Journal of Transport Geography 107 (2023) 103546

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

The effect of high-speed rail connectivity and accessibility on tourism seasonality

David Boto-García *, Levi Pérez

Department of Economics, University of Oviedo, Spain

A R T I C L E I N F O

Keywords:
Tourism seasonality
Demand concentration
High-speed rail
Accessibility
Connectivity
Dose-response framework

A B S T R A C T

The improvement of regional transport connectivity and accessibility is a key determinant of tourism demand. This paper evaluates whether high-speed rail (HSR) connectivity and accessibility affects tourism seasonality. Using a panel dataset of 50 Spanish provinces in the period 2004–2019, we study the impact of variations in a synthetic indicator of province's connectivity and accessibility to the HSR network on the concentration of both domestic and international tourism demand. Tourism seasonality is measured using Gini and coefficient of variation indicators. Based on two-way panel fixed effects regressions and a dose-response framework, we show that access to HSR mitigates tourism seasonality in both the domestic and international segments. The effect is found to be greater at high levels of connectivity and accessibility. In the advent of forthcoming expansions of the HRS network in Spain, our findings suggest that improvements in regional connectivity and accessibility through HSR are a promising strategy to alleviate the seasonality of tourism arrivals.

1. Introduction

It is widely acknowledged that transport infrastructure and regional connectivity and accessibility are important drivers of long-run economic growth (Farhadi, 2015; Crescenci and Rodríguez-Pose, 2012), since they produce positive economic effects for local areas in terms of increased manufacturing location (Holl, 2004) and regional productivity (Matas et al., 2015), among many others. As for the tourism sector, transport connectivity (ease of communication from one area to another) and potential- (travel demand the infrastructure can serve) and locational-based accessibility (ease of reaching a node) improve the attractiveness of destinations (Khadaroo and Seetanah, 2008), tourist firm's value (Zhang et al., 2020), hotel occupancy rates (Deng et al., 2021) and can even produce spatial spillovers on neighbouring regions (Tian et al., 2022; Zhou et al., 2022).

A large body of research has studied the effect of the provision of HSR services across regional areas on several different tourism outcomes like the number of arrivals (Li et al., 2019; Pagliara et al., 2017), total overnight stays (Albalate and Fageda, 2016; Albalate et al., 2022), average length of stay (Albalate et al., 2017), expenditure (Yang and Li, 2020; Yao et al., 2022), revenues (Campa et al., 2016), revenues per arrival (Gao et al., 2019) or occupancy rates (Deng et al., 2021). Most of this literature has analysed the cases of Spain (Albalate and Fageda, 2016; Albalate et al., 2017, 2022; Campa et al., 2016, 2018) and China (Bo and Ningqiao, 2018; Chen and Haynes, 2015; Gao et al., 2019; Li et al., 2019; Yang and Li, 2020; Zhang et al., 2020; Zhou et al., 2022), which stand as the countries with the largest HSR networks in the world. In general, these works typically find that the introduction of the HSR services has been positive in terms of increased visitors and overnight stays for Chinese regions (Chen and Haynes, 2015; Yang and Li, 2020; Zhang et al., 2020); in the case of Spain, the effect of HSR on the number of tourists is, on the contrary, found to be weak (Albalate et al., 2017; Campa et al., 2016, 2018) and even negative (Albalate and Fageda, 2016; Albalate et al., 2022). Nonetheless, important heterogeneous effects across provinces in both countries are also encountered (Albalate et al., 2017; Bo and Ningqiao, 2018; Gao et al., 2019).

The impact of HSR services is generally modelled using difference-in-differences research designs that consider binary indicators for whether a region/city has at least one HSR station (Albalate et al., 2017; Bo and Ningqiao, 2018; Gao et al., 2019; Yao et al., 2022; Zhang et al., 2020). However, one limitation of this approach is that it does not consider the heterogeneity in the number of lines or stations across regions or the type and size of the provinces they connect to. Moreover, although the effect of HSR on tourism demand has been widely examined, the impact

* Corresponding author at: Oviedo Efficiency Group, Department of Economics, Faculty of Business and Economics, Avenida del Cristo s/n 33006, Oviedo, Asturias, Spain.

E-mail addresses: botodavid@uniovi.es (D. Boto-García), lperez@uniovi.es (L. Pérez).

https://doi.org/10.1016/j.jtrangeo.2023.103546
Received 24 October 2022; Received in revised form 13 January 2023; Accepted 30 January 2023
Available online 7 February 2023

on its concentration within the year (tourism seasonality) has not been studied so far. This is relevant because, apart from increasing demand levels, regional authorities and stakeholders are interested in smoothing demand over the year.

This paper studies the effect of HSR connectivity and accessibility on tourism demand concentration (seasonality). Seasonality has been a central topic in the tourism literature and has regained more interest recently due to its implications for sustainable goals (Martín-Martín et al., 2020) and the difficulties it imposes on the management of solid waste (Caponi, 2022), among others. Several studies indicate that seasonality has increased in recent years in advanced economies in general (Duro and Turrión-Prats, 2019) and in Mediterranean countries specialized in sun and beach tourism like Spain in particular (Duro, 2016). Accordingly, it seems policy relevant to study whether HSR has the capacity to mitigate tourism seasonality by spreading the number of tourists throughout the year. If so, the promotion and development of HSR services could be a promising mechanism to alleviate the intrinsic seasonality of certain areas.

Tourism seasonality is measured using the coefficient of variation and the Gini indexes, which are constructed based on the monthly number of hotel guests per province in Spain (NUTS 3). The Spanish case is of interest because of being one of the countries with the highest seasonality rates (Turrión-Prats and Duro, 2019) and the European country with the largest HSR network (Albalate et al., 2017). Since the literature has pointed to distinct effects of HSR introduction on tourism demand between domestic and international tourists (Li et al., 2019; Yang and Li, 2020; Zhang et al., 2020), the seasonality indicators are separately computed for domestic, international and total tourists. HSR connectivity and accessibility is measured using a synthetic index (denoted by HSRI) proposed by Xu et al. (2018) that incorporates (i) the degree of connectivity of the HSR network within a particular province, (ii) the number of areas that are reachable from the province, and (iii) the population of reachable areas reflecting the potential accessibility of the HSR.

This work contributes to two streams of literature that have been developed in parallel. It expands the literature on tourism seasonality (Cisneros-Martínez et al., 2018; Duro, 2016, 2018; Fernández-Morales and Cisneros-Martínez, 2019; Fernández-Morales et al., 2016; Turrión-Prats and Duro, 2017, 2019) and research on the effect of HSR development on tourism outcomes (Albalate and Fageda, 2016; Albalate et al., 2017, 2022; Chen and Haynes, 2015; Gao et al., 2019; Pagliara et al., 2017) by investigating for the first time how variations in HSR connectivity and accessibility impact the concentration of demand. Although there are some works on the factors that mitigate tourism seasonality (Rosselló et al., 2004; Turrión-Prats and Duro, 2018), to the best of our knowledge there are no studies that examine the role played by HSR connectivity and accessibility. In this vein, the closest paper to ours is that by Vergori and Arima (2022). These authors investigate how the time variation in the use of different modes of transport (cars, planes, ships and trains) by inbound tourists affects tourism seasonality in Italy. They show that seasonality decreases when a greater proportion of international tourists comes by plane. The current study differs from prior research in that we look at the effect of a synthetic index of HSR connectivity and accessibility on both domestic and international tourism seasonality.

The rest of the paper is structured as follows. Section 2 reviews the related literature. Section 3 presents the dataset and describes the construction of the connectivity and accessibility indicator. Section 4 outlines the two alternative but complementary empirical strategies used in the analysis. Section 5 presents the main results together with some extensions and robustness checks. Finally, Section 6 summarizes the main findings and offers some policy implications.

2. Literature review

2.1. Tourism seasonality

The uneven concentration of tourism demand at certain months of the year has been traditionally a topic of interest for tourism scholars and policy makers. Existing empirical evidence shows that tourism seasonality exerts important economic effects on the financial performance of hotels (Zhang et al., 2021), the labour market through the associated temporality of contracts in the hospitality industry (Krakover, 2000), regional productivity (Saito and Romao, 2018), the share of high-growth firms (Stojcic et al., 2022), the welfare of residents through overcrowding, crime and noise (Meleddu, 2014), the real estate market through increases in rents (Mikulic et al., 2021), and sustainable goals related to solid waste (Caponi, 2022) and the environmental degradation of natural resources (Martín-Martín et al., 2020), among others.

Most attention has been paid to annual seasonality taking months as the basic seasonal units (Martín-Martín et al., 2014; Rosselló et al., 2004), although some works have also examined monthly and weekly seasonality (Rosselló and Sansó, 2017; Sainaghi et al., 2019). With respect to the measurement of seasonality, the Gini index has been the most used synthetic indicator (Coshall et al., 2015; Fernández-Morales and Cisneros-Martínez, 2019; Fernández-Morales et al., 2016; Lozano et al., 2021; Saito and Romao, 2018). Nonetheless, recent studies have also adopted the coefficient of variation and entropy-based measures like the Theil index (Duro, 2016; Rosselló and Sansó, 2017; Sainaghi et al., 2019).

Spain is possibly the country that has received the greatest attention because of being one of the most relevant tourist destinations worldwide together with experiencing a high degree of seasonality (Turrión-Prats and Duro, 2019). Existing studies show that the Balearic Islands exhibit the highest concentration whereas Madrid and the Canary Islands the least (Duro, 2016). It is found that the overall seasonality is mostly driven by the foreign segment, which has both increased its share over total demand and monthly concentrated over time (Duro, 2016). Turrión-Prats and Duro (2018) estimate that two-thirds of the tourism concentration is attributed to the United Kingdom, France and Germany. In contrast, the domestic market exhibits comparatively lower concentration (Rosselló and Sansó, 2017). From a supply perspective, high-quality hotels located in urban areas are less seasonal and are more likely to remain open more time (Capó-Parrilla et al., 2007). A recent work by Lozano et al. (2021) develops an integrated analysis of seasonality in hotel prices and stays showing that differences in seasonality in quantities depend on hotel's price flexibility. Notwithstanding this, their study reveals substantial heterogeneity in the relationship between the concentration of prices and quantities over regions, which depends on the performance of competing destinations, the business cycle and managerial ability.

Part of the literature has been concerned with decomposing seasonality as a sum of the weights attributable to each market (Cisneros-Martínez and Fernández-Morales, 2013; Fernández-Morales, 2003; Fernández-Morales and Cisneros-Martínez, 2019; Fernández-Morales and Mayorga-Toledano, 2008; Fernández-Morales et al., 2016). A common result is that the closer the countries of origin, the less seasonally concentrated they tend to be. In addition, visiting friends and relatives and business trips are the least seasonal segments.

Another stream of research has tried to uncover the factors that explain seasonality differences across areas and periods. This literature has shown that prices, exchange rates, income levels, climatic aspects, institutional factors and cultural dimensions are the most relevant predictors (Duro and Turrión-Prats, 2019; Rosselló et al., 2004; Turrión-Prats and Duro, 2017, 2018, 2019). The study by Cisneros-Martínez et al. (2018) goes beyond and evaluates the capacity of a Spanish social tourism programme that promotes tourism travelling for older people (IMSERSO) to de-seasonalize tourism demand. Their results indicate that this social tourism programme actually generates greater hotel

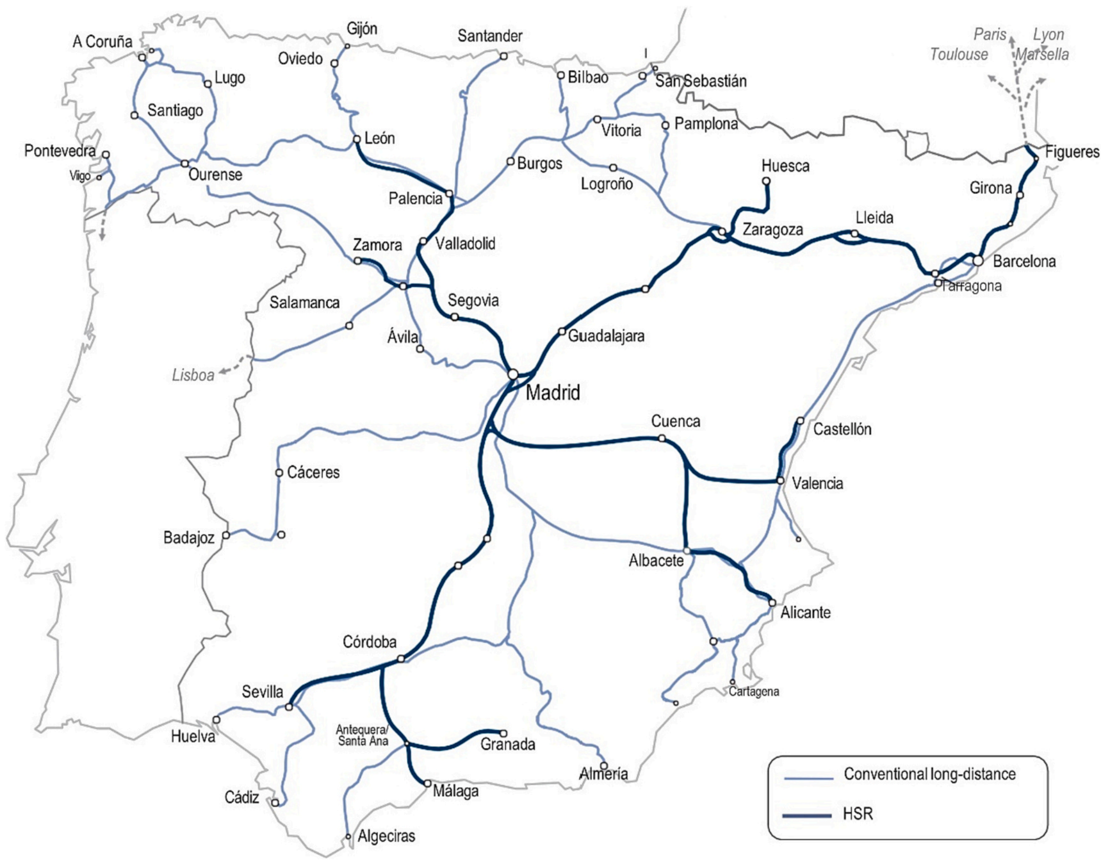

Fig. 1. The Spanish HSR and conventional long-distance network by the end of 2020.
Source: adapted from *Informe 2020. Observatorio del Ferrocarril en España.*

occupancy during the low season, contributing to alleviate seasonality and preventing the closure of tourist establishments. Nonetheless, the effect is found to be modest in those areas who exhibit an extremely high concentration of demand.

The analysis of spatial differences in seasonality is another research area of growing interest. Using a worldwide panel dataset for the period 2008–2013, Duro and Turrión-Prats (2019) show that tourism seasonality is significantly greater in countries in high latitudes. Ferrante et al. (2018) study the characteristics and amplitude of tourism seasonality in 21 European countries and identify six clusters of countries with similar seasonal trends. Interestingly, their work reveals that the clusters exhibit a clear geographical pattern, which the authors interpret in terms of common institutional factors and habits. Using data for Scotland, Coshall et al. (2015) find that more urbanised and accessible core regions present the lowest levels of seasonal concentration of international tourism demand. The opposite applies to rural areas, illustrating that accessibility is a relevant factor for explaining seasonality.

Despite the extant research on the topic, we know little about the role played by transport connectivity and accessibility in explaining tourism demand concentration. In this respect, the works by Chung and Whang (2011) and Donzelli (2010) present mixed evidence on the influence of low-cost carriers on tourism seasonality. On the one hand, Donzelli (2010) documents that the advent of low-cost air services in Italy has resulted in more balanced traffic flows throughout the year. On the contrary, Chung and Whang (2011) report that low-cost carriers have had little impact in reducing tourism seasonality in Korea. In a recent work, Vergori and Arima (2022) postulate that tourists travelling in different modes of transport might have distinct seasonal patterns associated with mode frequency and prices, so that variations in the distribution of travellers across modes might change the concentration of demand throughout the season. Their empirical analysis for the Italian case shows that tourists travelling by car contribute to increase tourism concentration whereas the opposite applies to airplane travellers. To the best of our knowledge, no previous study has evaluated the ability of HSR connectivity and accessibility in alleviating tourism arrivals concentration. The present study aims to fill this research gap.

2.2. *High-speed rail and tourism outcomes*

The analysis of how the development of new transport infrastructures affects tourism demand has also received great attention in the tourism literature in recent years. Whereas some works have focused on the increase in arrivals following the construction of new airports (e.g., Doerr et al., 2020), most of the literature has evaluated the effects of the provision of HSR services. In this section, we briefly review existing evidence on the topic.

From a theoretical perspective, the development of transport infrastructures increases the accessibility of some destinations, thereby expanding the basket of tourists' choices (Prideaux, 2000). Microeconomic models of destination choice attach negative utility to distance and travel time, which act as important deterrence factors (Khadaroo and Seetanah, 2008; Papatheodorou, 2001). Everything else being equal, improvements in transport connectivity and the associated drops in travel times are predicted to increase tourism demand. Although this likely depends on the relationship between ticket prices and travel time savings, high-speed trains offer the advantage of being more on time than other modes of transport and are unaffected by weather conditions. Nonetheless, some authors argue that new induced demand is rather small and that the HSR mainly produces transfer demand due to shifts in transportation modes (Sun and Lin, 2018). Each time a new HSR line is opened, tourists might change air transportation for the HSR (Albalate and Fageda, 2016).

Most empirical works have studied the effects of the introduction of the high-speed rail on tourism demand in China. This literature shows that China's HSR has increased tourism firms' value (Zhang et al., 2020) and hotel occupancy rates (Deng et al., 2021), decreased intercity travel times and boosted the number of one-day and weekend trips (Jin et al., 2020), and enhanced market competition, inducing a reallocation or urban tourism centres (Wang et al., 2012). Ex-post examinations of the introduction of the HSR at the province/city/prefecture level indicate that areas connected with HSR services experience increases in the number of international (Chen and Haynes, 2015; Li et al., 2019) and domestic tourists (Yao et al., 2022) and revenues (Bo and Ningqiao, 2018). Some works nonetheless report that the being connected to the HSR network matters more for increasing the number of tourists than for revenues (Gao et al., 2019).

Although the evidence for the Chinese case is rather robust, studies in Europe report inconclusive findings. For the EU-28, Castillo-Manzano et al. (2018) find that HSR services boost domestic tourism, whereas air transportation is more relevant for the international market. In the case of Italy, Pagliara et al. (2017) show that domestic visitors and overnight stays increased in Italian municipalities after being connected to the HSR network. As for Spain, empirical studies show that tourism has barely increased after provinces received new HSR lines (Albalate and Fageda, 2016). Although the HSR seems to have increased the number of foreign visitors and revenues (Campa et al., 2016), in the domestic market the HSR mainly benefits coastal destinations; its impact on inland regions is negligible or even negative (Albalate et al., 2017, 2022; Campa et al., 2018). Regarding whether HSR matters more for domestic or for international tourism, empirical evidence is also contradictory. Whereas some works document that HSR has a stronger effect for international than for domestic arrivals (Li et al., 2019; Yang and Li, 2020), others find just the opposite (Bo and Ningqiao, 2018).

A common finding is that the impact of HSR on tourism is heterogeneous across regions. For instance, Yao et al. (2022) and Gao et al. (2019) show that the effect of HSR is stronger in less developed and central regions. Campa et al. (2018) indicate the HSR reinforces previously attractive destinations and central inland locations. Yan et al. (2014) report that *Guangdong* and *Hunan* benefited more from HSR than *Hubei* due to the small number of stations in the later province. Similarly, Yan et al. (2014) find that the impact of HSR on arrivals is higher in inland regions. Bo and Ningqiao (2018) indicate that the positive impacts of HSR occur in regions with low accessibility. Other scholars identify winners and losers from being connected to the HSR network because of agglomeration and dispersion forces. For instance, Masson and Petiot (2009) show that the opening of a HSR line between *Barcelona* and *Perpignan* benefited the former at the expense of the later. Relatedly, Zhou et al. (2021) document that tourism economies experience distinct agglomeration and dispersion effects depending on the intensity of HSR daily operations.

The documented heterogeneity is likely to be due to the *degree* of connectivity and accessibility the provision of HSR rail conveys to each region. A given region might benefit more from having access to HSR depending on their connectivity to the network, the number of stations and the number and population size of source markets (Jin et al., 2020).

3. Data and methods

In this section, we first explain the two indicators of tourism seasonality considered in the analysis and the data source. Second, we describe the high-speed rail network in Spain and how we construct the synthetic index of connectivity and accessibility. Next, we discuss some control variables to be considered in the empirical analysis. Finally, we conduct a descriptive analysis of the link between high-speed connectivity and accessibility and tourism seasonality and present some summary statistics.

3.1. Dependent variable: tourism seasonality synthetic indicators

We retrieve a monthly panel dataset of the number of hotel guests for the 50 Spanish provinces (NUTS 2) during the period 2004:1–2019:12 ($tourists_{itm}$).[1] The autonomous cities of *Ceuta* and *Melilla* are excluded from the analysis. The data is drawn from the *Hotel Occupancy Survey* developed by the Spanish National Statistics Institute. Related studies on the impacts of HSR services also use the number of tourists lodged at hotel establishments as the demand indicator (Albalate and Fageda, 2016; Albalate et al., 2017; Campa et al., 2016).

We select the well-known Gini index as a synthetic indicator of tourism seasonality per year. This index is defined based on the Lorenz curve and has a long tradition as a measure of demand concentration (Coshall et al., 2015; Fernández-Morales and Cisneros-Martínez, 2019; Fernández-Morales et al., 2016; Lozano et al., 2021; Saito and Romao, 2018). Although there are different equivalent formulas to compute the Gini index, we follow the one used by Martín-Martín et al. (2014) and Coshall et al. (2015):

$$Gini_{it} = 1 + \frac{1}{12} - \left(\frac{2}{12^2 \times \frac{\sum_{m=1}^{12} tourists_{itm}}{12}} \right) \times (tourists_{1it} + 2 \times tourists_{2it} + \ldots + 12 \times tourists_{12it}) \tag{1}$$

where $tourists_{1it}$, $tourists_{2it}$, …, $tourists_{12it}$ are the number of hotel guests ordered in descending order of magnitude for each province *i* and year *t*. The Gini index ranges between zero (no seasonality, implying an equal distribution of tourists in all months) and one (highest level of concentration).

Duro (2016) notes that the Gini is sensitive to variations in months with demand close to the average. Turrión-Prats and Duro (2018) and Duro and Turrión-Prats (2019) advocate for the Coefficient of Variation (CV), which is "insensitive to the place where the monthly changes occur, and hence treats the changes that take place in the different months homogenously, regardless of their location on the monthly ranking" (Turrión-Prats and Duro, 2018, p.25). For this reason, we also consider CV (calculated as the within year standard deviation of $tourists_{itm}$ divided by the mean) as an alternative indicator.

As discussed in Section 2, tourism seasonality differs between domestic and international markets. For this reason, the Gini index and the Coefficient of Variation are calculated for the total number of tourists and also distinguishing between domestic and international tourists.

3.2. High-speed rail connectivity and accessibility

As aforementioned, we follow Xu et al. (2018) and use a synthetic index (denoted by HSRI) to measure the connectivity and accessibility of a province within the HSR Spanish network during the sample period. The HSRI is constructed as a linear combination of one connectivity and two accessibility indexes as follows:

$$HSRI_{it} = w_1 Beta_{it} + w_2 LBA_{it} + w_3 PBA_{it} \tag{2}$$

where $Beta_{it}$ controls for the connectivity degree of a certain province *i* within the HSR network in year *t* (see below), LBA_{it} considers the number of provinces in year *t* that can be reached by the HSR network

[1] Although different indicators of tourism demand have been used, we focus on the total number of tourists as a quantity indicator. Duro (2018) decompose the monthly concentration of tourism revenues into the contribution of average daily revenues, lengths of stay and tourist numbers. This author shows that tourists numbers is the variable that best explains the seasonality of tourism outcomes.

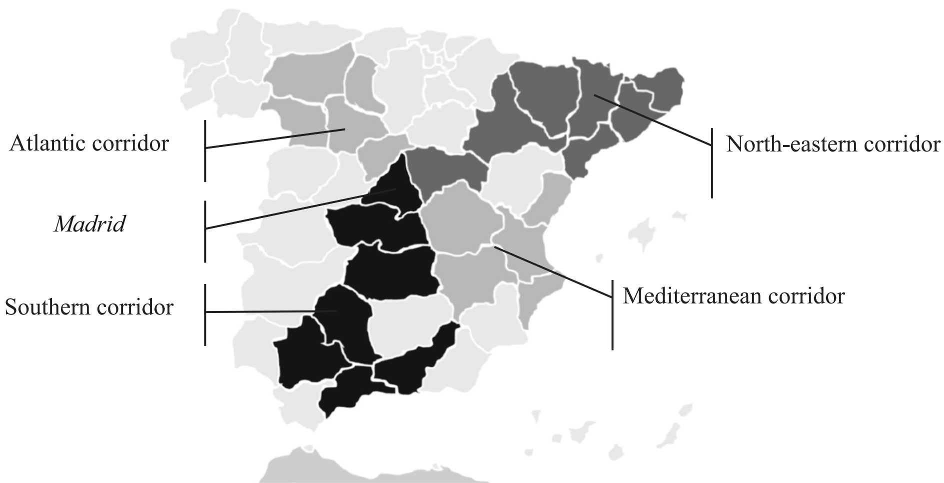

Fig. 2. Provinces in each corridor.
Note: Atlantic corridor: *Segovia* (2008), *Valladolid* (2008), *León* (2015), *Palencia* (2015), *Zamora* (2016); North-eastern corridor: *Guadalajara* (2004), *Huesca* (2004), *Lleida* (2004), *Zaragoza* (2004), *Barcelona* (2008), *Tarragona* (2008), *Girona* (2011); Mediterranean corridor: *Albacete* (2011), *Cuenca* (2011), *Valencia* (2011), *Alicante* (2013), *Castellón* (2018); Southern corridor: *Toledo* (2002), *Ciudad Real* (2002), *Córdoba* (2002), *Sevilla* (2002), *Málaga* (2007), *Granada* (2019). Each province first year with at least one HSR line in parentheses.

from province *i*, PBA_{it} represents the potential population that might take HSR to get to a certain province *i* in year *t*, and w_1, w_2 and w_3 are the corresponding weighs. The three subindexes are independent of one another but subject to the structure of the HSR network and the distribution of the residential population.

The data for the calculation of the three subindexes is gathered from *Informe 2020 Observatorio del Ferrocarril en España*, the latest annual report on Spanish railway published by the Ministry of Transport, Mobility and Urban Agenda (https://www.mitma.gob.es/ferrocarriles/observatorios/observatorio-del-ferrocarril-en-espana). Fig. 1 maps the distribution of HSR (maximum speed is over 200 km per hour and average speed is over 150 km per hour) and conventional long-distance train lines in Spain by the end of 2020.

As can be seen in Fig. 1, HSR in Spain has a radial structure with the capital city of Madrid being the centre of the network. It consists of four main corridors (Atlantic, Southern, North-eastern and Mediterranean) departing from the province of *Madrid* to the coast or to inland provinces (see Fig. 2 below). Importantly, travellers cannot switch from one corridor to another without passing through Madrid and changing of train. The provinces in light grey color do not have any HSR connection and therefore HSRI takes zero values for all the periods.

We now briefly explain how each of the three subindexes and the weights in Eq. (2) are defined.

3.2.1. Beta connectivity index

An adaptation of the Beta index (Taaffe et al., 1996; Rodrigue, 2020) is proposed, in which the connectivity of each province is measured by the ratio of the number of HSR lines (links) over the number of stations (nodes) within its corresponding boundaries as follows:

$$Beta_{it} = \frac{S_{it}}{N_{it}} \tag{3}$$

where $Beta_{it}$ is the beta index for province *i* in year *t*, S_{it} is the number of HSR lines passing through the boundary of province *i* during *t*, and N_{it} is the number of HSR stations within the boundary of province *i* in year *t*.

3.2.2. Accessibility indexes

The proposed accessibility indexes aim to capture both the number of provinces that can be reached from a given province through the HSR network (denoted by LBA_{it}) and the population that can potentially get to that province using the HSR network (PBA_{it}). Since Moyano et al. (2018) indicate that the influencing distance for HSR is between 400 km and 600 km, the LBA index is calculated as the number of provinces that can be directly reached from province *i* by HSR within a particular corridor divided by 100 in each period *t*. This closely follows Xu et al. (2018). The PBA accessibility index for a certain province *i* reflects the total population of other provinces within the same HSR corridor, thereby measuring the size of potential source markets.

3.2.3. Weights (w_1, w_2 and w_3)

The definition of the weights of each of the previously described connectivity and accessibility indexes in (2) is subject to debate. Xu et al. (2018) consider $w_1 = w_2 = w_3 = 1$ so that all subindexes are given the same importance. However, when there are scale differences among the subindexes, this makes the HSRI to mainly reflect the subindex that is greater in magnitude. We opt instead for estimating the weights based on regression analysis without restrictions. In particular, we regress the (log of) the total number of passengers boarding and alighting from long distance and high-speed trains per province and period ($ln\ passengers_{it}$) on the three connectivity and accessibility indexes ($Beta_{it}$, LBA_{it} and PBA_{it}) for each of the four corridors presented in Fig. 2.[2] Implicitly, this procedure assumes that HSR demand likely reflects the connectivity and accessibility of the network so that a linear regression is able to capture the relative influence of each of the three subindexes. The regressions use province fixed effects. Table A1 in Appendix presents the estimation results. Because the weights are corridor-specific, note that this method allows for geographic heterogeneity in the relative importance of each of the three subindexes in the overall HSRI.[3] The mean values of the resulting HSRI per province are shown in Table A2 in Appendix.

Before moving on, it is important to stress that, in line with related studies for the Spanish case (Albalate and Fageda, 2016; Albalate et al., 2017; Campa et al., 2018), the HSRI is taken as exogenous. This is because the development of the HSR network in Spain has not responded to traffic congestion problems and mobility demand but to political reasons and centralization strategies from the central Government (Albalate et al., 2012). Several works have discussed that the regional allocation of transport infrastructure in Spain has been driven by political interests, resulting in oversupply and mismatch with demand (Albalate et al., 2015). As such, provinces' HSR connectivity and accessibility can be taken as exogenous to demand concentration conditional on economic factors. See Albalate and Fageda (2016) for a discussion on this.

[2] The data for the number of passengers per province and year is obtained from the annual reports of *Observatorio del Ferrocarril en España*. Unfortunately, it does not distinguish between long distance and high-speed train passengers.

[3] We disregard the use of the entropy method as used by Xu et al. (2018). The reason is that this procedure requires to ex-ante define the support of the coefficients (weights). Given the differences in scale among the subindexes, imposing the coefficients to lie within, for instance, the unit interval would artificially force the subindex with the greatest mean to get more weight, regardless of its relevance in explaining HSR passenger demand.

3.3. Control variables

To examine the effect of HSR connectivity and accessibility on demand concentration, we consider the following variables at the province level as controls: population size (in thousands of people), GDP (in thousands of euros measured at basic prices), the Consumer Price Index for accommodation services (base 2016) and the number of airport passengers.[4] The first three variables are obtained from the Spanish National Statistics Institute based on the annual population census, regional accounting statistics and the Consumer Price Index, respectively. Similar control variables have been used in related empirical studies (e.g., Albalate et al., 2017; Li et al., 2019). Airport passengers are drawn at the airport level and aggregated at the province level from the annual statistics of traffic provided by AENA (https://www.aena.es/es/estadisticas/informes-anuales.html). Although this variable includes both passengers' arrivals and departures (including the trips by residents), we believe the use of total passengers is a suitable proxy of the size, transit, flight frequency and connectivity of the airports in a particular region. Its use as a control for the effect of HSR on tourism outcomes follows Pagliara et al. (2017) and Albalate and Fageda (2016).

3.4. Descriptive analysis of tourism seasonality

Panels A and B in Fig. 3 present kernel density plots of the Gini and CV indexes, separately for domestic, international and total tourists. It is shown that the overall seasonality closely reproduces the pattern of the international segment. This is because, as discussed in Duro (2016, 2018), international tourists represent a major percentage of total tourists, especially in the Spanish provinces with highest seasonality rates. Note that although the shapes of the two indicators for each type of segment appear to be very similar, there are relevant scale differences (see Fig. A1 in Appendix).

Table 1 presents descriptive statistics of the Gini and CV indexes, the connectivity and accessibility index (*HSRI*) defined before and its subcomponents (*BETA*, *LBA* and *PBA*), and the control variables (*POPULATION*, *GDP*, *CPI_ACCOM* and *AIR_PASSENGERS*). The pairwise correlation matrix is presented in Table A3 in Appendix. International demand is notably more seasonal, on average, than the domestic market. The concentration of international tourists has also more variability over time and among regions according to the larger standard deviations. Table A4 in Appendix breaks down the mean values of the seasonal indexes per province. The provinces of *the Balearic Islands* and *Tarragona* are the ones with the highest seasonality in both domestic and international segments, in line with Duro (2016). Other provinces like *Lugo*, *Girona* or *Almería* also present high demand concentration. On the contrary, *Barcelona*, *Granada* and *Valladolid* are the least seasonal for the domestic segment; *Madrid*, *Las Palmas* and *Tenerife* present the lowest demand concentration in relation to international tourists, also as in Duro (2016). Fig. A2 in Appendix plots the time evolution of the mean values of seasonality. The concentration of both domestic and international tourists smoothly increased until 2014. From then onwards, it seems to have slightly decreased. The corresponding values are presented in Appendix, Table A5.

The HSRI index ranges between zero (67% of the sample) and 2.026, with a mean value of 0.19. A panel linear regression with period and province fixed effects of *HSRI* on the control variables in log form suggests that HSR connectivity and accessibility is positively correlated with GDP, population size and prices and inversely associated with the number of airport passengers (Appendix, Table A6). This suggests that these controls are relevant to make HSRI conditionally exogenous to tourism seasonality.

Figs. 4, 5 and 6 present binned scatterplots of Gini_DOMESTIC, Gini_INTERNATIONAL and Gini_TOTAL on HSRI. The corresponding plots for the Coefficient of Variation are presented in Figs. A3, A4 and A5 in Appendix. We see that there is an unconditional negative relationship between HSRI and tourism seasonality in the three segments. However, this association might be confounded by other factors. To properly examine the effect of HSRI on the seasonality indicators, we move to regression analysis.

4. Empirical strategy

This section presents our empirical strategy. First, we propose a canonical two-way panel fixed effects (TWFE) with some time-varying controls at the province level. Next, we move to a dose-response framework (Cerulli, 2015) to examine heterogeneity in the Average Treatment Effect (ATE) of the HSR index over its domain.

4.1. Two-way fixed effects panel regression

We first propose a canonical two-way fixed effects (TWFE) panel data model as follows:

$$y_{it} = \alpha + \beta HSRI_{it} + \gamma X_{it} + T_t + \mu_i + \varepsilon_{it} \tag{4}$$

where i indexes provinces for $i = 1, \ldots, 50$ and t years for $t = 2004, \ldots, 2019$; y_{it} is a tourism seasonality indicator; $HSRI_{it}$ is the HSR connectivity and accessibility index defined before; X_{it} is the set of time-varying controls; T_t are year fixed effects, μ_i province fixed effects; and ε_{it} is the random error term.[5]

The model specification in (4) can be seen as a type of difference-in-differences (DiD) setup with a continuous treatment. Since $HSRI_{it}$ can be considered ‘as good as randomly assigned’ conditional on controls, the parameter β captures the effect of HSR connectivity and accessibility on tourism seasonality.[6] That is, β is the estimand of interest (Average Treatment Effect on the Treated, ATT), obtained as the weighted sums of the ATTs in each province and period. However, recent developments in the DiD literature have proved that in the presence of treatment effect heterogeneity (i.e., the role of $HSRI_{it}$ likely varies across provinces and periods), β might be a biased estimate of the ATT (de Chaisemartin and D'Haultfœuille, 2020). In particular, aside from within and between heterogeneity in the role of HSR accessibility documented in previous works (Campa et al., 2018; Gao et al., 2019; Yao et al., 2022), marginal increases in $HSRI_{it}$ might exert different effects on seasonality depending on the level of connectivity and accessibility. To inspect this closer, we move to a dose-response framework.

4.2. Dose-response methodology (DRM)

In this subsection we briefly describe the dose-response methodology (DRM) originally developed by Cerulli (2015), which can be seen as an extension of the Regression Adjustment model developed by Wooldridge (2010) to a continuous treatment setting. We closely follow the notation used by Cerulli (2015). The purpose is to allow the impact of the dose of HSRI on seasonality to vary depending on whether HSRI is high or low.

[4] Ideally, it would be better to control for the Hotel Price Index. Unfortunately, the lowest regional disaggregation of this index is the Autonomous Community level (NUTS 2). Since we work at the province level, we use instead the subindex for accommodation services of the Consumer Price Index.

[5] All the controls except CPI_ACCOM are included in the regression in logs. Since our dependent variable is demand concentration rather than demand levels, we do not envisage problems of endogeneity bias due to the simultaneous determination of air traffic and tourism. To deal with the fact that some provinces do not have airport (and therefore air passengers), the variable is transformed using the inverse sine transformation to keep the zeroes.

[6] To test the plausibility of this assumption, we regress HSRI on lags on annual tourism demand (Appendix, Table A7). The lags are never found to be significant, suggesting that the levels of connectivity and accessibility have not reacted to tourism demand and can be taken as conditionally exogenous.

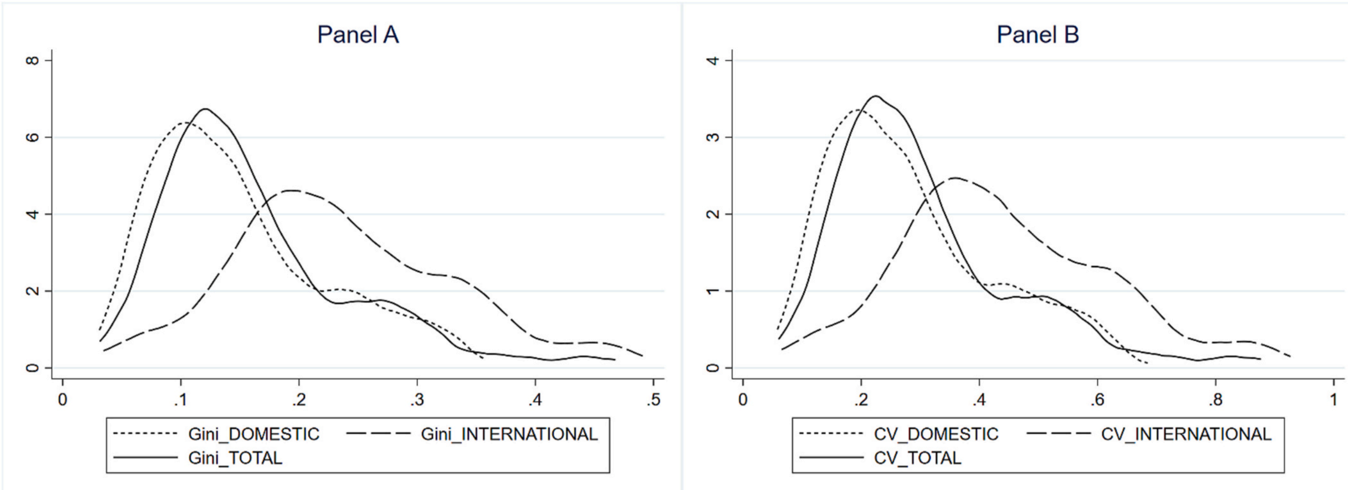

Fig. 3. Kernel density plots of Gini (Panel A) and CV (Panel B) indexes for domestic, international and total tourists.

Table 1
Summary statistics of the variables.

Variable	Mean	SD	Min	Max
Gini_DOMESTIC	0.149	0.073	0.031	0.358
Gini_INTERNATIONAL	0.236	0.094	0.034	0.490
Gini_TOTAL	0.163	0.082	0.031	0.467
CV_DOMESTIC	0.282	0.138	0.058	0.684
CV_INTERNATIONAL	0.440	0.177	0.065	0.926
CV_TOTAL	0.307	0.153	0.060	0.876
HSRI	0.194	0.343	0	2.026
BETA	0.401	0.657	0	2.5
LBA	0.015	0.027	0	0.22
PBA	4.538	7.222	0	16.852
POPULATION	900,895	1,149,254	65,488	6,663,394
GDP	2.13e+07	3.34e+07	1,858,201	2.41e-08
CPI_ACCOM	90.046	9.108	67.072	119.387
AIR_PASSENGERS	4,172,583	1.01e+07	0	6.17e-07

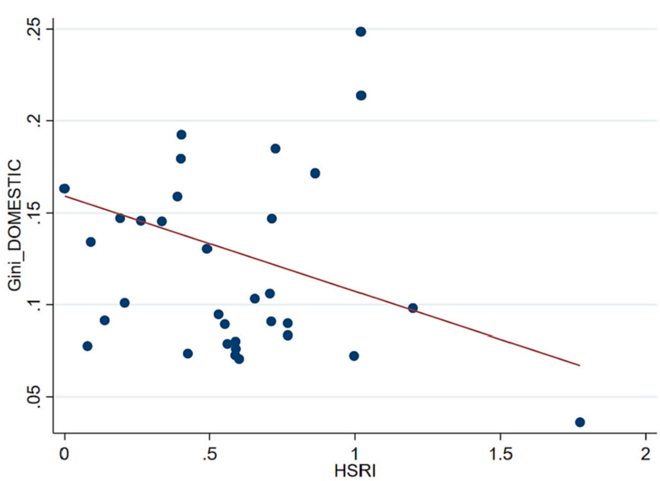

Fig. 4. Binned scatterplot (100 quantiles) of Gini_DOMESTIC on HSRI.

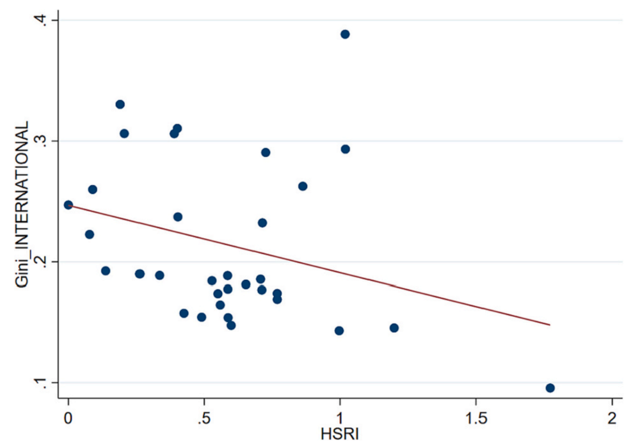

Fig. 5. Binned scatterplot (100 quantiles) of Gini_INTERNATIONAL on HSRI.

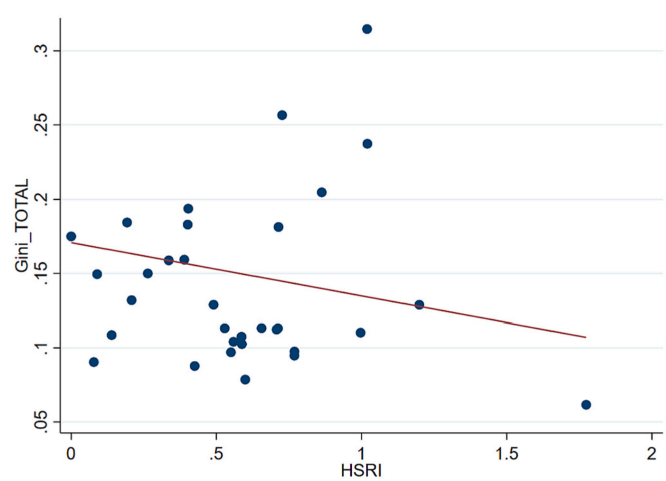

Fig. 6. Binned scatterplot (100 quantiles) of Gini_TOTAL on HSRI.

Let us define ω_{it} as a binary indicator that takes value 1 if $HSRI_{it} > 0$ and 0 otherwise (treatment indicator) and ACI_{it} denote $HSRI_{it}$ normalized to take values on the [0,100] interval (re-scalation). Assuming a parametric form of the potential outcomes of tourism seasonality with additive separability for 'treated' and 'non-treated' provinces, we have:

$$\begin{cases} \omega_{it} = 1 : y_{1it} = \alpha_1 + g(ACI_{it}) + \gamma_1 X_{it} + T_{1t} + \mu_{1i} + \varepsilon_{1it} \\ \omega_{it} = 0 : y_{0it} = \alpha_0 + \gamma_0 X_{it} + T_{0t} + \mu_{0i} + \varepsilon_{0it} \end{cases} \quad (5)$$

where y_{1it} and y_{0it} are the potential outcomes of tourism seasonality depending on the binary accessibility indicator ω_{it}, and $g(ACI_{it})$ is a general derivative function of the normalized index ACI_{it}. Under the potential outcomes framework (Rubin, 1974), the conditional-on-covariates Average Treatment Effect (ATE) can therefore be defined as follows:

$$ATE = \omega_{it} \times \left\{ \widetilde{\alpha} + g(ACI_{it}) + \widetilde{\gamma}X_{it} + \widetilde{T}_t + \widetilde{\mu}_i \right\} + (1 - \omega_{it}) \times \left\{ \widetilde{\alpha} + \widetilde{\gamma}X_{it} + \widetilde{T}_t + \widetilde{\mu}_i \right\} \tag{6}$$

where $\widetilde{\alpha} = \alpha_1 - \alpha_0$, $\widetilde{\gamma} = \gamma_1 - \gamma_0$; $\widetilde{T}_t = T_{1t} - T_{0t}$ and $\widetilde{\mu}_i = \mu_{1i} - \mu_{0i}$. The ATE can be alternatively rewritten as the weighted sum of the Average Treatment Effect on the Treated (ATET) and the Average Treatment Effect on the non-treated provinces (ATENT):

$$ATE = \omega_{it} \times ATET + (1 - \omega_{it}) \times ATENT \tag{7}$$

The Dose Response Function (DRF) is obtained by averaging the ATE over the set of explanatory variables as follows:

$$ATE(ACI_{it}) = \begin{cases} ATET + \left\{ g(ACI_{it}) - \overline{g}_{ACI_{it}>0} \right\} \text{ if } ACI_{it} > 0 \\ ATENT \text{ if } ACI_{it} = 0 \end{cases} \tag{8}$$

As mentioned before, we assume HSRI is conditionally exogenous to the seasonality indicators given controls (*unconfoundedness*). Under this mean independence assumption, the ATE in (7) and therefore of the DRF in (8) can be consistently estimated using the following OLS regression:

$$y_{it} = \alpha + \tau\omega_{it} + \gamma_0 Z_{it} + \omega_{it} \times \gamma(Z_{it} - \overline{Z}) + \omega_{it} \times \{ g(ACI_{it}) - \overline{g} \} + \eta_{it} \tag{9}$$

where $\tau = ATE$; Z_{it} gathers all the explanatory variables so that $Z_{it} = (X_{it}, T_t, \mu_i)$; $\eta_{it} = \varepsilon_{1it} + \omega_{it} \times (\varepsilon_{1it} - \varepsilon_{0it})$; $\overline{g}$ is the sample mean of $g(ACI_{it})$ and $g(ACI_{it})$ is assumed to be a third-degree polynomial parametric function given by the following expression:

$$g(ACI_{it}) = \lambda_1 ACI_{it} + \lambda_2 ACI_{it}^2 + \lambda_3 ACI_{it}^3 \tag{10}$$

The parameter estimates from (9) are subsequently used for the computation of the ATET and the ATENT by plugging their values in the above formulas. Finally, the ATE for each potential value of the accessibility index is expressed as:

$$\begin{aligned} ATE(ACI_{it}) = & \omega_{it} \left\{ \widehat{ATET} + \widehat{\lambda}_1 \left(ACI_{it} - \frac{1}{N} \sum_{j=1}^{J} ACI_{it} \right) \right. \\ & \left. + \widehat{\lambda}_2 \left(ACI_{it}^2 - \frac{1}{N} \sum_{j=1}^{J} ACI_{it}^2 \right) + \widehat{\lambda}_3 \left(ACI_{it}^3 - \frac{1}{N} \sum_{J=1}^{J} ACI_{it}^3 \right) \right\} \\ & + (1 - \omega_{it}) \widehat{ATENT} \end{aligned} \tag{11}$$

where subindex *j* is used in place of *it* because of the panel structure and $J = NxT$. If we plot $ATE(ACI_{it})$ against ACI_{it}, we have a graphical representation of the DRF in (8).

As discussed by Cerulli (2015), the DRM offers several advantages over other approaches like the generalized propensity score method proposed by Hirano and Imbens (2004). First, it does not rely on the normality assumption. Second, it is well-suited for cases where a high proportion of units have zero values of the treatment. The reader is referred to Cerulli (2015) for further details. Empirical applications of this methodology in different setting include Cusimano et al. (2021), Cerulli and Ventura (2021), Filippetti and Cerulli (2018) and Prifti et al. (2019).

5. Results

5.1. Main findings

Table 2 presents the estimation results for the panel linear regression in (4) using both Gini and CV indicators of domestic, international and total demand seasonality. We find that, on average, marginal increases in HSRI are negatively and significantly associated with demand concentration. This implies that demand concentration decreases as the high-speed rail connectivity and accessibility of the province improves. Auxiliary regressions using the share of tourists in the high (June to September) and low (October to May) seasons suggest that HSRI increases demand shares in the low season, thereby decreasing concentration in the high season (see Appendix, Table A8).

If we divide the coefficient estimate for the Gini index in Table 2 by the sample mean and express in percentage terms, the effect size of HSRI is greater in magnitude for international (−0.025/0.236*100 = −10.6%) than for domestic tourists (−0.010/0.149*100 = −6.7%). The same percentage drop is obtained for the CV (−0.046/0.440*100 = −10.4% for international tourists and − 0.019/0.282*100 = −6.7% for domestic tourists). This result is likely explained by the following mechanism. The radial structure of the HSR network in Spain (Fig. 1) makes all the provinces in each corridor to be linked to the capital city. Madrid has the most important airport in the country (Adolfo Suárez-Madrid Barajas), which in 2019 received 61.7 million passengers (22.5% of total air passengers in the country). Our of them, 44 million were international travellers (16 million domestic travellers). Previous works analysing long-distance travelling have shown that tourists prioritize travel time savings and are not highly sensitive to travel costs (Becken and Achiff, 2011), especially foreign tourists that tend to be highly time constrained.[7] In the low season, foreign tourists arriving by plane to Madrid from abroad might prefer faster modes of transport (e.g., HSR) over conventional trains or buses, which therefore could explain the relatively greater effect of improving HSR connectivity and accessibility for this segment.

With regard to the control variables, domestic seasonality is negatively correlated with GDP, implying that tourism demand in more economically developed provinces is less seasonal. However, the population size of the province, the price index for accommodation services and the number of air passengers are not significant predictors of tourism seasonality in none of the regressions.[8] The mean VIF is 2.36, implying that the estimates do not suffer from collinearity problems.

The regressions in Table 2 also detect relevant temporal and cross-sectional heterogeneity in seasonality. Figs. A9 and A10 report the year fixed effects for the different regressions. In line with the descriptive evidence presented in Subsection 3.4, the seasonality of the international segment exhibits an inverted U-shaped pattern: it increased up to 2014. From then onwards, it has decreased since there are no significant differences with 2004 (reference category). In contrast, although domestic seasonality has also slightly decreased since 2014, it is still higher than the corresponding values at the beginning of the study period. As regards province heterogeneity, Figs. A11-A16 in Appendix map the province fixed effects obtained from the regressions in Table 2. The Balearic Islands and Tarragona stand as the provinces with greater seasonality in both international and domestic tourists. Other Catalan provinces like Barcelona and Girona, North-western provinces like Asturias, Cantabria and A Coruña, and the Andalusian provinces of Cádiz, Málaga and Almería algo exhibit high concentration in the domestic segment.

We now turn to the DRF. Figs. 7, 8 and 9 plot the ATE and 95% confidence interval of the normalized HSRI (ACI) over its domain following Eq. (8) for Gini_DOMESTIC, Gini_INTERNATINAL and Gini_TOTAL, respectively. Standard errors have been bootstrapped using 100 replications. The corresponding graphs and point estimates for the CV are presented in Figs. A17-A19 and Table A9 in Appendix. Interestingly, we document that the effect of HSRI on seasonality is non-linear and varies depending on the level of connectivity and accessibility. It is U-

[7] Tourists' willingness to pay for travel time savings is highly dependent on income (Van Can, 2013), which tends to be higher among foreign travellers.

[8] The latter is corroborated using (unconditional) descriptive binned scatterplots (see Figures A6-A8 in Appendix), which clearly indicate that annual airport passengers are uncorrelated with tourism seasonality.

Table 2
Coefficient estimates for panel linear regression.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Gini_DOM	Gini_INTERN	Gini_TOTAL	CV_DOM	CV_INTERN	CV_TOTAL
HSRI	−0.010**	−0.025***	−0.019***	−0.019**	−0.046***	−0.046***
	(0.005)	(0.007)	(0.005)	(0.009)	(0.012)	(0.009)
Ln GDP	−0.052**	0.028	−0.039*	−0.117***	0.055	−0.093**
	(0.021)	(0.029)	(0.021)	(0.040)	(0.055)	(0.041)
Ln POPULATION	−0.020	−1.6e-04	−0.016	−0.050	0.001	−0.042
	(0.020)	(0.027)	(0.020)	(0.038)	(0.052)	(0.039)
CPI_ACCOM	−3.9e-04	3.5e-05	−1.1e-04	−0.001*	8.6e-05	−7.8e-05
	(2.7e-04)	(3.7e-04)	(2.7e-04)	(0.001)	(0.001)	(0.001)
Ln AIR_PASSENGERS	−2.7e-04	−1.0e-04	−3.4e-05	−0.001	−3.1e-04	1.3e-04
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Constant	1.270***	−0.233	1.007**	2.892***	−0.476	2.349***
	(0.414)	(0.558)	(0.407)	(0.786)	(1.069)	(0.808)
Year fixed effects	YES	YES	YES	YES	YES	YES
Province fixed effects	YES	YES	YES	YES	YES	YES
Observations	800	800	800	800	800	800
Number of provinces	50	50	50	50	50	50

Standard errors in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

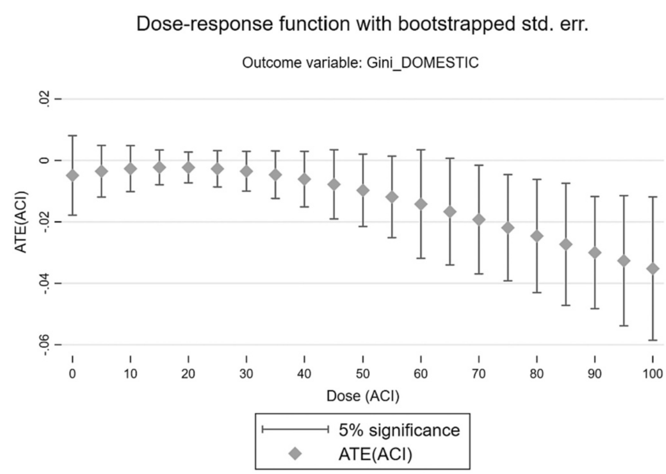

Fig. 7. Average Treatment Effect of ACI on Gini_DOMESTIC.

shaped for the international segment but inverted U-shaped for the domestic one. In both cases, the size effect is larger at high levels of HSRI. This suggests that high levels of high-speed connectivity are required to produce significant decreases in yearly seasonality, especially for domestic tourists.

5.2. Robustness checks

We have performed a battery of robustness checks and extensions to our main findings. First, we have inspected the sensitivity of our estimates to omitted variable bias using the methods proposed by Oster

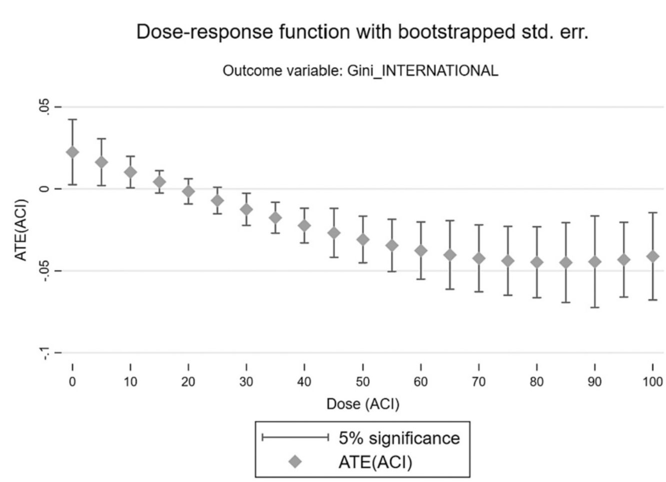

Fig. 8. Average Treatment Effect of ACI on Gini_INTERNATIONAL.

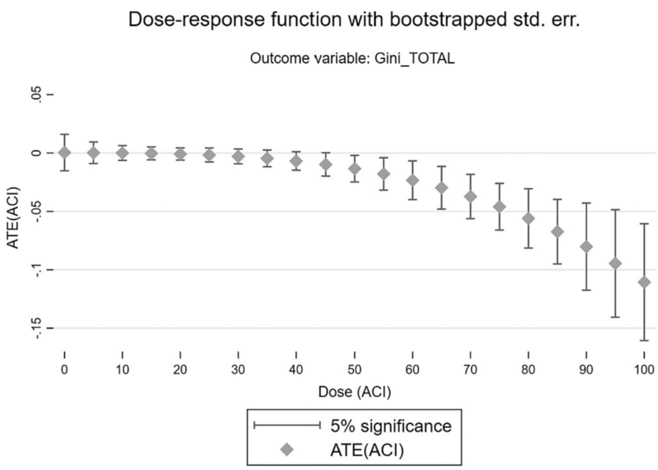

Fig. 9. Average Treatment Effect of ACI on Gini_TOTAL.

(2019). For the case of GINI_TOTAL, selection in unobservables must be over 1.80 times selection in observables to make the effect of HSRI non-significant (see Appendix, Table A10). Although we cannot completely rule out the possibility of bias from omitted factors, under the assumption that selection in observables is proportional to selection in unobservables (Oster, 2019), the results are quite robust to omitted confounders. Second, we have estimated the model in (4) (i) with province and time fixed effects but without controls, and (ii) with controls but without HSRI (Appendix, Tables A11-A12). We find that the lack of statistical significance of some of the controls is preserved when HSRI is excluded from the regression. Third, because the Balearic and the Canary Islands stand as the regions with the highest and lowest seasonality together with their insular character, we re-estimated the model excluding them from the sample (Appendix, Table A13). The results are robust. Fourth, since the Gini index is restricted to lie in the unit interval, we have estimated panel beta regressions with a logistic link function for the conditional mean as done by Vergori and Arima (2022). The results remain similar (Appendix, Table A14), although in this case the parameters must be interpreted in terms of the change in the log odds of the seasonality indicator. Fifth, we have used alternative synthetic indicators of seasonality as dependent variables like the modified Theil measure and the Herfindahl-Hirschman index. The formulas for the calculation of these alternative indexes together with summary statistics are presented in Appendix (Table A15). The results appear to be robust to the seasonality indicator (Appendix, Table A16). Sixth, we have considered alternative specifications of the control variables like GDP per capita (in logs), a binary indicator for having an airport instead of the number of passengers and the General Consumer Price Index in place of the subindex for accommodation services (Table A17 in Appendix). The results remain largely unchanged.

With regard to the sensitivity of the findings to the HSRI index construction, we have re-estimated the models (i) entering the three subindexes separately in the specification, (ii) using standardized (zero mean and standard deviation equal to one) values of $Beta_{it}$, LBA_{it} and PBA_{it} in the construction of HSRI, (iii) without the 1/100 adjustment in the definition of LBA_{it}. The estimation results are very similar, although we detect that the three subindexes have distinct effects on seasonality when included separately (Tables A18-A22 in Appendix).

Our analysis considers different measures of intra-annual seasonality in the number of tourists (lodged at hotel accommodations). However, some works have paid attention to the effects of the HSR on other demand indicators like overnight stays (Albalate and Fageda, 2016; Albalate et al., 2017). Table A23 in Appendix replicates the estimates in Table 2 using CV and Gini indicators of seasonality in overnights stays, also distinguishing between domestic, international and total tourists. Similarly, Figs. A20-A25 present the corresponding dose-response estimates. The results are pretty similar to the main analysis; although HSRI is now not significant for explaining seasonality in the domestic market, the dose-response estimates point again to non-linear inverse relationship between HSR connectivity and accessibility and the seasonality of overnight stays.

As a final extension of the analysis, we have constructed a municipality-level panel dataset on domestic, international and total tourists in the most important cities of the country.[9] Based on that, we have computed the corresponding CV and Gini indexes for each segment in each city. Cities with missing values in some months have been discarded because the calculation of the seasonality indicators requires information for all the months within the year. As a result, we end with a balanced panel dataset consisting of 46 cities, most of which are capital cities (see Appendix, Table A24).[10] Table A25 in Appendix reports the TWFE estimates and Figs. A26-A31 plot the dose-response results. The negative effect of HSRI on the seasonality of the international segment and total demand is preserved, showing great robustness. Nonetheless, in this case the effect on domestic seasonality becomes non-significant.

6. Conclusions

6.1. Summary of findings

This study has evaluated for the first time the effect that high-speed rail connectivity and accessibility has on regional tourism seasonality. As opposed to previous research that mainly considers a binary indicator for whether the region has at least a high-speed rail line (Albalate et al., 2017; Bo and Ningqiao, 2018; Gao et al., 2019; Yao et al., 2022; Zhang et al., 2020), we have used a synthetic index built upon Xu et al. (2018). Our HSRI index considers the number of stations and lines, the provinces that can be reached through the HSR network and the population size of the potential source markets. Using a panel dataset at the province level for the period 2003–2019, two-way panel fixed effects regressions indicate that marginal improvements in HSR connectivity and accessibility decrease the seasonality of both domestic and international tourism demand, on average. The effect size is found to be larger for the international segment, which could be explained by the radial structure of the network with Madrid at the centre (where the main airport for international arrivals is located) and the likely greater preference for travel time savings among foreign travellers. Better HSR connections are found to significantly increase (decrease) the share of arrivals in the low (high) season, thereby smoothing seasonality. By applying a dose-response methodology, we further document that the capacity of HSR to decrease demand concentration is non-linear and greater at high levels of connectivity and accessibility. Seasonality exhibits a U-shaped pattern for the international segment but inverted U-shaped for the domestic segment during the study period. This suggests that although improvements in HSR connectivity and accessibility are more relevant quantitatively speaking for the seasonality of international demand, marginal improvements at high connectivity levels seems to be more relevant for the domestic market. Importantly, the findings are robust to the use of different seasonality indicators, controls and omitted

[9] This dataset is drawn from INE (Hotel Occupancy Survey) for the period 2005–2019. INE offers information about hotel tourism demand in selected municipalities (cities) along the country (*Puntos de Interés Turístico*) that concentrate most of the tourism demand of their provinces.

[10] We disregarded this municipality-level dataset for the main analysis because it is affected by selection issues: the discarded areas are those with missing values, which is the result of statistical confidentiality associated to low demand in some periods of the year. The provinces of Asturias (Oviedo and Gijón), A Coruña (Coruña and Santiago), Alicante (Alicante and Benidorm), Málaga (Málaga and Marbella) and Murcia (Murcia and Cartagena) have two cities each. On the contrary, the municipality-level dataset does not contain any city located in the provinces of Álava, Badajoz, Ciudad Real, Palencia, Jaén, Huesca, Huelva, Guadalajara and Girona.

confounders.

6.2. Contribution and policy implications

Our results complement existing knowledge about the effects of transport infrastructure development on tourism outcomes in general and in Spain in particular. The literature about the effects of HSR on tourism demand levels in Spain has found weak or non-significant effects and relevant heterogeneity across areas (Albalate and Fageda, 2016; Albalate et al., 2017, 2022; Campa et al., 2016, 2018). This paper has deepened into the potential source of this discrepancy by considering the different degree of connectivity and accessibility to the network over time and across provinces, showing the link between the HSR and seasonality.

The findings of this study have important policy implications for Spanish authorities. Tourism demand in some Spanish regions (particularly those specialized in sun and beach) is highly seasonal (Duro, 2016), which has important implications in terms of carrying capacity, sustainable goals, the temporality of hospitality workers' contracts or the welfare of residents. Our results show that the development of HSR projects that convey improved connectivity and accessibility are a promising mechanism to alleviate the concentration of tourism demand. This is economically relevant because it distributes tourism revenues throughout the year, improves destination image and attractiveness through word-of-mouth effects, and increases tourist firms' revenues by making profits not only during the peak season.

Although Spain is already the European country with the largest HSR network, new lines are currently under construction with the aim of avoiding transportation inequalities among regions. The provinces of Asturias, Burgos, Murcia, Cáceres and Badajoz are expected to join the HSR network soon. According to our estimates, tourism seasonality is predicted to decrease in these regions; however, this will be contingent on the degree of connectivity and accessibility conveyed by the new infrastructure. The main takeaway from our analysis for transport planning is therefore that the impact of HSR projects on tourism demand strongly depends on how well the new lines connect with source markets and specially with Madrid.

Since travelling through HSR is more efficient in terms of carbon emissions per kilometre as compared to air transportation (Sun and Lin, 2018), the promotion of railway transportation within Spanish regions could be also positive for climate change mitigation goals. Furthermore, hospitality managers and regional governments should develop marketing and promotional campaigns to foster tourists' awareness about the availability of HSR connections to reach their travel destinations.

From a broader perspective, since our HSRI index measures how well a province is connected to and accessible from the rest of the network, it could be applied in transport planning for understanding transport regional inequalities. Xu et al. (2018) use this measure to identify winners and losers of HSR construction plans among Chinese regions. Its time-varying nature allows policy makers and transport planners to track and evaluate the regional changes in high-speed railway connectivity and accessibility to identify regions that are left behind. From this viewpoint, this index could be a valuable indicator of the shrinkage of space HSR systems provide to a region, complementing the findings by Moyano et al. (2018) using alternative measures.

6.3. Limitations

The study has some limitations that must be acknowledged. First, the seasonality indexes are computed using data on the number of tourists lodged at hotels. Although hotels are still the most important accommodation for tourists, the effects of HSR connectivity and accessibility could be different for other accommodation segments. In this regard, Airbnb has emerged as a relevant competitor, and its staggered entrance throughout Spanish cities has resulted in important increases in the volume of tourists hosted (e.g., Jiménez et al., 2022). Although the effects of the HSR on peer-to-peer accommodation markets is beyond the scope of this paper, it would constitute an interesting research question for future studies. Second, our main analysis has been performed at the province level. Future works might expand our work by considering data at more disaggregated levels like municipalities. Although we have done a check on this direction as a robustness check, data availability problems limit the conclusions drawn from this analysis. It seems worthwhile to examine the potential heterogeneous effects of HSR infrastructure improvements between urban and rural areas or between sun & beach and nature-based destinations.

From a methodological standpoint, HSR connectivity and accessibility is modelled using the index proposed by Xu et al. (2018), but alternative indicators could be feasible. Authors like Gutiérrez (2001), Gutiérrez et al. (2010), López et al. (2009), Moyano et al. (2018) and Ortega et al. (2014) have made much progress on this respect, and the use of travel times, gravity-based indicators, economic potential accessibility or potential accessibility dispersion indexes could offer complementary insights. On the other hand, even though Oster's method (Oster, 2019) suggests that the significant negative effect we document is robust to omitted confounders, we cannot completely rule out potential bias in the size effect through omitted time-varying province-specific shocks. Finally, related studies employ marginal decompositions of the Gini index to examine the contribution of different market segments in explaining their values (e.g., Fernández-Morales et al., 2016; Vergori and Arima, 2022). We have disregarded this approach since we are not interested in examining how the different source markets affect demand concentration but in an inferential analysis of the relationship between seasonality and HSR connectivity and accessibility. Nonetheless, this could be a valuable avenue for future research.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

David Boto-García: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Levi Pérez:** Conceptualization, Data curation, Formal analysis, Investigation, Writing – review & editing.

Declaration of Competing Interest

None.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jtrangeo.2023.103546.

References

Albalate, D., Fageda, X., 2016. High speed rail and tourism: empirical evidence from Spain. Transp. Res. A 85, 174–185.
Albalate, D., Bel, G., Fageda, X., 2012. Beyond the efficiency-equity dilemma: centralization as a determinant of government investment in infrastructure. Pap. Reg. Sci. 91 (3), 599–615.
Albalate, D., Bel, G., Fageda, X., 2015. When supply travels far beyond demand: causes of oversupply in Spain's transport infrastructure. Transp. Policy 41, 80–89.
Albalate, D., Campos, J., Jiménez, J.L., 2017. Tourism and high speed rail in Spain: does the AVE increase local visitors? Ann. Tour. Res. 65, 71–82.

Albalate, D., Campos, J., Jiménez, J.L., 2022. Local tourism effects on HSR in small cities: three synthetic control case studies. Curr. Issue Tour. https://doi.org/10.1080/13683500.2022.2086106.
Becken, S., Achiff, A., 2011. Distance models for New Zealand international tourists and the role of transport prices. J. Travel Res. 50 (3), 303–320.
Bo, Z., Ningqiao, L., 2018. The impact of high-speed trains on regional tourism economies: empirical evidence from China. Tour. Econ. 24 (2), 187–203.
Campa, J.L., López-Lambas, M.E., Guirao, B., 2016. High speed rail effects on tourism: Spanish empirical evidence derived from China's modelling experience. J. Transp. Geogr. 57, 44–54.
Campa, J.L., Arce, R., López-Lambas, M.E., Guirao, B., 2018. Can HST improve the mobility of international tourists visiting Spain? Territorial evidence derived from the Spanish experience. J. Transp. Geogr. 73, 94–107.
Caponi, V., 2022. The economic and environmental effects of seasonality of tourism: a look at solid waste. Ecol. Econ. 192, 107262.
Capó-Parrilla, J., Riera-Font, A., Rosselló-Nadal, J., 2007. Accommodation determinants of seasonal patterns. Ann. Tour. Res. 34 (2), 422–436.
Castillo-Manzano, J.I., Castro-Nuño, M., López-Valpuesta, L., Pedregal-Tercero, D.J., Garrido-Michó, J.M., 2018. High speed rail: fast tracking tourism in the EU? Ann. Tour. Res. 71, 64–66.
Cerulli, G., 2015. *Ctreatreg*: command for fitting dose-response models under exogenous and endogenous treatment. Stata J. 15 (4), 1019–1045.
Cerulli, G., Ventura, M., 2021. A dose-response approach to evaluate the effects of different levels of partial credit guarantees. Appl. Econ. 53 (12), 1418–1434.
Chen, Z., Haynes, K.E., 2015. Impact of high-speed rail on international tourism demand in China. Appl. Econ. Lett. 22 (1), 57–60.
Chung, J.Y., Whang, T., 2011. The impact of low-cost carriers on Korean Island tourism. J. Transp. Geogr. 19, 1335–1340.
Cisneros-Martínez, J.D., Fernández-Morales, A., 2013. Cultural tourism as tourist segment for reducing seasonality in a coastal area: the case study of Andalusia. Curr. Issue Tour. 18 (8), 765–784.
Cisneros-Martínez, J.D., McCabe, S., Fernández-Morales, A., 2018. The contribution of social tourism to sustainable tourism: a case study of seasonally adjusted programmes in Spain. J. Sustain. Tour. 26 (1), 85–107.
Coshall, J., Charlesworth, R., Page, S.J., 2015. Seasonality of overseas tourism demand in Scotland: a regional analysis. Reg. Stud. 49 (10), 1603–1620.
Crescenci, R., Rodríguez-Pose, A., 2012. Infrastructure and regional growth in the European Union. Pap. Reg. Sci. 91 (3), 487–513.
Cusimano, A., Mazzola, F., Barde, S., 2021. Place-based policy in southern Italy: evidence from a dose-response approach. Reg. Stud. 55 (8), 1442–1458.
de Chaisemartin, C., D'Haultfœuille, X., 2020. Two-way fixed effects estimators with heterogeneous treatment effects. Am. Econ. Rev. 111 (9), 2964–2996.
Deng, T., Gan, C., Hu, Y., 2021. Do hotel business benefit from increased tourist accessibility? Evidence from China's high-speed railway program. Tour. Econ. 27 (7), 1357–1374.
Doerr, L., Dorn, F., Gaebler, S., Potrafke, N., 2020. How new airport infrastructure promotes tourism: evidence from a synthetic control approach in German regions. Reg. Stud. 54 (10), 1402–1412.
Donzelli, M., 2010. The effect of low-cost air transportation on the local economy: evidence from southern Italy. J. Air Transp. Manag. 16, 121–126.
Duro, J.A., 2016. Seasonality of hotel demand in the main Spanish provinces: measurements and decomposition exercises. Tour. Manag. 52, 52–63.
Duro, J.A., 2018. Seasonality of tourism: a new decomposition. Tour. Econ. 24 (5), 615–621.
Duro, J.A., Turrión-Prats, J., 2019. Tourism seasonality worldwide. Tour. Manag. Perspect. 31, 38–53.
Farhadi, M., 2015. Transport infrastructure and long-run economic growth in OECD countries. Transp. Res. A 74, 73–90.
Fernández-Morales, A., 2003. Decomposing seasonal concentration. Ann. Tour. Res. 30 (4), 942–956.
Fernández-Morales, A., Cisneros-Martínez, J.D., 2019. Seasonal concentration decomposition of cruise tourism demand in southern Europe. J. Travel Res. 58 (8), 1389–1407.
Fernández-Morales, A., Mayorga-Toledano, M.C., 2008. Seasonal concentration of the hotel demand in Costa del Sol: a decomposition by nationalities. Tour. Manag. 29, 940–949.
Fernández-Morales, A., Cisneros-Martínez, J.D., McCabe, S., 2016. Seasonal concentration of tourism demand: decomposition analysis and marketing implications. Tour. Manag. 56, 172–190.
Ferrante, M., Lo Magno, G.L., De Cantis, S., 2018. Measuring tourism seasonality across European countries. Tour. Manag. 68, 220–235.
Filippetti, A., Cerulli, G., 2018. Are local public services better delivered in more autonomous regions? Evidence from European regions using a dose-response approach. Pap. Reg. Sci. 97 (3), 801–826.
Gao, Y., Su, W., Wang, K., 2019. Does high-speed rail boost tourism growth? New evidence from China. Tour. Manag. 72, 220–231.
Gutiérrez, J., 2001. Location, economic potential and daily accessibility: an analysis of the accessibility impact of the high-speed line Madrid-Barcelona-French border. J. Transp. Geogr. 9, 229–242.
Gutiérrez, J., Condeço-Melhordo, A., Martín, J.C., 2010. Using accessibility indicators and GIS to assess spatial spillovers of transport infrastructure investment. J. Transp. Geogr. 18, 141–152.
Hirano, K., Imbens, G.W., 2004. The propensity score with continuous treatments. In: Gelman, A., Meng, X.-L. (Eds.), Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives. Wiley, Chichester, UK, pp. 73–84.
Holl, A., 2004. Manufacturing location and impacts of road transport infrastructure: empirical evidence from Spain. Reg. Sci. Urban Econ. 34, 341–363.
Jiménez, J.L., Ortuño, A., Pérez-Rodríguez, J.V., 2022. How does Airbnb affect local Spanish tourism markets? Empir. Econ. 62 (5), 2515–2545.
Jin, S., Yang, J., Wang, E., Liu, J., 2020. The influence of high-speed rail on ice-snow tourism in northeastern China. Tour. Manag. 78, 104070.
Khadaroo, J., Seetanah, B., 2008. The role of transport infrastructure in international tourism development: a gravity model approach. Tour. Manag. 29, 831–840.
Krakover, S., 2000. Partitioning seasonal employment in the hospitality industry. Tour. Manag. 21, 461–471.
Li, L.S.Z., Yang, F.X., Cui, C., 2019. High-speed rail and tourism in China: an urban agglomeration perspective. Int. J. Tour. Res. 21, 45–60.
López, E., Monzón, A., Ortega, E., Mancebo-Quintana, S., 2009. Assessment of cross-border spillover effects of National Transport Infrastructure Plans: an accessibility approach. Transp. Rev. 29 (4), 515–536.
Lozano, J., Rey-Maquieira, J., Sastre, F., 2021. An integrated analysis of tourism seasonality in prices and quantities, with an application to the Spanish hotel industry. J. Travel Res. 60 (7), 1581–1597.
Martín-Martín, J.M., Jiménez-Aguilera, J.D., Molina-Moreno, V., 2014. Impacts of seasonality on environmental sustainability in the tourism sector based on destination type: an application to Spain's Andalusia region. Tour. Econ. 20 (1), 123–142.
Martín-Martín, J.M., Salinas-Fernández, J.A., Rodríguez-Martín, J.A., Ostos-Rey, M.S., 2020. Analysis of tourism seasonality as a factor limiting the sustainable development of rural areas. J. Hosp. Tour. Res. 44 (1), 45–75.
Masson, S., Petiot, R., 2009. Can the high speed rail reinforce tourism attractiveness? The case of the high speed rail between Perpignan (France) and Barcelona (Spain). Technovation 29, 611–617.
Matas, A., Raymond, J.L., Roig, J.L., 2015. Wages and accesibility: the impact of transport infrastructure. Reg. Stud. 49 (7), 1236–1254.
Meleddu, M., 2014. Tourism, residents' welfare and economic choice: a literature review. J. Econ. Surv. 28 (2), 376–399.
Mikulic, J., Vizek, M., Stojcic, N., Payne, J.E., Casni, A.C., Barbic, T., 2021. The effect of tourism activity on housing affordability. Ann. Tour. Res. 90, 103264.
Moyano, A., Martínez, H.S., Coronado, J.M., 2018. From network to services: a comparative accessibility analysis of the Spanish high-speed rail system. Transp. Policy 63, 51–60.
Ortega, E., López, E., Monzón, A., 2014. Territorial cohesion impacts of high-speed rail under different zoning systems. J. Transp. Geogr. 34, 16–24.
Oster, E., 2019. Unobservable selection and coefficient stability: theory and evidence. J. Bus. Econ. Stat. 37 (2), 187–204.
Pagliara, F., Mauriello, F., Garofalo, A., 2017. Exploring the interdependences between high speed rail systems and tourism: some evidence from Italy. Transp. Res. A 106, 300–308.
Papatheodorou, A., 2001. Why people travel to different places. Ann. Tour. Res. 28 (1), 164–179.
Prideaux, B., 2000. The role of the transport system in destination development. Tour. Manag. 21, 53–63.
Prifti, E., Estruch, E., Daidone, S., Davis, B., 2019. How much is too much: does the size of income support transfers affect labor supply? J. Policy Model 41, 179–196.
Rodrigue, J.P., 2020. The Geography of Transport Systems, 5th ed. Routledge.
Rosselló, J., Sansó, A., 2017. Yearly, monthly and weekly seasonality of tourism demand: a decomposition analysis. Tour. Manag. 60, 379–389.
Rosselló, J., Riera-Font, A., Sansó-Rosselló, A., 2004. The economic determinants of seasonal patterns. Ann. Tour. Res. 31 (3), 697–711.
Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. J. Educ. Psychol. 66, 688–701.
Sainaghi, R., Mauri, A., d'Angella, F., 2019. Decomposing seasonality in an urban destination: the case of Milan. Curr. Issue Tour. 22 (16), 1919–1924.
Saito, H., Romao, J., 2018. Seasonality and regional productivity in the Spanish accommodation sector. Tour. Manag. 69, 180–188.
Stojcic, N., Mikulic, J., Vizek, M., 2022. High season, low growth: the impact of tourism seasonality and vulnerability to tourism on the emergence of high-growth firms. Tour. Manag. 89, 104455.
Sun, Y.Y., Lin, Z.W., 2018. Move fast, travel slow: the influence of high-speed rail on tourism in Taiwan. J. Sustain. Tour. 26 (3), 433–450.
Taaffe, E., Gauthier, H., Kelly, M., 1996. Geography of Transportation, Second edition. Prentice Hall, Upper Saddle River, NJ, US.
Tian, F., Yang, Y., Jiang, L., 2022. Spatial spillover of transport improvement on tourism growth. Tour. Econ. 28 (5), 1416–1432.
Turrión-Prats, J., Duro, J.A., 2017. Tourist seasonality in Catalonia: the relevance of demand factors. Tour. Econ. 23 (4), 846–853.
Turrión-Prats, J., Duro, J.A., 2018. Tourist seasonality and the role of markets. J. Destin. Mark. Manag. 8, 23–31.
Turrión-Prats, J., Duro, J.A., 2019. Seasonality determinants for the main markets in Spain. Int. J. Tour. Res. 21, 462–477.
Van Can, V., 2013. Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model. Transp. Res. A 49, 149–159.
Vergori, A.S., Arima, S., 2022. Transport modes and tourism seasonality in Italy: by air or by road? Tour. Econ. 28 (3), 583–598.
Wang, X., Huang, S., Zou, T., Yan, H., 2012. Effects of the high-speed rail network on China's regional tourism development. Tour. Manag. Perspect. 1, 34–38.
Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*. Chap. 18. MIT Press, Cambridge.
Xu, W., Zhou, J., Yang, L., Li, L., 2018. The implications of high-speed rail for Chinese cities: connectivity and accessibility. Transp. Res. A 116, 308–326.

Yan, Y.Q., Zhang, H.Q., Ye, S.H., 2014. Assessing the impacts of the high-speed train on tourism demand in China. Tour. Econ. 20 (1), 157–169.
Yang, Z., Li, T., 2020. Does high-speed rail boost urban tourism economy in China? Curr. Issue Tour. 23 (16), 1973–1989.
Yao, S., Yan, X., Lei, C.K., Wang, F., 2022. High-speed railway and tourism development in China. Tour. Econ. 28 (6), 1520–1544.
Zhang, A., Liu, L., Liu, G., 2020. High-speed rail, tourism mobility, and firm value. Econ. Model. 90, 108–116.
Zhang, D., Xie, J., Sikveland, M., 2021. Tourism seasonality and hotel firms' financial performance: evidence from Norway. Curr. Issue Tour. 24 (21), 3021–3039.
Zhou, B., Wen, Z., Yang, Y., 2021. Agglomerating or dispersing? Spatial effects of high-speed trains on regional tourism economies. Tour. Manag. 87, 104392.
Zhou, B., Wen, Z., Sutherland, I., Lee, S.K., 2022. The spatial heterogeneity and dynamics of tourism-flow spillover effect: the role of high-speed train in China. Tour. Econ. 28 (2), 300–324.