



# Is peer-to-peer demand cointegrated at the listing level?

Jorge V. Pérez-Rodríguez<sup>1</sup> · Heiko Rachinger<sup>2</sup> · Rafael Suárez-Vega<sup>1</sup>

Received: 28 December 2022 / Accepted: 20 October 2023  
© The Author(s) 2023

## Abstract

In this paper, we analyse the long-run equilibrium demand of the peer-to-peer sharing economy. Our panel data demand model relates occupancy rates to relative prices of Airbnb and HomeAway listings, prices of competitors (hotels and apartments) and a proxy for income of tourists visiting the Canary Islands (Spain). We use a fractional heterogeneous panel data model which allows for a more general persistence and cointegration relationship and incorporates individual and interactive fixed effects. We find some evidence for (fractional) cointegration in P2P at the listing level. Regarding elasticities, classic cointegration methods give larger estimates for individual slopes and mean group coefficients than the fractional integrated heterogeneous model. Finally, own-price elasticities are inelastic, and the cross-price elasticity indicates that P2P listings and hotels are substitute goods. Income elasticity is lower than 1 and is not statistically significant, indicating that the demand for tourism in the Canary Islands is not sensitive to the economic conditions in the origin countries.

**Keywords** Demand modelling · Peer-to-peer accommodation · Cointegration · Fractionally integrated heterogeneous panel data analysis

## 1 Introduction

Economic theory suggests that tourism demand is influenced by different price (e.g., product price, prices of other products) and non-price factors (e.g., income, size of the market, advertising and promotion, tastes, among others) (see Song and Witt 2000, and Dwyer et al. 2010, for an overview over tourism demand modelling). Therefore,

---

✉ Jorge V. Pérez-Rodríguez  
jv.perez-rodriguez@ulpgc.es

<sup>1</sup> Department of Quantitative Methods for Economics and Management, University of Las Palmas de Gran Canaria, Las Palmas, Canary Islands, Spain

<sup>2</sup> Department of Applied Economics, University of the Balearic Islands, Palma de Mallorca, Balearic Islands, Spain

estimated tourism demand elasticities (e.g., own-price, cross-prices, income, marketing, among others) are useful not only for tourism managers to conduct an insightful demand sensitivity analysis based on prices and income, but also for short- and long-term policy recommendations.

To improve the understanding of short- and long-run properties of tourism data and to provide a framework in which hypotheses regarding growth and fluctuations in tourism can be tested, cointegration analysis has been frequently used to assess the degree of interdependence of tourism markets and tourism demand and its determinants in the long-term (e.g., Dritsakis 2004; Narayan 2004; Bonham et al. 2006; or Massidda and Mattana 2013; among others).

Particularly, the tourism demand analysis distinguishes between the demand for travel to a destination (e.g., visitor arrivals or expenditure) and the demand for related products or services (e.g., hotels, peer-to-peer accommodation) using time series analysis. In the first case, the empirical literature on the tourism industry has used cointegration methods to analyse the long-run equilibrium between tourism demand using tourism arrivals and several determinants such as the relative price of destination compared to the country of origin of tourists, the relative price for substitute destinations, and real income, among others (e.g., Webber 2001; Lim and McAleer 2001; Li et al. 2004, 2005; Dritsakis 2004). In addition, there are papers using panel data approaches such as Seetanah et al. (2010), Çalışkan et al. (2019) or Dogru et al. (2021). In general, results have shown that demand is cointegrated with its classic determinants both in the time series and panel data framework (see the literature review section below for an overview). In the second case, there are also studies focusing on the analysis of the long-run equilibrium link between tourism demand for products and services and their explanatory factors such for hotel demand (e.g., Lee 2010; Bianchi and Chen 2020; Lee and How 2019, 2022a, 2022b; among others) or restaurants (e.g., Herrington and Bosworth 2016), using time series models. However, there are no studies of tourism demand in a panel data context investigating the long-term relationship in both hotel or P2P accommodation markets, except for Gómez et al. (2021)'s analysis of the relationship between Airbnb and hotels in Mexico.

Panel data are a useful approach to investigate observations for different individuals at different moments in time (Baltagi 2001). The panel data framework allows us to consider a cross-section of listings along time which accounts for unobserved time-invariant heterogeneity. Therefore, this paper's main contribution to the empirical literature is the analysis of the long-run equilibrium relationship between P2P tourism demand and its determinants at the listing level; that is using microeconomic information for active listings in the P2P accommodation market in a panel data framework which allows us to analyse the economic behaviour of cross-section of listings over time.

More specifically, we consider the heterogeneity of listings' demand in a panel data context. Demand of listings in P2P accommodation markets has been studied by Gunter et al. (2020), Jiménez et al. (2023), and Suárez-Vega et al. (2023). These papers have studied the relevance of elasticities of demand, cross-prices and income using spatial panel data methods. They have reported that the classical accommodation market (hotels) and peer-to-peer (P2P) accommodation market such as Airbnb

are substitute goods (Günter et al. 2020; Jiménez et al. 2023), and therefore, competitors. Hence, if they increase prices disproportionately, the business might lose sales and revenues as customers transfer to local competitors, including neighbours' listings (Gunter et al. 2020). However, they do not account for unobserved dependence between cross-sectional units (e.g., between neighbour properties and, therefore, between their demands). Therefore, we use two heterogeneous panel data cointegration models based on cross-sectional dependence: first, the standard panel cointegration model with a multi-factor error structure, as proposed by Pesaran (2006), based on the common correlated effects (CCE) estimator (which is robust to error cross-section dependence of different types, possible slope heterogeneity, and unit roots in factors), and second, the fractionally integrated heterogeneous panel data model recently proposed by Ergemen and Velasco (2017), which permits contemporaneous correlation of factors and innovations.

Our focus lies on the fractional cointegration panel data framework. In particular, the existence of fractional cointegration could indicate that the long-run equilibrium recovers slowly after deviations from it. This could have important policy implications, insofar as the transition speed to the long-run equilibrium might require the implementation of different policies. Therefore, listings' managers and tourism policymakers would have to take this slow (rather than fast) transition into account to define demand-driven adjustments to changes in prices and income in formulating their demand policies.

The database used is a balanced panel from January 2016 to September 2021 (57 months) for Airbnb listings in the Canary Islands (Spain). The data was provided by AirDNA. We use, as previously mentioned, both standard panel cointegration and a fractionally integrated heterogeneous panel data model.

The rest of the paper is organised as follows. Section 2 defines the theoretical microeconomic background. Section 3 reviews the empirical literature on panel data demand models and cointegration and describes the P2P market in the Canary Islands. Section 4 focuses on the fractional cointegration panel data methodology. In Sect. 5, we describe the variables in the database and main results. Finally, in Sect. 6 we conclude.

## 2 Theoretical background

The analysis of the preferences underlying the tourism demand at the listing level is based on the microeconomic theory of Marshallian demand assuming that tourism is an ordinary and a normal good (Gunter et al. 2020). Following Bonham et al. (2009), empirical models of tourism demand borrow heavily from consumer theory which predicts that the optimal consumption level depends on the consumer's income, the price of the tourism good, the prices of related goods (substitutes and complements), and other demand shifters.

A model of tourism demand at the listing level to be estimated and tested can be written as:

$$D_i = f\left(Y_i, P_i, P_{ik}^l, P_i^h, Z\right), \quad (1)$$

where  $D_i$  is the tourism demand of the  $i$ -th listing;  $Y_i$  is the income of the visitors at the  $i$ -th listing;  $P_i$  denotes the price of  $i$ -th listing;  $P_{ik}^l$  is the price of others listings near  $i$ -th listing at the  $k$ -th radius;  $P_i^h$  is the price of the tourism product in the competing accommodation market for the  $i$ -th listing, for example, hotels ( $h$ ); and  $Z$  are other tourism demand factors which can influence on tourism demand (e.g., tastes, marketing, among others).

Marshallian demand expresses demand as the quantity vector which maximises an individual customer's utility subject to their income, as well as the prevailing price vector. In this context, there is an inverse relationship between own price and demand (ordinary good), income has a positive effect on the demand and there is a substitution effect (that is tourism is a normal good). The algebraic signs for own-price, cross-prices and income elasticities are negative, positive and positive, respectively.

Estimation of price elasticity of demand is important for tourism managers seeking to maximize sale revenues. Knowledge of the income elasticity can help managers identify potential markets for their products. Cross-price elasticity can help tourism managers determine if their products have substitutes, complements or are independent. This information is essential for formulating pricing strategy and analysis of risk associated with various products.

Demand is under homogeneity a function of real income and relative destination and substitute prices,

$$D_{ij} = f\left(Y_i / P_i, P_{ik}^l / P_i, P_i^h / P_i, Z\right),$$

and will be estimated in the empirical demand modelling.

### 3 Literature review

#### 3.1 Demand models in a panel data context

There is a large empirical literature on tourism demand for several countries and regions using both static and dynamic panel data models, in both standard and spatial contexts.

Using panel data, several authors have analysed the economic determinants of travel demand. For example, Sakai et al. (2000) considered the effects of Japanese demographic change, while Garín-Muñoz and Amaral (2000) considered the number of tourists, income per capita, exchange rate and real prices for Spain. For their part, Ledesma-Rodríguez et al. (2001) analysed demand for the Spanish island of Tenerife, Naudé and Saayman (2005) for tourist arrivals in 43 African countries, while Roget and González (2006) investigated rural tourism demand in Galicia (Spain). In general, the factors influencing demand that have been analysed include prices of tourism products at the destination relative to the origin (relative prices), the prices of tourism products in competing destinations (substitute prices), tourist's income levels, and exchange rates (Li et al. 2005; Song and Li 2008). Other important determinants include marketing policies Ledesma-Rodríguez et al. (2001), attacks on tourists (Bonham et al. 2006),

financial crises (Song et al. 2011), climate change (Moore 2010), and political stability (Saha and Yap 2014).

Several papers modelling tourism demand are based on dynamic panel data models with lagged demand as an explanatory variable (e.g., Garín-Muñoz 2006; Garín-Muñoz and Montero-Martín, 2007). The main intuition of dynamics in tourism demand is the occurrence of repeat visits, word-of-mouth recommendations, time lags in implementing a decision, information asymmetry, supply rigidities, and long-term adjustments (see Morley 2009; Song et al. 2012). Tourism demand is most commonly represented by tourism arrivals (e.g., Garín-Muñoz 2006; Naudé and Saayman 2005; Seetaram 2010), but also by tourism expenditure (e.g., Li et al. 2004; Lyssiotou 2000; Wu et al. 2012) and occupancy rates (Jiménez et al. 2022). These have obtained mixed estimated coefficients of the lagged dependent variable: Some are highly significant and positive (Song and Witt 2003), suggesting a strong positive willingness to return and/or a high word-of-mouth effect. However, others are negative (Naudé and Saayman 2005), suggesting a negative willingness. Further studies have even indicated that the lagged dependent variable is the principal tourism demand determinants (Song et al. 2010; Song and Witt 2003).

Only Jiménez et al. (2022) have studied the habit persistence effect in the P2P accommodation sector. They use a dynamic panel data model at the Spanish city level for 2014–2017 to analyse the effects of Airbnb on the size of local tourism markets and find a positive habit persistence effect for both Airbnb occupancy rates and for hotel overnight stays.

Finally, tourism demand research has also been conducted with respect to the spatial dimension, considering the relationships between distinct units due to their locations and the tourism demand determinants of units of interest, such as destinations, regions, or lodging units, over time. This literature is based on spatial spillover effects (i.e., a higher mutual influence of units closer together compared to those farther apart).

Papers considering the spatial dimension have studied the effect of several determinants, such as inbound tourism, prices, and supply, among others. Inbound tourism has been analysed for Australia (Deng and Athanasopoulos 2011) and for China (e.g., Zhang 2009; Yang and Wong 2012; Ma et al. 2015; Yang and Zhang 2019). With respect to online P2P lodging markets such as Airbnb, price determinants have been studied for the United Kingdom (Voltes-Dorta and Inchausti-Sintes 2021), Estonia (Önder et al. 2019) and Spain (Gutiérrez et al. 2017; Adamiak et al. 2019; Eugenio-Martín et al. 2019; Boto-García et al. 2021). Gunter et al. (2020) is the first to analyse Airbnb demand in a spatial econometric model with occupancy rates as dependent variable. Among other results, these authors have found that Airbnb demand in New York City is price inelastic and that the city's traditional lodging businesses and nearby Airbnb listings are substitutes for the studied Airbnb listings. More recently, Suárez-Vega et al. (2023) proposed a dynamic spatial panel data model to estimate direct, indirect, and total marginal effects of several variables on tourism demand.

## 3.2 Demand models and cointegration

Tourism demand studies have also used time series econometric models to investigate cointegration between tourism demand and its determinants (see Song et al., 2019, for an overview). Such studies analyse tourism demand by testing for cointegration and estimating the cointegrating vectors, via error correction models (ECM).

### 3.2.1 Time series models and cointegration

Lathiras and Sriopoulos (1998) identified economic factors that were interrelated in the long-run with tourism demand and identified a statistically adequate short-run dynamic specification with appropriate forecasting properties. Kim and Song (1998) used ECMs to study the long- and short-term inbound tourism demand in South Korea with respect to Japan, US, UK, and Germany, as the four main tourist origin countries. Salman (2003) used the Engle-Granger cointegration approach to estimate the long-run relationship between monthly tourist flows to Sweden and income, price, and exchange rate. Kadir and Karim (2009) analysed cointegration and used error correction models to estimate a tourism demand model for Malaysia with respect to US and UK tourists. They obtained a long-run relationship between tourist arrivals and several explanatory factors and have found that US and UK tourist arrivals in Malaysia are significantly affected by income and the relative price of tourism.

More recently, Husein and Kara (2020) estimated the long-run tourism demand for Puerto Rico from the USA for 1970–2016. They accounted for the asymmetric impact of income changes on tourism demand, because of potentially asymmetric income elasticity over business cycles. In particular, they explored asymmetric cointegration with the nonlinear autoregressive distributed lag (ARDL) model of Shin et al. (2014).

Several papers used Johansen's maximum likelihood method to test for cointegration and to estimate the cointegrating vectors within a vector autoregressive (VAR) model (e.g., Webber 2001, Lim and McAleer 2001, Dritsakis 2004, Bonham et al. 2006, 2009, or Massidda and Mattana 2013). Dritsakis (2004), for example, investigated changes in the long-term demand for tourism of Germans and Britons to Greece. He explained the demand for tourism with tourism prices in Greece, income in origin countries (Germany and Great Britain), and transportation cost and exchange rates between the three countries. Bonham et al. (2006) used vector error correction models (VECM) to make dynamic visitor predictions, in their specific case to determine whether tourism in Hawaii had completely recovered from 9/11 and other tragic international events. Their paper also considered policy alternatives to facilitate the recovery of international tourism to the US. Bonham et al. (2009) also estimated a fully identified VECM of Hawaii tourism, with relevant demand and supply-side effects. They identified reasonable long-run equilibrium relationships and demonstrated satisfactory forecasting performance using Diebold–Mariano forecast-accuracy tests.

Massidda and Mattana (2013) used a structural VECM for the Italian economy to investigate the contemporaneous and short- and long-run relationships across per capita international tourism arrivals, real GDP, and total international commercial transactions. They found bidirectional causality between real GDP and tourism arrivals

and unidirectional long-run causality running from real GDP to total international commercial transactions and from total international commercial transactions to arrivals. From a structural model and the impulse-response functions of “meaningful” shocks they obtained valuable information for the design of policies and business strategies.

On the other hand, some papers analysed the long-term relationship between demand for different tourism related products. Lee (2010), for example, studied interactions between hotel room rates and the number of international inbound tourists for Singapore. He tested for the existence of any cointegrating relationship between them using the bounds testing approach to cointegration and found no evidence that Singapore’s hotel room rates and international inbound tourists are cointegrated. Also, Bianchi and Chen (2020) used a vector error correction model of hotel demand to incorporate both the short-run demand fluctuations and the long-run tourism growth. They found no evidence of a long-run market equilibrium between all three endogenous variables in the model, namely hotel nights, Swiss real gross domestic product, and real exchange rate of the Swiss franc. Also, Lee and How (2022a, b) investigated the influences of Singapore Economic Policy Uncertainty (EPU) on hotel room rates of two types of hotels, luxury and economy, in Singapore using the cointegration technique and error correction model. Their findings showed that the national EPU had a negative impact on economy hotel room rates, but no impact on luxury hotel room rates; and the global financial crisis had a negative effect on luxury hotel room rates but had no effect on economy hotel room rates. Finally, Herrington and Bosworth (2021) studied the demand for restaurants. In particular, they tested for cointegration between average unit sales and advertising and found that of the 184 individual restaurant chains examined, 135 were co-integrated.

### 3.2.2 Panel data models and cointegration

Regarding the panel data approach, Narayan (2004) used cointegration techniques and error correction models to examine the short- and long-term relationship between tourist arrivals in Fiji, real disposable incomes and relative hotel and substitute prices for the period 1970–2000. The paper used a cointegration method – the bounds testing methodology developed within the autoregressive distributed lag (ARDL) model (Pesaran et al. 1999). In the long-run, relative hotel and substitute prices affect visitor arrivals negatively, and income in the origin countries affect positively. Seetanah et al. (2010) modelled inbound tourism demand for South Africa with a gravity model. They estimated price and income sensitivity and the influence of other important determinants of tourist flows, such as socio-political factors and the location of markets. Using Pedroni’s panel cointegration estimation, they found that tourist flows are sensitive to tourism price changes in competing destinations and also to price changes in South Africa.

Çalışkan et al. (2019) analysed the link between incoming tourists and export and import volumes of Turkey and 13 Silk Road countries using panel ARDL. They found a link between international trade and tourist flows, which indicates the importance of international trade for the development of tourism.

More recently, Dogru et al. (2021) presented theoretical and methodological inaccuracies in tourism demand modelling and the use of panel data in this context. In a

guide of the use of panel data, they, among others, recommend nonlinear empirical approaches or Bayesian methods.

To our knowledge, the tourism demand long-run relationship has not been studied allowing for cross-sectional heterogeneity in a panel data context, except for the paper of Gómez et al (2021) which analyses Airbnb's influence on hotel occupancy in Mexico in a single approach with hotels grouped according to their category. However, these authors study the cross-sectional dependence, cross-sectional unit roots, and causality between these variables, but then analyse the cointegration relationship using Kao and Pedroni's cointegration tests without employing recent advances such as the Common Correlated Effect (CCE) estimator considering the cross-sectional dependence. Given the rise of the sharing economy, tourism demand models should incorporate alternative online lodging platforms, such as Airbnb or HomeAway (Dogru et al. 2021), and analyse cointegration in the P2P accommodation market using panel data cointegration techniques. For this purpose, we therefore propose use of a cross-sectional dependence model allowing for slope heterogeneity in a fractional cointegration panel data framework, which extends the methodological framework of the CCE estimator.

#### 4 Econometric methodology

To estimate the tourism demand model in a panel data context, we account for cross-sectional correlation and persistence of the variables. First, we assume this persistence being governed by a unit-root. To estimate such model, we use the CCE estimator of Pesaran (2006) (see Appendix 1). Second, we assume a more general persistence governed by long-range dependence. To estimate the resulting model, we use the fractionally integrated heterogeneous panel data model proposed by Ergemen and Velasco (2017). It extends the factor structure and the CCE estimation method of Pesaran (2006) to a fractional context. Unlike Pesaran (2006), both common factors and innovations can be fractionally integrated. The resulting persistence is thus more general than a unit root. Finally, the model allows for fixed effects. We, further estimate Rodríguez-Caballero (2022) approach which extends Ergemen and Velasco (2017)'s methodology by allowing for multi-level cross-sectional dependence.

Specifically, Ergemen and Velasco (2017)'s panel data model for  $N$  individual observations and  $T$  periods is defined as:

$$\begin{aligned} y_{it} &= \alpha_i + \beta'_{i0}x_{it} + \gamma'_i f_t + \Delta_t^{d_{i0}} e_{1it}, \quad i = 1, \dots, N, t = 1, \dots, T, \\ x_{it} &= \mu_i + \Gamma'_i f_t + \Delta_t^{d_{i0}} e_{2it}, \end{aligned} \quad (2)$$

where  $y_{it}$  depends on potentially endogenous covariates  $x_{it}$  and  $m$  unobserved common factors  $f_t$ . These are fractionally integrated of order  $\omega_i$  ( $f_t \sim I(\omega_i)$ ),  $i = 1, \dots, m$ .  $y_{it}$  and  $x_{it}$  depend on these factors according to their factor loadings  $\lambda_i$  and  $\Gamma_i$  respectively. In addition to idiosyncratic covariance stationary shocks  $e_{1it}$  and  $e_{2it}$ , there are covariate-specific fixed effects  $\alpha_i$  and  $\mu_i$ . These fixed effects would cause the mentioned endogeneity if not properly accounted for.  $\Delta_t^\delta$  denotes the truncated fractional



filter defined as  $\Delta_t^\delta = \sum_{j=0}^{t-1} \pi_j(\delta)L^j$ , with  $\pi_j(\delta) = \Gamma(j - \delta) / [\Gamma(j + 1)\Gamma(-\delta)]$ ,  $\Gamma(\tau) = \infty$ ,  $\tau = 0, -1, -2, \dots$ , but  $\Gamma(0) / \Gamma(0) = 1$ , and  $L$  denoting the lag operator.

The memory parameters of interest are the residual integration order,  $d_{i0}$ , and the memory of the defactored (unobserved) explanatory variable,  $\vartheta_{i0}$ . It is worth noting that the memory of  $y_{it}$  is  $\max\{\vartheta_{i0}, d_{i0}, \max_i \omega_i\}$  and that of  $x_{it}$  is  $\max\{\vartheta_{i0}, \max_i \omega_i\}$ . Besides  $\vartheta_{i0}$  and  $d_{i0}$ , we are interested in the potential fractional cointegrating vector  $\beta'_{i0}$ . The cointegrating error  $y_{it} - \alpha_i - \beta'_{i0}x_{it} - \lambda_i f_t$  is asymptotically stationary for  $d_{i0} < 0.5$  and asymptotically nonstationary for  $d_{i0} \geq 0.5$ . (Fractional) cointegration requires  $d_{i0} < \vartheta_{i0}$  which can be tested with the  $t$ -test,  $t = (\hat{\vartheta}_{i0} - \hat{d}_{i0}) / s.e.(\hat{\vartheta}_{i0} - \hat{d}_{i0})$ , where  $s.e.(\hat{\vartheta}_{i0} - \hat{d}_{i0})$  is the standard error of the difference of memory estimates.

To estimate (2), fixed effects are removed by first differences:

$$\begin{aligned} \Delta y_{it} &= \beta'_{i0} \Delta x_{it} + \gamma'_{i0} \Delta f_t + \Delta_t^{1-d_{i0}} e_{1it}, \quad i = 1, \dots, N, t = 1, \dots, T, \\ \Delta x_{it} &= \Gamma'_i \Delta f_t + \Delta_t^{1-\vartheta_{i0}} e_{2it}. \end{aligned}$$

After projecting the individual time series on their (fractionally) differenced cross-section averages, the heterogeneous slope and memory parameters are estimated by generalized least squares. Finally, an equation-by-equation conditional-sum-of-squares approach renders estimates of the relevant memory parameters. Note that, contrary to the unit root case, estimates and test statistics have standard properties.

The common-correlation mean-group (CCMG) estimate is obtained as average of the individual slope coefficients  $\hat{\beta}_{i0}(\hat{d}_i, \hat{\vartheta}_i)$  depending on the individual memory parameters  $\hat{d}$  and  $\hat{\vartheta}$ :

$$\hat{\beta}_{CCMG}(\hat{d}, \hat{\vartheta}) = \left[ \frac{1}{N} \sum_{i=1}^N \hat{\beta}_{i0}(\hat{d}_i, \hat{\vartheta}_i) \right].$$

The hypothesis  $\beta_{CCMG}(d, \vartheta) = \beta_0$  can then be tested by the  $t$ -test:

$$t_{CCMG} = \sqrt{N} \frac{(\hat{\beta}_{CCMG}(\hat{d}, \hat{\vartheta}) - \beta_0)}{\hat{\Omega}_w(\hat{d}, \hat{\vartheta})^{1/2}},$$

where  $\hat{\Omega}_w(\hat{d}, \hat{\vartheta})$  is an estimate of the asymptotic variance–covariance matrix:

$$\hat{\Omega}_w(\hat{d}, \hat{\vartheta}) = \frac{1}{N-1} \sum_{i=1}^N (\hat{\beta}_{i0}(\hat{d}_i, \hat{\vartheta}_i) - \hat{\beta}_{CCMG}(\hat{d}, \hat{\vartheta})) \times (\hat{\beta}_{i0}(\hat{d}_i, \hat{\vartheta}_i) - \hat{\beta}_{CCMG}(\hat{d}, \hat{\vartheta}))'.$$

Rodríguez-Caballero (2022) extends Ergemen and Velasco (2017)’s methodology by dividing the panel into blocks and allowing for different factors in each of these blocks (in addition to general factors for the whole panel). This leads to a multi-level cross-sectional dependence. In particular, (2) changes to:

$$y_{r,it} = \alpha_{r,i} + \beta'_{r,i0} x_{r,it} + \kappa'_i g_t + \gamma'_{r,i} f_{r,t} + \Delta_t^{d_{r,i0}} e_{1r,it}, \quad i = 1, \dots, N_r, t = 1, \dots, T, r = 1, \dots, R,$$

$$x_{r,it} = \mu_{r,i} + K_i' g_t + \Gamma_{r,i}' f_{r,t} + \Delta_i^{\vartheta_{r,i0}} e_{2r,it},$$

where  $R$  denotes the number of blocks,  $N_r$  the number of units in block  $r$ , and where all variables, factors, and parameters from (2) are now block  $r$  specific. In addition to the block-specific factors  $f_{r,t}$ , there are pervasive factors  $g_t$  that affect all blocks.  $\kappa_i$  and  $K_i$  are their respective factor loadings.

The author highlights that neglecting the presence of this non-pervasive cross-sectional dependence may cause possible biases. To estimate the model, it suffices to apply Ergemen and Velasco (2017)'s methodology to each of the blocks separately.

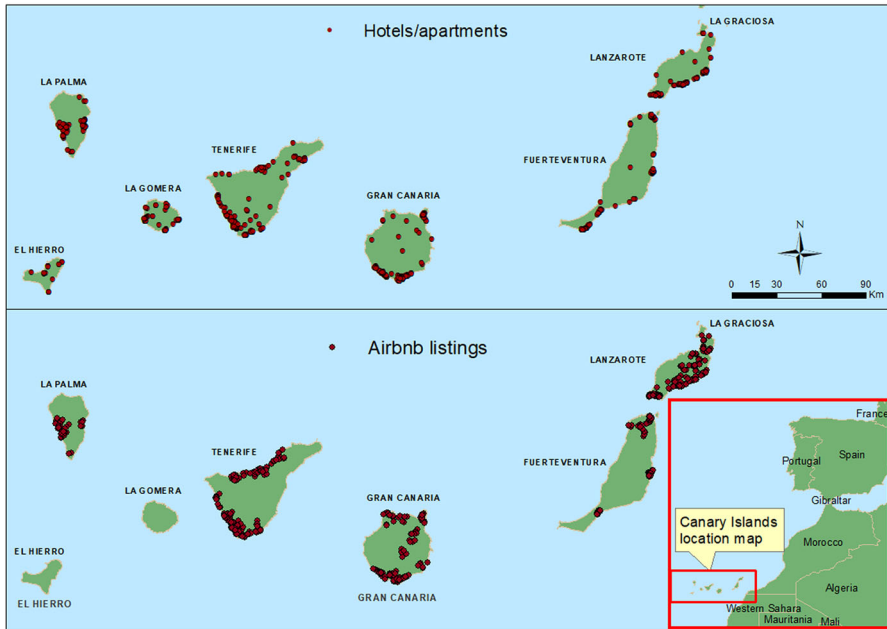
Before estimating these panel models with an underlying factor structure, we motivate this methodology by applying Bailey et al. (2016)'s Exponent of Cross-Sectional Dependence test. This allows estimating the degree of dependence  $\alpha$  between the different lodgings. Only for strong or semi-strong dependence, i.e. for  $\alpha > 0.5$ , a factor structure is sensible (see Bailey et al 2016, for further details and Chakraborty and Mazzanti (2020) and Rodríguez-Caballero and Vera-Valdés, 2021, for applications).

## 5 Empirical analysis

### 5.1 Data and variables

The Canary Islands is a Spanish archipelago located 2000 km from Peninsular Spain in the Atlantic Ocean. It has been a popular holiday destination for Europeans – especially UK, Germany, and mainland Spain—since the 1960s, with more than 15 million visitors in the pre-pandemic year 2019. The tourism is varied: sandy beaches, nature (including three National Parks) and urban areas, especially, in the two main islands Gran Canaria and Tenerife. Its sharing accommodation market has grown rapidly over the last decade, with above 100,000 listings in the Airbnb and HomeAway online platforms. These include apartments (most abundant type with over 50% of the total), houses, villas (the highest priced type of accommodation), condominiums (jointly with apartments the lowest in average price), bungalows, cottages, and townhouses. These types of accommodation represent more than 90% of the total and are managed by hosts both with one and with multiple properties with the former obtaining lower revenue per property. An increasing number of properties per host has continuously increased the revenue. Since the 2010s, the sharing accommodation sector's market share has continuously increased, with around 8.6% in 2016 and 11.9% in 2020 (ISTAC 2021).

This paper uses monthly observations of Airbnb properties in the Canary Islands (Spain) from January 2017 to September 2021. We obtain a balanced panel by only considering properties with data available for the whole period. For reasons of consistency, only establishments classified as “Entire home/apt” have been included in the sample. In this group of listings, the three most frequent types were considered: Villa (49.44%), Apartment (41.01%), and House (9.55%). The resulting database after applying these selection criteria contained 1068 listings with information over 57 months, equivalent to a total of 60,876 observations. Figure 1 illustrates the spatial distribution of the Airbnb listings in the study period in the Canary Islands and the



**Fig. 1** Spatial distribution of the sample of Airbnb listings and hotels/apartments in the Canary Islands during the study period (Source: AirDna and ISTAC)

spatial distribution of the hotels and apartments operating in 2020 in the region. The distributions of hotels and Airbnb houses are very similar, being concentrated principally in coastal areas mainly associated with beach tourism. However, the arrival of Airbnb has increased the supply of accommodation in the inland areas of the islands of Gran Canaria and Lanzarote. It should be added that there are no properties that have been operational during the entire period on the two smaller islands (La Gomera and El Hierro).

In our empirical study, we work with the following variables from the general model structure defined by Eq. (1) in Sect. 2. The most popular dependent variable in demand studies since 2000 has been visitor arrivals (Song and Li 2008), although there are other variables such as the number of tourist nights spent by tourists or the tourism expenditure by visitors in a destination which have also been used in several studies (Song and Witt 2000). In this paper, we use the occupancy rate (OCR), instead, which depends on the number of days spent, as a proxy for demand following recent papers on tourism demand such as Gunter et al. (2020), Suárez-Vega et al. (2023), and Jiménez et al. (2023). Proxies for demand determinants tend to differ substantially between studies (Song and Witt 2000; Li et al. 2005). They generally incorporate variables such as income, relative prices, substitute prices, but also factors such as travel costs, exchange rates, deterministic trends, and dummy variables for some special events.<sup>1</sup> We consider the demand for competitors in the vicinity – Airbnb listings and hotels –

<sup>1</sup> It should be noted that several potentially relevant explanatory variables have not been used. In particular, exchange rate adjusted relative prices are not used since we do not know the origin market of tourists visiting

instead. Their prices represent the costs of substitute products. Finally, the real GDP for the main tourist country visitors serves as proxy for tourist income.

The empirical demand model is based on the following variables:

- **OCR:** average monthly occupancy of Airbnb housing. This information is provided monthly by Airbnb through AirDNA (<https://www.airdna.co/>) and it is measured as the Total Booked Days/(Total Booked Days + Total Available Days). This variable varies both temporally (monthly) and spatially at the level of the listing location. Calculation only includes vacation rentals with at least one booked night.
- **RADR:** the relative average daily rate of the listing for a given month. The ADR (average daily rate) is also provided by AirDNA and represents the average price per day of rent. The RADR for a given listing in each month is then calculated as the proportion representing the own ADR with respect to the average of the ADR of the Airbnb listings in a 5 km surrounding area. Values greater than 1 imply that the own ADR exceeds the average ADR in the surrounding area. RADR has a similar temporal and spatial distribution to that of the OCR.
- **HADR:** the average daily rate of hotels in the municipality where the house is located. This information is supplied by the Canary Institute of Statistics (known by its initials in Spanish as ISTAC). The variable varies monthly (in time) and the spatial variation is at the municipal level.
- **GDP:** the average real GDP of visitors to the island where the house is located. It is calculated as the weighted average (in function of number of tourists) of the real GDPs of the seven countries that provide the most visitors to the islands (Germany, Belgium, Denmark, Great Britain, Holland, Sweden and Spain). Tourists from these countries make up 78.56% of the total number of tourists visiting the Canary Islands in the study period (January 2017–September 2021). Data concerning the number of visitors to each island by origin and month were provided by ISTAC (2021). Quarterly real GDP series (in million euros) for these countries were obtained from EUROSTAT (2021). The quarterly GDP was equally distributed among the three months that make up the quarter. Finally, the values were updated by taking 2017 as the base year and using the corresponding harmonized index of consumer prices (HICP). GDP varies spatially by island.<sup>2</sup>

The variables are described in Table 1 (Panel A). The average occupancy (OCR) for the properties in the sample is 38%, with some houses having no occupancy for a month while others are occupied the whole month. The relative average daily rate (RADR) is almost 1, which would suggest a balance of the ADR of the listings with respect to those in the surrounding area. However, there are cases where the differences are significant, reaching almost 36 times the average price supplied in the neighbourhood. Finally, the ADR for hotels in the region is €80.70, with fluctuations up to almost €153.

---

Footnote 1 continued

each listing. We also do not use transportation costs or other deterministic effects such as time trends to capture evolving consumer tastes, dummies to account for various one-off events such as oil crises, travel and foreign currency restrictions, or modifications in the data collection. These, if neglected, could lead to biased parameter estimates (e.g., Crouch et al. 1992). The factor structure should account at least partially for several of these factors. This should make their non-inclusion rather innocuous.

<sup>2</sup> Note that since in April 2020 there were no visitor data, we use the proportions of visitors for April 2019 instead.

**Table 1** Descriptive statistics for the whole and restricted sample (not seasonally adjusted)

| Variables   | Mean  | SD    | Minimum | Maximum |
|---|-------|-------|---------|---------|
| <b>Panel A:</b> Whole sample with N = 1068 listings     |       |       |         |         |
| OCR   | 0.376 | 0.362 | 0.0     | 1.0     |
| RADR  | 1.002 | 1.275 | 0.0     | 35.948  |
| HADR  | 80.70 | 19.58 | 2.35    | 152.92  |
| GDP*  | 1.473 | 0.237 | 0.845   | 2.050   |
| <b>Panel B:</b> Restricted sample with N = 199 listings |       |       |         |         |
| OCR   | 0.522 | 0.313 | 0       | 1       |
| RADR  | 1.730 | 1.503 | 0       | 16.13   |
| HADR  | 80.42 | 18.34 | 22.56   | 152.92  |
| GDP*  | 1.495 | 0.245 | 0.845   | 2.050   |

\* In 100 billion euros

Especially for OCR and RADR many values are missing. Therefore, we drop all listings in which more than 10 observations are missing in at least one of the variables. For the remaining listings, we replace missing values with 0. This reduces the number of listings to N = 199. Table 1 (Panel B) shows the descriptive statistics of this new sample. While HADR and GDP are similar, OCR and RADR are larger in the restricted sample, reflecting the fact that missing data is concentrated in these two variables.

To avoid seasonal “noise” in the standard and fractional cointegration analysis, all data are seasonally adjusted using the X-11 decomposition method initially developed by the US Census Bureau in 1965.<sup>3</sup> The X-11 method uses weighted averages over a moving window of the time series. Data for each listing in the panel was converted into a time series. The series which featured seasonality, as detected by the QS test (Maravall 2012), were replaced with their seasonally adjusted version.

As mentioned before, to further motivate the factor structure with strong dependence between the underlying variables, we estimate the cross-sectional exponent. Table 2 confirms that the dependence is either semi-strong or strong. In particular, for all four variables (both in levels and the more relevant first-differences), the 95% confidence intervals lie above 0.5. Thus, we conclude that a factor structure can account for the dependence.

## 5.2 Estimation results

In this section, we present the standard and the fractional cointegration analysis.

To obtain a linear relationship between the variables they are transformed to natural logarithms. The regression coefficients are then the price elasticity of demand of

<sup>3</sup> The seasonal adjustment was done using function *seas* belonging to the “seasonal” R package. Alternatively, we could have worked with not seasonally adjusted data since in our context the seasonality might be captured by the factor structure (see Camacho et al. (2015) for a simulation exercise on the forecasting performance of both approaches).

**Table 2** Estimation of cross-sectional exponent and test of Bailey et al (2016)

| Variables                                   | $\hat{\alpha}$ | Std. Error | [95% confidence interval] |       |
|---|----------------|------------|---------------------------|-------|
| <b>Panel A:</b> Series in levels            |                |            |                           |       |
| OCR   | 0.850          | 0.0258     | 0.800                     | 0.901 |
| RADR  | 0.939          | 0.0781     | 0.786                     | 1.092 |
| HADR  | 0.991          | 0.0603     | 0.873                     | 1.109 |
| GDP*  | 1.002          | 0.0350     | 0.933                     | 1.070 |
| <b>Panel B:</b> Series in first differences |                |            |                           |       |
| OCR   | 0.802          | 0.0154     | 0.771                     | 0.832 |
| RADR  | 0.687          | 0.0134     | 0.661                     | 0.714 |
| HADR  | 0.992          | 0.0477     | 0.899                     | 1.086 |
| GDP*  | 1.001          | 0.0393     | 0.925                     | 1.079 |

Estimates of the exponent of cross-sectional dependence for the four variables. The bias adjusted exponents are calculated using four principal components

the listing relative to the neighbour listings (RADR), price elasticity for substitute accommodation in form of hotels and apartments (HADR), and income elasticity (GDP).

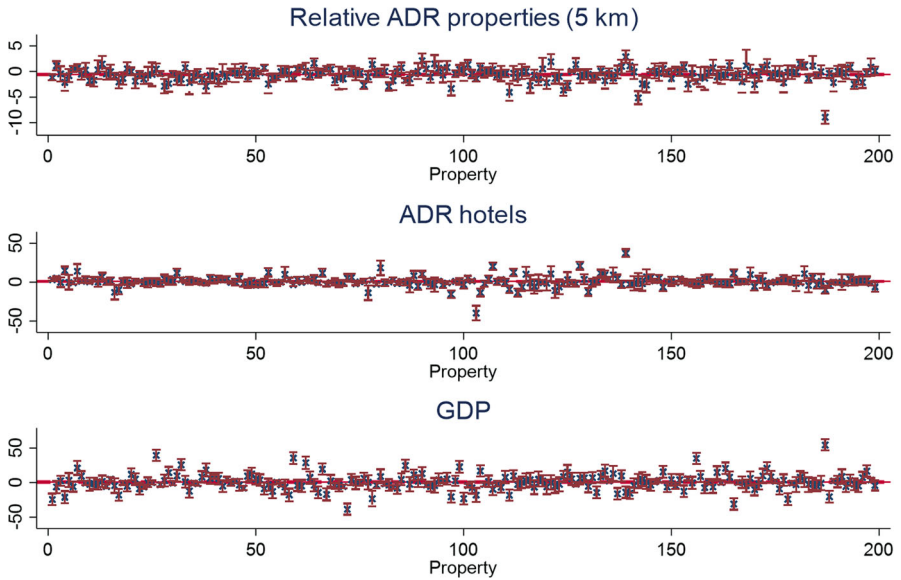
All the tests and estimations presented in this paper were done using STATA v.17 for the CCE estimator and MATLAB R2018b for the fractional integration and cointegration models. It should be noted that the CCE results are obtained with three cross-sectional lags. Also, the number of variables in the mean group regression is 597 and the number of variables partialled out is 3383 in the CCE method.

Next, we briefly describe the individual slope and mean group estimates.

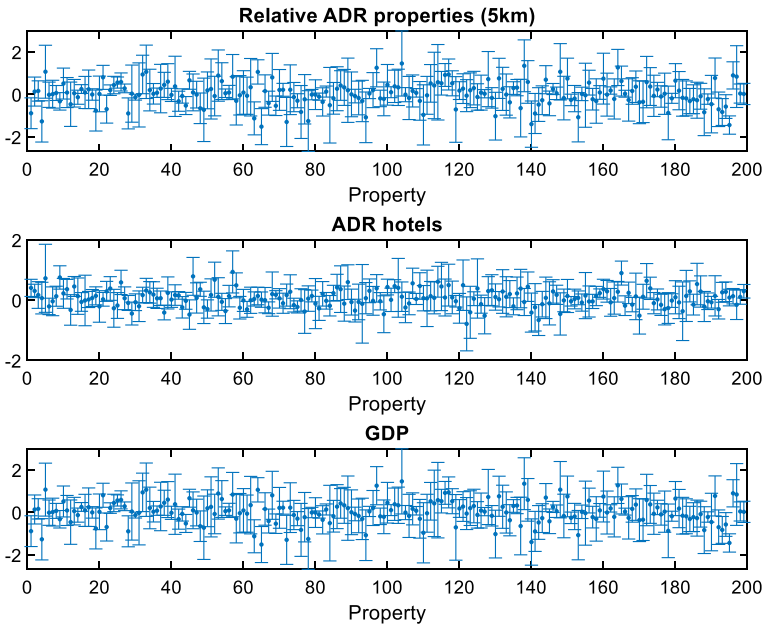
### 5.2.1 Individual slope estimates and fractional cointegration analysis

Because the number of listings is high ( $N = 199$ ), for the sake of brevity we do not show the estimated coefficients for all properties. Instead, we plot the individual CCE and fractional integrated heterogeneous panel data estimates in Figs. 2 and 3, respectively. These figures include point estimates and their 95% confidence intervals for each variable: the relative ADR of properties in a radius of 5 km, the ADR of hotels in the municipal demarcation of the listings and GDP.

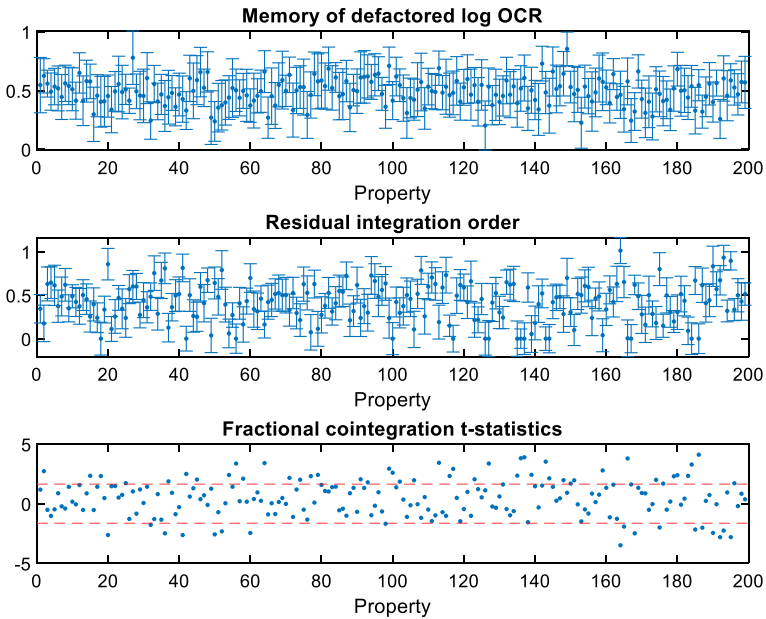
In both cases, the estimates of the three slope coefficients are quite variable, with positive and negative estimates for different listings. For example, 40 listings feature negative RADR coefficients, while 17 feature positive HADR coefficients and 8 positive GDP coefficients, all statistically significant (at the 5% significant level). On the other hand, we observe that the sample variability of the CCE estimates is higher than that of the fractional integrated heterogeneous panel data estimates. In the latter case, 98 listings have a negative RADR coefficient, and 25 and 17 listings have positive coefficients for HADR and GDP, respectively, all statistically significant at the 5% significance level.



**Fig. 2** Individual slope estimates and confidence intervals from CCE. *Note.* Individual slope estimates (together with the corresponding confidence intervals at the 95% confidence level)

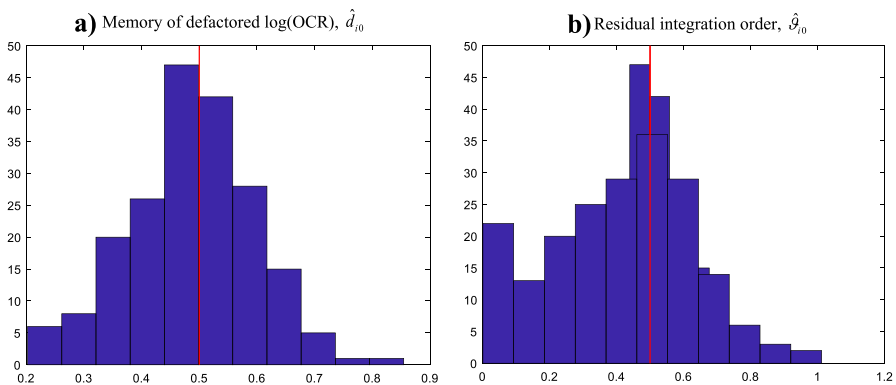


**Fig. 3** Individual slope estimates from the fractionally integrated heterogeneous panel data model. *Note.* Individual slope estimates (together with the corresponding confidence intervals at the 95% confidence level)



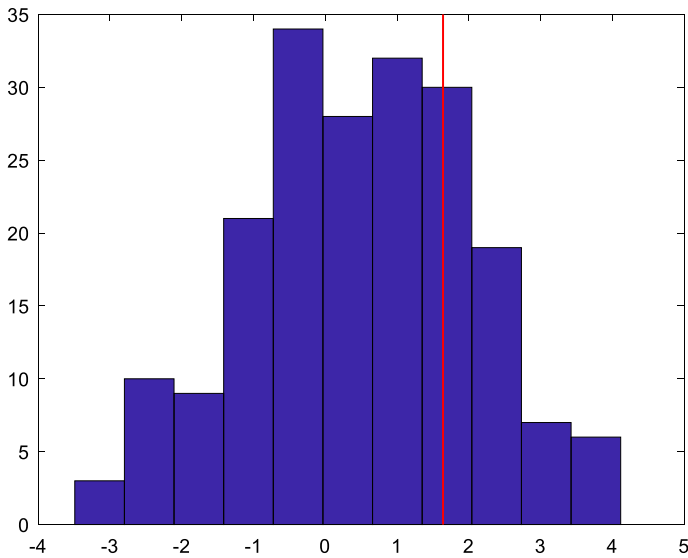
**Fig. 4** Memory of defactored log OCR, residual integration order and cointegration t-statistics. *Note.* Point estimates together with the 90% confidence intervals of the memory of defactored log OCR and residual integration order. t-statistics for testing cointegration together with the 90% critical values

Another interesting aspect concerns the memory estimates in the fractional integrated heterogeneous panel data model. Figure 4 shows the point estimates together with the 90% confidence intervals of the memory of defactored log OCR,  $\hat{\vartheta}_{i0}$ , (panel A) and the residual (cointegration error) integration order,  $\hat{d}_{i0}$ , (panel B) by lodging. Figure 5 shows the corresponding distributions of the memory estimates of the log OCR and of the residual integration order. Clearly, for most listings both are between 0



**Fig. 5** Distribution of the memory of defactored log OCR and the residual





**Fig. 6** Distribution of the  $t$ -statistics for the test of fractional cointegration. *Note:* Frequency distribution over the  $t$ -test statistics for fractional cointegration. The vertical line at 1.64 indicates the region in which we find cointegration at the 5% significance level

and 1, giving support for the fractional analysis. Thus, the OCR show a nonstationary mean-reverting long-memory behaviour with long-lasting effects of shocks. Similarly, the cointegration errors (i.e., the deviations from the long-run equilibrium) are also quite persistent.

Next, we formally analyse the fractional cointegration properties. As previously mentioned, fractional cointegration requires that  $\hat{d}_{i0} < \hat{\vartheta}_{i0}$ , which can be tested using the  $t$ -test,  $t = (\hat{\vartheta}_{i0} - \hat{d}_{i0})/s.e.(\hat{\vartheta}_{i0} - \hat{d}_{i0})$ . Figure 4 panel C shows the  $t$ -statistics for testing cointegration together with the 90% critical values by lodging. Figure 6 shows the distribution over these  $t$ -statistics. It turns out that in 45 out of the 199 listings (around 23%) fractional cointegration is confirmed (the mass on the right of the vertical line at 1.64 in Fig. 6, which represents the 5% significance level for a one-sided test and the number of dots above the critical value in Fig. 4, panel C). We take this as evidence for a long-run relationship between the occupancy rate and the prices and income. In fact, the listings could be sufficiently similar, and we would expect a similar long-run relationship for most of them. Given that the time dimension is quite short, we further would expect that the cointegration detection mechanism does not have too high power. Finally, since cointegration is found in considerably more than 5% of the properties, we argue that there is indeed a long-run relationship at the listings level.

We also have applied the standard panel cointegration tests by Pedroni (1999, 2004) and by Westerlund (2005). These test the null of no-cointegration against either homogenous (with all units behaving the same) or heterogeneous alternatives. They further deal with serial correlation and cross-sectional dependence (Westerlund 2005).

The tests were performed in Stata using the command *xtcointtest*, with the autoregressive order (to deal with serial correlation) chosen by AIC in the Pedroni panel cointegration test, and a time trend and time demeaning in both tests. Finally, the alternative is cointegration in at least some panels in the Westerlund cointegration test. All tests clearly find cointegration (with  $p$  values of 0.00). In consequence, evidence of cointegration is stronger with the standard methods. However, while these methods have found cointegration for all lodgings, with the fractional method we have found cointegration for a subset of the lodgings.

### 5.2.2 Mean group estimates

The cross-sectional and slope heterogeneity estimates in the previous section show evidence of cointegration for some cross-sectional units. In this sense, the long-term parameters of the tourism demand models can be estimated because the variables have a cointegration relationship (Dogru et al. 2021).

Table 3 shows the mean group estimates from the CCE and the fractionally integrated heterogeneous panel data model. Both show a significant negative effect of the relative ADR of the facility and a positive significant effect of the ADR for the hotels. For example, an increase of 1% in the relative listing prices (RADR) implies a decrease of 0.25% in OCR, while an increase of 1% in the HADR implies an increase of 0.07% of OCR, both in the fractionally integrated case. The estimated elasticities are notably larger with the CCE. Therefore, demand is own-price-inelastic and traditional accommodation and P2P in the Canary Islands are substitute goods. Furthermore, in both, the GDP of the origin country is insignificant, indicating that the income of the origin country of tourists does not affect the demand at the destination. Therefore, economic policies need not account for this variable.

These results are in line with the empirical literature such as Gunter et al. (2020) for the city of New York, and Jiménez et al. (2023) for Spanish cities. For example, Gunter et al. (2020) showed that Airbnb demand to New York City is price-inelastic,

**Table 3** MG estimation results

|                | CCE         |           | Fractionally integrated heterogeneous panel data model |           |
|----------------|-------------|-----------|--|-----------|
|                | Coefficient | $p$ value | Coefficient  | $p$ value |
| Log RADR       | -0.5772     | 0.00      | -0.2515  | 0.00      |
| Log HADR       | 0.9954      | 0.04      | 0.0693   | 0.00      |
| Log GDP        | 0.8461      | 0.31      | 0.0226   | 0.28      |
| R-squared (MG) | 0.43        |           |  |           |
| CD statistic   | -1.06       | 0.29      |  |           |

R-squared (MG) is the mean group  $R^2$ . CD statistic is the Pesaran (2015) test for weak cross-sectional dependence. A heterogeneous constant is partialled out. The jack-knife bias correction technique is used in all cases except when it is negative in which case the R-squared (MG) is used instead

which is consistent with earlier findings for Vienna (Gunter and Önder 2018). Also, their results suggest that Airbnb listings in New York City are substitutes for the traditional accommodation industry. Other authors such as Guttentag and Smith (2017) also showed that Airbnb is a substitute for hotels and other accommodations. However, Gunter and Önder (2018) argue that the city's Airbnb offer is a complement, rather than as a substitute, for the traditional accommodation industry in Vienna. Nevertheless, with regard to income, our results indicate that there is no influence. But, Gunter et al. (2020) showed that Airbnb demand in New York City is income-elastic, i.e., Airbnb accommodation is a luxury good. Finally, Jiménez et al. (2023) studying several Spanish cities where Airbnb works, have found that there is not only a substitute relationship between Airbnb and hotels, but also that the greater number of beds offered by Airbnb has led to an increase in the total number of visitors received by this type of firms (due to higher Airbnb occupancy rates).

Finally, our paper complements the recent work of Suárez-Vega et al. (2023), who analysed the dynamic and spatial properties of the substitution effect between P2P and traditional accommodation (hotels and apartments) using a dynamic spatial demand model with monthly data. For the period before COVID-19, they find positive autocorrelation in the occupancy rates, own-price-elastic demand, and significant substitution with its competitors (hotels and apartments) in the short-run. Both price sensitivity and substitution effect increase in the long-run. Whereas they analysed the dynamic properties in a spatial context, we focus on the long-run equilibrium demand in a time-series panel context. Note that the COVID-19 outbreak considerably distorted price and income elasticities in their study, an aspect which we have not studied in this paper but leave for future research.

## 5.3 Robustness checks

### 5.3.1 Results for non-seasonal adjusted time series

In this section, as a robustness check, we repeat the preceding analysis with data which is not seasonally adjusted. As mentioned in footnote 3, we might expect that the factors collect the seasonal noise (see Camacho et al 2015). Table 4 shows that the estimated MG effects are comparable to the ones obtained with the seasonal adjusted data. In particular, with the CCE approach results are negative for all coefficients and statistically significant at 5% significance level for Log RADR and Log GDP, while with the fractional approach we obtain significant negative results for Log RADR and positive for Log HADR.

Therefore, seasonal and non-seasonal adjusted data lead to similar results for the fractional panel model but not for the classic CCE. In any case, recall that the mean group estimates of the CCE are affected by the high variability of the individual slope estimates.

**Table 4** MG estimation results (data without seasonal adjustment)

|                | CCE         |                | Fractionally integrated heterogeneous panel data model |                |
|----------------|-------------|----------------|--|----------------|
|                | Coefficient | <i>p</i> value | Coefficient  | <i>p</i> value |
| Log RADR       | - 1.0131    | 0.00           | - 0.4051   | 0.00           |
| Log HADR       | - 0.3322    | 0.19           | 0.0455   | 0.05           |
| Log GDP        | - 1.5531    | 0.05           | 0.0085   | 0.80           |
| R-squared (MG) | 0.24        |                |  |                |
| CD statistic   | - 2.32      | 0.02           |  |                |

R-squared (MG) is the mean group  $R^2$ . CD statistic is the Pesaran (2015) test for weak cross-sectional dependence. A heterogeneous constant is partialled out. The jack-knife bias correction technique is used in all cases except when it is negative in which case the R-squared (MG) is used instead

**Table 5** MG estimation results by province

|          | Santa Cruz de Tenerife |                | Las Palmas  |                | Both provinces |                |
|----------|------------------------|----------------|-------------|----------------|----------------|----------------|
|          | Coefficient            | <i>p</i> value | Coefficient | <i>p</i> value | Coefficient    | <i>p</i> value |
| Log RADR | - 0.1965               | 0.00           | - 0.2683    | 0.00           | - 0.2506       | 0.00           |
| Log HADR | 0.0124                 | 0.77           | 0.0746      | 0.00           | 0.0593         | 0.01           |
| Log GDP  | - 0.1258               | 0.58           | 0.0586      | 0.15           | 0.0132         | 0.84           |

Results obtained with the Rodríguez-Caballero (2022) approach which consists of employing the Ergemen and Velasco (2017) approach separately to the two provinces Santa Cruz de Tenerife (Tenerife and La Palma) and Las Palmas (Fuerteventura, Gran Canaria, and Lanzarote)

### 5.3.2 Analysis by provinces

In this section we repeat the fractional analysis for seasonal adjusted series but now applying Rodríguez-Caballero (2022)'s approach which allows for multi-level cross-sectional dependence. For doing so, we split the sample into two blocks of lodgings: the ones belonging to the province Santa Cruz de Tenerife which comprises of the islands Tenerife and La Palma (with 49 observations) and the province Las Palmas which comprises of the islands Fuerteventura, Gran Canaria, and Lanzarote (with 150 observations).<sup>4</sup> Table 5 gives the mean group estimates by province and for both provinces combined.

These results confirm that also separated by the provinces Log RADR keeps being negative and significant, however, Log HADR stops being so for the province of Las Palmas. Most importantly, for both provinces together, but allowing for different

<sup>4</sup> We also tried defining the lower layer as islands rather than provinces and thus disaggregating the lodgings even further. However, especially Fuerteventura and La Palma have very few lodgings and results are too volatile and therefore unreliable.

factors in each province, the mean group estimators resemble the ones found in Table 2 for the Ergemen and Velasco (2017) approach.

## 6 Conclusions

We analyse the long-run equilibrium relationship (cointegration) at the listing level between P2P occupancy rates (as proxy of their demand) and several exogenous variables, in particular, own prices, price of competitors (hotel and apartments) and visitor income. The econometric methodology is based on the fractional heterogeneous panel data cointegration model of Ergemen and Velasco (2017), which allows for heterogeneity of listings, dependency between them and fractional integration to accommodate possible slow reversion in the adjustment process to the equilibrium. For comparison, we also apply the classic common correlated effects (CCE) estimator.

We use information on the P2P accommodation market in the Canary Islands on the basis of 5 years of monthly data and analyse the degree of substitution relative to hotels and apartments as most important types of lodging in this archipelago.

In general, there is some evidence of cointegration at the listing level (in particular, there is (fractional) cointegration for 23% of listings), indicating that the market could have a long-run equilibrium relationship between occupancy rates and own prices, cross prices, and income for some listings. Also, mean group estimates for all slopes show that demand is own-price-inelastic, while cross-price elasticity indicates that traditional accommodation and P2P in the Canary Islands are substitute goods. Income elasticity is lower than 1 and is not statistically significant, indicating that the demand for tourism in the Canary Islands is insensitive to the economic situation in the origin countries.

### 6.1 Theoretical implications

From the theoretical point of view, both fractional and standard cointegration analyse long-run equilibria. Standard cointegration assumes series with unit roots that are connected in a stationary long-run relationship. However, both assumptions are rather restrictive and relaxed within the fractional cointegration framework. This framework allows both original series and cointegrating errors to be fractionally integrated instead. Hence, the fractional cointegration analysis allows a more general long-run relationship in which short-term system shocks are slowly accommodated. The analysis conducted in our study confirms that a fractional analysis might indeed be more appropriate. Therefore, managers could reach erroneous conclusions if they did not account for fractionally integrated framework in their empirical study.

Both the CCE approach and the fractional panel model rely on a rank condition which requires the number of factors be not too large. More specifically, this number must not exceed the rank of the matrix of averaged factor loadings which in practice is unobserved (see De Vos et al. 2021, for a classifier aiming at testing this condition). In our context, we assume that the variables in our model suffice to guarantee that this condition holds. In fact, this could be an additional motivation

for Rodríguez-Caballero (2022)'s approach which via the assumed multi-level cross-sectional dependence might reduce the overall number of factors for each province.

## 6.2 Practical implications

From a managerial point of view, our results are consistent with the standard demand theory which points out that own and competitor prices play a key role in determining long-term demand. Therefore, estimating such models can prove important for demand strategies.

In particular, our analysis allows managers to use such models to conduct insightful studies on price sensitivity and income and to design long-term policy recommendations in the P2P industry. On the one hand, managers or policy makers can use the estimated price elasticities to analyze their impact on total revenue (Song and Witt 2000) or predict the effectiveness of policies implemented in reducing or increasing demand (Dwyer et al. 2010). For example, following Song and Witt (2000), managers seeking to calculate total revenue for listings could focus on the price elasticity and price of tourism to calculate the marginal revenue (i.e., extra revenue generated by a one-unit increase in sales). If P2P demand were price elastic (in absolute value it will exceed unity), marginal revenue will be positive. However, for our data, the price elasticity of demand is negative and below one (in absolute value) or inelastic, implying that an increase in the own price listings will result in a less than proportional change in listing demand. Therefore, as a result, total tourist revenue would decrease.<sup>5</sup>

On the other hand, the estimated cross-price elasticity can help managers determine whether their listings are substitutes, complements or are independent as regards other products (e.g., other neighbouring Airbnb listings, hotels). This information is essential for formulating pricing strategies and for analysing risk associated with various products (e.g., multi-unit hosts managing several types of listings where important substitute relations could exist among them). For example, in our case, Airbnb listings are substitutes for the traditional hotel accommodation industry. However, the low cross-price elasticity implies a low sensitivity to price changes in the hotel accommodation industry in the long-run and a rather low level of competition between Airbnb listings and hotels. Therefore, managers need not to respond to a competitor's price reduction as fast as they would have to if the cross-price elasticity was high. Finally, the estimated income elasticity carries little information. In fact, our results cannot help managers to implement marketing strategies for Airbnb listings such as identifying potential markets for their products when changes in income are expected.<sup>6</sup>

Another interesting implication of the cointegrated model is that it allows managers to perform sensitivity analyses on demand in the long-run. For example, managers or

<sup>5</sup> Following Song and Witt (2000, page 12) and based on tourism revenue formula (e.g.,  $P$  (= tourist price at the destination)  $\times$   $Q$  (= quantity of units sold)), marginal revenue is equal to  $P \times (1 + 1/\text{price elasticity of demand})$ . For example, if we assume that price elasticity of demand is  $-0.25$  (e.g., fractional integrated case in Table 2), marginal revenue equals  $-3 \times P$ . Therefore, it will imply that a price decrease, holding other variables constant, will result in a less than proportionate increase in tourist demand, leading to a revenue decrease.

<sup>6</sup> For example, if the tourism product was a normal good an income increase would increase the demand, while for an inferior good the demand would decrease.

policy makers can use the proposed model to analyze the effects of own and substitution price elasticities on demand using subsamples of listings, e.g., Airbnb listings located in different provinces. We, also, could analyze demand differences between the type of hosts (e.g., single-unit- and multi-unit hosts), which can represent different levels of professionalism in the P2P industry.

Certain limitations should be acknowledged. One is technical in nature, namely the dependency of our proposed method on balanced panel data which reduces considerably the sample data. Another methodological limitation is that we do not incorporate the lagged dependent variable and/or weakly exogenous variables as regressors because there is no econometric model for this specification. In particular, to our knowledge, whereas for the standard model there is the model of Chudik and Pesaran (2015), there is no equivalent model for the fractional case. The inclusion of the lagged dependent variable would allow for a dynamic panel structure. This could account for certain dynamic features of the lodgings. In particular, the occupancy rate could not only depend on the relative average daily rate of the listing, the average daily rate of hotels in the municipality and GDP, but also in a general manner on its own past. In any case, in the fractional model, the memory parameter allows to capture at least some of this dynamic behaviour.

Future research could try to incorporate the lagged dependent variable as regressor in the fractional panel model and include the impacts of other economic factors on demand such as exchange rates and marketing policies. With more available data, it would also be interesting to investigate the effect of COVID-19 on the analysed long-run demand relationship. Alternatively, to the employed CCE approach, one could model the cross-sectional dependence using Bai (2009)'s approach of interactive fixed effects. This model uses principal components analysis to estimate both the structural parameters and the common component. The latter might contain information on common features in the data.

**Acknowledgements** This study was supported by the Universidad de Las Palmas de Gran Canaria (ULPGC) Program together with *Ministerio de Ciencia, Innovación y Universidades* under grant GOB-ESP2019-07, the ULPGC Program through grant COVID 19-04 and by the Spanish Ministry of Science and Innovation—Agencia Estatal de Investigación [grant number PID2020-114646RB-C43]. The views expressed here are those of the authors and not necessarily those of the institution with which they are affiliated.

**Funding** Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study was supported by the Universidad de Las Palmas de Gran Canaria (ULPGC) Program together with *Ministerio de Ciencia, Innovación y Universidades* under grant GOB-ESP2019-07, the ULPGC Program through grant COVID 19-04, and by the Spanish Ministry of Science and Innovation—Agencia Estatal de Investigación [grant number PID2020-114646RB-C43].

## Declarations

**Conflict of interest** Authors declare that we have no conflict of interest.

**Ethical approval** This article does not contain any studies with human participants performed by any of the authors.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adamiak C, Szyda B, Dubownik A, García-Álvarez D (2019) Airbnb offer in Spain-Spatial analysis of the pattern and determinants of its distribution. *ISPRS Int J Geo Inf* 8(3):155. <https://doi.org/10.3390/ijgi8030155>
- Bai J (2009) Panel data models with interactive fixed effects. *Econometrica* 77:1229–1279
- Bailey N, Kapetanios G, Hashem Pesaran M (2016) Exponent of cross-sectional dependence: Estimation and inference. *J Appl Economet* 31:929–960
- Baltagi BH (2001) *Econometric analysis of panel data*, 2nd edn. Wiley, New York
- Bianchi G, Chen Y (2020) The short- and long-run hotel demand in Switzerland: a weighted macroeconomic approach. *Journal of Hospitality & Tourism Research* 44(5):835–857
- Bonham C, Edmonds C, Mak J (2006) The impact of 9/11 and other terrible global events on tourism in the United States and Hawaii. *J Travel Res* 45(1):99–110
- Bonham C, Gangnes B, Zhou T (2009) Modeling tourism: a fully identified VECM approach. *Int J Forecast* 25(3):531–549
- Boto-garcía D, Mayor M, De P (2021) Spatial price mimicking on Airbnb: Multi-host vs single-host. *Tourism Manag* 87(March 2020):104365. <https://doi.org/10.1016/j.tourman.2021.104365>
- Çalışkan U, Saltik IA, Ceylan R, Bahar O (2019) Panel cointegration analysis of relationship between international trade and tourism: case of Turkey and silk road countries. *Tourism Management Perspectives* 31:361–369
- Camacho M, Lovcha Y, Quiros GP (2015) Can we use seasonally adjusted variables in dynamic factor models? *Stud Nonlinear Dyn Econ* 19(3):377–391
- Chakraborty SK, Mazzanti M (2020) Energy intensity and green energy innovation: checking heterogeneous country effects in the OECD. *Struct Change Econ Dyn* 52:328–343
- Crouch G, Schultz L, Valerio P (1992) Marketing international tourism to Australia: a regression analysis. *Tour Manag* 13:196–208
- De Vos I, Everaert G, Sarafidis V (2021) A method for evaluating the rank condition for CCE estimators. Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 21/1013, Ghent University, Faculty of Economics and Business Administration
- Deng M, Athanasopoulos G (2011) Modelling Australian domestic and international inbound travel: a spatial-temporal approach. *Tour Manag* 32(5):1075–1084. <https://doi.org/10.1016/j.tourman.2010.09.006>
- Dogru T, Bulut U, Sirakaya E (2021) Modeling tourism demand: theoretical and empirical considerations for future research. *Tour Econ* 27(4):874–889
- Dritsakis N (2004) Cointegration analysis of German and British tourism demand for Greece. *Tour Manag* 25(1):111–119
- Dwyer L, Forsyth P, Dwyer W (2010) *Tourism economics and policy*. Channel View Publications, Bristol
- Ergemen YE, Velasco C (2017) Estimation of fractionally integrated panels with fixed effects and cross-section dependence. *J Econom* 196(2):248–258
- Eugenio-Martin JL, Cazorla-Artiles JM, González-Martel C (2019) On the determinants of Airbnb location and its spatial distribution. *Tour Econ*. <https://doi.org/10.1177/1354816618825415>
- EUROSTAT (2021) GDP and main aggregates-selected international quarterly data. Retrieved from [https://ec.europa.eu/eurostat/databrowser/product/page/NAIDQ\\_10\\_GDP](https://ec.europa.eu/eurostat/databrowser/product/page/NAIDQ_10_GDP)
- Garín-Muñoz T (2006) Inbound international tourism to Canary Islands: a dynamic panel data model. *Tour Manag* 27(2):281–291. <https://doi.org/10.1016/j.tourman.2004.10.002>
- Garín-Muñoz T, Amaral TP (2000) An econometric model for international tourism flows to Spain. *Appl Econ Lett* 7(8):525–529



- Garín-Muñoz T, Montero-Martín LF (2007) Tourism in the Balearic Islands: a dynamic model for international demand using panel data. *Tour Manag* 28(5):1224–1235. <https://doi.org/10.1016/j.tourman.2006.09.024>
- Gómez M, Tinoco NS, Tinoco LM (2021) The influence of Airbnb on hotel occupancy in Mexico: A Big Data analysis (2007–2018). *Revista CIMEXUS*, Vol. XVI, No. 1, 9–32
- Gunter U, Önder I (2018) Determinants of Airbnb demand in Vienna and their implications for the traditional accommodation industry. *Tour Econ* 24:270–293
- Gunter U, Önder I, Zekan B (2020) Modeling Airbnb demand to New York city while employing spatial panel data at the listing level. *Tour Manag* 77:104000. <https://doi.org/10.1016/j.tourman.2019.104000>
- Gutiérrez J, García-Palomares JC, Romanillos G, Salas-Olmedo MH (2017) The eruption of Airbnb in tourist cities: comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tour Manag* 62:278–291. <https://doi.org/10.1016/j.tourman.2017.05.003>
- Guttentag DA, Smith SL (2017) Assessing Airbnb as a disruptive innovation relative to hotels: Substitution and comparative performance expectations. *Int J Hosp Manag* 64:1–10
- Herrington JD, Bosworth C (2016) The short- and long-run implications of restaurant advertising. *J Foodserv Bus Res* 19(4):325–337. <https://doi.org/10.1080/15378020.2016.1178052>
- Husein J, Kara SM (2020) Nonlinear ARDL estimation of tourism demand for Puerto Rico from the USA. *Tour Manag*. <https://doi.org/10.1016/j.tourman.2019.103998>
- ISTAC (2021) Encuestas de alojamiento turístico [tourist accommodation survey]. Retrieved from <http://www.gobiernodecanarias.org/istac/> September 1, 2021
- Jiménez JL, Ortuño A, Pérez-Rodríguez JV (2022) How does AirBnb affect local Spanish tourism markets? *Empir Econ* 62:2515–2545
- Jiménez JL, Ortuño A, Pérez-Rodríguez JV (2023) How does Airbnb affect local Spanish tourism markets? *Empir Econ* 62:2515–2545
- Kadir N, Karim M (2009) Demand for Tourism in Malaysia by UK and US Tourists: A Cointegration and Error Correction Model Approach. Chapter 4, In Á. Matias et al. (eds.), *Advances in Tourism Economics*, 51–70. [https://doi.org/10.1007/978-3-7908-2124-6\\_4](https://doi.org/10.1007/978-3-7908-2124-6_4)
- Kim S, Song H (1998) Analysis of inbound tourism demand in south korea: a cointegration and error correction approach. *Tour Anal* 3(1), 25–41(17)
- Lathiras P, Siriopoulos C (1998) The demand for tourism to Greece: a cointegration approach. *Tour Econ* 4(2):171–185. <https://doi.org/10.1177/135481669800400204>
- Ledesma-Rodríguez F, Navarro-Ibáñez M, Pérez-Rodríguez J (2001) Panel data and tourism: a case study of Tenerife. *Tour Econ* 7(1):75–88. <https://doi.org/10.5367/000000001101297748>
- Lee CG (2010) The dynamic interactions between hotel room rates and international inbound tourists: evidence from Singapore. *Int J Hosp Manag* 29(4):758–760
- Lee CG, How S-M (2019) Long-run causality between customer satisfaction and financial performance: the case of Marriott. *Curr Issue Tour* 22(14):1653–1658
- Lee CG, How S-M (2022a) The impacts of domestic and global economic policy uncertainties on the hotel room demand: evidence from Singapore. *Tour Econ*. <https://doi.org/10.1177/13548166221086062>
- Lee CG, How S-M (2022b) Determining hotel room rates: domestic economic policy uncertainty and crises. *Tour Anal*. <https://doi.org/10.3727/108354222X16571659728574>
- Li G, Song H, Witt SF (2004) Modeling tourism demand: a dynamic linear AIDS approach. *J Travel Res* 43(2):141–150. <https://doi.org/10.1177/0047287504268235>
- Li G, Song H, Witt S (2005) Recent developments in econometric modeling and forecasting. *J Travel Res* 44(1):82–99
- Lim C, McAleer M (2001) Cointegration analysis of quarterly tourism demand by Hong Kong and Singapore for Australia. *Appl Econ* 33(12):1599–1619
- Lyssiotou P (2000) Dynamic analysis of British demand for tourism abroad. *Empir Econ* 25(3):421–436. <https://doi.org/10.1007/S001810000025>
- Ma T, Hong T, Zhang H (2015) Tourism spatial spillover effects and urban economic growth. *J Bus Res* 68(1):74–80. <https://doi.org/10.1016/j.jbusres.2014.05.005>
- Maravall A (2012) Update of seasonality tests and automatic model identification in TRAMO-SEATS. Banco de España. November 2012 draft
- Massidda C, Mattana P (2013) A SVECM analysis of the relationship between international tourism arrivals, GDP and trade in Italy. *J Travel Res* 52(1):93–105
- Moore WR (2010) The impact of climate change on Caribbean tourism demand. *Curr Issue Tour* 13(5):495–505

- Morley CL (2009) Dynamics in the specification of tourism demand models. *Tour Econ* 15(1):23–39
- Narayan P-K (2004) Fiji's tourism demand: the ARDL approach to cointegration. *Tour Econ* 10(2):193–206
- Naudé WA, Saayman A (2005) Determinants of tourist arrivals in Africa: a panel data regression analysis. *Tour Econ* 11(3):365–391. <https://doi.org/10.5367/000000005774352962>
- Önder I, Weismayer C, Gunter U (2019) Spatial price dependencies between the traditional accommodation sector and the sharing economy. *Tour Econ* 25(8):1150–1166. <https://doi.org/10.1177/1354816618805860>
- Pesaran MH, Shin Y, Smith RP (1999) Pooled mean group estimation of dynamic heterogeneous panels. *JASA* 94:621–634
- Pedroni P (1999) Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economic and Statistics* 61:653–670
- Pedroni P (2004) Panel cointegration: Asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econom Theory* 20:597–625
- Pesaran MH (2006) Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica* 74:967–1012
- Rodríguez-Caballero CV (2022) Energy consumption and GDP: a panel data analysis with multi-level cross-sectional dependence. *Econom Stat* 23:128–146
- Rodríguez-Caballero CV, Vera-Valdés JE (2021) Air pollution and mobility, what carries COVID-19? *Econometrics* 9(4):37
- Roget FM, González XAR (2006) Rural tourism demand in Galicia, Spain. *Tour Econ* 12:21–31
- Saha S, Yap G (2014) The moderation effects of political instability and terrorism on tourism development: a cross-country panel analysis. *J Travel Res* 53(4):509–521
- Sakai M, Brown J, Mak J (2000) Population aging and Japanese international travel in the 21st century. *J Travel Res* 38:212–220
- Salman AK (2003) Estimating tourist demand through cointegration analysis: Swedish Data. *Curr Issue Tour* 6(4):323–339
- Seetanah B, Durbarry R, Ragodoo JFN (2010) Using the panel cointegration approach to analyse the determinants of tourism demand in South Africa. *Tour Econ* 16(3):715–729
- Seetaram N (2010) Use of dynamic panel cointegration approach to model international arrivals to australia. *J Travel Res* 49(4):414–422. <https://doi.org/10.1177/0047287509346992>
- Shin Y, Yu B, Greenwood-Nimmo M (2014) Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. In: Sickles R, Horrace W (eds) *Festschrift in honor of Peter Schmidt: econometric methods and applications*. Springer, New York, pp 281–314
- Song L, Li G (2008) Tourism demand modelling and forecasting—a review of recent research. *Tour Manag* 29(2):203–220
- Song H, Witt S (2000) *Tourism demand modelling and forecasting. Modern econometric approaches*. Pergamon, Amsterdam
- Song H, Witt SF (2003) Tourism forecasting: The general-to-specific approach. *J Travel Res* 42(1):65–74. <https://doi.org/10.1177/0047287503253939>
- Song H, Li G, Witt SF, Fei B (2010) Tourism demand modelling and forecasting: how should demand be measured? *Tour Econ* 16(1):63–81. <https://doi.org/10.5367/000000010790872213>
- Song H, Lin S, Witt SF, Zhang X (2011) Impact of financial/economic crisis on demand for hotel rooms in Hong Kong. *Tour Manag* 32(1):172–186
- Song H, Dwyer L, Li G, Cao Z (2012) Tourism economics research: a review and assessment. *Ann Tour Res* 39(3):1653–1682
- Song H, Qiu RTR, Park J (2019) A review of research on tourism demand forecasting: Launching the *Annals of Tourism Research Curated Collection on tourism demand forecasting*. *Ann Tour Res* 75:338–362
- Suárez-Vega R, Pérez-Rodríguez JV, Hernández JM (2023) Substitution among hotels and P2P accommodation in the COVID era: a spatial dynamic panel data model at the listing level. *Curr Issue Tour* 26(17):2883–2899
- Voltes-Dorta A, Inchausti-Sintes F (2021) The spatial and quality dimensions of Airbnb markets. *Tour Econ* 27(4):688–702. <https://doi.org/10.1177/1354816619898075>
- Webber A (2001) Exchange rate volatility and cointegration in tourism demand. *J Travel Res* 39:398–405
- Westerlund J (2005) New simple tests for panel cointegration. *Econom Rev* 24:297–316
- Wu DC, Li G, Song H (2012) Economic analysis of tourism consumption dynamics. A time-varying parameter demand system approach. *Ann Tour Res* 39(2):667–685. <https://doi.org/10.1016/j.annals.2011.09.003>

- Yang Y, Wong KKF (2012) A spatial econometric approach to model spillover effects in tourism flows. *J Travel Res* 51(6):768–778. <https://doi.org/10.1177/0047287512437855>
- Yang Y, Zhang H (2019) Spatial-temporal forecasting of tourism demand. *Ann Tour Res* 75:106–119. <https://doi.org/10.1016/j.annals.2018.12.024>
- Zhang J (2009) Spatial distribution of inbound tourism in China: determinants and implications. *Tour Hosp Res* 9(1):32–49. <https://doi.org/10.1057/thr.2008.41>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.