



Article A Regression Tree Approach for Investigating the Impact of High Speed Rail on Tourists' Choices

Francesca Pagliara *^(D), Filomena Mauriello^(D) and Lucia Russo^(D)

Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy; filomena.mauriello@unina.it (F.M.); luciarusso993@gmail.com (L.R.)

* Correspondence: fpagliar@unina.it

Received: 17 December 2019; Accepted: 24 January 2020; Published: 26 January 2020



Abstract: This paper provides a contribution to the international literature by applying regression tree methods to the analysis of the expected effects of the High Speed Rail project in Italy on the tourism market. This approach, as far as the author knows, has never been applied in this context. Tourism and transport information have been gathered for 99 Italian provinces during the 2006–2016 period. Tree-structured methods have been chosen as an application of regression models in which some explanatory variables are used as covariates to predict the dependent variable values on the basis of some decision rules. This approach establishes a casual effect between dependent and independent variables. The dependent variables chosen are the Italian and foreign tourists, and the number of overnights spent by Italians and foreigners. Among the independent variables are the presence of HSR, the presence of first-level airport hubs and the number of attractions in a given province, the presence of the sea, the population and the percentage of unemployment. The main outcome of this study is that HSR affects the tourism market.

Keywords: high speed rail; tourists' behavior; regression tree-methods

1. Introduction

There are many studies analyzing the factors affecting tourists' destination choice and Hunt can be considered as the pioneer of this approach [1]. A tourist destination can be considered as a product made up of natural resources, infrastructures and services, cultural events, etc. [2]. The attributes "shaping" this product are related to its attractiveness, its facilities and to the accessibility provided by transportation system [3]. Specifically, the attractiveness factors generate flows to a given destination. The tourist facilities are fundamental, since their absence prevents tourists from travelling to enjoy these attractions. In general, the accessibility to given tourist destinations is related to the different transport modes available to reach them [4].

New interventions in the transportation system bring an increase in accessibility, which in turn fosters tourism development [5]. Khadarooa and Seetanah stated that the "provision of suitable transport has transformed dead centers of tourist interest into active and prosperous places attracting multitudes of people" [6]. Alternative transport modes have been changing over the centuries according to the development of technology. Since 1964, with the Shinkansen in Japan, the revolution in the transportation sector has been represented by High Speed Rail (HSR). The focus of this manuscript is to investigate whether HSR can induce changes in tourists' behaviour. Guirao and Campa [7] demonstrated that the high costs of new HSRs require a selection methodology to define which HSR corridors, within a network, should be built first, and the most suitable evaluation tool is the multi-criteria approach. Indeed, they showed that, in any corridor-ranking methodology, and especially in countries with high tourism attractiveness, the impacts of HSR on tourism should be considered

as a variable. The High Speed/High Capacity (HS/HC) Rail project in Italy has been chosen as a case study, where the first HSR line was inaugurated in 1992 between Florence and Rome [8]. The new generation of HSRs, running at 300 km/h, started in December 2005 between Rome and Napoli and Milan and Bologna. Later, in December 2009, the project was extended with the Milan–Turin and the Bologna–Florence lines. In 2010, the Italian HSR network was operational and other projects are still a work in progress. There are also the High Capacity (HC) Rail lines (see Figure 1), which speed up and increase the capacity of the existing rail lines.



Figure 1. The High-Speed/High-Capacity Rail system in Italy. Source: Delaplace et al. (2014).

In Table 1, the total number of tourists visiting Italy during the period 2006–2017 has been reported.

Table 1. Trend of tourism in Italy (in millions) (2006–2017).

Italians 172.4	171.2	166.5	165.5	165.5	166.4	158.2	155.0	180.9	190.2	203.5	209.3
Foreigners 142.3	148.3	147.8	145.8	151.6	161.1	164.9	167.9	176.8	182.6	199.4	210.7
Total 314.7	319.5	314.3	311.3	317.1	327.5	323.1	322.9	357.7	372.8	402.9	420.7

Source: Italian Census.

There are several papers in the international literature investigating the impact of HSR on tourism, following different approaches [9–11]. However, in this paper the methodological approach developed is new to this context, as far as the authors know. A panel dataset for 99 Italian provinces, built for the 2006–2016 period, has been collected, and the impact of the HS/HC rail project in Italy on tourism has been studied through the specification of regression trees.

The structure of the paper is as follows. In Section 2, a literature review on the impacts of HSR on tourism is reported. Section 3 describes the methodology. In Section 4, the case study is presented together with the main results. Section 5 deals with the results, while, in Section 6, the conclusions and further perspectives are shown.

2. HSR and Tourism: Is There a Link?

For Della Corte et al. [12], a tourist destination is represented by accessibility to a given destination; by attraction factors, by the presence of hotels, by the amenities, by the activity of tour operators and by the agencies providing tours.

According to Kaul [5], destinations which are characterised by an efficient transportation system, in terms of infrastructure and services, can experience a development of tourism. Several authors have contributed to the analysis of the relationship between HSR and the tourism market [13–20].

It is important to state, from the very beginning, that the contributions present in the literature provide information on tourism statistics that are sometimes aggregated (the sources are commonly represented by the Census); this means that is not easy to infer the real impact of HSR from them. However, only direct surveys of tourists can actually be considered the right way to get information concerning tourists' chosen travel mode for reaching a given destination. In the following, some international experiences are reported.

Bazin et al. [21] assessed the effects of HSR on urban and business tourism, choosing French case studies. The main outcome was that HSR was chosen for urban tourism, especially for short-stay tourism. Moreover, HSR proved to be cheaper when travelling in groups and is more accessible with respect to other alternative transport modes if the rail station is located in the city centre.

Chen and Haynes [22] demonstrated, for the case study of China, that HSR services significantly affected the tourism market. Kuriharaa and Wu [23] studied the tourism variation in Japan. Specifically, tourism arrivals increased in cities served by the network.

Delpalace et al. [10] analysed the link between HSR and theme parks, i.e., Disneyland Paris and Futuroscope Parks, served by an HSR station. In the case study of Disneyland, HSR had an impact on tourists' behaviour. On the other hand, for the case study of Futuroscope, HSR was not important.

Giurao and Soler [24] and Campa et al. [25] highlighted a positive relationship in Spain between the increase in tourism outputs and HSR deployment.

Albalate and Fageda [26] and Albalate et al. [27], for the same case study, showed a negative impact of HSR on tourist arrivals and revenues. This behaviour might be attributed to a network design that does not correspond to the riders' needs.

The impact of HSR on the travel behaviour of tourists in Taiwan in relation to time, space and carbon emissions was investigated by Sun and Lin [28]. In general, HSR had a weak effect on travel distance and length of stay, but a 10% reduction in transport carbon emissions through intermodal substitution was registered. For the same case study of China, Wang et al. [29] showed that HSR increased tourism-based economic relationships between cities.

There are also some contributions in the current literature which focus on the probability of re-visiting a given destination by HSR. This aspect has been discussed by Seddighi and Theocharous [30], who studied the probability of revisiting Cyprus in terms of socio-demographic and destination characteristics, as well as by Barros and Assaf [31], who concluded that the probability of revisiting Lisbon "increases significantly with accommodation range, events, food quality, expected weather, beach, overall quality, nightlife, reputation, and safety". Delaplace et al. [32] studied the factors that have an impact on destination choice for tourism purposes and the role of HSR systems in affecting the choice to revisit Rome and Paris. Pagliara et al. [33,34] followed the same approach by comparing the factors influencing the choice to revisit Madrid and Paris.

In general, the gap that can be found in these contributions is that the general approach is an econometric one. On the other hand, the methodological approach proposed in this manuscript is based on the regression-tree approach, which has never been applied in this context.

3. The Methodology

In this paper, regression trees methods have been proposed with objective of analysing the impacts of HSR projects on tourism. Other case studies have proposed the use of these methods for making inference on tourism, mainly for analysing tourism-based regional economic development [35,36]. The dataset collected contains information both on tourism and transport for 99 Italian cities during the 2006–2013 period. These data are panel data, i.e., a combination of cross-sectional and time series data, where all information for each city is observed during the period under analysis. In the literature, there are several approaches dealing with these data, like the mixed effects or generalized

estimating equation models [37]. The regression models have their basic assumptions and pre-defined underlying relationships between dependent and independent variables. If these assumptions are violated, the model could lead to erroneous estimation [38]. The use of tree-structured methods, like regression trees, can be seen as an application of a regression model in which some explanatory variables are used as covariates to predict the dependent variable values on the basis of some decision rules [39–42]. This approach establishes a casual effect between dependent and independent variables. These tree-structured methods do not require a priori probabilistic knowledge of the phenomena under study, and no assumption is required [43,44]. This can be considered as one of the main advantages of these methods with regards to standard econometric analysis [45]. A tree is a hierarchical and graphical representation of interactions between variables, formed by a finite number of nodes departing from the root node or father node (see Figure 2).



Figure 2. An example of a flow-chart structure of a regression tree. Each node contains means, standard deviations, number of observations, percentage of observations w.r.t. the total number of observations, and predicted values of the dependent variable. Source: Authors' elaboration adapted from IBM SPSS Advanced Statistics 20.

The tree building approach is based on the implementation of the Classification And Regression Tree (CART) algorithm proposed by Breiman et al. [46]. The tree is a binary recursive splitting algorithm, where each parent node is linked to two children nodes: the left and the right nodes. The children nodes can be classified into internal and terminal nodes. An internal node is recursively treated as a parent node, and the whole process continues for all the subsequent nodes. The internal node is

connected to its parent node at the top. At the bottom, a terminal node has no children nodes and represents the final result of a combination of decisions or events.

The splitting algorithm is based on increasing the internal homogeneity of the dependent variable *Y*. In order to perform recursive binary splitting, it is necessary to select the predictor $X_j \in \{X_1, ..., X_p\}$, and the cutpoint *c*, i.e., split the predictor space into the regions $\{X|X_j < c\}$ and $\{X|X_j \ge c\}$. This approach leads to the greatest reduction in impurity. Specifically, all predictors $X_1, ..., X_p$, and all possible values of the cutpoint *c* for each of the predictors are considered, and then the predictor and cutpoint are chosen by considering the resulting tree with the lowest impurity [47].

In the regression tree algorithm, the impurity of a node is measured by the Least-Squared Deviation (LSD) R(t), which is the within variance of the dependent variable for the node t. It can be expressed as follows

$$R(t) = \frac{1}{N_t} \sum_{t} \left[y_{i(t)} - \overline{y}_{(t)} \right]^2$$
(1)

where N_t is the number of observations in the node t, $y_{i(t)}$ are the individual values of the dependent variable at node t and $\overline{y}_{(t)}$ is the mean of the dependent variable at node t. Given the impurity function, R(t), the split s determines the two subsequent nodes, i.e., the left t_L (left) and t_R (right) nodes. The goodness of a split is measured by a decrease in impurity $\Delta I(s, t)$

$$\Delta I(s,t) = R(t) - p_L R(t_L) - p_R R(t_R)$$
⁽²⁾

where $R(t_L)$ is the sum of squares of the left child node and $R(t_R)$ is the sum of squares of the right child node generated by the split *s*, while p_L and p_R are the portions of units assigned to the left and right children nodes [48].

A criterion has been selected to arrest the tree building, i.e., the minimum decrease in impurity equal to 0.01 [49].

However, in addition to the analysis of the link between the dependent and independent variables, it has been possible to compute a predictor ranking (also known as variable importance) based on the contribution of the predictors in building up the tree. The ranking considers that an important variable might not appear in any split in the final tree when the tree includes another surrogate variable. The latter is defined when two variables X' and X'' are highly correlated and are strong competitors. In this case, if variable X' is first selected in the tree building process, this might prevent variable X'' from being selected, masking the influence of the variable itself [50]. If the masking variable is removed, this variable could show up in a prominent split in a new tree that is almost as good as the original [51]. The importance of the score of a given variable X_j is the sum of the improvement in impurity measures across all the nodes in the tree when it acts as a primary or surrogate splitter, as defined by Breiman et al. [40]

$$M(X_j) = \sum_{t \in T} \Delta I(\tilde{s}_j, t)$$
(3)

The Variable Importance $VI(X_j)$ is equal to

$$VI(X_j) = \frac{M(X_j)}{max\{M(X_j)\}} \times 100$$
(4)

where $VI(X_j)$ are the importance values, given by $M(X_j)$ divided by the largest importance values $max\{M(X_j)\}$ and expressed as percentages. This allows the identification of the masking variable and non-linear correlation among attributes [52].

4. The Case Study

The dataset deals with information regarding the tourists' arrivals as well as transport modes for the 99 Italian provinces, observed during the 2006–2016 time period. Therefore, 1089 observations

(99 provinces x 11 years) have been collected. The dependent variables considered are listed in the following:

- ItalianTourist: no. of Italian Tourists, i.e., number of arrivals from other Italian provinces travelling for both tourism and business tourism purposes (Italian Census, ISTAT, www.istat.it);
- ForeignTourist: no. of Foreign Tourists; number of arrivals from other countries travelling for both tourism and business tourism purposes (Italian Census, ISTAT, www.istat.it);
- Overnights_Italian: no. of nights spent in tourist installations by Italian tourists (nights—hundreds of thousands—Census data);
- Overnights_Foreign: no. of nights spent in tourist installations by Foreign tourists (hundreds of thousands—Census data).

The independent variables are:

Transportation systems variables

- HSR is a dummy variable assuming Value 1 if the HSR is present, 0 if otherwise;
- HUB2 is a dummy variable assuming Value 1 if the airport is not a first level hub; 0 if otherwise;
- LowCost: no. of operating bases of low-cost airlines.

Attractiveness variables

- GDP is the Gross Domestic Product of the province (Italian Census, ISTAT, www.istat.it);
- Attraction: is the no. of activities in a given province (sum of museums, historical sites, etc., information collected through different websites);
- Sea is a dummy variable assuming Value 1 if the province is close to the sea; 0 if otherwise;
- POP is the number of inhabitants in a given province (hundreds of thousands—Census data);
- Unemployment: percentage of unemployed in a given province (Census data).

The descriptive statistics of the variables are reported in Table 2, taking into account that the HSR, HUB2 and SEA are dummies, i.e., continuous variables.

Variable	Minimum	Maximum	Mean	Std. Dev.
ItalianTourist	0.33	122.90	19.59	23.12
ForeignTourist	0.08	252.92	17.26	38.44
Overnights_Italian	0.08	894.21	87.24	129.33
Overnights_Foreign	0.10	1,463.61	79.99	204.19
HSR				
POP	0.86	43.40	5.87	6.29
Low-Cost	0.00	19.00	2.00	4.64
HUB2				
GDP	1.60	138.40	15.84	19.49
Unemployment	2.10	269.58	22.55	29.99
Sea				
Attract	49.00	1,981.00	526.77	372.88

Table 2.	Variables	descriptive	statistics.
----------	-----------	-------------	-------------

The choice of these independent variables is in line with the literature. Binary variables to describe the transport supply are, for example reported, in the papers of Albalate and Fageda [26]; Pagliara et al. [9,11] and Albalate et al. [27]. Moreover, in the paper by Albalate et al. [27]), just one variable, i.e., the hotel price index, was introduced by the authors to describe the attraction of a destination.

It is important to highlight that the unavailability of data for the whole period of analysis (i.e. ten years) represented a limitation of this manuscript, and therefore dummy variables have been reported to try to resolve the issue.

5. Results and Discussion

In this section, the regression trees representative of the collected data, relating to the analysis period corresponding to the years 2006–2016, will be presented.

In the regression trees, the following dependent variables have been assessed: (1) number of Italian Tourists (Italian Tourist); (2) number of Foreign Tourists (Foreign Tourist); (3) nights spent in tourist installations by Italian Tourists (Overnights_Italian); and (4) nights spent in tourist installations by Foreign Tourists (Overnights_Foreign).

The *Italian Tourist* variable, as shown in Figure 3, presents 26 nodes and the first split variable is the *HSR* one. If the provinces are served by the HSR, they attract a higher number of tourists, who are even more attracted if the number of *LowCost* companies is high, otherwise the presence of an *HUB2* influences their choice. In the other branch of the tree, in the provinces not served by HSR, the tourist flow is influenced by the presence of a hub of a network carrier, and if there are a low number of low cost companies, their choice is influenced also by the unemployment rate and the number of inhabitants in the province. On the other hand, if there is not a second level hub, the variable that influences the tourist flow is the presence of attractions, followed by the presence of the sea.



Figure 3. Regression Tree for the dependent variable ItalianTourist.

By analyzing the importance values (see Figure 4), the most important one is the *Attraction*, indicating that, in cities with many attractions, there is a greater flow. The *POP* and *LowCost* variables are important as well. The value of the *LowCost* variable shows how the Italian tourist flow is influenced by the travel cost.



Figure 4. Variable importance for the dependent variable ItalianTourist.

The variable *ForeignTourist*, as shown in Figure 5, presents eight nodes and the first partition variable is represented by the *HSR* variable. As expected, cities with HSR attract more tourists and the tourist flow goes to the provinces with more attractions (such as the number of museums, archeological sites, etc.) present in a given province.

For provinces where HSR is not presented, tourists choose destinations served by second level hubs (HUB2) and without the presence of the sea.

Considering the importance variable values (see Figure 6), for foreign tourists, the variable *Attraction* is the most important. The *LowCost* variable is important as well. Among the transport variables, the most important one is the HSR, indicating that, in those provinces served by HSR, there is a greater flow. Less important, but with an impact on tourist destination choice, is the variable *GDP*, indicating the Gross Domestic Product of the province.

As shown in Figure 7, the tree representative of the independent variable *Overnights_Italian* is positively influenced by the transport variable *HSR*. If there is an HSR station, tourist flow is influenced by the attractions of the provinces and where there are less attractions for tourists, the latter are mainly influenced by the presence of a second level hub (*HUB2*). In the other branch of the tree, the presence of a HUB2 strongly influences tourists' choices, on the other hand the attractions of the provinces play a fundamental role.

By analysing the importance variables (see Figure 8), the most important variable is the one linked to the attractiveness of a given province, considering that an attractive province requires more nights to be spent at the destination. The *GDP* and *LowCost* variables are important as well.

Figure 9 shows that the Overnights Foreign variable is influenced by the presence of *HSR*. If HSR is present, the most visited provinces are those with high attractions, while, in the other case, the presence of a *HUB2* is very important. Foreign tourists choose provinces with a high *GDP*.

By analyzing the importance values (see Figure 10), the most important variable is the *Attraction*, followed by the *GDP* and the *LowCost* variables. The transport variable that involves a higher flow of foreign tourists is *HSR*, while the presence of *HUB2* is not very significant.







Figure 6. Variable importance for the dependent variable ForeignTourist.



Figure 7. Regression Tree for the dependent variable Overnights Italian.



Figure 8. Variable importance for the dependent variable Overnights Italian.



Figure 9. Regression Tree for the dependent variable Overnights Foreign.



Figure 10. Variable importance for the dependent variable Overnights Foreign.

The results obtained in this manuscript are consistent with other methodological approaches, proposed for the same case study of Italy, present in the literature. In the article of Pagliara et al. [9], the impact of HSR on the tourism market was studied with a database containing information both on tourism and transport for 77 Italian provinces, during the 2006–2013 period. Through the specification and estimation of a panel model, it was demonstrated that the effects of HSR on the number of visitors and the number of nights spent at destination are positive in all the provinces served by an HSR line. A different approach was proposed by Pagliara and Mauriello [11], where the impact of HSR on the tourism market was analysed through the specification of Geographically Weighted Regression techniques, embedded within a Poisson model. This approach can measure the relationship between independent and dependent variables with respect to space. The main outcome of the analysis, based on the same number of Italian provinces, i.e., 99, presented in this manuscript and the same the time period, i.e., the 2006–2016 one, was that HSR affects tourists' choices of a given destination.

6. Conclusions and Further Perspectives

HSR systems can have impacts on the tourist areas they serve, thanks to the increased accessibility they bring to the served areas. Indeed, this manuscript has found consistent evidence in favor of a positive relationship between HSR and tourist outcomes. Several approaches are present in the international literature to analyse whether this effect exists, both qualitative and quantitative, as described in Section 2.

In this manuscript, an analysis has been carried out with the aid of a dataset containing information on both tourism and transport for 99 Italian provinces during the 2006–2016 period.

The methodology proposed is the CART analysis, which, according to the authors' knowledge, has never been applied in this context before. CART provides both theoretical and applied advantages relative to the parametric models. Indeed, from the theoretical perspective, the advantage of the CART method is that it does not require the specification of the functional form of the model in advance or the assumption of an additive relationship between dependent and independent variables. Another advantage is that the CART analysis can effectively handle collinearity problems. When a serious correlation between independent variables exists, the variability of the estimated coefficients will be inflated. It follows that an interpretation of the relationship between an independent and dependent variable is difficult to define. On the other hand, regression tree methods are also not sensitive to outliers, since the splitting is based on the sample's proportion within the split ranges and not on the absolute values.

From the applied perspective, the regression tree methods are very intuitive and easy to explain. Moreover, they have the advantage of giving each variable the chance to appear in different contexts with different covariates, and thus better reflect its potential impact on the dependent variable. However, unlike a linear regression model, a variable in the CART algorithm can be considered highly important even if it never appears as a node splitter.

Further research is required on the use of HSR variables, which should describe the connectivity and territorial distribution of the HSR network, and the service conditions offered by the operating companies (e.g., fares, timetables, frequency). Specifically, considering that one of the limitations of the current literature is represented by aggregate data on tourism [53,54], authors suggest to employ ad hoc surveys in order to directly interview tourists and asking them their travel mode chosen to reach a given destination for their holidays. More disaggregate analyses should be represented in the international literature to fill this gap.

Moreover, the application of the same methodology to other case studies will be taken into account in order to make a comparison.

Author Contributions: F.P. has supervised, written and edited the whole paper. F.M. has collected the data set and developed the tree-methodology. L.R. has contributed to the development of Section 5. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: No conflict of interest.

References

- 1. Hunt, J.D. Image as a factor in tourism development. J. Travel. Res. 1975, 13, 1–7. [CrossRef]
- 2. Kim, H. Perceived attractiveness of Korean destinations. Ann. Tourism Res. 1998, 25, 340–361.
- 3. Jha, S.M. Tourism Marketing, Bombay; Himalaya Publishing House: Delhi, India, 1995.
- 4. Das, D.; Sharma, S.K.; Mohapatra, P.K.; Sarkar, A. Factors influencing the attractiveness of a tourist destination: A case study. *J. Serv. Res.* **2007**, *7*, 103–134.
- 5. Kaul, R.N. Dynamics of Tourism: A Trilogy; Sterling Publishers Private Limited: New Delhi, India, 1985.
- 6. Khadaroo, J.; Seetanah, B. The role of transport infrastructure in international tourism development: A gravity model approach. *Tour. Manag.* **2008**, *29*, 831–840. [CrossRef]
- 7. Guirao, B.; Campa, J.L. The effects of tourism on HSR: Spanish empirical evidence derived from a multi-criteria corridor selection methodology. *J. Transp. Geogr.* **2015**, *47*, 37–46. [CrossRef]
- 8. Cascetta, E.; Papola, A.; Pagliara, F.; Marzano, V. Analysis of mobility impacts of the high speed Rome–Naples rail link using withinday dynamic mode service choice models. *J. Transp. Geogr.* **2011**, *19*, 635–643. [CrossRef]
- 9. Pagliara, F.; Mauriello, F.; Garofalo, A. Exploring the interdependences between High Speed Rail systems and tourism: Some evidence from Italy. *Transp. Res. Part A: Policy Pr.* **2017**, *106*, 300–308. [CrossRef]
- Delaplace, M.; Pagliara, F.; La Pietra, A. Does high-speed rail affect destination choice for tourism purpose? Disneyland Paris and Futuroscope case studies. *Belgeo* 2016, *3*, 1–23.
- 11. Pagliara, F.; Mauriello, F. Modelling the impact of High Speed Rail on tourists with Geographically Weighted Poisson Regression. *Transp. Res. Part A: Policy Pr.* **2020**, *132*, 780–790. [CrossRef]
- 12. Della Corte, V.; Piras, A.; Zamparelli, G. Brand and image: The strategic factors in destination marketing. *Int. J. Leis. Tour. Mark.* **2010**, *1*, 358. [CrossRef]
- Coronado, J.M.; Garmendia, M.; Moyano, A.; Ureña, J.M. Assessing Spanish HSR network utility for same-day tourism. *Rech. Transp. Secur.* 2013, 29, 161–175. [CrossRef]
- 14. Mimeur, C.; Facchinetti-Mannone, V.; Carroue, G.; Berion, P. Les stratégies de développement touristique des territoires de l'espace Rhin-Rhône: Une nouvelle cohérence impulsée par le TGV? *Recherche Transport et Sécurité* **2013**, *29*, 193–210. [CrossRef]
- 15. Delaplace, M.; Perrin, J. Multiplication des dessertes TGV et Tourismes urbains et d'affaires, Regards croisés sur la Province et l'Ile de France. *Recherche Transport et Sécurité* **2013**, *29*, 177–191. [CrossRef]
- Bazin, S.; Beckerich, C.; Delaplace, M. Desserte TGV et villes petites et moyennes, une illustration par le cas du tourisme à Arras, Auray, Charleville-Mézières et Saverne. *Les Cahiers Scientifiques du Transport* 2013, 63, 33–62.

- 17. Wang, X.; Huang, S.; Zou, T.; Yan, H. Effects of the high speed rail network on China's regional tourism development. *Tour. Manag. Perspect.* **2012**, *1*, 34–38. [CrossRef]
- 18. Chen, X. Assessing the Impacts of High Speed Rail Development in China's Yangtze River Delta Megaregion. *J. Transp. Technol.* **2013**, *3*, 113–122. [CrossRef]
- 19. Bazin, S.; Delaplace, M. Desserte ferroviaire à grande vitesse et tourisme: Entre accessibilité, image et outil de coordination. *Teoros* **2013**, *2*, 37–46.
- 20. Cvelbar, L.K.; Dwyer, L.M.; Koman, M.; Mihalič, T. Drivers of Destination Competitiveness in Tourism: A Global Investigation. *J. Travel Res.* **2016**, *8*, 1041–1050. [CrossRef]
- 21. Bazin, S.; Beckerich, C.; Delaplace, M. High Speed Railway, Service Innovations and Urban and Business Tourisms Development. In *Economics and Management of Tourism: Trends and Recent Developments*; Collecçao Manuais; Sarmento, M.A., Matias, A., Eds.; Universidade Luisiada Editora: Lisbon, Portugal, 2011.
- 22. Chen, Z.; Haynes, K.E. Tourism Industry and High Speed Rail Is There a Linkage: Evidence from China's High Speed Rail Development. GMU School of Public Policy Research Paper No. 2012-14. Available online: https://ssrn.com/abstract=2130830 (accessed on 25 January 2020).
- 23. Kurihara, T.; Wu, L. The Impact of High Speed Rail on Tourism Development: A Case Study of Japan. *Open Transp. J.* **2016**, *10*, 35–44. [CrossRef]
- 24. Giurao, B.; Soler, F. Impacts of the new high-speed service on small touristic cities: The case of Toledo, in The Sustainable City V. Urban Regeneration and Sustainability. *Wessex Trans. Ecol. Environ.* **2008**, *117*, 465–473.
- 25. Campa, J.L.; López-Lambas, M.E.; Guirao, B. High speed rail effects on tourism: Spanish empirical evidence derived from China's modelling experience. *J. Transp. Geogr.* **2016**, *57*, 44–54. [CrossRef]
- 26. Albalate, D.; Fageda, X. High speed rail and tourism: Empirical evidence from Spain. *Transp. Res. Part A: Policy Pr.* **2016**, *85*, 174–185. [CrossRef]
- 27. Albalate, D.; Campos, J.; Jiménez, J.L. Tourism and high speed rail in Spain: Does the AVE increase local visitors? *Ann. Tour. Res.* **2017**, *65*, 71–82. [CrossRef]
- 28. Sun, Y.-Y.; Lin, Z.-W. Move fast, travel slow: The influence of high-speed rail on tourism in Taiwan. *J. Sustain. Tour.* **2017**, *26*, 433–450. [CrossRef]
- 29. Wang, D.-G.; Niu, Y.; Qian, J. Evolution and optimization of China's urban tourism spatial structure: A high speed rail perspective. *Tour. Manag.* 2018, *64*, 218–232. [CrossRef]
- 30. Seddighi, H.; Theocharous, A. A model of tourism destination choice: A theoretical and empirical analysis. *Tour. Manag.* **2002**, *23*, 475–487. [CrossRef]
- 31. Barros, C.P.; Assaf, A.G. Analyzing Tourism Return Intention to an Urban Destination. *J. Hosp. Tour. Res.* **2012**, *36*, 216–231. [CrossRef]
- 32. Delaplace, M.; Pagliara, F.; Perrin, J.; Mermet, S. Can High Speed Rail Foster the Choice of Destination for Tourism Purpose? *Procedia Soc. Behav. Sci.* **2014**, *111*, 166–175. [CrossRef]
- 33. Pagliara, F.; Delaplace, M.; Vassallo, J.M. High-speed trains and tourists: What is the link? In *Evidence from the French and Spanish Capitals. In proceedings of the conference Urban Transport XX—Urban Transport and the Environment in the 21st century, England, UK*; WIT Press: Southampton, UK, 2014; Volume 138, pp. 17–27.
- 34. Pagliara, F.; La Pietra, A.; Gomez, J.; Vassallo, J.M. High Speed Rail and the tourism market: Evidence from the Madrid case study. *Transp. Policy* **2015**, *37*, 187–194. [CrossRef]
- Curtis, P.G.; Kokotos, D.X. A Decision Tree Application in Tourism-based Regional Economic Development. MPRA paper no. 25302. Available online: http://mpra.ub.uni-muenchen.de/25302/.2008 (accessed on 25 January 2020).
- 36. Legoherel, P.; Wong, K.K.F. Market Segmentation in the Tourism Industry and Consumers' Spending. *J. Travel Tour. Mark.* **2006**, *20*, 15–30. [CrossRef]
- 37. Fitzmaurice, G.M.; Laird, N.M.; Ware, J.H. *Applied Longitudinal Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
- Chang, L.-Y.; Chen, W.-C. Data mining of tree-based models to analyze freeway accident frequency. J. Saf. Res. 2005, 36, 365–375. [CrossRef] [PubMed]
- Galimberti, G.; Montanari, A. Regression Trees for Longitudinal Data with Time-Dependent Covariates. In *Studies in Classification, Data Analysis, and Knowledge Organization*; Jajuga, K., Sokolowski, A., Bock, H.-H., Eds.; Springer: Berlin/Heidelberg, Germany, 2002; pp. 391–398.
- Sela, R.J.; Simonoff, J.S. RE-EM trees: A data mining approach for longitudinal and clustered data. *Mach. Learn.* 2012, *86*, 169–207. [CrossRef]

- 41. Barros, R.C.; De Carvalho, A.C.; Freitas, A.A. *Automatic Design of Decision-Tree Induction Algorithms*; Springer Science and Business Media LLC.: Berlin/Heidelberg, Germany, 2015.
- 42. D'Ambrosio, A.; Heiser, W.J. A Recursive Partitioning Method for the Prediction of Preference Rankings Based Upon Kemeny Distances. *Psychometrika* **2016**, *81*, 774–794. [CrossRef] [PubMed]
- 43. D'Ambrosio, A.; Aria, M.; Siciliano, R. Robust Tree-Based Incremental Imputation Method for Data Fusion. In Proceedings of the Computer Vision—ECCV 2012, Florence, Italy, 7–13 October 2012; Springer Science and Business Media LLC.: Berlin/Heidelberg, Germany, 2007; Volume 4723, pp. 174–183.
- 44. Han, J.; Pei, J.; Kamber, M. Data Mining: Concepts and Techniques; Elsevier: Waltham, MA, USA, 2011.
- 45. Gatnar, E. Tree-Based Models in Statistics: Three Decades of Research; Springer: New York, NY, USA, 2002.
- 46. Gordon, A.D.; Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.
- 47. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013.
- 48. Yohannes, Y.; Webb, P. Classification and Regression Trees, CART: A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity; The International Food Policy Research Institute: Washington, DC, USA, 1999; Volume 3.
- 49. Montella, A.; Aria, M.; D'Ambrosio, A.; Mauriello, F. Analysis of powered two-wheeler crashes in Italy by classification trees and rules discovery. *Accid. Anal. Prev.* **2012**, *49*, 58–72. [CrossRef] [PubMed]
- 50. Ishwaran, H. Variable importance in binary regression trees and forests. *Electron. J. Stat.* **2007**, *1*, 519–537. [CrossRef]
- 51. Pham, H. Springer Handbook of Engineering Statistics; Springer Science & Business Media: Berlin, Germany, 2006.
- 52. Therneau, T.M.; Atkinson, E.J. An Introduction to Recursive Partitioning Using the RPART Routines. Available online: https://r.789695.n4.nabble.com/attachment/3209029/0/zed.pdf (accessed on 25 January 2020).
- 53. Ping, Y.; Pagliara, F.; Wilson, A. How Does High-Speed Rail Affect Tourism? A Case Study of the Capital Region of China. *Sustainability* **2019**, *11*, 472. [CrossRef]
- Campa, J.L.; Pagliara, F.; Lopez-Lambas, M.E.; Arce, R.; Guirao, B. Impact of High-Speed Rail on Cultural Tourism Development: The Experience of the Spanish Museums and Monuments. *Sustainability* 2019, 11, 5845. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).