

Article

Analyzing the Impact of High-Speed Rail on Tourism with Parametric and Non-Parametric Methods: The Case Study of China

Francesca Pagliara ^{1,*} , Filomena Mauriello ¹  and Yin Ping ² 

¹ Department of Civil, Architectural and Environmental Engineering, University of Naples Federico II, 80125 Napoli, Italy; filomena.mauriello@unina.it

² Department of Tourism Management, School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China; pyin@bjtu.edu.cn

* Correspondence: fpagliar@unina.it

Abstract: High-speed rail (HSR) and tourism are closely related activities since improved mobility is perceived to facilitate tourist behavioral changes. The interest in research is very high and this contribution tries to provide an insight into this topic by making a comparison between the estimation of the parametric Generalized Estimating Equation (GEE) approaches with the non-parametric Classification and Regression Tree (CART). A dataset containing information both on tourism and transport for thirty Chinese provinces, during the 2001–2017 period, has been collected. The finding of this paper shows that the presence of HSR has value in the explanation of tourist arrivals.

Keywords: high-speed rail; tourism market; generalized estimating equation; classification and regression tree; Chinese provinces



Citation: Pagliara, F.; Mauriello, F.; Ping, Y. Analyzing the Impact of High-Speed Rail on Tourism with Parametric and Non-Parametric Methods: The Case Study of China. *Sustainability* **2021**, *13*, 3416. <https://doi.org/10.3390/su13063416>

Academic Editor: Marc A. Rosen

Received: 8 March 2021

Accepted: 17 March 2021

Published: 19 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Transport for tourism or tourism for transport? This question, which apparently appears as a dilemma, has only an answer, i.e., the two systems are strictly related to each other. In his book, Kaul [1] stated that the development of tourism was mainly influenced by that of means of transport. On the other hand, tourism was a mass phenomenon which was needed for transport infrastructures and services.

The role of transport is therefore crucial for the increase of tourist flow. The integration of domestic and international transport systems and parallel co-ordination with other countries contributes to the ease of tourism flow and growth of domestic and international tourism. An important aspect is related to the technological developments which would exercise a deep impact on the means and patterns of transportation in both developing and developed societies, with the result that a more efficient, faster and safer transport system would be beneficial to the growth and expansion of tourism improvement.

Significant innovation in transportation technology and continuous construction of High-Speed Rail (HSR) jointly trigger the “second railway age” [2,3]. The tourism sector can benefit from the expansion of the HSR network, which, in turn, improves the transport accessibility of the cities served. An interesting body of literature reports the effects of HSR on tourism using different methodologies. For example, in Europe, the inauguration of the HSR line linking Barcelona and Perpignan increased the daily trip in the market of Barcelona [4]. Cuenca and Toledo experienced an increase in the tourism industry due to HSR [5]. Ureña et al. [6] showed that large intermediate cities served by HSR, such as Lille, Zaragoza and Cordoba, benefited from a growth in urban and business tourism. Studies in France demonstrated that HSR increased the attraction of Disneyland Paris [7] and visitors’ re-visiting Paris and Madrid [8].

The objective of this paper is to study the impact of HSR on tourism using both parametric and non-parametric methodologies. Among the first ones, the Generalized

Estimating Equations (GEEs) was chosen and among the second ones, the Classification and Regression Tree (CART) model was proposed.

Generalized Estimating Equations (GEEs) are usually chosen, assuming that the relations are constant across all provinces, i.e., the associations are spatially stationary [1]. The limited ability of this technique to take spatial effects into account can be overcome through the specification of spatial regression models. In general, the commonly used spatial models are the Geographically Weighted Poisson Regression (GWPR); the Spatial Autoregressive Regression (SAR) model and the Conditional Autoregressive Regression (CAR) model. The latter has been considered here. The rest of this paper is organized as follows. The case study together with data set and the methodology are reported in Section 2. In Section 3 the CART model is specified. Conclusions and further perspectives are reported in Section 4.

2. The Methodology

2.1. The Case Study

In the past decade, China has experienced a remarkable growth of the HSR network, changing the spatial structure of cities within the transportation system.

Indeed, China has the longest HSR in the world (see Figure 1). In 2019, the national railway operating mileage was more than 139,000 km, including 35,000 km of HSR; with 5474 km new mileage constructed and went into operation in 2019 (See Figure 1). The coverage of the HSR network to cities with a population of more than 500,000 expands from 28% in 2012 to 86% in 2019. All provincial capital cities, except Lhasa in the mainland, have been connected with HSR, and Hong Kong has entered the national HSR network. The Capital Area (Beijing–Tianjin–Hebei), Yangtze River Delta, and other urban agglomerations can be traversed by HSR in under two hours; Beijing, Shanghai, and other big cities with distance of 1000 km can be reached in four hours and distance of 2000 km in eight hours [9].



Figure 1. The High-Speed Rail network in China.

In August 2020, China National Railway Group Co., Ltd. issued and released The Outline of Railway Development Plan for A Powerful Transportation Country, defining the development blueprint of China Railway in the next 15 years and 30 years. According to the plan, in 15 years, i.e., in 2035, China will have built about 200,000 km of national

railway network, including about 70,000 km of HSR. The HSRs mileage will be doubled with respect to the year 2019, and all provincial cities and cities with a population of more than 500,000 will be accessible by HSR. The three-hour HSR circle is mostly realized between the provincial capitals in adjacent regions.

The analysis of the factors affecting tourists' choices for the case study of thirty Chinese provinces has been taken into account, where the impact of HSR has been analyzed. In the literature, studies can be found mainly concentrating on accessibility and mobility impacts of HSR in China [10–12], as well as on the effect of HSR on regional tourism development [13] and on international tourists [14]. However, the causal effect of HSR and tourism growth has been less highlighted. Among the few contributions, Chinese scholars did not reach an agreement on the same results, considering that in some cases they found a positive impact [15,16], while in other cases the effect was not significant [17]. For example, Chen and Haynes [14] demonstrated that the Chinese provinces served by HSR experienced an increase in the number of foreign tourists of 20% and an increase of tourism revenue of almost 25%.

Taotao et al. [18] studied the impact of HSR on tourism in the Yangtze River Delta. The increase of regional tourism in the Yangtze River Delta showed an “HSR effect”, and the demand and supply of tourist products improved considerably. Yuhua and Jun [19] showed that the opening of the HSR had an effect on the tourism development of the cities served by HSR.

Lei et al. [20] analyzed the tourism spatial structure of Huangshan city before the opening of HSR, just after its opening and two years after its inauguration, and found that the impact of the opening of the HSR on the tourism in Huangshan city was not significant.

Another study demonstrated that the extension of the HSR network in the Capital Region of China significantly increased the total tourism spatial interaction and reduced the spatial difference. Smaller cities connected through HSR to core cities benefited the most from the HSR network inauguration [9]. HSR also affected tourist flows and spatial relationships of the two cities, Beijing and Tianjin, served by it [21,22]. Ziyang et al. [23] considered Xiamen City as an example and verified that HSR had a correlation with tourism. Yongze et al. [24] demonstrated that the opening of HSR played a significant role in promoting regional tourism, but this role gradually decreased from the East to the West parts of the country.

2.2. The Data Set

The dataset collected for this study contains information concerning thirty out of the thirty-four Chinese provinces (excluding Hong Kong, Macao, Taiwan and Tibet), and deals with the time period 2001–2017. For each province, for each year, eight variables have been identified. Two dependent and six independent ones (See Table 1).

Table 1. Variables.

Dependent Variables	
Chinese Tourists	no. of Chinese tourists
Foreign Tourists	no. of foreign tourists
Independent Variables	
Hotels	no. of hotels present in the province
Resorts	no. of resorts present in the province
GDP	Gross Domestic Product of the province
Passengers	no. of rail passengers
HSRstations	dummy variable equal to 1 if in the province there is a HSR station; 0 otherwise
IntAirport	dummy variable equal to 1 if in the province there is an international airport; 0 otherwise

Table 1. Cont.

Dependent Variables	
Hotels	no. of hotels present in the province
Resorts	no. of resorts present in the province
GDP	Gross Domestic Product of the province

Regression models assume that the relationship between a dependent variable and any independent variable is stationary in space. To avoid that the autocorrelation between variables seriously affects the estimated coefficients, leading to inadequate statistical inferences, the data considered involve measurements over time for the same provinces (panel data). To take into account these possible unknown correlations, a further extension of the GLM (Generalized Linear Model) was considered, namely, the Generalized Estimating Equation (GEE).

Count data have been modeled through a Poisson distribution, where the probability of a given province i having y_{it} tourists every year is equal to the following expression [25]:

$$P(y_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \quad (1)$$

where $P(y_i)$ is the probability of province i with a number of y_i tourists every year and λ_i represents the parameter for province i of the Poisson distribution. The relationship between count data and the dependent variables, above listed, has been identified through the specification of Generalized Linear Models (GLMs). The information collected involves measurements over time for the same cities and the problem connected with the serial correlation was avoided thanks to the introduction of panel data regression models, which provide a variety of correlation patterns. To deal with possible unknown correlations problems, Generalized Estimating Equations (GEEs) have been specified, being an extension of the GLMs.

The student's t -test statistic was chosen in order to test the significance of each variable. Therefore, a coefficient is significant when t is greater than 1.96.

Specifically, the first GEE model estimated through the software STATA (version 16), whose results are reported in Table 2, considers as dependent variable the number of Chinese tourists (using the ChineseTourists variable).

Table 2. Regression models results: ChineseTourists.

	Coef.	Std. Err.	t-Value
GDP	5.60×10^{-6}	5.66×10^{-8}	99.1
Passengers	1.94×10^{-5}	1.40×10^{-7}	138.94
HSRstations	0.2053	0.0008	262.53
IntAirport	0.3736	0.0024	156.09
Resorts	0.0058	1.26×10^{-5}	462.11
(Intercept)	8.6672	0.0028	3139.96

Analyzing the results, it is clear that almost all the independent variables are significant, with the exception of the Hotels variable, which is not shown in the table for this reason.

Variables Resorts, GDP and Passengers have a positive impact, even though a bit smaller, on the number of domestic tourists. The variables that have the greatest impact on the value of the dependent variable are IntAirport, which has an estimated coefficient of 0.3736, and HSRstations, whose coefficient is 0.2053. It is, therefore, clear that Chinese tourists are strongly influenced, in the choice of destinations, both by the presence of international airports and by the presence of HSR connections. The second GEE model estimated

takes into account as dependent variable the number of foreign tourists (ForeignTourists), and the results obtained are shown in the following Table 3.

Table 3. Regression models results: ForeignTourists.

	Coef.	Std. Err.	t-Value
Hotels	0.0007	2.35×10^{-6}	312.64
GDP	9.89×10^{-6}	5.12×10^{-8}	192.99
Passengers	1.48×10^{-5}	1.52×10^{-7}	97.62
HSRstations	0.3665	0.0010	363.79
IntAirport	0.1166	0.0021	56.57
Resorts	0.0005	0.000012	39.27
(Intercept)	6.9855	5.3×10^{-3}	1318.62

For this model, all the variables have a *t*-value greater than 1.96, thus resulting significant. All the variables have a positive coefficient, in particular, the IntAirport and HSRstations variables, which assume a higher coefficient value compared to other independent variables. Comparing the two models, it is possible to note that the Hotels variable is significant only in the case of the ChineseTourists model. This result shows that foreign tourists prefer as a tourist destination one in which there are a large number of Hotels, which are not so fundamental for domestic tourists. The coefficients relating to the Resorts, GDP and Passengers variables are almost constant. The variable HSRstations has a higher coefficient in the second model, which implies that the presence of a HSR network has a higher impact on the choices of foreign tourists compared to those of the domestic ones.

The coefficient relating the IntAirport variable has a higher value in the first model than in the second. Therefore, the presence of international airports seems to have a higher impact on Chinese tourists than on the foreign ones.

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. 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. This approach establishes a casual effect between dependent and independent variables. These tree-structured methods do not require a priori probabilistic knowledge about the phenomena under study, and no assumption is required [26,27]. This can be considered as one of the main advantages of these methods with regard to standard econometric analysis [28]. 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.

3. The CART Model

The application of the non-parametric Classification and Regression Tree (CART) model does not require an a priori probabilistic knowledge of the phenomenon under study and the fulfilment of strict hypotheses, neither on the type of relationship, nor on the form of distribution of the dependent variable. These aspects represent the main advantages over parametric techniques. It was decided to apply this methodology, in addition to those previously described, to collect information not highlighted and left out in the parametric models.

Each node of the tree indicates the predicted value, the number of experimental units contained in the node and its descriptive percentage. The regression tree relating to the ChineseTourists dependent variable has 14 nodes and the first split variable is the one relating to Resorts (see Figure 2). Results were obtained through the R-Part software.

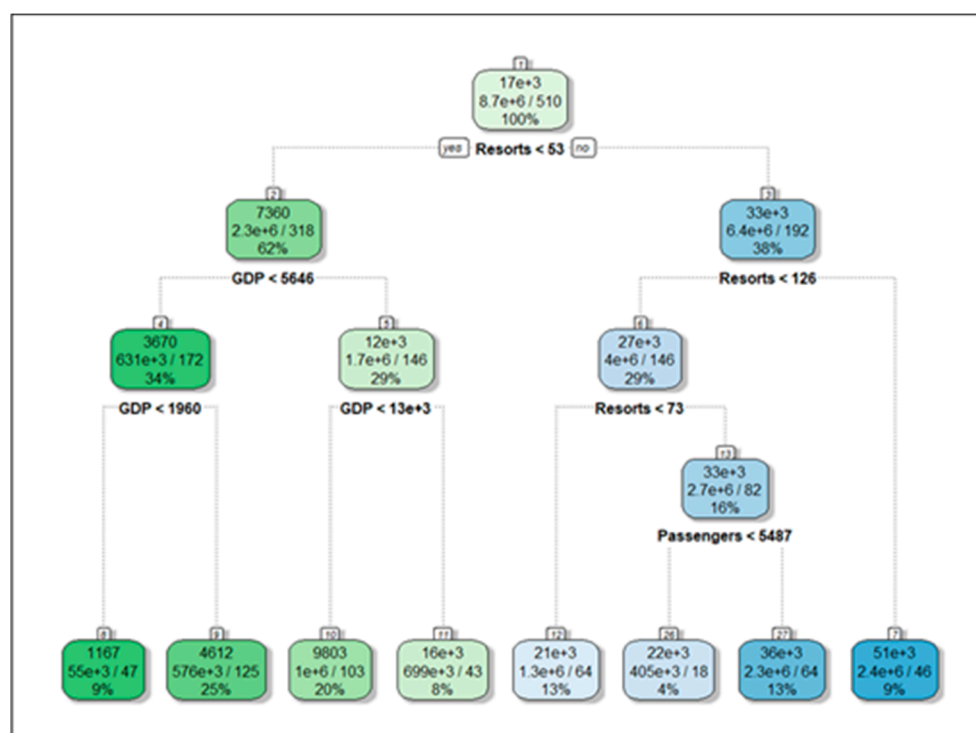


Figure 2. Regression tree for the Resorts variable: ChineseTourists.

By analysing the tree, it is clear that, as the number of resorts increases, the number of foreign tourists increases. Another variable that strongly influences the tourist flow is the GDP variable. In the case of richer cities, the number of foreign tourists increases. In this tree, the HSRstations variable never appears as a split variable, like the IntAirport variable.

By analysing the values of importance of the variables (see Figure 3), it is clear that the most important variable is Resorts, followed by GDP, Passengers and HSRstations.

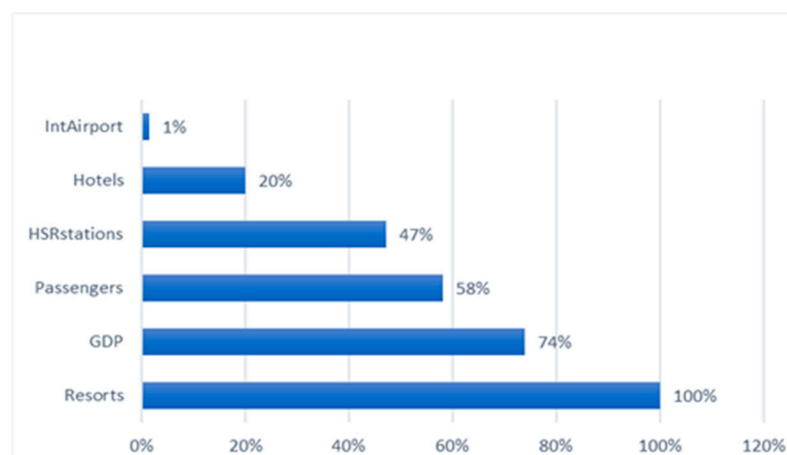


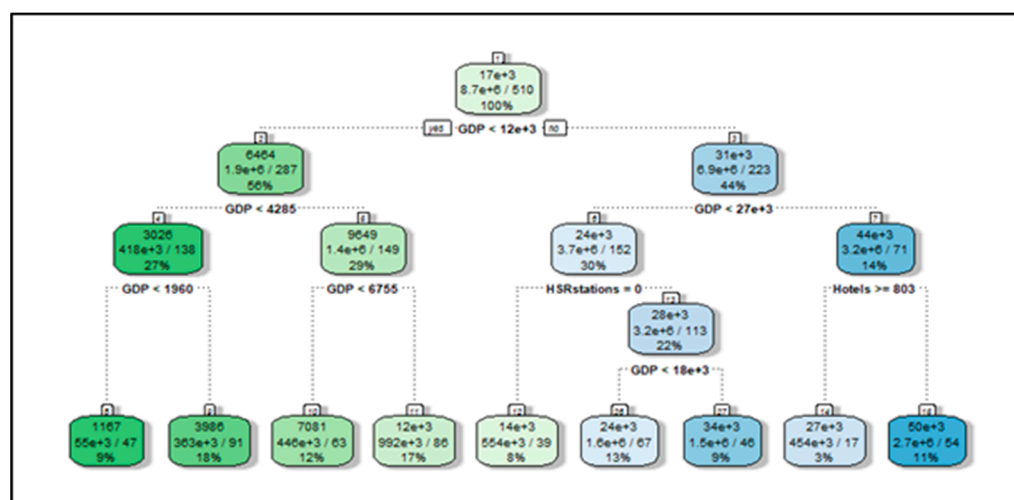
Figure 3. Importance of variables: ChineseTourists.

As previously stated, the HSRstations variable never appears as a split variable; this can be explained by the “masking” effect of the Resorts variable with respect to it. Indeed, analysing the correlation matrix, the two variables are strongly correlated (see Table 4).

Table 4. Correlation matrix.

Variables	Hotels	GDP	Passengers	HSRstations	IntAirport	Resorts
Hotels	1	0.488	0.422	0.116	0.282	0.366
GDP	0.488	1	0.842	0.568	0.248	0.850
Passengers	0.422	0.842	1	0.554	0.298	0.752
HSRstations	0.116	0.568	0.554	1	0.206	0.658
IntAirport	0.282	0.248	0.298	0.206	1	0.242
Resorts	0.366	0.850	0.752	0.658	0.242	1

Without considering the Resorts variable, the resulting regression tree presents the HSRstations variable as a split variable (see Figure 4).

**Figure 4.** Regression tree for the HSRstations variable: ChineseTourists.

Through the analysis of this tree, it is clear that the presence or absence of HSRstations affects the value of the ChineseTourists variable. The child nodes (12 and 13) have highly differentiated values, a predicted value of 554,000 and a much higher one, i.e., 3,200,000. Therefore, the presence of HSR stations has an impact on the number of domestic tourists.

The tree relating to the dependent variable ForeignTourists has 16 nodes and also in this case, there are no HSRstations and IntAirport variables. The first variable of the split is Hotels, in addition to it, the independent variable that most influences the number of Foreign tourists is the GDP variable (see Figure 5).

Considering the importance values of the variables, there is an increase in the importance of the IntAirport variable and a decrease in the relative importance of the HSRstations variable (see Figure 6).

The analysis of the regression trees shows that the HSRstations and IntAirport variables are not considered as split variables. This can be justified by the fact that the two variables, being dichotomous variables (i.e., dummies), can only assume values 0 and 1, and therefore they provide poor information and do not fully express the variability of information itself. It is likely that the HSRstations variable would have a higher explanatory power if the number of stations in the territory was quantified, given the different sizes of the provinces and the regions considered.

A comparison between the two methodologies adopted in this paper follows. The R2 marginal was adopted (see Table 5).

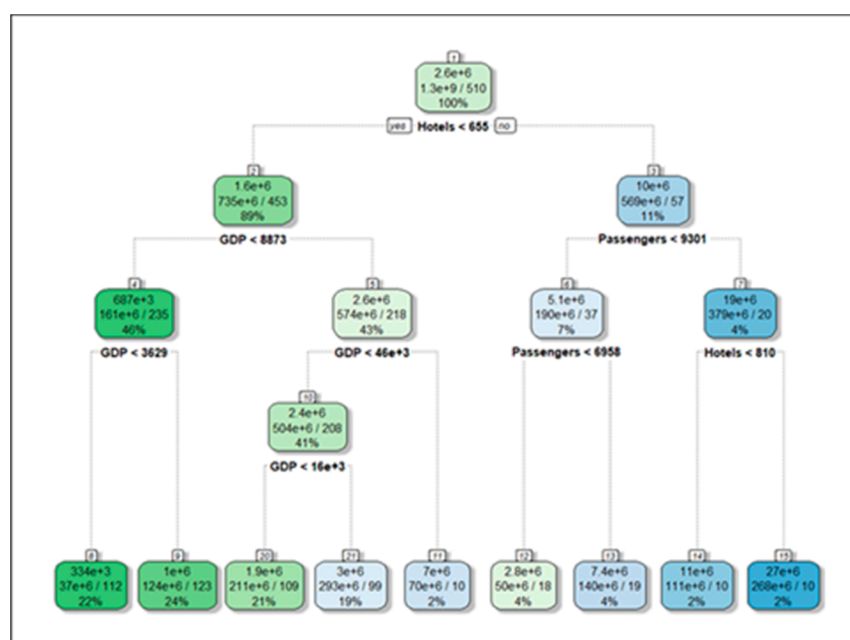


Figure 5. Regression tree for the Hotels variable: ForeignTourists.

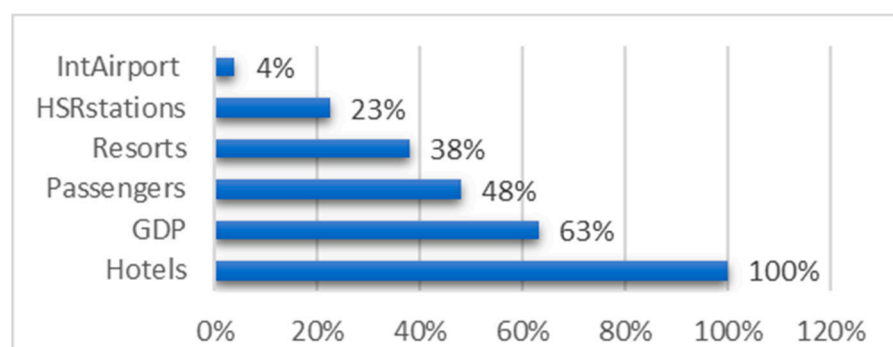


Figure 6. Importance of the variables: ForeignTourists.

Table 5. R^2 marginal comparison.

R^2 Marginal			
GEE		CART	
Chinese Tourists	Foreign Tourists	Chinese Tourists	Foreign Tourists
0.5178	0.2842	0.7923	0.5617

4. Conclusions and Further Research

In this paper the objective is that of analysing the impact that HSR systems can have on the tourist areas they serve, thanks to the increased accessibility they bring. Indeed, this study found consistent evidence in favour of a positive relationship between HSR and tourist outcomes for the case study of China. Estimation of parametric and non-parametric models have been proposed to analyse tourist behaviours, both Chinese and foreign ones. Results show that HSR has affected the tourism industry and that therefore investments in these systems have significant effects on regional development.

Very few are the contributions in the literature applying the CART methodology in this context. 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 and

the assumption of the 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 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 samples 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 of appearing in different contexts with different covariates, and thus better reflecting 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 split.

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, and so on). Specifically, considering that one of the limitations of the current literature is represented by aggregate data on tourism, the authors suggest proposing direct surveys to tourists and ask them the travel mode chosen to reach a given destination for their holidays. More disaggregate analysis should be represented in international literature to fill this gap.

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

Author Contributions: F.P. was involved in supervising the work, writing and editing the whole paper; F.M. was involved in the methodology and results analysis; Y.P. was involved in data collection. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaul, R.N. *Dynamics of Tourism: The Phenomenon*; Sterling: New York, NY, USA, 1985.
2. 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 Pract.* **2017**, *106*, 300–308. [\[CrossRef\]](#)
3. Pagliara, F.; Mauriello, F. Modelling the impact of High Speed Rail on tourists with Geographically Weighted Poisson Regression. *Transp. Res. Part A Policy Pract.* **2020**, *132*, 780–790. [\[CrossRef\]](#)
4. Masson, S.; Petiot, R. Can the high speed rail reinforce tourism attractiveness? The case of the high speed rail between Perpignan (France) and Barcelona (Spain). *Technovation* **2009**, *29*, 611–617. [\[CrossRef\]](#)
5. Varela, V.C.; Martínez, J.N. High-Speed Railway and Tourism: Is There an Impact on Intermediate Cities? Evidence from Two Case Studies in Castilla-La Mancha (Spain). *J. Urban Reg. Anal.* **2016**, *82*, 133–158.
6. Ureña, J.M.; Menerault, P.; Garmendia, M. The high-speed rail challenge for big intermediate cities: A national, regional and local perspective. *Cities* **2009**, *26*, 266–279. [\[CrossRef\]](#)
7. 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.
8. Pagliara, F.; Delaplace, M.; Vassallo, J.M. High-speed rail systems and tourists' destination choice: The case studies of paris and madrid. *Int. J. Sustain. Dev. Plan.* **2015**, *10*, 395–405. [\[CrossRef\]](#)
9. Yin, P.; Pagliara, F.; Wilson, A. How Does High-Speed Rail Affect Tourism? A case study of the capital region of China. *Sustainability* **2019**, *11*, 472.
10. Cao, J.; Liu, X.C.; Wang, Y.; Li, Q. Accessibility impacts of China's high-speed rail network. *J. Transp. Geogr.* **2013**, *28*, 12–21. [\[CrossRef\]](#)
11. Jiao, J.; Wang, J.; Jin, F. Impacts of high-speed rail lines on the city network in China. *J. Transp. Geogr.* **2017**, *60*, 257–266. [\[CrossRef\]](#)
12. Shaw, S.-L.; Fang, Z.; Lu, S.; Tao, R. Impacts of high speed rail on railroad network accessibility in China. *J. Transp. Geogr.* **2014**, *40*, 112–122. [\[CrossRef\]](#)
13. Wang, W.-C.; Chang, Y.-J.; Wang, H.-C. An Application of the Spatial Autocorrelation Method on the Change of Real Estate Prices in Taitung City. *ISPRS Int. J. Geoinf.* **2019**, *8*, 249. [\[CrossRef\]](#)

14. Chen, Z.; Haynes, K.E. Impact of high-speed rail on international tourism demand in China. *Appl. Econ. Lett.* **2014**, *22*, 57–60. [[CrossRef](#)]
15. Degen, W.; Jia, Q.; Yu, N. Spatial pattern and evolution of urban tourism field strength in China under high speed railway network. *Acta Geogr. Sin.* **2016**, *71*, 1784–1800.
16. Xin, W.; Tongqiao, Z. The influence of high speed railway network on the development and layout of China's regional tourism industry. *Econ. Geogr.* **2010**, *30*, 1189–1194.
17. Feng, F.; Linhao, C. The opening of high-speed railway and the development of tourism in station cities: "Engine" or "Corridor"? *Econ. Manag.* **2020**, *42*, 175–191.
18. Taotao, D.; Lei, Z.; Mulan, M. Research on the impact of Yangtze River Delta high speed railway network on Urban Tourism Development. *Econ. Manag.* **2016**, *38*, 137–146.
19. Yuhua, Z.; Jun, C. Heterogeneous impact of high speed rail on tourism development of station cities: A study based on double difference method. *Tour. Sci.* **2018**, *32*, 79–92.
20. Lei, L.; Lin, L.; Chenglin, M.; Xiaolong, S. Evolution of spatial structure of tourism flow in typical tourism cities in the era of high speed railway network: A case study of Huangshan City. *Econ. Geogr.* **2019**, *39*, 207–216.
21. Yin, P.; Lin, Z.; Prideaux, B. The impact of high-speed railway on tourism spatial structures between two adjoining metropolitan cities in China: Beijing and Tianjin. *J. Transp. Geogr.* **2019**, *80*, 102495. [[CrossRef](#)]
22. Yin, P.; Tonghao, Z.; Hanyan, Y. Study on the change characteristics of urban tourism spatial function of Bei-jing Tianjin Hebei under the influence of high speed railway. *Tour. Guide* **2020**, *4*, 30–44.
23. Ziyang, W.; Dan, L.; Yuting, Z. Research on the impact of high speed railway on tourism development based on grey correlation analysis: A case study of Xiamen City. *Logist. Eng. Manag.* **2020**, *42*, 163–165.
24. Yongze, Y.; Yu, F.; Haitao, Z. The impact of the opening of high speed railway on the development of re-gional tourism. *Res. Financ. Issues* **2020**, 31–38.
25. Washington, S.P.; Karlaftis, M.G.; Mannering, F. *Statistical and Econometric Methods for Transportation Data Analysis*; CRC Press: Boca Raton, FL, USA, 2010.
26. D'Ambrosio, A.; Aria, M.; Siciliano, R. Robust tree-based incremental imputation method for data fusion. In *Advances in Intelligent Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 174–183.
27. Han, J.; Pei, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Elsevier: Waltham, MA, USA, 2011.
28. Gatnar, E. Tree-based models in statistics: Three decades of research. In *Classification, Clustering and Data Analysis: Recent Advances and Applications*; Springer: Berlin/Heidelberg, Germany, 2002; pp. 399–408.