Annals of Tourism Research 78 (2019) 102738

Contents lists available at ScienceDirect

Annals of Tourism Research

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

Scientific value of econometric tourism demand studies

Ulrich Gunter, Irem Önder, Egon Smeral*

MODUL University Vienna, Department of Tourism and Service Management, Am Kahlenberg 1, A-1190 Vienna, Austria

ARTICLE INFO

Associate Editor: Haiyan Song

Keywords:
Substantive significance
Statistical significance
Study limitations
Justification of chosen methods

ABSTRACT

The objective of this paper was to evaluate the scientific value of econometric tourism demand studies. Based on a questionnaire answered by ourselves we analyzed articles published in Annals of Tourism Research, Journal of Travel Research, Tourism Management, and Tourism Economics during the period 2007 to 2017. The evaluation showed that current scientific practice generally failed to differentiate between substantive (economic) significance and statistical significance, and used these terms interchangeably in many cases. In line with these flaws, most authors avoided discussing the estimation results in terms of their size and their reliability, as well as failing to adequately address the limitations of their studies and to justify the chosen methods.

Introduction

First, we have to explain why we started this project: after reading many econometric tourism demand studies over the last few years, we gained the impression that the focus of the publications is predominantly on the statistical significance of estimated coefficients and that less – or no – attention is paid to their substantive (economic) significance. Authors appeared rather vague about the distinction between substantive and statistical significance, with the two terms being used interchangeably in many cases, and questions regarding the size of the measured effects playing only an unimportant role or being overlooked altogether. In the end, we asked ourselves what is the scientific value of such studies or, in other words, what are we supposed to take away as key messages.

One possible reason for the observed tendency to focus predominantly on statistical significance, at the expense of substantive significance, might be related to publication pressures in the academic world, whereby authors constantly face the risk of submissions with statistically insignificant results being rejected. Rosenthal (1979) pointed out that the over-representation of statistically significant results in published works leads to an overall bias, through which the size of an effect is likely to be over- or under-estimated, because those studies that failed to achieve statistically significant results are less likely to be published. Oakes (1986), in turn, stated that the significance (in the sense of importance) of a result depends on the size of the effect found and whether it can be replicated.

We became further alerted as we discovered that a good part of econometric tourism demand studies celebrate the increasing number of statistical significance tests, yet often fail to justify their choice of estimation approaches and frequently avoid a discussion of the results in a broader context. Our awareness about the observed problems increased as we read the general recommendations provided in most standard textbooks as to how one should carry out the general-to-specific modeling approach:

- First, authors recommend the construction of a general demand model in the form of an Autoregressive Distributed Lag Model (ADLM). This model should consider a large number of explanatory variables, including the lagged dependent and lagged explanatory variables. Economic theory suggests the possible variables to be included, and the nature of the data suggests the lag length.

* Corresponding author.
E-mail addresses: ulrich.gunter@modul.ac.at (U. Gunter), irem.onder@modul.ac.at (I. Önder), egon.smeral@modul.ac.at (E. Smeral).

https://doi.org/10.1016/j.annals.2019.06.005
Received 1 April 2019; Received in revised form 29 May 2019; Accepted 12 June 2019
Available online 27 June 2019

- Second, different tests are recommended to test various restrictions in order to achieve a parsimonious but statistically significant specification.
- Third, the normal diagnostics tests should be used to examine whether the final model is statistically acceptable or not.
- Fourth, the final model can be used for policy evaluation and/or forecasting.

Following these guidelines, it is entirely possible that, in the worst case, we end up with a statistically significant model incorporating unimportant variables and a non-replicable approach. In this sense, Ronald A. Fisher (1925) was wrong, as the quality of fit is not the same thing as scientific findings and/or the size/importance of the impact (Ziliak & McCloskey, 2014). We have to point out at this stage that in statistical usage/practice, 'significant' means signifying a characteristic of the population from which the sample is drawn, regardless of whether the characteristic is important. Wonnacott and Wonnacott (1990) stated that there is a problem with the term "statistically significant" as it is a technical phrase that simple means enough data have been collected to establish that a difference does exist, irrespective of whether this difference is important or not. Additionally, Wooldridge (2000); Wooldrige (2004) agreed strongly with the notion that statistical significance is neither a necessary nor a sufficient condition for economic significance. Similarly, Wolpin (2013) pointed out the existing limitations of statistical inference without theory (Rust, 2014).

The literature has already revealed that in many cases scientists assume that statistical significance equals substantive (economic) significance and that they fail to focus on both criteria: The findings of Ziliak and McCloskey (2014) showed that 43% of the papers published in American Economic Review (AER) in the 1990s did not distinguish between statistical and economic significance in their conclusions. Although are this figure is alarming, it actually represents an improvement over the corresponding evaluation of published papers in AER during the 1980s when the figure stood at 70%, thus showing that the scientific practice improved over the course of one decade (Ziliak & McCloskey, 2014). Given these facts we hope that the scientific practice will continue to improve, according to the general guideline that the **CON** should stay in e**CON**ometrics and we do not lose it (Leamer, 1983; McAleer, Pagan, & Volker, 1985).

The impressions and observations mentioned above, together with the specific research outcomes, were drivers to ask: what is the current standing of tourism econometrics? To address this question, which, to the best of our knowledge has not been done to date, we designed a project to analyze the scientific value of econometric tourism demand studies. To construct a database, we analyzed studies from four tourism journals (Annals of Tourism Research, Journal of Travel Research, Tourism Management, and Tourism Economics) published during the period 2007 to 2017, selected according to certain criteria. In order to address our research objectives, we developed a questionnaire appraising the scientific value of econometric tourism demand studies. After explaining the questionnaire, we present and discuss our findings based on the evaluation of our own answers to the different questions. Finally, some overall conclusions are drawn.

A survey of scientific practice

The major objective of the study is to determine whether extant scientific practice acknowledges the difference between substantive (economic) and statistical significance in a clear way, or if both concepts are used as equivalent terms. To construct a database we analyzed all full-length papers on econometric tourism demand modeling published in Annals of Tourism Research, Journal of Travel Research, Tourism Management, and Tourism Economics during the period 2007 to 2017 as an unbiased selection of best-practice examples. We selected 115 papers and considered only studies employing causal multivariate models and economic explanatory variables that were published in any of the four tourism journals.

As noted by Kirilenko and Stepchenkova (2018), Annals of Tourism Research, Journal of Travel Research, and Tourism Management have been regarded as the "Big Three" journals in the tourism discipline for > 40 years. On the other hand, the percentage of studies on the topic of tourism demand published in the three aforementioned journals has fallen over time (Kirilenko & Stepchenkova, 2018), which is due to the emergence of more specialist journals, with Tourism Economics being a particularly suited outlet for (econometric) tourism demand studies. In selecting the sample according to these guidelines, we disregarded the following types of papers: research notes, pure forecasting studies without sections presenting the demand modeling approach and the estimation results, studies without economic explanatory variables (e.g., those using web-based indicators such as Google Analytics, Google Trends, etc.), studies based on non-causal univariate models (e.g., those using exponential smoothing methods or some form of ARIMA modeling, etc.), studies based on questionnaires and surveys, purely conceptual/theoretical contributions, studies relying exclusively on simulations rather than using real data, papers exclusively focusing on certain statistical testing procedures (e.g., unit root and co-integration testing, Granger causality testing, etc.), as well as meta studies and review papers.

Each paper from the sample was evaluated by ourselves based on the same 15 "yes" or "no" questions to guarantee their consistent appraisal, thus also showing if the authors worked out the differences between statistical and substantive significance exactly enough. In doing so, we also highlighted the indirectly related criteria indicating if the authors of the investigated papers dealt properly with the main objective of our own study.

From these 15 questions, the first three were related to data use, questions four to seven dealt with modeling issues, questions eight to twelve focused on the results, and questions thirteen to fifteen evaluated the discussion and interpretation of the results. Another block of five questions focused on some descriptive statistical information such as other contents of the study, the estimation technique, or the scope of the study.

The evaluated questions were as follows:

Q1 Are the descriptive statistics of the variables included (including their units)? In our opinion, empirical work should include the units of the variables (including their means). This is necessary as readers cannot judge whether something is large or small when it is reported without units or scale.
Q2 Are the data graphed and inspected for trends, seasonal patterns, structural breaks, etc.? A careful analysis of the data is essential for choosing the optimal treatment in order to guarantee an unbiased estimation.
Q3 Has there been proper data adjustment (unit root testing, first differencing, seasonal adjustment, logarithms, structural break testing, etc.)? Careful data adjustment is important to obtain reliable and interpretable results.
Q4 Do the employed models conform to the data characteristics? In our view, specific data characteristics (e.g., long-term vs. short-term data, length of time series, cross section situations, etc.) also inform the choice of appropriate analytic approaches.
Q5 Do the authors justify the method(s) chosen? An appropriate justification of the method(s) chosen, including discussion of the pros and cons of the approach used in comparison with competing solutions, is an important part of the scientific process.
Q6 Is there a discussion of why certain variables in the chosen approach are considered as important? This discussion could be based on theoretical considerations, own research experiences, or results from other studies.
Q7 Is there a discussion of why certain variables are not considered in the chosen approach? This issue is important, because it fortifies the justification of the method used (see **Q5**).
Q8 Are the coefficients in elasticity form (or in some other interpretable form)? The estimated coefficients should be in a problem-relevant form so that the reader can discern the impact of the variables. This is also an important issue as sometimes authors neglect to state the actual effect and publish only tables of coefficients with asterisks indicating the levels of significance (so-called “asterisks econometrics”; Ziliak & McCloskey, 2014).
Q9 Are the estimated coefficients carefully interpreted? This question summarizes whether or not the authors discuss the algebraic signs and the direction of the effects according to the principles of economic theory.
Q10 Are the results validated? Here we analyze whether the paper employs simulations and/or robustness checks.
Q11 Are specific significance tests avoided when they are not relevant? This question determines whether the paper celebrates statistics by reporting many test results, even though most/all of the significance tests used are not appropriate to the situation.
Q12 Is statistical significance used as a criterion of scientific importance? This key question reflects if the authors separate substantive and statistical significance in a clear way as well as if they interpret the economic impact of their estimates. Furthermore, we also look for cases in which statistical significance is used as an argument to avoid discussions about the relevance of the results.
Q13 Does the paper mention the similar work of others, and are the results discussed in a broader context? This important question reveals whether the authors consider the ongoing scientific discussion within which an estimated impact could be judged as being “strong”, “weak”, or reliable at all.
Q14 Are the limitations of the study results discussed? In this case, we evaluate if the authors are aware about the limitations of the method chosen, the data used, and the variables considered as well as the consequences on the estimated coefficients.
Q15 Does the study suggest topics for future research efforts? This point indicates whether the authors truly acknowledge the limitations of their own results, by recognizing the value of future research and providing sage advice in that direction.

As mentioned before, the last block of five questions focused on some descriptive statistical information.

Q16 Does the study also cover tourism demand forecasting?
Q17 Does the study employ panel-data techniques?
Q18 Is the scope of the study short-term oriented?
Q19 Is the scope of the study long-term oriented?
Q20 Does the study analyze destination data at the country level?

The data

In total, 115 papers were analyzed for the present study. For the list of the journal titles, article titles, and more bibliographic information on all studies in chronological order see Table 1. The titles and keywords of the articles were analyzed using NVivo 12, and Figs. 2 and 3 were produced using this software. The other figures were created by the authors using Microsoft Excel.

Publications based on the institutions

The descriptive statistics of the articles, such as the institutions associated with the authors of each study, were analyzed. Fig. 1 shows the number of publications per institution, while an accompanying footnote discloses the institutions' names in detail. Out of the 121 institutions in the study sample, the most frequently occurring ones are: The Hong Kong Polytechnic University (16.5%), University of Surrey (12.1%), University of Bologna (5.8%), The Austrian Institute of Economic Research (5%), and Victoria University (5%).

Keywords analysis

We investigated the keywords of the studies to figure out the most common themes in the research area of econometric tourism

Table 1
Journals and titles in chronological order.
Source: Authors' own elaboration.

Journal title	Article title	Author names	Year	Vol. (Iss.)	Pages
Tourism Management	Tourism in the Balearic Islands: A dynamic model for international demand using panel data	Garín-Muñoz, T., Montero-Martín, L. F.	2007	28	1224–1235
Tourism Management	German demand for tourism in Spain	Garín-Muñoz, T.	2007	28	12–22
Tourism Economics	Measuring the economic impact of Australian tourism marketing expenditure	Kulendran, N., Divisekera, S.	2007	13(2)	261–274
Tourism Economics	Analysis of foreign tourism demand for Croatian destinations: long-run elasticity estimates	Mervar, A., Payne, J. E.	2007	13(3)	407–420
Journal of Travel Research	Analyzing and Forecasting Tourism Demand: A Rough Sets Approach	Goh, C., Law, R., Mok, H. M. K.	2008	46	327–338
Journal of Travel Research	The Impact of Distance on International Tourist Movements	McKercher, B., Chan, A., Lam, C.	2008	47(2)	208–224
Tourism Management	Assessing impacts of SARS and Avian Flu on international tourism demand to Asia	Kuo, H-I., Chen, C-C., Tseng, W-C. Ju, L-F., Huang, B-W.	2008	29	917–928
Tourism Management	Co-integration analysis of quarterly European tourism demand in Tunisia	Ouerfelli, C.	2008	29	127–137
Tourism Economics	A model of tourism demand for Tunisia: inclusion of the tourism investment variable	Choyakh, H.	2008	14(4)	819–838
Tourism Economics	The role of economic development in tourism demand	Eugenio-Martín, J. L., Martín-Morales, N., Sinclair, M. T.	2008	14(4)	673–690
Tourism Economics	Determinants of inbound tourism to South Africa	Saayman, A., Saayman, M.	2008	14(1)	81–96
Journal of Travel Research	The Impact of the Financial and Economic Crisis on European Tourism	Smeral, E.	2009	48(1)	3–13
Journal of Travel Research	Impacts of the Financial and Economic Crisis on Tourism in Asia	Song, H., Lin, S.	2009	49(1)	16–30
Tourism Management	The analysis of the relationships of Korean outbound tourism demand: Jeju Island and three international destinations	Seo, J. H., Park, S. Y., Yu, L.	2009	30	530–543
Tourism Management	The impact of crisis events and macroeconomic activity on Taiwan's international inbound tourism demand	Wang, Y.-S.	2009	30	75–82
Tourism Economics	Estimation of outbound Italian tourism demand: a monthly dynamic EC-LAIDS model	Cortés-Jiménez, I., Durbarry, R., Pulina, M.	2009	15(3)	547–565
Tourism Economics	Ex post demand for Australian tourism goods and services	Divisekera, S.	2009	15(1)	153–180
Tourism Economics	Quality, quantity and duration decisions in household demand for vacations	Fleischer, A., Rivlin, J. (Byk)	2009	15(3)	513–530
Tourism Economics	Tourism in Galicia: domestic and foreign demand	Garín-Muñoz, T.	2009	15(4)	753–769
Tourism Economics	Estimating the impact of avian flu on international tourism demand using panel data	Kuo, H-I., Chang, C-L., Huang, B-W., Chen, C-C., McAleer, M.	2009	15(3)	501–511
Tourism Economics	Dynamics in the specification of tourism demand models	Morley, C. L.	2009	15(1)	23–39
Tourism Economics	An analysis of the relationship between transport capital and tourism development in a dynamic framework	Seetanah, B., Khadaroo, J.	2009	15(4)	785–802
Tourism Economics	Impacts of climate change on domestic tourism in the UK: a panel data estimation	Taylor, T., Ortiz, A. R.	2009	15(4)	803–812
Annals of Tourism Research	Confidence Intervals for Tourism Demand Elasticities	Song, H., Kim, J., Yang, S.	2010	37(2)	377–396
Journal of Travel Research	Use of Dynamic Panel Cointegration Approach to Model International Arrivals to Australia	Seetaram, N.	2010	49(4)	414–422
Journal of Travel Research	Impacts of the World Recession and Economic Crisis on Tourism: Forecasts and Potential Risks	Smeral, E.	2010	49(1)	31–38
Tourism Management	Climate in the region of origin and destination choice in outbound tourism demand	Eugenio-Martín, J., L., Campos-Soria, J., A.	2010	31	744–753
Tourism Management	A dynamic panel data analysis of snow depth and winter tourism	Falk, M.	2010	31	912–924
Tourism Management	Tourism demand for Italy and the business cycle	Guizzardi, A., Mazzocchi, M.	2010	31	367–377
Tourism Economics	Disney's return to theme park dominance in Florida	Braun, B. M., Soskin, M. D.	2010	16(1)	235–250
Tourism Economics	Economics of leisure and non-leisure tourist demand: a study of domestic demand for Australian tourism	Divisekera, S.	2010	16(1)	117–136
Tourism Economics	Factors affecting international tourism flows to Turkey: a gravity model approach	Eryigit, M., Kotil, E., Eryigit, R.	2010	16(3)	585–595
Tourism Economics	An econometric analysis of the aggregate outbound tourism demand of Turkey	Halicioglu, F.	2010	16(1)	83–97
Tourism Economics	Exchange rate regimes and tourism	Santana-Gallego, M., Ledesma-Rodríguez, F. J., Pérez-Rodríguez, J. V.	2010	16(1)	25–43
Tourism Economics	Using the panel cointegration approach to analyse the determinants of tourism demand in South Africa	Seetanah, B., Durbarry, R., Ragodoo, J. F. N.	2010	16(3)	715–729
Tourism Economics	Tourism demand modelling and forecasting: how should demand be measured?	Song, H., Li, G, Witt, S. F., Fei, B.	2010	16(1)	63–81
Journal of Travel Research	Determinants versus Composite Leading Indicators in Predicting Turning Points in Growth Cycle	Kulendran, N., Wong, K., F., K.	2011	50(4)	417–430
Journal of Travel Research	Analyzing Tourist Consumption: A Dynamic System-of-Equations Approach	Wu, D., C., Li, G., Song, H.	2011	50(1)	46–56
Tourism Management	Modelling Australian domestic and international inbound travel: a spatial-temporal approach	Denga, M., Athanasopoulos, G.	2011	32	1075–1084

(*continued on next page*)

Table 1 (*continued*)

Journal title	Article title	Author names	Year	Vol. (Iss.)	Pages
Tourism Management	Estimating time series and cross section tourism demand models: Mainland United States to Hawaii data	Nelson, L. A., Dickey, D. A., Smith, J. M.	2011	32	28–38
Tourism Management	Demand elasticity estimates for New Zealand tourism	Schiff, A., Becken, S.	2011	32	564–575
Tourism Management	Impact of financial/economic crisis on demand for hotel rooms in Hong Kong	Song, H., Lin, S., Witt, S. F., Zhang, X.	2011	32	172–186
Tourism Economics	Impact of Approved Destination Status on Chinese travel abroad: an econometric analysis	Arita, S., Edmonds, C., Lacroix, S., Mak, J.	2011	17(5)	983–996
Tourism Economics	Crowding-out effects of cruise tourism on stay-over tourism in the Caribbean: non-parametric panel data evidence	Bresson, G., Logossah, K.	2011	17(1)	127–158
Tourism Economics	Macroeconomic factors influencing Malaysian tourism revenue, 2002–2008	Jayaraman, K., Lin, S. K., Haron, H., Ong, W. O.	2011	17(6)	1347–1363
Tourism Economics	International tourism and trade flows: a causality analysis using panel data	Keum, K.	2011	17(5)	949–962
Tourism Economics	Quantile elasticity of international tourism demand for South Korea using the quantile autoregressive distributed lag model	Li, H., Park, S. Y., Seo, J. H.	2011	17(5)	997–1015
Tourism Economics	What's pushing international tourism expenditures?	Vietze, C.	2011	17(2)	237–260
Annals of Tourism Research	Exploring Impact of Climate Tourism Demand	Goh, C.	2012	39(4)	1859–1883
Annals of Tourism Research	Forecasting Tourist arrivals in Greece and impact of macroeconomic shocks from the countries of tourists' origins	Gounopoulos, D., Petmezas. D., Santamaria, D.	2012	39(2)	641–666
Annals of Tourism Research	International Tourism Demand and the Business Cycle	Smeral, E.	2012	39(1)	379–400
Annals of Tourism Research	Economic Analysis of Tourism Consumption Dynamics: a Time Varying Parameter Demand System Approach	Wu, D., Li, G., Song, H.	2012	39(2)	667–685
Journal of Travel Research	The Impact of the EU Emissions Trading System on Air Passenger Arrivals in the Caribbean	Blanc, E., Winchester, N.	2012	52(3)	353–363
Journal of Travel Research	Assessing the Impacts of the Global Economic Crisis and Swine Flu on Inbound Tourism Demand in the United Kingdom	Page, S., Song, H., Wu, D. C.	2012	51(2)	142–153
Journal of Travel Research	A Spatial Econometric Approach to Model Spillover Effects in Tourism Flows	Yang, Y., Wong, K., K., F.	2012	51(6)	768–778
Tourism Management	The determinants of Italian domestic tourism: A panel data analysis	Massidda, C., Etzo, I.	2012	33	603–610
Tourism Management	Estimating the long-run effects of socioeconomic and meteorological factors on the domestic tourism demand for Galicia (Spain)	Otero-Giráldez, M. S., Álvarez-Díaz, M., González-Gómez, M.	2012	33	1301–1308
Tourism Management	Academic tourism demand in Galicia, Spain	Rodríguez, X. A., Martínez-Roget, F., Pawlowska, E.	2012	33	1583–1590
Tourism Economics	Assessing the impact of shocks on international tourism demand for Portugal	Daniel, A. C. M., Rodrigues, P. M. M.	2012	18(3)	617–634
Tourism Economics	Visits to national parks and hiking areas: a panel data analysis of their sociodemographic, economic and site quality determinants	Nerg, A., Uusivuori, J., Mikkola, J., Neuvonen, M., Sievänen, T.	2012	18(1)	77–93
Tourism Economics	Modelling international tourism demand for the Caribbean	Onafowora, O. A., Owoye, O.	2012	18(1)	159–180
Tourism Economics	Estimating demand elasticities for Australia's international outbound tourism	Seetaram, N.	2012	18(5)	999–1017
Tourism Economics	Cultural effects on inbound tourism into the USA: a gravity approach	Vietze, C.	2012	18(1)	121–138
Annals of Tourism Research	Economic Impacts of Visa Restrictions on Tourism: a Case of Two Events in China	Shina, L., Song, H.	2013	43	251–271
Tourism Management	Impacts of exported Turkish soap operas and visa-free entry on inbound tourism to Turkey	Balli, F., Ozer Balli, H., Cebeci, K.	2013	37	186–192
Tourism Economics	Estimating dynamic asymmetries in demand at the Munich Oktoberfest	Süssmuth, B., Woitek, U.	2013	19(3)	653–674
Annals of Tourism Research	Estimating demand elasticities in non-stationary panels: The case of Hawaii tourism	Fuleky, P., Zhao, Q., Bonham, C.	2014	44	131–142
Journal of Travel Research	Habit Persistence in Air Passenger Traffic Destined for Florida	Cazanova, J., Ward W. R., Holland, S.	2014	53(5)	638–655
Journal of Travel Research	The Moderation Effects of Political Instability and Terrorism on Tourism Development: A Cross-Country Panel Analysis	Saha, S., Yap, G.	2014	53(4)	509–521
Journal of Travel Research	Air Passenger Duty and Outbound Tourism Demand from the United Kingdom	Seetaram, N., Song, H., Page, S., J.	2014	53(4)	476–487
Tourism Management	Impacts of seasonal patterns of climate on recurrent fluctuations in tourism demand: Evidence from Aruba	Ridderstaat, J., Oduber, M., Croes, R., Nijkamp, P., Martens, P.	2014	41	245–256
Tourism Management	Domestic tourism demand of urban and rural residents in China: Does relative income matter?	Yang, Y., Liu, Z.-L., Qi, Q.	2014	40	193–202
Tourism Economics	The determinants of international academic tourism demand in Europe	Cerdeira Bento, J. P.	2014	20(3)	611–628
Tourism Management	Demographic change, tourism expenditure and life cycle behaviour	Bernini, C., Cracolici, M. F.	2015	47	191–205
Tourism Economics	How does advertising affect the price elasticity of lodging demand? Evidence from Taiwan	Chen, C-M., Lin, Y-C., Tsai, Y-C.	2015	21(5)	1035–1045
Tourism Economics	Dynamic impacts of income and the exchange rate on US tourism, 1960–2011	Chi. J.	2015	21(5)	1047–1060
Tourism Economics	The demand for winter sports: empirical evidence for the largest French ski-lift operator	Falk, M.	2015	21(3)	561–580
Tourism Economics	Demand for international medical travel to the USA	Johnson, T. J., Garman, A. N.	2015	21(5)	1061–1077

(*continued on next page*)

Table 1 (*continued*)

Journal title	Article title	Author names	Year	Vol. (Iss.)	Pages
Tourism Economics	Does internal migration affect Italian domestic tourism? A panel data analysis	Massidda, C., Piras, R.	2015	21(3)	581–600
Tourism Economics	A dynamic analysis of tourism determinants in Sicily	Provenzano, D.	2015	21(3)	441–454
Tourism Economics	An ARDL bounds test approach to modelling tourist expenditure in South Africa	Saayman, A., Saayman, M.	2015	21(1)	49–66
Tourism Economics	Tourism demand analysis of Chinese arrivals in Thailand	Untong, A., Ramos, V., Kaosa-Ard, M., Rey-Maquieira, J.	2015	21(6)	1221–1234
Annals of Tourism Research	Measuring price elasticities of demand for outbound tourism using competitiveness indices	Seetarama, N., Forsyth, P., Dwyer, L.	2016	56(4)	65–79
Journal of Travel Research	Why Can Package Tours Hurt Tourists? Evidence from China's Tourism Demand in Hong Kong	Chen, Y., Schuckert, M., Song, H., Chon, K.	2016	55(4)	427–439
Journal of Travel Research	Modeling the Impact of Australia's Mining Boom on Tourism: A Classic Case of Dutch Disease	Dwyer, L., Pham, T., Jago, L., Bailey, G., Marshall, J.	2016	55(2)	233–245
Journal of Travel Research	Modeling Caribbean Tourism Demand: An Augmented Gravity Approach	Lorde, T., Li, G., Airey, D.	2016	55(7)	946–956
Tourism Management	Nonconstant reputation effect in a dynamic tourism demand model for Spain	Albaladejo, I. P., González-Martínez, M. I., Martínez-García, M. P.	2016	53	132–139
Tourism Management	Seasonal concentration of tourism demand: Decomposition analysis and marketing implications	Fernández-Morales, A., Cisneros-Martínez, J. D., McCabe, S.	2016	56	172–190
Tourism Economics	Effect of terrorism on demand for tourism in Kenya	Buigut, S., Amendah, D. D.	2016	22(5)	928–938
Tourism Economics	Interdependencies of demand for international air transportation and international tourism	Divisekera, S.	2016	22(6)	1191–1206
Tourism Economics	The decline of tourism income elasticities in a global context	Gunter, U., Smeral, E.	2016	22(3)	466–483
Tourism Economics	The economic dimension of climate change impacts on tourism: The case of Portugal	Pintassilgo, P., Rosselló, J., Santana-Gallego, M., Valle, E.	2016	22(4)	685–698
Tourism Economics	The influence of formal trade agreements and informal economic cooperation on international tourism flows	Saayman, A., Figini, P., Cassella, S.	2016	22(6)	1274–1300
Annals of Tourism Research	Modelling the interdependence of tourism demand: The global vector autoregressive approach	Cao, Z., Li, G., Song, H.	2017	67	1–13
Annals of Tourism Research	Tourism expenditures and crisis transmission: A general equilibrium GVAR analysis with network theory	Konstantinos, N. K., George, S., Panayotis, G. M.	2017	66(4)	74–94
Journal of Travel Research	Tourism Demand Modeling by Purpose of Visit and Nationality	Cortés-Jiménez, I., Blake, A.	2017	50(4)	408–416
Journal of Travel Research	Asymmetric Business Cycle Effects and Tourism Demand Cycles	Croes, R., Ridderstaat, J., Rivera, M.	2017	Online first	
Journal of Travel Research	Relative Climate Index and Its Effect on Seasonal Tourism Demand	Li, H., Goh, C., Hung, K., Li Chen, J.	2017	Online first	
Journal of Travel Research	A Dynamic Panel Data Analysis of Climate and Tourism Demand: Additional Evidence	Li, H., Song, H., Li, L.	2017	56(2)	158–171
Journal of Travel Research	The Puzzle of Tourism Demand at Destinations Hosting UNESCO World Heritage Sites: An Analysis of Tourism Flows for Italy	Ribaudo, G., Figini, P.	2017	56(2)	187–205
Journal of Travel Research	The Link between Money Supply and Tourism Demand Cycles: A Case Study of Two Caribbean Destination	Ridderstaat, R., Croes, R.	2017	56(2)	187–205
Journal of Travel Research	Does Political and Economic Freedom Matter for Inbound Tourism? A Cross-National Panel Data Estimation	Saha, S., Su, S., Campbel, N.	2017	56(2)	221–234
Journal of Travel Research	Variations in Seasonal Outbound Travel across the Business Cycles	Smeral, E.	2017	57(7)	936–946
Journal of Travel Research	Designing a Tourism Spillover Index Based on Multidestination Travel: A Two-Stage Distance-Based Modeling Approach	Yang, Y., Fik, T., F., Zhang, H.	2017	56(3)	317–333
Journal of Travel Research	The Impact of Climate Variables on Seasonal Variation in Hong Kong Inbound Tourism Demand	Zhang, H., Q., Kulendran, N.	2017	56(1)	94–107
Tourism Management	Remodeling international tourism demand: Old theory and new evidence	Dogru, T., Sirakaya-Turk, E., Crouch, G.	2017	60	47–55
Tourism Management	An empirical analysis of the influence of macroeconomic determinants on World tourism demand	Martins, L. F., Gan, Y., Ferreira-Lopes, A.	2017	61	248–260
Tourism Management	The determinants of Chinese visitors to Australia: A dynamic demand analysis	Pham, T. D., Nghiem, S., Dwyer, L.	2017	63	268–276
Tourism Economics	Turkish tourism, exchange rates and income	Akay, G. H., Cifter, A., Teke, O.	2017	23(1)	66–77
Tourism Economics	The spillover effect of attractions: Evidence from Eastern China	Bo, Z., Bi, Y, Hengyun, L., Hailin, Q.	2017	23(4)	731–743
Tourism Economics	The effects of business cycles on tourism demand flows in small island destinations	Croes, R., Ridderstaat, J.	2017	23(7)	1451–1475
Tourism Economics	The main determinants effecting international visitor arrivals in New Zealand: Some empirical evidence	Gani, A., Clemes, M. D.	2017	23(5)	921–940
Tourism Economics	Golf, supply and demand: The influence of economic factors	Barcena-Martín, E., Rodríguez-Fernández, M., Borrego-Domínguez, S.	2017	23(6)	1220–1234
Tourism Economics	Fishing for answers? Impacts of marine ecosystem quality on coastal tourism demand	Otrachshenko, V., Bosello, F.	2017	23(5)	963–980
Tourism Economics	Assessing the determinants of local tourism demand: A simultaneous equations model for the Italian provinces	Zamparini, L., Vergori, A. S., Arima, S.	2017	23(5)	981–992

Fig. 1. Author institutions based on the number of publications. Source: Authors' own illustration.
Names of the institutions from the top to the bottom of the chart: The Hong Kong Polytechnic University, University of Surrey, University of Bologna, The Austrian Institute of Economics (WIFO), Victoria University, University of Hawaii at Manoa, University of Central Florida, University of Balearic Islands, North-West University, Universiti Sains Malaysia, University of Florida, Universidad de Málaga, MODUL University Vienna, Bournemouth University, University of South Carolina, Griffith University, Monash University, University of New South Wales, University of Mauritius, Nottingham University, The Chinese University of Hong Kong, UNED, North Carolina State University, National Technical University of Athens, National Chung Hsing University, Abant Izzet Baysal University, George Washington University.

demand modeling. Fig. 2 shows the most commonly used keywords in the study sample as a word cloud. As can be seen “tourism” and “demand” are the most commonly used keywords in the data sample followed by the methodological keywords “model” and “panel”. The fifth most commonly used keyword category is “elasticity” which includes price and income elasticities of demand. The results show that keywords mostly focus on the topic of the publication such as tourism demand analysis and also include the employed model (e.g. gravity model) and the data type (e.g. panel data). Since the sensitivity of tourism demand to variations in explanatory variables is typically explained in terms of elasticities this is also one of the most common keywords used in the publications' keywords section.

Fig. 2. Keyword analysis word cloud.
Source: Authors' own illustration.

Titles of the studies

The titles of the studies were analyzed to see if any themes emerged other than those captured by the keywords. The results were similar to the keyword analysis and mainly grouped in two categories: “tourism” and “analysis” (see Table 2). The majority of the publications have tourism in their title, most include the analysis method, and some note the region that was investigated in the study (i.e., Italy, Europe, China, etc.). Fig. 3 shows graphically by size how often the titles are used as a tree map.

Descriptive statistics according to the questionnaire

We summarized the main outcomes of the questionnaire as follows (see Fig. 4a–d): While tourism demand studies at more granular geographical levels are becoming more popular, for instance due to the increasing importance of city tourism (Gunter, Önder, & Gindl, 2019), 63% of the sampled econometric tourism demand studies still concentrate on destination data at the country level. Only 13 of the 115 analyzed studies additionally considered tourism demand forecasting. This observation implies that only a small minority of authors in our sample investigated the out-of-sample applicability of their in-sample findings. Given the large amount of tourism demand forecasting studies in the literature (see, e.g., Song, Qiu, & Park, 2019, or Jiao & Chen, 2018, for recent surveys of the literature), this also may indicate that tourism demand studies typically do not put much emphasis on presenting and discussing the demand modeling approach and the estimation results.

Moreover, papers employing panel-data techniques (50 out of 115) and those covering tourism demand forecasting were mutually exclusive. In other areas of forecasting, panel-data techniques have been frequently applied for many years (e.g., Baltagi & Griffin, 1997; Brucker & Siliverstovs, 2006; Hoogstrate, Palm, & Pfann, 2000; Maddala, Trost, Li, & Joutz, 1997), mostly because pooling the data across entities increases the sample size while fewer parameters need to be estimated. This is advantageous because pooling results in a more parsimonious model specification with more degrees of freedom typically increases the efficiency and stability of the parameter estimates and attenuates potential collinearity concerns, thus also reducing parameter uncertainty in out-of-sample forecasting (Baltagi, 2008). However, panel-data techniques have only recently been applied in tourism demand forecasting, with the contributions by Assaf, Li, Song, and Tsionas (2019), Gunter (2018), and Yang and Zhang (2019) being some notable examples.

Table 2
Publication title analysis.
Source: Authors' own evaluation.

Codes	Number of coding words
Anaylsis	
Causality analysis	2
Co-integration analysis	2
Decomposition analysis	2
Demand analysis	1
Dynamic analysis	2
Dynamic demand analysis	1
Dynamic panel data analysis	2
Econometric analysis	3
Empirical analysis	2
Economic analysis	1
Panel data analysis	4
Tourism	
Academic tourism demand	2
Australian tourism	1
Australian tourism goods	1
Australian tourism marketing expenditure	1
Coastal tourism demand	1
Cross section tourism demand models	1
Cruise tourism	2
Domestic tourism	4
Domestic tourism demand	2
Dynamic tourism demand model for	1
European tourism demand	2
Foreign tourism demand	1
For outbound tourism	1
Inbound tourism	3
Inbound tourism demand	2
International outbound tourism	1
International tourism	4
International tourism demand	6
International tourism expenditures	1
International tourism flows	1
Italian tourism demand	1
Local tourism demand	1
Malaysian tourism revenue	1
Outbound tourism demand	4
Seasonal tourism demand	1
Stay-over tourism	2
Tourism demand	11
Tourism demand flows	1
Tourism demand models	1
Tourism determinants	2
Tourism development	1
Tourism expenditure	1
Tourism income elasticities	1
Tourism investment variable	2
Turkish tourism	1
Winter tourism	2

The findings

In general, the results were mixed. However, the first overall impression supports the judgement that most of the papers analyzed are flawed in terms of the chosen assessment criteria:

- $>$ 70% of the 115 papers did not include descriptive statistics (or only $<$ 30% included descriptive statistics), so that evaluations of whether an estimated coefficient is large or small becomes difficult (see Table 3). The 82 papers without descriptive statistics had relatively lower scores for data treatment efforts and discussions of why certain variables in the chosen approach are considered as important, relative to those which included descriptive statistics. Furthermore, papers without descriptive statistics also performed worse in terms of successfully comparing and discussing their results with other related research outcomes as well as pointing out the limitations of the study and highlighting necessary future research efforts.
- Only 15% of the articles adequately justify the method chosen (see Table 4). Related to that critical outcome is also the finding that only 6% of the papers discuss why certain important variables were not considered. Similar to the papers lacking descriptive statistics, the papers which failed to justify their research approach also tended to feature additional flaws. These are relatively

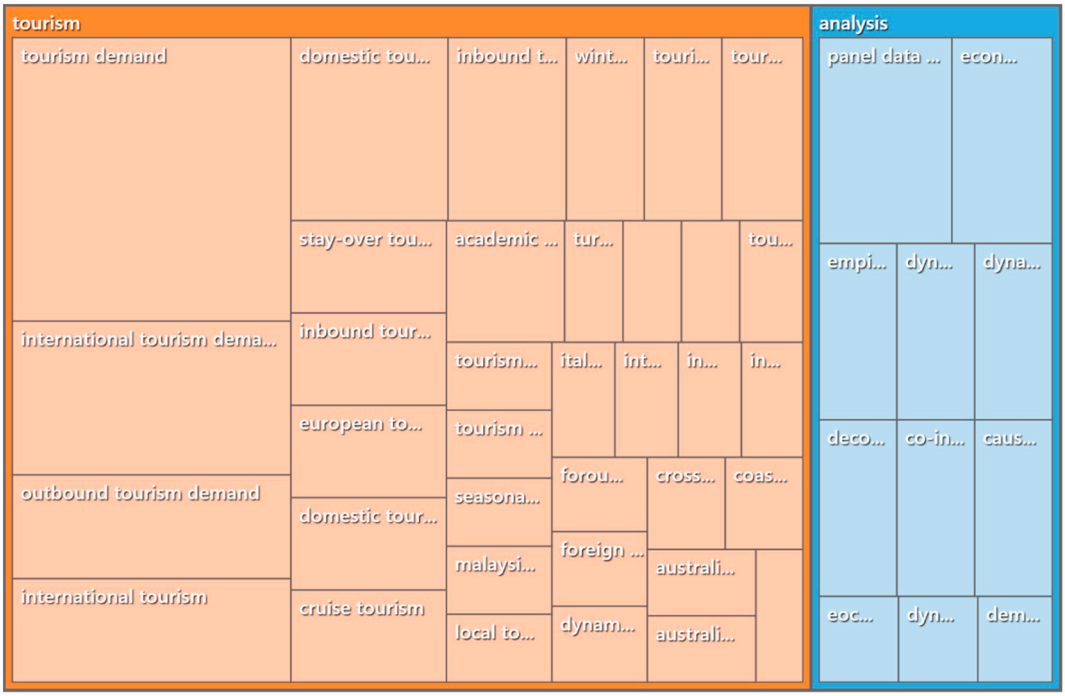
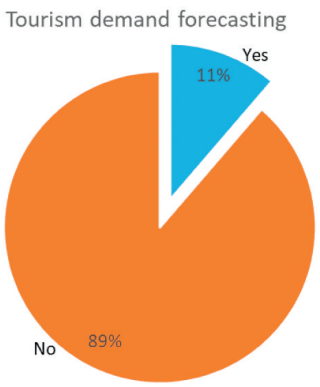


Fig. 3. Publication title tree map. Source: Authors' own illustration.
The subcategories of tourism include: tourism demand, international tourism demand, domestic tourism, international tourism, outbound tourism demand, inbound tourism, academic tourism demand, cruise tourism, domestic tourism demand, European tourism demand, inbound tourism demand, stay-over tourism, tourism determinants, tourism investment variable, winter tourism, Australian tourism, Australian tourism goods, Australian tourism marketing expenditure, Coastal tourism demand, cross section tourism demand models, dynamic tourism demand model, foreign tourism demand, for outbound tourism, international outbound tourism, international tourism expenditures, international tourism flows, Italian tourism demand, local tourism demand, Malaysian tourism revenue, seasonal tourism demand, tourism demand flows, tourism demand models, tourism development, tourism expenditure, tourism income elasticities, Turkish tourism. The subcategories of analysis include: panel data analysis, econometric analysis, causality analysis, co-integration analysis, decomposition analysis, dynamic analysis, dynamic panel data analysis, empirical analysis, demand analysis, dynamic demand analysis, economic analysis.

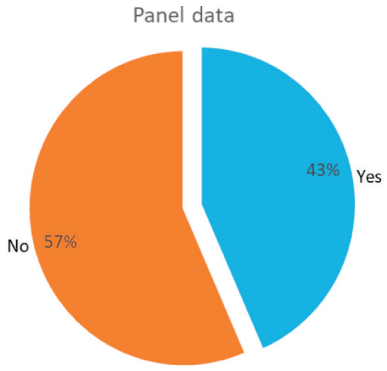
lower shares of data treatment efforts and a relatively lower frequency of highlighting the limitations of the study. However, it has to be pointed out that relatively higher shares of these studies discussed the consideration of included and excluded variables.

- > 90% of all papers used statistical significance as the sole criterion of scientific importance or did not discuss the results because the estimated coefficients were statistically significant. In these cases, statistics are used means of avoiding discussion.
- Related to the latter finding is that only 28% of the studies actually discuss their research outcomes in a broader context to demonstrate the reliability and to allow interpretation of the size of their results (see Table 5). Interestingly, the studies which omit a discussion on their outcomes in comparison to the work of other authors had relatively higher frequencies in excluding descriptive statistics and the limitations of their research. These papers also tended to avoid a discussion as to why they considered certain variables in their chosen approaches.
- Only < 30% of the articles discussed their limitations (see Table 6). These papers had relatively high shares of excluding descriptive statistics, the justification of their chosen methods, discussions about the incorporation of certain variables in their approaches, as well as of failing to consider specific potentially explanatory factors. They also exhibited lesser discussion of their results in a broader context based on the research outcomes of other authors.
- Accompanying the finding of omitting a discussion of the results of other similar/related publications and pointing out the limitations of their own studies, we could not find any suggestions for necessary future research efforts in > 50% of the papers.

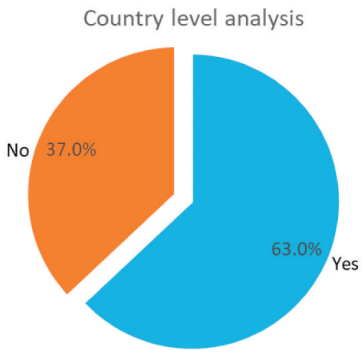
On the other hand, we have to emphasize that we observed some positive tendencies with respect to the other criteria, although these were not always sufficient to redeem the scientific quality of the studies. On this positive note, we found evidence that in > 50% of the papers (see Tables 3 to 6: all studies):



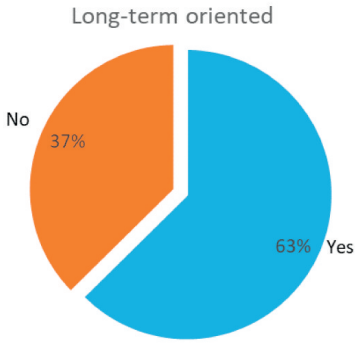
a: Tourism demand forecasting availability in the sample. Source: Authors' own elaboration and calculation.



b: Usage of panel-data methods in the sample. Source: Authors' own elaboration and calculation.



c: Destination data at the country level in the sample. Source: Authors' own elaboration and calculation.



d: Studies based on data utilization in the sample (short-term: less than three years). Source: Authors' own elaboration and calculation.

Fig. 4. a: Tourism demand forecasting availability in the sample.
b: Usage of panel-data methods in the sample.
c: Destination data at the country level in the sample.
d: Studies based on data utilization in the sample (short-term: less than three years).
Source: Authors' own elaboration and calculation.

- The data were graphed and/or inspected for trends, seasonal patterns, structural breaks, etc.;
- Some data treatment was performed;
- The employed models conformed with the data characteristics;
- Discussions were elaborated of why certain variables in the chosen approaches were considered as important;
- Estimated coefficients were in an interpretable form and carefully interpreted (although impact sizes and their reliability were often not specifically discussed);
- Simulations and robustness checks were done;
- And unnecessary significance tests were avoided.

Table 3
Evaluation of all studies and studies excluding descriptive statistics (Q1).
Source: Authors' own calculations.

Question	All studies		Q1 = Yes	
	Absolute number of studies with "Yes"	Percentage of studies with "Yes"	Absolute number of studies with "Yes"	Percentage of studies with "Yes" (relative to studies with Q1 = Yes)
Q1	33	29%	33	100%
Q2	67	58%	19	58%
Q3	101	88%	27	82%
Q4	90	78%	25	76%
Q5	17	15%	5	15%
Q6	85	74%	21	64%
Q7	7	6%	1	3%
Q8	114	99%	33	100%
Q9	95	83%	26	79%
Q10	83	72%	27	82%
Q11	115	100%	33	100%
Q12	107	93%	31	94%
Q13	32	28%	8	24%
Q14	31	27%	6	18%
Q15	51	44%	13	39%

Table 4
Evaluation of all studies and studies without methodological justification (Q5).
Source: Authors' own calculations.

Question	All studies		Q5 = Yes	
	Absolute number of studies with "Yes"	Percentage of studies with "Yes"	Absolute number of studies with "Yes"	Percentage of studies with "Yes" (relative to studies with Q5 = Yes)
Q1	33	29%	5	29%
Q2	67	58%	13	76%
Q3	101	88%	14	82%
Q4	90	78%	16	94%
Q5	17	15%	17	100%
Q6	85	74%	14	82%
Q7	7	6%	5	29%
Q8	114	99%	17	100%
Q9	95	83%	15	88%
Q10	83	72%	13	76%
Q11	115	100%	17	100%
Q12	107	93%	15	88%
Q13	32	28%	7	41%
Q14	31	27%	3	18%
Q15	51	44%	10	59%

Table 5
Evaluation of all studies and studies avoiding a discussion in a broader context (Q13).
Source: Authors' own calculations.

Question	All studies		Q13 = Yes	
	Absolute number of studies with "Yes"	Percentage of studies with "Yes"	Absolute number of studies with "Yes"	Percentage of studies with "Yes" (relative to studies with Q13 = Yes)
Q1	33	29%	8	25%
Q2	67	58%	18	56%
Q3	101	88%	29	91%
Q4	90	78%	26	81%
Q5	17	15%	7	22%
Q6	85	74%	22	69%
Q7	7	6%	4	13%
Q8	114	99%	31	97%
Q9	95	83%	31	97%
Q10	83	72%	24	75%
Q11	115	100%	32	100%
Q12	107	93%	30	94%
Q13	32	28%	32	100%
Q14	31	27%	7	22%
Q15	51	44%	17	53%

Table 6
Evaluation of all studies and studies excluding a discussion on their limitations (Q14).
Source: Authors' own calculations.

Question	All studies		Q14 = Yes	
	Absolute number of studies with "Yes"	Percentage of studies with "Yes"	Absolute number of studies with "Yes"	Percentage of studies with "Yes" (relative to studies with Q14 = Yes)
Q1	33	29%	6	19%
Q2	67	58%	18	58%
Q3	101	88%	29	94%
Q4	90	78%	25	81%
Q5	17	15%	3	10%
Q6	85	74%	20	65%
Q7	7	6%	1	3%
Q8	114	99%	30	97%
Q9	95	83%	26	84%
Q10	83	72%	22	71%
Q11	115	100%	31	100%
Q12	107	93%	29	94%
Q13	32	28%	7	23%
Q14	31	27%	31	100%
Q15	51	44%	23	74%

Discussion

We certainly have to discuss whether our sample size is sufficiently large that we can dare to state that a large part of the econometric tourism demand studies published in the period 2007 to 2017 in leading tourism journals is scientifically doubtful. Extending the sample would have meant considering more journals and/or a longer time period. Both are possible, but in case of more journals we would have risked operating with a biased sample as we would have mixed different qualities of publications. In the case of a period extension, we have to be aware that the research styles change over time period as statistical/econometric software improves in efficiency and user-friendliness, and the availability of data and different econometric techniques increases. Evaluating the advantages and disadvantages of our approach, we concluded that it is likely that using a smaller sample and a shorter time period minimizes the risk of bias compared to a bigger sample.

Furthermore, we have to ask how big or small as well as how good or bad our results are. Unfortunately, to the best of our knowledge, a comparable study has not been done until today in the tourism discipline. One comparable source for benchmarking figures is the book of Ziliak and McCloskey (2014), in which the authors summarized their earlier studies about the scientific qualities of papers published in AER (McCloskey & Ziliak, 1996; Ziliak & McCloskey, 2004). Another source for comparisons is the contribution by Seth, Carlson, Hatfield, and Lan (2009), who analyzed the papers published in the Strategic Management Journal during 2007 using regression methodology.

In both contributions mentioned above, we found three questions which are approximately comparable to our own criteria. These are:

1. Are the descriptive statistics of the variables included?
2. Is statistical significance used as a criterion of scientific importance?
3. Does the paper mention the work of others and are the results discussed in a broader context (i.e., if measured effects are small or large)?

Below we summarize the results of the evaluation for a comparison:

1. 66% of the studies in AER included descriptive statistics (sampled tourism journals: 29%; Strategic Management Journal: 63%).
2. 43% of the articles in AER used statistical significance as a criterion of scientific importance (sampled tourism journals: 93%; Strategic Management Journal: 90%).
3. 54% of the papers in AER mention the work of others and discuss the results in a broader context, i.e., whether the measured effects are small or large (sampled tourism journals: 28%; Strategic Management Journal: 0%).

Taking the study results from Ziliak and McCloskey (2004) from the published papers in the 1990s as a benchmark, the relative scientific quality of econometric tourism demand studies is not so convincing. On the other hand, based on the evaluation of the Strategic Management Journal, the position of the tourism journals is better in terms of the presence of discussions of results in a broader context (i.e., whether the measured effects are small or large). However, both results fall short of a critical threshold, with the 0% in the Strategic Management Journal especially alarming. The tourism journals and the Strategic Management Journal are also similar in case of question 2. The evaluation of question 1 showed similarities and more or less sufficient results for both AER and the Strategic Management Journal, whereas the performance of the tourism journals lies in a critical range.

We have to recognize that discussion about the scientific quality of econometric tourism demand studies is only just starting with this article, whereas in some other disciplines this type of discussion has progressed rather further:

Political science is gradually moving away from an exclusive focus on statistical significance towards the greater consideration of the magnitude and importance of effects (McCaskey & Rainey, 2015). These authors argue that the focus on point estimates hides uncertainty behind a veil of statistical significance and recommend that researchers should explicitly take account of uncertainty by interpreting the range of values contained in the confidence interval. This recommendation was based on analyses of published articles: An evaluation of all 316 articles published in the American Political Science Review and the American Journal of Political Science in the period 2011 to 2013 showed that 73% present empirical analyses, and from those only about half contained a judgement about the substantive importance of the estimated effects (McCaskey & Rainey, 2015).

In social psychology, the Journal of Basic and Applied Social Psychology banned the use of significance testing altogether from papers it publishes, requiring authors to use other measures to evaluate hypotheses and impact (Novella, 2015; Woolston, 2015). Furthermore, the human behavior scientists Amrhein and Greenland (2017) state clearly that significance and non-significance are often equated with falsity and truth of hypotheses in their disciplines, reflecting overconfidence about mathematical results and ignoring uncertainties not captured in models (Benjamin, 2017). There is also a growing consensus that results must be published regardless of statistical significance, as the substantive significance is the important matter (Amrhein & Greenland, 2017). This supports our argument that worshipping stringent statistical thresholds will impede scientific progress, as following false-negative conclusions precludes statements about substantive significance to a non-negligible extent.

A strong signal about the use of *p*-values and the term statistical significance was sent by the American Statistical Association (ASA): the ASA point out very clearly that the widespread use of statistical significance – to be understood as a 5% p-value threshold – as a justification for scientific findings leads to a biased perception of the scientific process (Wasserstein & Lazar, 2016). The ASA-statement was an impulse to the scientific community to move further towards a world beyond $p < 0.05$ (Wasserstein, Schirm, & Lazar, 2019). Supporting these developments, Hubbard, Haig, and Parsa (2019) as well as Ziliak (2019) state that it is necessary to

recognize that statistical inference is not equivalent to scientific inference.

Following the literature, we found several recommendations regarding what one should consider to prove/interpret findings in empirical research (Greenland, 2019; Ioannidis, 2019; Wasserstein et al., 2019). One rule is: do not conclude that an effect is present/absent or important/unimportant based purely on the level of statistical significance. Moreover, scientific work has to consider that the calculated *p*-values do not allow statements about the probability that the relevant test hypothesis is true.

What could be done to improve scientific work? Goodman (2019) states that social change is needed in academic institutions, in the policy of the scientific journals, and among private and public funding. Statistical education as well as the scientific practice, including publication style, has to move beyond $p < 0.05$ to avoid further damage to science caused by a publish-or-perish culture and the related doubtful metrics (Colquhoun, 2019; Maurer, Hudiburgh, Werwinski, & Bailer, 2019; Steel, Liermann, & Guttorp, 2019).

Conclusions

The unsatisfying phenomenon of many publications focusing predominantly on the statistical significance of their estimated coefficients and less, or not at all, on the substantive (economic) significance of their results triggered the design of a research project with the major objective of asking: What is the standing of tourism econometrics and what is the scientific value of all these studies? To fill this research gap, we used the econometric tourism demand studies of four leading tourism journals published during the period 2007 to 2017 as our database and analyzed these studies according to a checklist of specific criteria. Each paper was confronted with the same “yes” or “no” questions, which together reveal an impression of how carefully the papers' authors worked out the differences between statistical and substantive significance. In doing so, we also incorporated indirectly related criteria, thus allowing for a comprehensive evaluation of the articles according to the main objective of the study.

Our main findings allow us to conclude that most of the papers analyzed contain flaws in terms of our key criteria. Moreover, we conclude that the manner of the scientific discussion and argumentation has to be improved. The analysis showed that > 90% of the papers used statistical significance as the sole criterion of scientific importance or did not discuss the results any further because the estimated coefficients were statistically significant. A majority of the papers did not justify the methods chosen and did not discuss why certain important variables were not considered. Many studies also failed to include descriptive statistics, so that any evaluation of whether an estimated coefficient is large or small becomes difficult. These “unit-free” papers also widely avoided discussing their results in the context of other related research outcomes to demonstrate the reliability of their findings, as well as pointing out any limitations and necessary future research efforts. In total, more than two thirds of the papers omitted discussions about their limitations and, of these, a relatively high share also excluded justification of their chosen methods and discussion about the incorporation or omission of certain variables in the chosen approach.

To compare the outcome of this evaluation with previous results is difficult, as we found only three questions which are approximately comparable to our criteria, and these were addressed in two non-tourism journals. Based on these benchmarks, we have to say that in general the results from the evaluation of the tourism journals are worse than those of the others. However, one has to point out at this point that the evaluation results of the two benchmark journals only report an absolute view and are also far from great.

Appraising the state of discussion about the scientific quality of econometric tourism demand studies, we have to recognize that it has only just started with this article, whereas in some other disciplines this type of discussion has already progressed rather further. A clear trend in these fields is a move away from an exclusive focus on statistical significance towards greater consideration of the substantive significance of the estimates. This we found for political science, human behavioral sciences, and for social psychology. In the latter discipline, two core journals banned the use of significance testing altogether in papers it publishes, requiring authors to use other measures to evaluate hypotheses and impacts. A statement from the American Statistical Association about the celebrated use of the 5% *p*-value as a threshold for the claim of scientific findings should sound a warning for all authors, reviewers, and editors (Wasserstein & Lazar, 2016). In line with this statement from the American Statistical Association, Amrhein and Greenland (2017) concluded that significance and non-significance should not be equated with falsity and truth of hypotheses.

At the end, we have to ask what can now be done to improve the scientific quality of econometric tourism demand studies. Evidently, contributions to raising scientific quality must come from all parties involved: the authors, the reviewers, the editors, the publishers, and the involved institutions such as universities and corporations. Statistical education has to change such that students should follow the principles that the logic and execution of the studies should come before interpreting results and significance testing. Furthermore, journal editors should make sure that every author and reviewer knows the ASA-statement on *p*-values and statistical significance. The major focus of scientific practice should move away from concentrating solely on statistical significance towards the substantive significance of the estimates, meaning that in order to call an effect significant it must also have a certain size to be recognized as meaningful.

In addition, to judge the reliability of results and whether the estimated impact is “strong” or “weak”, future papers must also discuss the estimation results in a broader context, thereby considering the work of others. Here, thorough discussions of the employed approach and its justification, the data and the variables considered, as well as the limitations of the study play an important role. That means, on the other hand, that studies mostly focusing on results should be avoided as in most cases the process leading to the results is the true scientific value of the paper.

Clearly, we also have to discuss the limitations of our own study. Such a limitation could be our relatively small sample size. However, extending the sample by considering more journals and/or a longer time period has some risks. In the case of more journals we could end up with a biased sample as we would mix different qualities of the publications, and in the case of a period extension,

we have to be aware that the research styles change over time as statistical/econometric software improves. Being conscious of these trade-offs, we decided that keeping the risk of bias low is more likely when using a smaller sample and a shorter time period compared to a bigger sample.

In a future research project, one could compare and discuss different time periods and journal groups in terms of result stabilities and differences. A possible outcome of such a research effort could be to gain more insight into the dynamics of the scientific development process. Although our study dealt only with econometric tourism demand studies, the results also indicate the value of analyzing the scientific value of other empirical published research projects focusing on tourism modeling in non-econometric contexts.

References

Amrhein, V., & Greenland, S. (2017). Remove, rather than redefine, statistical significance. *Nature Human Behaviour, 02*(1), 24. https://doi.org/10.1038/s41562-017-0224-0.

Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2019). Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model. *Journal of Travel Research, 58*, 383–397.

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of Forecasting, 27*, 153–173.

Baltagi, B. H., & Griffin, J. M. (1997). Pooled estimators vs their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics, 77*, 303–327.

Benjamin, D. (2017). Redefine statistical significance. *Nature Human Behaviour, 1*, 0189. https://doi.org/10.1038/s41562-017-0189-z.

Brucker, H., & Siliverstovs, B. (2006). On the estimation and forecasting of international migration: How relevant is heterogeneity across countries. *Empirical Economics, 31*, 735–754.

Colquhoun, D. (2019). The false positive risk: A proposal concerning what to do about p-values. *The American Statistician, 73*, 192–201.

Fisher, R. A. (1925). *Statistical methods for research workers.* Edinburgh, UK: Oliver and Boyd.

Goodman, S. (2019). Why is getting rip of p-values so hard? Musings on science and statistics. *The American Statistician, 73*, 26–30.

Greenland, S. (2019). Valid p-values behave exactly as they should: Some misleading criticisms of p-values and their resolutions with s-values. *The American Statistician, 73*, 106–114.

Gunter, U. (2018). Conditional forecasts of tourism exports and tourism export prices of the EU-15 within a global vector autoregression framework. *Journal of Tourism Futures, 4*, 121–138.

Gunter, U., Önder, I., & Gindl, S. (2019). Exploring the predictive ability of LIKES of posts on the Facebook pages of four major city DMOs in Austria. *Tourism Economics, 25*, 375–401.

Hoogstrate, A. J., Palm, F. C., & Pfann, G. A. (2000). Pooling in dynamic panel-data models: An application to forecasting GDP growth rates. *Journal of Business and Economic Statistics, 18*, 274–283.

Hubbard, R., Haig, B. D., & Parsa, R. A. (2019). The limited role of formal statistical inference in scientific inference. *The American Statistician, 73*, 91–98.

Ioannidis, J. (2019). What have we (not) learned from millions of scientific papers with p-values? *The American Statistician, 73*, 20–25.

Jiao, E. X., & Chen, J. (2019). Tourism forecasting: A review of methodological developments over the last decade. *Tourism Economics, 25*, 469–492.

Kirilenko, A., & Stepchenkova, D. (2018). Tourism research from its inception to present day: Subject area, geography, and gender distributions. *PLoS One, 13*(11), e0206820. https://doi.org/10.1371/journal.pone.0206820.

Leamer, E. (1983). Let's take the con out of econometrics. *American Economic Review, 73*, 31–43.

Maddala, G. S., Trost, R. P., Li, H., & Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators. *Journal of Business and Economic Statistics, 15*, 90–100.

Maurer, K., Hudiburgh, L., Werwinski, L., & Bailer, J. (2019). Content audit for p-value principles in introductory statistics. *The American Statistician, 73*, 385–391.

McAleer, M., Pagan, A. P., & Volker, P. (1985). What will take out the con of econometrics? *American Economic Review, 75*, 293–307.

McCaskey, K., & Rainey, C. (2015). Substantive importance and the veil of statistical significance. *Statistics, Politics, and Policy, 6*, 77–96.

McCloskey, D. N., & Ziliak, T. S. (1996). The standard error of regressions. *Journal of Economic Literature, XXXIV*(March), 97–114.

Novella, S. (2015). Psychology journal bans significance testing. *Science-Based Medicine*. (February 25) https://sciencebasedmedicine.org/psychology-journal-bans-significance-testing/, Accessed date: 21 June 2019.

Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences.* New York: Wiley.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Rust, J. (2014). The limits of inference with theory: A review of Wolpin (2013). *Journal of Economic Literature, 52*, 820–850.

Seth, A., Carlson, K., Hatfield, D., & Lan, H. W. (2009). So what? Beyond statistical to substantive significance in strategy research. In D. Bergh, & D. Ketchen (Vol. Eds.), *Research methodology in strategy and management. Vol. 5. Research methodology in strategy and management* (pp. 3–27). Emerald.

Song, H., Qiu, R. T. R., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research, 75*, 338.362.

Steel, A., Liermann, M., & Guttorp, P. (2019). Beyond calculations: A course in statistical thinking. *The American Statistician, 73*, 392–401.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician, 70*, 129–133.

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p $<$ 0.05". *The American Statistician, 73*, 1–19.

Wolpin, K. (2013). *The limits of inference without theory.* Cambridge, Mass and London: MIT Press.

Wonnacott, T., & Wonnacott, R. (1990). *Introductory statistics for business and economics* (4 ed.). New York: John Wiley & Sons.

Wooldridge, J. (2000). *Introductory econometrics, a modern approach.* The South-Western College Publishing, Thomson Learning.

Wooldrige, J. (2004). Statistical significance is o.k. too: Comments on size matters. *Journal of Socio-Economics, 33*, 577–580.

Woolston, C. (2015). Psychology journal bans P values. *Nature, 519*(7541), 9. https://doi.org/10.1038/519009f.

Yang, Y., & Zhang, H. (2019). Spatial-temporal forecasting of tourism demand. *Annals of Tourism Research, 75*, 106–119.

Ziliak, S. (2019). How large are your G-values? Try Gosset's guinnessometrics when a little "p" is not enough. *The American Statistician, 73*, 281–290.

Ziliak, T. S., & McCloskey, D. N. (2004). Size matters: The standard error of regressions in the American economic review. *Journal of Socio-Economics, 33*, 665–675.

Ziliak, T. S., & McCloskey, D. N. (2014). *The cult of statistical significance.* Ann Arbor: University of Michigan Press.

Ulrich Gunter is an Associate Professor at the Department of Tourism and Service Management at MODUL University Vienna.

Irem Önder is an Associate Professor at the Department of Tourism and Service Management at MODUL University Vienna.

Egon Smeral is a Full Professor at the Department of Tourism and Service Management at MODUL University Vienna.