Annals of Tourism Research 120 (2026) 104268

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.journals.elsevier.com/annals-of-tourism-research

Research note

An endogenous spatial stochastic frontier model for tourism

Luojia Wang [a], Kerui Du [a], Federica Galli [b,*]

[a] School of Management, China Institute for Studies in Energy Policy, Xiamen University, Xiamen, 361005, Fujian, China
[b] Department of Statistical Sciences "Paolo Fortunati", University of Bologna, 40124, Bologna, Italy

ARTICLE INFO

Handling editor: Yang Yang

Keywords:
Spatial stochastic frontier analysis
Endogeneity
Tourism

Introduction

The economic impact of tourism has long been debated between two seemingly competing hypotheses. On one hand, the tourism-led growth hypothesis argues that tourist inflows stimulate the economy by bringing new information, and promoting local innovation (Balaguer & Cantavella-Jordá, 2002; Marrocu & Paci, 2011). On the other hand, the tourism resource curse hypothesis suggests that an expanding tourism sector can raise non-tradable goods prices, divert capital and labor away from manufacturing, and exacerbate capital misallocation, thereby crowding out other productive sectors (Copeland, 1991).[1]

Yet, little empirical evidence has investigated both of these distinct channels. In particular, most tourism-growth studies concentrate exclusively on tourism's effect on the production frontier (Song et al., 2012), while only a few papers suggest that tourist flows may also affect economic efficiency (Marrocu & Paci, 2011). To test both channels simultaneously, it is natural to treat tourism as a determinant of both the production frontier and the inefficiency term within stochastic frontier analysis (SFA), particularly the one-step SFA framework[2].[3] Theoretically, SFA defines efficiency as the gap between actual output and the technological frontier, reflecting an economy's ability to maximize production given its available resources.

* Corresponding author.
E-mail addresses: ljwang@stu.xmu.edu.cn (L. Wang), kerrydu@xmu.edu.cn (K. Du), federica.galli14@unibo.it (F. Galli).

[1] See Appendix A for an overview of the two hypotheses.

[2] SFA methods for studying efficiency determinants fall into two main approaches: the "two-step method" and the "one-step method." The "two-step method" involves first estimating the frontier equation to obtain the inefficiency, and then regressing the inefficiency or its variations on multiple influencing factors. However, this strategy assumes inefficiency terms are identically distributed in the first step, while explaining inefficiency variations using influencing factors in the second step, which leads to model misspecification. To overcome this technical problem, Battese and Coelli (1995) proposed a "one-step method" estimation strategy that estimates both the frontier and efficiency influencing factors, avoiding assumption contradictions and ensuring theoretical consistency and efficiency of the estimators.

[3] Empirical models that introduce an inefficiency equation related to the mean of the inefficiency error term or adopt a scaling function allow the same covariates to appear in both the frontier and inefficiency equations without raising identification concerns (see Battese & Coelli, 1995 and Kutlu et al., 2020 for technical details). For other SFA specifications, researchers should exercise caution when including identical variables in both the frontier and the inefficiency function, as identification could theoretically rely on efficiency functional form assumptions.

https://doi.org/10.1016/j.annals.2026.104268
Received 22 November 2025; Received in revised form 13 April 2026; Accepted 24 July 2026
Available online 27 July 2026
0160-7383/

However, traditional SFA relies on the independence, exogeneity, and homogeneity assumptions, which are often violated in real-world applications. Such inadequate empirical designs may explain the conflicting evidence on the economic effects of tourism (Del P. Pablo-Romero & Molina, 2013). While recent studies on tourism performance have considered spatial spillovers and competitive dynamics (Pérez-Rodríguez et al., 2024, 2026; Tian et al., 2025), the issue of endogeneity remains unresolved (Assaf & Tsionas, 2019). Addressing this gap is essential because neglecting endogeneity between frontier and environmental variables, and idiosyncratic noise within a spatial framework can bias efficiency estimates and distort causal interpretations.

In this study, we correct for endogeneity by incorporating the control function approach within the spatial SFA framework of Galli (2023). To implement this, we construct a Bartik instrument specifically tailored for tourism performance assessment and use a panel dataset of 278 cities in China from 2011 to 2020. Our results reveal a dual effect of tourism: it negatively affects the production frontier, yet it significantly reduces inefficiency. Mechanism tests show that tourism expansion exacerbates capital misallocation and crowds out the secondary industry, thereby lowering the frontier, while simultaneously stimulating local innovation through knowledge spillovers, which enhances economic efficiency. The net impact on GDP depends on the relative magnitude of these two opposing forces. This framework addresses both spatial spillovers and endogeneity, offering empirical estimates of tourism's economic consequences that account for these methodological challenges.

Methodology

A spatial stochastic frontier model

Tian et al. (2025) highlighted the pervasive spatial interactions in tourism SFA research, yet focusing exclusively on tourism's effect on the technological frontier may overlook other important economic channels. Following Marrocu and Paci (2011), we posit that tourist flows can promote efficiency by facilitating information spillovers. Thus, we simultaneously treat tourism as a frontier and an environmental variable to test its influence on both the technical frontier and economic efficiency. We adopt the one-step spatial SFA framework of Galli (2023) as follows:

$$y_{it} = \alpha_1 + \rho \sum_{j=1}^{N} w_{ij} y_{jt} + \beta_k k_{it} + \beta_l l_{it} + \frac{1}{2}\beta_{kk} k_{it}^2 + \frac{1}{2}\beta_{ll} l_{it}^2 + \beta_{kl} k_{it} l_{it} + \beta_t t + \beta_{tk} t\, k_{it} + \beta_{tl} t\, l_{it} + \frac{1}{2}\beta_{tt} t^2 + \beta_1 \mathrm{Tourism}_{it} + v_{it} - u_{it} \tag{1}$$

$$u_{it} \sim N^+\left(\mu_{it}, \sigma_u^2\right) \tag{2}$$

$$\mu_{it} = \alpha_2 + \beta_2 Tourism_{it} + \gamma_2 x_{2it} \tag{3}$$

where y_{it} represents the logarithmic output (GDP) of city i in year t, input factors k_{it} and l_{it} are capital stock and labor force, respectively, and t represents the time trend. Tourism is our key explanatory variable, defined as the logarithm of the total number of domestic and foreign tourists. A positive (negative) β_1 indicates that tourism benefits (harms) the economic frontier. v_{it} is the idiosyncratic error term, $v_{it} \sim i.i.d. N\left(0, \sigma_v^2\right)$. The Translog production frontier also includes a spatial autoregressive term, with the parameter ρ capturing productivity spillovers. The spatial weight matrix W is defined as a squared inverse distance matrix truncated at 800 km.

The inefficiency error term u_{it} measures the gap between the theoretical optimum and the actual output level, and follows a truncated normal distribution (Eq. 2). To capture the role of tourism on cities' efficiency performance, we model the mean of u_{it} based on *Tourism*. A negative β_2 coefficient indicates that tourism decreases inefficiency, thus improving overall economic efficiency. Moreover, x_{2it} is a vector of log-transformed control variables that includes financial efficiency (*Fratio*), city size (*Pop*), and marketization level (*Market*).[4]

Correcting for endogeneity with a control function

Assessing the economic effects of tourism raises concerns about potential reverse causality. Identification is further challenged by omitted factors, such as visa policies or subsidies, which may simultaneously influence both tourism activity and economic performance. To address potential correlations between frontier and/or environmental variables with the two-sided error term v_{it}, we adopt the control function approach of Kutlu et al. (2020).

Specifically, the residuals from the auxiliary regressions used to instrument the endogenous variables (originating from the frontier and/or the inefficiency equations) are incorporated in the frontier equation as correction terms. These residuals capture the endogenous components of the regressors that are correlated with the original error term v_{it}. While the correction term is explicitly included in the frontier, this procedure, within the simultaneous estimation framework, effectively purifies the entire system, including the inefficiency equation. In other words, by conditioning on the part of v_{it} that is correlated with the endogenous regressors in the frontier equation, the tourism variable entering the inefficiency equation becomes orthogonal to the remaining disturbance. As a result, this approach yields consistent parameter estimates and efficiency scores via maximum likelihood estimation (MLE), without requiring the same correction terms in the inefficiency function.

[4] See Appendix B for details of data and background.

Table 1
Main results of different specifications.

	(1)	(2)	(3)	(4)
	SF	SFE	SSF	SSFE
Frontier Function				
Tourism	−0.4454***	−0.3585***	−0.4071***	−0.3205***
	(0.0637)	(0.0635)	(0.0702)	(0.0733)
ρ			0.1392***	0.1369***
			(0.0137)	(0.0136)
k	0.5721***	0.5585***	0.5731***	0.5597***
	(0.0158)	(0.0157)	(0.0154)	(0.0153)
l	0.1640***	0.1450***	0.1356***	0.1174***
	(0.0084)	(0.0087)	(0.0087)	(0.0089)
$k \times l$	0.0295	−0.0045	0.0424	0.0089
	(0.0332)	(0.0331)	(0.0326)	(0.0325)
k^2	0.1167***	0.1289***	0.1407***	0.1523***
	(0.0244)	(0.0242)	(0.0241)	(0.0238)
l^2	0.0037	0.0167	−0.0116	0.0012
	(0.0167)	(0.0166)	(0.0164)	(0.0164)
t	0.0234***	0.0221***	0.0117***	0.0107***
	(0.0028)	(0.0028)	(0.0030)	(0.0029)
$t \times k$	−0.0479***	−0.0489***	−0.0511***	−0.0520***
	(0.0084)	(0.0083)	(0.0082)	(0.0081)
$t \times l$	0.0024	0.0039	−0.0003	0.0012
	(0.0061)	(0.0060)	(0.0060)	(0.0059)
t^2	0.0062***	0.0060***	0.0068***	0.0066***
	(0.0016)	(0.0016)	(0.0016)	(0.0016)
α_1	1.3326***	1.2457***	1.3673***	1.2710***
	(0.1573)	(0.1535)	(0.1982)	(0.1997)
Inefficiency Mean Function				
Tourism	−0.3876***	−0.3723***	−0.3474***	−0.3304***
	(0.0636)	(0.0625)	(0.0700)	(0.0723)
Fratio	0.0749***	0.0954***	0.0516**	0.0718***
	(0.0217)	(0.0221)	(0.0214)	(0.0218)
Market	−0.3455***	−0.3162***	−0.3512***	−0.3227***
	(0.0155)	(0.0160)	(0.0152)	(0.0157)
Pop	−0.5226***	−0.5354***	−0.4778***	−0.4910***
	(0.0174)	(0.0176)	(0.0176)	(0.0178)
α_2	1.4079***	1.3349***	1.4122***	1.3300***
	(0.1571)	(0.1533)	(0.1977)	(0.1993)
Bias Correction Term				
η		−0.8701***		−0.9374**
		(0.2781)		(0.4396)
Prediction function				
Bartik		0.5404***		0.5404***
		(0.0116)		(0.0116)
Fratio		0.3671***		0.3670***
		(0.0409)		(0.0409)
Market		0.2598***		0.2597***
		(0.0202)		(0.0202)
Pop		−0.0383		−0.0382
		(0.0272)		(0.0272)
α_3		0.2354***		0.2354***
		(0.0120)		(0.0120)
$ln\sigma_v$	−2.0570***	−2.0298***	−2.2106***	−2.1314***
	(0.3300)	(0.2963)	(0.5529)	(0.4548)
$ln\sigma_u$	−1.4889***	−1.5141***	−1.4726***	−1.5078***
	(0.1080)	(0.1077)	(0.1277)	(0.1321)
Observations	2780	2780	2780	2780

Notes: ***, **, * represent p-value $<$ 0.01, 0.05, and 0.1, respectively.

First, we specify a prediction function for the endogenous variable $Tourism_{it}$:

$$Tourism_{it} = \alpha_3 + z_{it}\delta + \varepsilon_{it} \tag{4}$$

where α_3 is a constant, and Z_{it} is a vector of exogenous variables including the instrument and the exogenous inefficiency determinants.

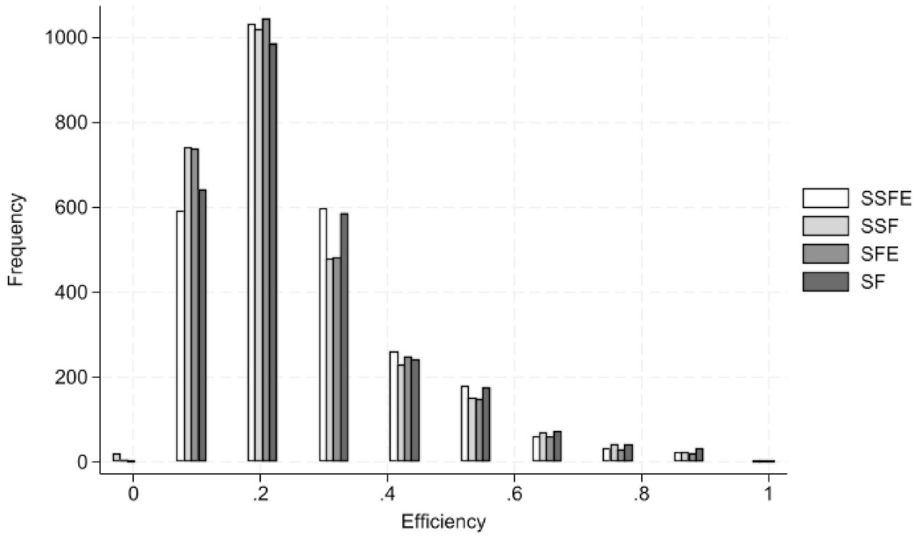

Fig. 1. Economic efficiency comparison.

The error term ε_{it} holds the endogenous component of $Tourism_{it}$ which is correlated with v_{it}. Now, v_{it} can be decomposed into a part correlated with the endogenous regressors (ε'_{it}) and an exogenous residual component (κ_{it}),[5] so that

$$v_{it} = \varepsilon'_{it}\eta + \kappa_{it} \tag{5}$$

where $\varepsilon'_{it} = (Tourism_{it} - \alpha_3 - Z_{it}\delta)'$ is the bias correction term. Substituting Eq. (5) into the frontier yields the exogenous form:

$$y_{it} = \alpha_1 + Tourism_{it}\beta_1 + \rho\sum_{j=1}^{N} w_{ij}y_{jt} + x_{1it}\gamma_1 + \varepsilon'_{it}\eta + \kappa_{it} - u_{it} \tag{6}$$

where the new error term κ_{it} is orthogonal to all regressors, ensuring consistent parameter estimation.

In addition to the exogenous control variables, Z_{it} includes a Bartik instrument. The endogenous regressor, $Tourism_{it}$, can be decomposed into a common national tourism shock and a city-specific share of national tourist flows, the latter being the source of endogeneity. Therefore, we construct the Bartik instrument using each city's pre-sample (2005) tourist share, which is predetermined and thus plausibly exogenous to subsequent local economic conditions:

$$Bartik_{it} = \ln\left[\left(\frac{Tourism_{i,2005}}{Tourism_{n,2005}}\right) \times Tourism_{nt}\right] \tag{7}$$

where $Tourism_{n,2005}$ is the total number of tourists in 2005, $Tourism_{i,2005}$ is the number of tourists of city i in 2005, $Tourism_{nt}$ is the total number of tourists visiting China in year t.

Results and discussion

We report the estimation results in Table 1 by comparing the model without the spatial term and endogeneity correction (SF), the model without the spatial autoregressive term but with a prediction function (SFE), the model with a spatially lagged term but without a prediction function (SSF), and the model with both spatial lagged term and endogeneity correction (SSFE).

The standard stochastic frontier model (SF) yields a significantly negative coefficient for tourism in the production function (-0.445). This estimate progressively moderates as we account for endogeneity (SFE) and spatial spillovers (SSF). The full SSFE model yields an estimate of -0.321, implying that omitting spatial and endogenous biases overstates tourism's negative impact. The statistically significant bias correction term (η) confirms that endogeneity poses a substantial challenge to conventional estimation. However, in spatial specifications, the coefficient of the explanatory variables cannot be directly interpreted due to spatial feedback. Marginal effects are reported in Appendix D.

Meanwhile, *Tourism* shows a significantly negative coefficient in the inefficiency equation across all specifications, consistently demonstrating its efficiency-enhancing effect. Fig. 1 shows that models ignoring spatial spillover and endogeneity may overestimate efficiency.

[5] The complete technical details, including the likelihood function, are provided in Appendix C.

In summary, the negative coefficient of tourism in the frontier equation validates the tourism resource curse hypothesis, while the negative coefficient of tourism in the inefficiency equation supports the tourism-led growth hypothesis. This dual effect remains robust under various settings (see Appendix E). The net impact of tourism on GDP is not unidimensional but depends on the dynamic game between these two opposite mechanisms (see Appendix F for the calculation of the net effect).

Mechanism tests

This section discusses how tourism changes production to influence economic growth. First, according to the theory of Copeland (1991), if the expanding tourism industry raises the price of non-tradable goods, production factors will flow from manufacturing to non-tradable sectors, resulting in a relative shrinkage of the manufacturing industry. Manufacturing is notoriously capital-intensive. Therefore, we test whether expanding tourism is correlated with increasingly chaotic capital inputs, leading to a decreasing scale of the secondary industry. Second, according to Marrocu and Paci (2011), tourism, as a supplementary path for knowledge diffusion, promotes efficiency progress in the tourism industry. Local firms learn valuable information from tourists and improve production and service accordingly. We explore the relationship between travelers and innovation at the city level to assess whether this knowledge diffusion channel holds.

Specifically, we introduce three mechanism variables: the misallocation of production factors (*KMis*); the ratio of the added value of the secondary industry to GDP (*Second*); and the patent application (*Patent*). The first two are for the tourism resource curse, and the last one is for the innovation stimulated by tourism. Among them, *KMis* is derived from a complex set of estimates. Following Popp (2002), under the assumption of the C-D production function, simple OLS regression is used to estimate the capital-output elasticity β_{Ki} first. After that, the absolute distortion coefficient of capital factor price $\widehat{\gamma}_{Ki}$ is calculated as follows:

$$\widehat{\gamma}_{Ki} = \left(\frac{K_i}{K}\right) \Big/ \left(\frac{s_i \beta_{Ki}}{\beta_K}\right) \tag{8}$$

where $s_i = \frac{y_i}{Y}$ represents the ratio of the GDP of the city i to the national GDP Y; $\beta_{Ki} = \sum_i^N s_i \beta_{Ki}$ represents the weighted capital contribution. $\widehat{\gamma}_{Ki}$ is defined as the percent of the actual capital ratio ($\frac{K_i}{K}$) compared with the optimal capital ratio when effectively allocated ($\frac{s_i \beta_{Ki}}{\beta_K}$). Then, the misallocation of production factors *KMis* is defined as follows:

$$KMis = \left| \frac{1}{\widehat{\gamma}_{Ki}} - 1 \right| \tag{9}$$

where the larger the *KMis*, the more serious the capital mismatch. In addition, we additionally control the stock of patents (*Stock*) as an explanation of innovation since the knowledge application is a cumulative process. *Stock* is calculated by the following equation:

$$Stock_{it} = \sum_{g=1985}^{t} Patent_{ig} \times \left(e^{-\beta_1 (t-g)}\right) \times \left\{1 - e^{[-\beta_2 (t-g)]}\right\} \tag{10}$$

where $Stock_{it}$ is the sum of the present value of the patent applied before; g starts from 1985, the first year China reported urban patent application data; the depreciation rate β_1 and the diffusion rate β_2 are specified as 0.22 and 0.03, following Popp (2002). The data in this section is obtained from the CEIC database and CnOpenData. Out of concern for the two-way causality of tourism and mechanism variables, this section uses instrument estimation with the same Bartik instrument. The two-stage least squares estimation is given as:

$$\widehat{Tourism}_{it} = \alpha_4 + \beta_3 Bartik_{it} + x_{2it}\gamma_2 + \varepsilon_{it} \tag{11}$$

$$y_{it} = \alpha_5 + \beta_4 \widehat{Tourism}_{it} + x_{2it}\gamma_3 + \lambda_p + \varpi_t + \mu_{it} \tag{12}$$

where λ_p and ϖ_t are the province- and year-fixed effect, capturing the conflicting influence from regional common and time-varying factors, respectively; x_{2it} represents the same vector of determinants in the inefficiency function but additionally controls *Stock* when the innovation mechanism is tested.

The results of the mechanism test are reported in Table 2. Columns (1)–(2) prove the negative impact of tourism on manufacturing: The expanding tourism industry intensifies the degree of urban capital misallocation and reduces the relative output of the secondary sector, thereby reducing the production potential. In addition, we observed in Column (3) that tourism promotes innovation at a city level.

In summary, the results support the channels through which tourism affects production by exacerbating capital misallocation, manufacturing crowding out effects, and stimulating innovation. The first two mechanisms explain the negative impact of tourism expansion on the secondary industry. The latter indicates that travelers come with valid information, activating the destination's corporate innovation potential, and improving economic efficiency.

Conclusion

In this study, we handle possible endogeneity and spatial effects in tourism performance analysis by taking advantage of the control

Table 2
Results of the mechanism test.

	(1)	(2)	(3)
	KMis	*Second*	*Patent*
Tourism	0.1227**	−0.0685***	0.0937***
	(0.0575)	(0.0097)	(0.0276)
Fratio	−0.3195***	−0.1083***	0.1945***
	(0.1073)	(0.0185)	(0.0381)
Market	0.2145***	0.3833***	0.0424
	(0.0679)	(0.0125)	(0.0294)
Pop	−0.0841*	−0.0592***	0.1865***
	(0.0492)	(0.0087)	(0.0182)
Stock			0.8128***
			(0.0154)
Obs.	2780	2780	2780

Notes: ***, **, * represent p value < 0.01, 0.05 and 0.1, respectively.

function approach within a spatial SFA framework. Due to the incidental parameters problem, accounting for unit fixed effects in our model remains a methodological gap left for future research.

CRediT authorship contribution statement

Luojia Wang: Writing – review & editing, Writing – original draft, Validation, Data curation. **Kerui Du:** Writing – review & editing, Writing – original draft, Conceptualization. **Federica Galli:** Writing – review & editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to the editor and the anonymous reviewers for their valuable comments and suggestions, which significantly improved the quality of this paper. Luojia Wang acknowledges the support of the China Scholarship Council programme (project ID. 202506310067). Kerui Du thanks the financial support from the National Natural Science Foundation of China (grant nos. 72473119 and 72074184).

Appendix. Supplementary Materials

Supplementary data to this article can be found online at https://doi.org/10.1016/j.annals.2026.104268.

Data availability

Data will be made available on request.

References

Assaf, A. G., & Tsionas, M. G. (2019). A review of research into performance modeling in tourism research—Launching the Annals of Tourism Research curated collection on performance modeling in tourism research. *Annals of Tourism Research, 76*, 266–277. https://doi.org/10.1016/j.annals.2019.04.010

Balaguer, J., & Cantavella-Jordá, M. (2002). Tourism as a long-run economic growth factor: The Spanish case. *Applied Economics, 34*(7), 877–884. https://doi.org/10.1080/00036840110058923

Battese, G. E., & Coelli, T. J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics, 20*(2), 325–332. https://doi.org/10.1007/BF01205442

Copeland, B. R. (1991). Tourism, welfare and de-industrialization in a small open economy. *Economica, 58*(232), 515. https://doi.org/10.2307/2554696

Del P. Pablo-Romero, M., & Molina, J. A. (2013). Tourism and economic growth: A review of empirical literature. *Tourism Management Perspectives, 8*, 28–41. https://doi.org/10.1016/j.tmp.2013.05.006

Galli, F. (2023). A spatial stochastic frontier model introducing inefficiency spillovers. *Journal of the Royal Statistical Society Series C: Applied Statistics, 72*(2), 346–367. https://doi.org/10.1093/jrsssc/qlad012

Kutlu, L., Tran, K. C., & Tsionas, M. G. (2020). A spatial stochastic frontier model with endogenous frontier and environmental variables. *European Journal of Operational Research, 286*(1), 389–399. https://doi.org/10.1016/j.ejor.2020.03.020

Marrocu, E., & Paci, R. (2011). They arrive with new information. Tourism flows and production efficiency in the European regions. *Tourism Management, 32*(4), 750–758. https://doi.org/10.1016/j.tourman.2010.06.010

Pérez-Rodríguez, J. V., Suárez-Vega, R., & Hernández, J. M. (2024). Spatial spillover effects on the efficiency of P2P accommodation units. *Journal of Destination Marketing & Management, 34*, Article 100945. https://doi.org/10.1016/j.jdmm.2024.100945
Pérez-Rodríguez, J. V., Suárez-Vega, R., & Hernández, J. M. (2026). A spatial stochastic frontier model at the P2P listing level. *Tourism Management, 113*, Article 105339. https://doi.org/10.1016/j.tourman.2025.105339
Popp, D. (2002). Induced innovation and energy prices. *American Economic Review, 92*(1), 160–180.
Song, H., Dwyer, L., Li, G., & Cao, Z. (2012). Tourism economics research: A review and assessment. *Annals of Tourism Research, 39*(3), 1653–1682. https://doi.org/10.1016/j.annals.2012.05.023
Tian, F., Wen, Z., & Yang, Y. (2025). Spatial spillovers in tourism performance. *Annals of Tourism Research, 114*, Article 104005. https://doi.org/10.1016/j.annals.2025.104005

Luojia Wang research interests include efficiency analysis.

Kerui Du research interests include applied econometrics, energy, and environmental economics.

Federica Galli research interests include spatial and regional statistics, productivity and efficiency analysis, well-being and consumption behavior, and tourism.