Panel Data Models in Tourism Research: Innovative Applications and Methods

Published Version:
J.L. Eugenio-Martin, R. Patuelli (2022). Panel Data Models in Tourism Research: Innovative Applications and Methods. TOURISM ECONOMICS, 28(5), 1348-1354 [10.1177/13548166221115784].

Availability:
This version is available at: https://hdl.handle.net/11585/900379 since: 2022-11-06

Published:
DOI: http://doi.org/10.1177/13548166221115784

Panel data models in tourism research: Innovative applications and methods

Juan L. Eugenio-Martin

University of Las Palmas de Gran Canaria, Spain

Roberto Patuelli

University of Bologna, Italy

1. Introduction

Panel data consists of repeated observations on the same cross section over time (Wooldridge, 2002). In tourism research, the subjects of the cross sections have usually been countries or regions due to data availability. However, they can also be tourists, households or firms. Nowadays, they can be reached not only through the administration of multi-wave surveys (Seaton and Palmer, 1997), but more often through the access to big data. The possibility to trace the same tourists' behaviour over time is advantageous and it certainly enriches the analysis with respect to cross-sectional or time series analysis. However, this kind of datasets are usually unbalanced, which impedes the standard application of panel data methods.

In some instances, a panel data descriptive analysis may be the only way to obtain new insights (Bronner and de Hoog, 2008). Moreover, its econometric modelling also provides the ability to study dynamic relationships and to take into account the heterogeneity among subjects. Econometric modelling permits the identification of determinants, the quantification of their relevance, and the simulation of their effects with further post-estimation analyses.

The econometrics of panel data kept evolving during the last 30 years. However, the application of panel data methods to tourism research was still scarce by the beginning of the 2000s. The first papers criticised the limitations of pooled regressions, because they assume that heterogeneity does not exist, this being a very difficult assumption to meet (see, e.g., Garín-Muñoz and Pérez Amaral, 2000). One way to solve this problem is first differencing, which removes the latent heterogeneity from the model (see, e.g., Ledesma-Rodríguez et al., 2001). Similarly, fixed effects (FE) models are based on the idea that differences across groups can be captured in differences in the constant term. In both cases, any time-invariant variable is removed.

However, random effects (RE) models can consider time invariant regressors. Such an approach can be useful to estimate the determinants of international tourism demand for different sets of origins and destinations (Naudé and Saayman, 2005). In an RE model, we can work on an error components model that can be estimated with feasible generalized least squares (FGLS). Under heteroskedastic errors and/or autocorrelation, though, both FE and RE estimates are inconsistent. However, the FGLS procedure is sufficiently flexible to estimate more complex disturbance covariance matrices that can accommodate such issues. Actually, in tourism research, destinations or origins are very different in size, and correlations among them are likely to happen. Such correlations are also different for each pair of origin-destination, and close destinations are likely to be more correlated than further ones (see for instance, Eugenio-Martin et al., 2008).

Dynamic panels can also be of interest in tourism research, especially for estimating growth models. They have indeed become popular, for instance for testing the tourism-led growth hypothesis (TLGH). However, theoretical and empirical concerns need to be taken into account in order to avoid misleading results (Song and Wu, 2021). A quantity of interest is the long-run multiplier which reveals the impact on the equilibrium of a change in a determinant. The assumptions of the dynamic panel models are very restrictive, that is, strict exogeneity, homoscedasticity, non-autocorrelation and uncorrelated observations. The presence of the lagged dependent variable on the right hand side of the equation is problematic for pooled OLS or FGLS estimators, which are inconsistent for this case. Arellano and Bond (AB) (1991) and Arellano and Bover (1995) suggest employing a large set of instruments (moment conditions) to estimate dynamic models by generalized methods of moments (GMM). Several authors have employed this methodology to test the TLGH (see, e.g., Eugenio-Martin, Martin-Morales and Scarpa, 2004; Sequeira and Nunes, 2008; Seetanah, 2011) within the conventional augmented Solow growth model. In tourism demand studies, the interpretation of the lagged dependent variable relies on habit formation (Garín-Muñoz, 2006).

A natural extension of panel data econometrics, when analysing tourism, is the inclusion of spatial autocorrelation towards the estimation of spatial panel data models. For this class of methods, a spatial weights matrix (or more), describing the proximity relations between tourism destinations, firms or individuals, needs to be defined (Anselin, 1988). For instance, Gunter *et al*. (2020) test whether Airbnb listings and the traditional accommodation industry are substitutes in New York City. For this case, spatial effects are critical in order to control the degree of spatial overlapping patterns between hotels and Airbnb listings (Eugenio-Martin, Cazorla-Artiles and González-Martel, 2019). Moreover, tourism demand varies within a year by seasons or within a week. Thus, panel data can cover a wider time span, so that the analysis is more comprehensive. Spatial competition or complementarity based on destination attractions can also be studied within this framework. For instance, in a Poisson panel model, Patuelli *et al*. (2013) focus on the effects of UNESCO World Heritage Sites on tourism flows, and by means of spatial lags (averages over neighbouring units), assess not only the direct effect of UNESCO sites, but also how regions compete for the same tourists (the indirect effect, moderating the former one). More generally, spatial panel models have also been employed for testing for spatial spillover effects of the destinations (Yang and Wong, 2012; Ma, Hong and Zhang, 2015) or for tourism demand forecasting (Yang and Zhang, 2019). Overall, the ever-growing spatial family of panel econometric models may allow, in tourism research, for a wide set of potential new research questions and topics to investigate. Moreover, the emerging relevance of the big data revolution, which typically implies geo-referenced data distributed in space as point patterns, opens the door for a further set of *ad hoc* tools in spatial modelling for tourism. However, this new – but already current – direction of research poses various computational and methodological challenges for panel econometrics, due to the size and the typically unbalanced nature of the data at hand, which most likely will be leading to some of the most important developments in the field over the coming years.

This special issue banks on the above developments and targets the use of panel data in tourism research. 20 abstracts were submitted after the call for papers, out of which 13 fulfilled our main purpose, that is, to consider methodological contributions as well as innovative applications. Finally, four papers were accepted and are presented in this special issue. Two of them provide a new way to employ new panel data methods, whereas the other two provide new empirical applications.

2. Special issue methodological contributions

2.1 Fractionally integrated heterogeneous panel data analysis

Tourism economics literature has paid attention to the ability of the tourism sector to promote economic growth, also called as the Tourism Led Growth Hypothesis (TLGH). Since GDP and tourism growth rates may follow a common slope, many researchers have employed panel cointegration tests to prove such relationship (Lee and Chang, 2008; Seetaram, 2010; Dritsakis, 2012). The standard panel cointegration test developed by Pedroni (1999, 2004) and the estimator of the panel cointegration relationship (Pedroni, 2001) extends the Engle and Granger (1987) cointegration test to a panel framework. Although the whole error series is taken into account in the test, most of the relevance is driven by the one-period lagged error. The parameter associated with the former term is the key of the test.

Pérez-Rodríguez, Rachinger and Santana-Gallego (2022) explores TLGH by applying a fractionally integrated heterogeneous panel data method. The main purpose of this approach is to extend the role of the lagged errors, so that a longer range dependence can be considered. Fractional cointegration describes a long-run relationship between two trending variables such as GDP and tourism growth rates with the equilibrium error being persistent, but less so than the variables themselves. They employ Ergemen's (2019) model, which accommodates general stationary or nonstationary long-range dependence through individual and interactive fixed effects. It allows for contemporaneous correlation in model innovations. More precisely, it nests the standard $I(0)$ and $I(1)$ cases and eliminates the necessity of preliminary unit root testing.

The results depend on several key parameters rather than just one as in Pedroni's test. Furthermore, cointegration is proved when certain inequalities are met. A problem may rise when some of these parameters are not significant. Overall, in the paper most of the key parameters turned out to be significant (12 out of 14 countries). Moreover, two different time spans were considered and the cointegration results were consistent for both periods for most countries (10 out of 12 countries). Finally, it should be noted that the TLGH tests show heterogeneous responses among the European countries. However, as usually happens with this kind of tests, the underpinnings behind such heterogeneity are unknown. Further research may consider such understanding as well as a wider range of countries.

2.2 Semiparametric Age-Period-Cohort analysis

The analysis of panel data has considered the econometric approach in most papers. However, the application of non-parametric or semiparametric techniques is not commonly applied in tourism economics and it may enrich the understanding provided by the econometric analysis. Weigert *et al*. (2021) contribute to the tourism literature with a semiparametric technique that comprises the analysis according to the age of tourists, the year (period), and the birth cohort they belong to. It is known as the Age-Period-Cohort (APC) model, which deal with repeated cross-sectional data in a multiplicative way.

They have been applied intensively in health studies, especially for understanding cancer incidence among the population. However, they have not been applied in tourism. They make sense for the tourism analysis because the APC model has the ability to disentangle the effects by age, periods and cohort groups. They argue that age matters since tourists' decisions are conditioned by their own life cycle. Moreover, the period effect depends on the developments that have taken place, especially in terms of infrastructure and economic development. Finally, they consider that the tourists may also belong to certain generations that may share a common behaviour.

Additionally, Weigert *et al*. (2021) provide heatmaps and ridgeline matrices of results that enrich the understanding of APCs in relation to different key variables such as travel distance. They provide a good illustration of how the employment of APC can enrich the understanding of tourists' behaviour. More precisely, they consider five birth cohorts that correspond to the silent generation, baby boomer, generation X, Y and Z. They analyse their travelling distance decisions for German travellers between 1971 and 2018. They provide interesting and new insights of the traveling decision making in tourism and prove that the employment of APC is a useful tool for tourism studies.

3. Special issue empirical contributions

3.1 Dutch disease

Countries that rely heavily on the rents from natural resources may face marked turbulences on the exchange rates. Demand pressure on such resources may end up with larger prices but also with a currency appreciation. This is particularly problematic for tourism destinations which their real prices should increase eroding its international competitiveness. This phenomenon is known as Dutch disease (Inchausti-Sintes, 2015). The paper by Gholipour *et al*. (2022) studies the long-run and short-run effect of natural rents on inbound and outbound business travels in resource-abundant economies. It is expected that after an increase on natural resource rents, the currency appreciates, which causes a decrease in the inbound tourism flow and an increase in the outbound tourism flow.

The authors employed Pooled Mean Group (PMG) estimator for this study (see Pesaran, Shin and Smith, 1999). PMG allows for specific individual short-term adjustments and adjustment speeds while imposing cross-section homogeneity restrictions only on the long-run coefficients. If the long-run coefficients are equal across countries the PMG will be consistent and efficient. If slope homogeneity does not exist, the PMG will be inconsistent and estimates will be biased downwards. In the tourism literature, this methodology had proved to be successful. See for instance, Falk (2010) who employed the expected snow from previous season to estimate the contemporaneous demand with PMG estimation.

The paper focuses on the international business travels which are closely related with the imports and exports activities and the exchange rate. The results show that inbound tourism demand does not vary in the short run with the natural resource rents, but in the long run. Moreover, the results for the outbound tourism demand show that an increase in the rents imply a positive impact in the flow. However, for the long run, the estimates are the same as with the inbound traffic, i.e. higher rents imply lower outbound demand. It makes sense that such appreciation deteriorates the industrial competitiveness of the country which will reduce its exports and all kind of business travel associated with it. Further research on how it affects other kinds of tourism may provide additional insights into this issue.

3.2 Transport infrastructure

The relevance of transport infrastructure is crucial for the success of any tourism destination. Tourism destinations may be reached by different modes and such development may have happened smoothly or abruptly over time. It is not straightforward to disentangle the role that each transport mode has played over time in terms of tourism success. Moreover, transportation to certain nodes has also implications in nearby regions and such impact may vary by transportation mode.

Tian, Yang and Jiang (2022) develop a spatial panel data model to understand whether the development of transport infrastructure in China has also implied tourism growth. They distinguish 337 Chinese prefecture-level regions between 2007 and 2016. More precisely, they employ a spatial Durbin model that takes into account the spillover effects within each region. They distinguish arrivals and revenue for both the domestic and international markets.

This method has proved to be useful to disentangle the role played for each transport mode in relation to traffic and revenue generated, not only in the region of interest but also in nearby regions. Moreover, the method also shows the indirect effects caused by the different transportation modes. It shows the presence of wider economic effects that are required in cost-benefit analyses of transport infrastructure projects. Nevertheless, the paper is not exempt of limitations, but it shows a valid methodology to be employed for the assessment of new transportation modes, especially when they are integrated into a multi-modal transport infrastructure network.

References

Anselin, L. (1988) *Spatial econometrics: Methods and models*. Studies in Operational Regional Science, Vol. 4. Springer, Dordrecht.

Arellano, M. and S. Bond (1991) “Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations”, *The Review of Economic Studies*, Vol. 58(2): 277-297.

Arellano, M. and O. Bover (1995) “Another look at the instrumental variable estimation of error-components models”, *Journal of Econometrics*, Vol. 68(1): 29-51.

Bronner, F. and R. de Hoog (2008) “Agreement and disagreement in family vacation decision-making”, *Tourism Management*, Vol. 29(5): 967-979.

Dritsakis, N. (2012) “Tourism development and economic growth in seven Mediterranean countries: A panel data approach”, *Tourism Economics*, Vol. 18(4): 801-816.

Engle, R.F. and C.W.J. Granger (1987) “Co-integration and error correction: representation, estimation, and testing”, *Econometrica*, Vol. 55: 251-276.

Ergemen, Y.E. (2019) “System estimation of panel data models under long-range dependence”, *Journal of Business & Economic Statistics*, Vol. 37(1): 13-26.

Eugenio-Martin, J.L., J.M. Cazorla-Artiles and C. González-Martel (2019) “On the determinants of Airbnb location and its spatial distribution”, *Tourism Economics*, Vol. 25(8): 1224-1244.

Eugenio-Martin, J.L., N. Martín-Morales and R. Scarpa (2004) “Tourism and economic growth in Latin American countries: A panel data approach”, *FEEM Working paper*, 26.2004.

Eugenio-Martin, J.L., N. Martín-Morales and M.T. Sinclair (2008) “The role of economic development in tourism demand”, *Tourism Economics*, Vol. 14(4): 673-690.

Falk, M. (2010) “A dynamic panel data analysis of snow depth and winter tourism”, *Tourism Management*, Vol. 31(6): 912-924.

Garín-Muñoz, T. and T. Pérez-Amaral (2000) “An econometric model for international tourism flows to Spain”, *Applied Economics Letters*, Vol. 7(8): 525-529.

Garín-Muñoz, T. (2006) “Inbound international tourism to Canary Islands: a dynamic panel data model”, *Tourism Management*, Vol. 27(2): 281-291.

Gholipour, H.F., R. Tajaddini and U. Al-mulali (2022) “Dutch disease phenomenon and demand for international business travels: Panel ARDL/PMG estimation”, *Tourism Economics*, forthcoming.

Gunter, U., I. Önder and B. Zekan (2020) “Modeling Airbnb demand to New York City while employing spatial panel data at the listing level”, *Tourism Management*, Vol. 77: 104000.

Inchausti-Sintes, F. (2015) “Tourism: economic growth, employment and Dutch disease”, *Annals of Tourism Research*, Vol. 54: 172-189.

Ledesma-Rodríguez, F.J., M. Navarro-Ibáñez and J.V. Pérez-Rodríguez (2001) “Panel data and tourism: a case study of Tenerife”, *Tourism Economics*, Vol. 7(1): 75-88.

Lee, C.C. and C.P. Chang (2008) “Tourism development and economic growth: a closer look at panels”, *Tourism Management,* Vol. 29(1): 180-192.

Ma, T., T. Hong and H. Zhang (2015) “Tourism spatial spillover effects and urban economic growth”, *Journal of Business Research,* Vol. 68(1): 74-80.

Naudé, W.A. and A. Saayman (2005) “Determinants of tourist arrivals in Africa: a panel data regression analysis”, *Tourism Economics,* Vol. 11(3): 365-391.

Patuelli, R., M. Mussoni and G. Candela (2013) “The effects of World Heritage Sites on domestic tourism: a spatial interaction model for Italy”, *Journal of Geographical Systems,* Vol. 15(3): 369-402.

Pedroni, P. (1999) “Critical values for cointegration tests in heterogeneous panels with multiple regressors”, *Oxford Bulletin of Economics and Statistics,* Vol. 61: 653-670.

Pedroni, P. (2001) “Purchasing power parity tests in cointegrated panels”, *Review of Economics and Statistics*, Vol. 83: 727-731.

Pedroni, P. (2004) “Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis”, *Econometric Theory,* Vol. 20: 597-625.

Pérez-Rodríguez, J.V., H. Rachinger and M. Santana-Gallego (2022) “Does tourism promote economic growth? A fractionally integrated heterogeneous panel data analysis”, *Tourism Economics,* forthcoming.

Pesaran, M.H., Y. Shin and R.P. Smith (1999) “Pooled mean group estimation of dynamic heterogeneous panels”, *Journal of the American Statistical Association*, Vol. 94: 621-634.

Seaton, A.V. and C. Palmer (1997) “Understanding VFR tourism behaviour: the first five years of the United Kingdom tourism survey”, *Tourism Management,* Vol. 18(6): 345-355.

Seetanah, B. (2011) “Assessing the dynamic economic impact of tourism for island economies”, *Annals of Tourism Research,* Vol. 38(1): 291-308.

Seetaram, N. (2010) “Use of dynamic panel cointegration approach to model international arrivals to Australia”, *Journal of Travel Research,* Vol. 49(4): 414-422.

Sequeira, T.N. and P.M. Nunes (2011) “Does tourism influence economic growth? A dynamic panel data approach”, *Applied Economics*, Vol. 40(18): 2431-2441.

Song, H. and D.C. Wu (2021) “A critique of tourism-led economic growth studies”, *Journal of Travel Research,* Vol. 61(4): 719-729.

Tian, F., Y. Yang and L. Jiang (2022) “Spatial spillover of transport improvement on tourism growth”, *Tourism Economics,* forthcoming.

Yang, Y. and K.K.F. Wong (2012) “A spatial econometric approach to model spillover effects in tourism flows”, *Journal of Travel Research*, Vol. 51(6): 768-778.

Yang, Y. and H. Zhang (2019) “Spatial-temporal forecasting of tourism demand”, *Annals of Tourism Research*, Vol. 75: 106-119.

Weigert, M., A. Bauer, J. Gernert, M. Karl, A. Nalmpatian, H. Küchenhoff and J. Schmude (2022) “Semiparametric APC analysis of destination choice patterns: Using generalized additive models to quantify the impact of age, period, and cohort on travel distances”, *Tourism Economics*, forthcoming.

Woodridge, J.M. (2002) “Econometric analysis of cross section and panel data”, London: MIT Press.