173–1196, 2013.

[12] D. C. Wu, G. Li, and H. Song, "Economic analysis of tourism consumption dynamics: A time-varying parameter demand system approach," *Annals of Tourism Research*, vol. 39, no. 2, pp. 667–685, 2012.

[13] A. Bilgic, W. J. Florkowski, J. Yoder, and D. F. Schreiner, "Estimating fishing and hunting leisure spending shares in the United States," *Tourism Management*, vol. 29, no. 4, pp. 771–782, 2008.

[14] K.-L. Chang, C.-M. Chen, and T. J. Meyer, "A comparison study of travel expenditure and consumption choices between first-time and repeat visitors," *Tourism Management*, vol. 35, pp. 275–277, 2013.

[15] A. Deaton and J. Muellbauer, *Economics and consumer behavior*. Cambridge university press, 1980.

[16] M.-S. Oh, J. W. Choi, and D.-G. Kim, "Bayesian inference and model selection in latent class logit models with parameter constraints: an application to market segmentation," *Journal of Applied Statistics*, vol. 30, no. 2, pp. 191–204, 2003.

[17] S. S. Kim, B. Prideaux, and K. Chon, "A comparison of results of three statistical methods to understand the determinants of festival participants' expenditures," *International Journal of Hospitality Management*, vol. 29, no. 2, pp. 297–307, 2010.

[18] L. Wu, J. Zhang, and A. Fujiwara, "Tourism participation and expenditure behaviour: Analysis using a scobit based discrete–continuous choice model," *Annals of Tourism Research*, vol. 40, pp. 1–17, 2013.

[19] M. Belenkiy and D. Riker, "Modeling the international tourism expenditures of individual travelers," *Journal of Travel Research*, vol. 52, no. 2, pp. 202–211, 2013.

[20] A. A. Lew and P. T. Ng, "Using quantile regression to understand visitor spending," *Journal of Travel Research*, vol. 51, no. 3, pp. 278–288, 2012.

[21] E. Marrocu, R. Paci, and A. Zara, "Micro-economic determinants of tourist expenditure: A quantile regression approach," *Tourism Management*, vol. 50, pp. 13–30, 2015.

[22] J. Tobin, "Estimation of relationships for limited dependent variables," *Econometrica: journal of the Econometric Society*, pp. 24–36, 1958.

[23] H.-C. Lee, "Determinants of recreational boater expenditures on trips," *Tourism Management*, vol. 22, no. 6, pp. 659–667, 2001.

[24] S. S. Kim, H. Han, and K. Chon, "Estimation of the determinants of expenditures by festival visitors," *Tourism Analysis*, vol. 13, no. 4, pp. 387–400, 2008.

[25] F. Muñoz-Bullón, "The gap between male and female pay in the Spanish tourism industry," *tourism Management*, vol. 30, no. 5, pp. 638–649, 2009.

[26] B. Zheng and Y. Zhang, "Household expenditures for leisure tourism in the USA, 1996 and 2006," *International Journal of Tourism Research*, vol. 15, no. 2, pp. 197–208, 2013.

[27] W. G. Kim, T. Kim, G. Gazzoli, Y. Park, S. H. Kim, and S. S. Park, "Factors affecting the travel expenditure of visitors to Macau, China," *Tourism Economics*, vol. 17, no. 4, pp. 857–883, 2011.

[28] J. G. Cragg, "Some statistical models for limited dependent variables with application to the demand for durable goods," *Econometrica: journal of the Econometric Society*, pp. 829–844, 1971.

[29] J. J. Heckman, "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, pp. 153–161, 1979.

[30] R. O. Weagley and E. Huh, "Leisure expenditures of retired and near-retired households," *Journal of Leisure Research*, vol. 36, no. 1, pp. 101–127, 2004.

[31] J. L. Nicolau and F. J. Mas, "Stochastic modeling: a three-stage tourist choice process," *Annals of Tourism Research*, vol. 32, no. 1, pp. 49–69, 2005.

[32] S. Jang, S. Ham, and G.-S. Hong, "Food-away-from-home expenditure of senior households in the United States: A double-hurdle approach," *Journal of Hospitality & Tourism Research*, vol. 31, no. 2, pp. 147–167, 2007.

[33] S. S. Jang and S. Ham, "A double-hurdle analysis of travel expenditure: Baby boomer seniors versus older seniors," *Tourism management*, vol. 30, no. 3, pp. 372–380, 2009.

[34] S. M. D. Brandolini and M. Disegna, "Demand for the quality conservation of Venice, Italy, according to different nationalities," *Tourism Economics*, vol. 18, no. 5, pp. 1019–1050, 2012.

[35] J. Alegre, S. Mateo, and L. Pou, "Tourism participation and expenditure by Spanish households: The effects of the economic crisis and unemployment," *Tourism Management*, vol. 39, pp. 37–49, 2013.

[36] C. Bernini and M. F. Cracolici, "Demographic change, tourism expenditure and life cycle behaviour," *Tourism Management*, vol. 47, pp. 191–205, 2015.

[37] G.-S. Hong, J. X. Fan, L. Palmer, and V. Bhargava, "Leisure travel expenditure patterns by family life cycle stages," *Journal of Travel & Tourism Marketing*, vol. 18, no. 2, pp. 15–30, 2005.

[38] Y. Lee, M. Hickman, and S. Washington, "Household type and structure, time-use pattern, and trip-chaining behavior," *Transportation Research Part A: Policy and Practice*, vol. 41, no. 10, pp. 1004–1020, 2007.

[39] S. K. Lee, W. S. F. Jee, D. C. Funk, and J. S. Jordan, "Analysis of attendees' expenditure patterns to recurring annual events: Examining the joint effects of repeat attendance and travel distance," *Tourism management*, vol. 46, pp. 177–186, 2015.

[40] S. Divisekera, "Ex post demand for Australian tourism goods and services," *Tourism Economics*, vol. 15, no. 1, pp. 153–180, 2009.

[41] D. Chenguang Wu, G. Li, and H. Song, "Analyzing tourist consumption: A dynamic system-of-equations approach," *Journal of Travel Research*, vol. 50, no. 1, pp. 46–56, 2011.

[42] D. C. Wu, G. Li, and H. Song, "Economic analysis of tourism consumption dynamics: A time-varying parameter demand system approach," *Annals of Tourism Research*, vol. 39, no. 2, pp. 667–685, 2012.

[43] Y.-H. O. Chang and B. M. Ayyub, "Fuzzy regression methods–a comparative assessment," *Fuzzy sets and systems*, vol. 119, no. 2, pp. 187–203, 2001.

[44] R. Coppi and P. D'Urso, "Three-way fuzzy clustering models for LR fuzzy time trajectories," *Computational statistics & data analysis*, vol. 43, no. 2, pp. 149–177, 2003.

[45] P. D'Urso, "Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data," *Computational Statistics & Data Analysis*, vol. 42, no. 1–2, pp. 47–72, 2003.

[46] P. D'Urso and R. Massari, "Fuzzy clustering of human activity patterns," *Fuzzy Sets and Systems*, vol. 215, pp. 29–54, 2013.

[47] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[48] D. Dubois and H. Prade, "On fuzzy syllogisms," *Computational Intelligence*, vol. 4, no. 2, pp. 171–179, 1988.

[49] Y. Li, X. He, and X. Liu, "Fuzzy multiple linear least squares regression analysis," *Fuzzy Sets and Systems*, vol. 459, pp. 118–143, 2023.

[50] N. Chukhrova and A. Johannssen, "Fuzzy regression analysis: systematic review and bibliography," *Applied Soft Computing*, vol. 84, p. 105708, 2019.

[51] H. Asai, S. Tanaka, and K. Uegima, "Linear regression analysis with fuzzy model," *IEEE Trans. Systems Man Cybern*, vol. 12, pp. 903–907, 1982.

[52] K. J. Kim, H. Moskowitz, and M. Koksalan, "Fuzzy versus statistical linear regression,"

European Journal of Operational Research, vol. 92, no. 2, pp. 417–434, 1996.

[53] P. Diamond and H. Tanaka, "Fuzzy regression analysis," in *Fuzzy sets in decision analysis, operations research and statistics*, Springer, 1998, pp. 349–387.

[54] M. Hojati, C. R. Bector, and K. Smimou, "A simple method for computation of fuzzy linear regression," *European Journal of Operational Research*, vol. 166, no. 1, pp. 172–184, 2005.

[55] W.-L. Hung and M.-S. Yang, "Fuzzy entropy on intuitionistic fuzzy sets," *International journal of intelligent systems*, vol. 21, no. 4, pp. 443–451, 2006.

[56] A. Celmiņš, "Least squares model fitting to fuzzy vector data," *Fuzzy sets and systems*, vol. 22, no. 3, pp. 245–269, 1987.

[57] P. Diamond, "Fuzzy least squares," *Information sciences*, vol. 46, no. 3, pp. 141–157, 1988.

[58] M. J. Carneiro, C. Eusebio, and M. Pelicano, "An expenditure patterns segmentation of the music festivals' market," *International journal of sustainable development*, vol. 14, no. 3–4, pp. 290–308, 2011.

[59] M. S. Rosenbaum and D. L. Spears, "An exploration of spending behaviors among Japanese tourists," *Journal of Travel Research*, vol. 44, no. 4, pp. 467–473, 2006.

[60] S. Dolnicar, G. I. Crouch, T. Devinney, T. Huybers, J. J. Louviere, and H. Oppewal, "Tourism and discretionary income allocation. Heterogeneity among households," *Tourism Management*, vol. 29, no. 1, pp. 44–52, 2008.

[61] J. Lima, C. Eusébio, and E. Kastenholz, "Expenditure-based segmentation of a mountain destination tourist market," *Journal of Travel & Tourism Marketing*, vol. 29, no. 7, pp. 695–713, 2012.

[62] M. Saayman, A. Saayman, and E.-M. Joubert, "Expenditure-based segmentation of visitors to the Wacky Wine Festival," *Tourism recreation research*, vol. 37, no. 3, pp. 215–225, 2012.

[63] J. Lee and C. Beeler, "An investigation of predictors of satisfaction and future intention: Links to motivation, involvement, and service quality in a local festival," *Event management*, vol. 13, no. 1, pp. 17–29, 2009.

[64] E. Koc and G. Altinay, "An analysis of seasonality in monthly per person tourist spending in Turkish inbound tourism from a market segmentation perspective," *Tourism management*, vol. 28, no. 1, pp. 227–237, 2007.

[65] A. K. Kau and P. S. Lim, "Clustering of Chinese tourists to Singapore: An analysis of their motivations, values and satisfaction," *International Journal of tourism research*, vol. 7, no. 4-5, pp. 231–248, 2005.

[66] R.-H. Sun, J.-X. Hao, D. K. C. Fong, R. Law, and Y. Yu, "Exploring the Expenditure-based Profile of Macao Visitors: a Cluster Analysis.," in *PACIS*, 2016, p. 133.

[67] C. Eusébio, A. Lopes, and M. J. Carneiro, "Diverse expenditure patterns of international tourists on Santiago Island—Cape Verde," *Tourism Planning & Development*, vol. 14, no. 3, pp. 389–410, 2017.

[68] A. Rabasa, A. Pérez-Martín, and D. Giner, "Optimal clustering techniques for the segmentation of tourist spending. Analysis of tourist surveys in the Valencian community (Spain): a case study," *International Journal of Design & Nature and Ecodynamics*, vol. 12, no. 4, pp. 482–491, 2018.

[69] R. Coppi, P. D'Urso, and P. Giordani, "Fuzzy and possibilistic clustering for fuzzy data," *Computational Statistics & Data Analysis*, vol. 56, no. 4, pp. 915–927, 2012.

[70] N. Biasetton, M. Disegna, E. Barzizza, and L. Salmaso, "A new adaptive membership function with CUB uncertainty with application to cluster analysis of Likert-type data," *Expert Systems with Applications*, vol. 213, p. 118893, 2023.

[71] S. de la R. de Sáa, M. Á. Gil, G. González-Rodríguez, M. T. López, and M. A. Lubiano, "Fuzzy rating scale-based questionnaires and their statistical analysis," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 1, pp. 111–126, 2014.

[72] A. Rahmani, F. Hosseinzadeh Lotfi, M. Rostamy-Malkhalifeh, and T. Allahviranloo, "A new method for defuzzification and ranking of fuzzy numbers based on the statistical beta

distribution," *Advances in Fuzzy Systems*, vol. 2016, 2016.

[73] R. Coppi, P. D'Urso, P. Giordani, and A. Santoro, "Least squares estimation of a linear regression model with LR fuzzy response," *Computational Statistics & Data Analysis*, vol. 51, no. 1, pp. 267–286, 2006.