Annals of Tourism Research 76 (2019) 266–277

Contents lists available at ScienceDirect

Annals of Tourism Research

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

A review of research into performance modeling in tourism research - Launching the Annals of Tourism Research curated collection on performance modeling in tourism research

A. George Assaf[a,*,1], Mike G. Tsionas[b,1]

[a] Isenberg School of Management, University of Massachusetts-Amherst, 90 Campus Center Way, 209A Flint Lab, Amherst, MA, 01003, United States of America

[b] Lancaster University Management School, United Kingdom of Great Britain and Northern Ireland

ARTICLE INFO

Keywords:
Performance modeling
Frontier models
Tourism performance

ABSTRACT

This paper presents a review of performance modeling in tourism research, with special focus on frontier models. We discuss the current status of the literature, identify the gaps, and highlight directions for future improvements across both parametric and non-parametric methodologies. More specifically, we elaborate on key methodological issues including endogeneity, bad outputs, dynamic formulations, heterogeneity, Bayesian estimation, bootstrapping, and stochastic DEA. For each of these areas we discuss and introduce some recent methodological breakthroughs that have been largely ignored in the tourism literature.

This article launches a Curated Collection on performance modeling in tourism research, containing all articles published in Annals on the topic.

Introduction

Performance modeling is now a well-established research area in tourism economics. One of the most important breakthroughs over the last decade was the introduction of frontier modeling to measure tourism performance. These methods not only account for the multiple input-output settings of the tourism industry but also introduce a completely new philosophy in the estimation of tourism performance.

According to recent reviews and meta-analysis by Assaf and Josiassen (2016) and Assaf and Tsionas (2018a), more than 90 studies have appeared over the last decade using frontier methods to measure tourism performance. These frontier methods continue to gain popularity as is evidenced by some recent studies on the topic (Assaf & Tsionas, 2018a, 2018b). They will undoubtedly occupy the lion's share of future performance modeling in tourism research. Even studies in some mainstream management journals have recently called for more use of frontier methods to measure firm performance (Chen, Delmas, & Lieberman, 2015). Frontier methods simply provide better “vehicles for characterizing performance in ways that go beyond conventional analysis of common financial measures” (Chen et al., 2015, p.31).

Motivated by the above, the goal of this paper is to provide a full overview of the literature,[2] highlight some of the main gaps, and

* Corresponding author.
E-mail addresses: assaf@isenberg.umass.edu (A.G. Assaf), m.tsionas@lancaster.ac.uk (M.G. Tsionas).

[1] Both authors have contributed equally to the paper.

[2] We do not aim to duplicate recent reviews on the topic (e.g. Assaf and Josiassen, 2015). Rather, we focus more on recommendations and directions for future research.

https://doi.org/10.1016/j.annals.2019.04.010
Received 22 October 2018; Received in revised form 10 April 2019; Accepted 10 April 2019
Available online 01 May 2019
0160-7383/

suggest areas in which there is room for improvement. We illustrate the advantages of frontier methods by first providing a brief overview of some traditional performance methods, highlighting their main limitations, and comparing them to frontier methods. Frontier methods have the ability "to assess an individual firm's performance relative to the best performers in an industry and are therefore well suited to address the challenges of measuring competitive advantage" (Chen et al., 2015, p.20).

In this paper, we differentiate between parametric and non-parametric frontier methods, though we focus more on the parametric approach where important issues such as heterogeneity, dynamic formulation, and bad outputs, can be addressed in a more flexible manner. We discuss and introduce some recent methodological breakthroughs that have been largely ignored in the tourism literature. Some of these breakthroughs, in fact, have the ability to provide more realistic performance measures for tourism destinations.

The remainder of this paper proceeds as follows: First, we provide a background on frontier methods, then we elaborate on the status of these frontier methods in tourism research. Finally, we discuss gaps in the current literature and provide directions for future research.

Traditional performance measures and the evolution of the "efficiency" concept

Performance measurement is well connected to strategy formulation (Chen et al., 2015). Well-designed performance measurement systems can help organizations evaluate and monitor their performance improvements (Altin, Koseoglu, Yu, & Riasi, 2018). In the past, most performance measures in hospitality and tourism were financially driven and derived based on some simple accounting-based measures, such as occupancy rate, number of tourism arrivals and receipts, and labor productivity (Anderson, Fish, Xia, & Michello, 1999).

Such measures, however, have significant limitations. As the name indicates, they are simple and partial, and as such, they do not account for the multiple input/output settings of firms in the industry. Importantly, with such methods it is not possible to identify a relative benchmark for performance improvements. The balanced scorecard (BSC) is an important improvement over these simple methods (Evans, 2005; Kaplan & Norton, 1996). BSC covers broader dimensions of performance measures; such as financial performance, customer relations, internal business processes, learning, and growth (Lipe & Salterio, 2000). Various studies have also shown that "the Balanced Scorecard has a positive impact on organizational performance. More specifically, the BSC improves the integration of the management processes and empowers people" (De Geuser, Mooraj, & Oyon, 2009, p.93). One of the main limitations of this method is that it may lead to subjective selection of performance indicators under each of the above-mentioned dimensions. In turn, this will make comparison between firms in the industry more challenging. This method also requires a significant amount of calculation and data collection.

Chen et al. (2015, p.19) emphasized the importance of having performance measures that would "identify the leading competitors and assess the reasons for their superiority". For instance, one of the reasons behind the popularity of performance indicators provided by "Smith Travel Research" is that they assess the performance of hotels against a competitive set. The performance measures also need to be comprehensive. It is highly unlikely that one firm would top the industry in various performance measures (Assaf & Tsionas, 2018a). In addition, the ability to "identify firms that possess advantage over competitors is a straightforward exercise if performance can be succinctly captured by a single measure" (Chen et al., 2015, p.20).

A well-suited performance measure to address this issue is the "efficiency" metric (Luo & Homburg, 2008). Over the last decade, it has become highly popular to use efficiency-based approaches in measuring hotel and tourism performance (Altin et al., 2018; Deng, Gao, Liang, & Morrison, 2018; Mendieta-Peñalver, Perles-Ribes, Ramón-Rodríguez, & Such-Devesa, 2018; Herrero-Prieto & Gomez-Vega, 2017; Ohe & Peypoch, 2016; Assaf and Tsionas, 2015). Not only do they account for competitive advantage, but they also measure performance relative to a frontier of best practices. They are appropriate for assessing theories and testing strategies related to competitive advantage.

In this paper, we focus on both the Data Envelopment Analysis (DEA) and the Stochastic Frontier (SF) approaches. Both methods measure efficiency based on distance from the industry's efficiency frontier. The frontier represents the maximum level of outputs (e.g. tourism arrivals, tourism receipts, etc.) that can be achieved given a vector of inputs (labor, capital, etc.). DEA and SF are two alternative methods to measure the efficient frontier; the first uses a linear programming approach, while the latter uses an econometric approach. The DEA formulation can be expressed as follows:

$$\hat{\theta}_i = \max\left\{\theta \;\middle|\; x_i \geq \sum_{i=1}^{n} \lambda_i x_i, \theta_{yi} \leq \sum_{i=1}^{n} \lambda_i y_i, \sum_{i=1}^{n} \lambda_i = 1\right\} \tag{1}$$

where n is the number of decision making units (DMU), y_i is a vector of observed outputs, x_i is a vector of observed inputs, θ_i is a scalar quantity and λ is a vector of weights. The term $\sum_{i=1}^{n} \lambda_i = 1$ indicates that the production technology is defined with variable returns to scale (VRS). The obtained value of θ_i is the output-oriented efficiency score.

The SF approach estimates the same frontier but in a parametric fashion. A simple SF model can be expressed as follows:

$$y_{it} = f(x_{it}, \beta) + v_{it} - u_{it}, i = 1, ..., n, t = 1, ..., T, \tag{2}$$

where y_{it} is a single output[3], x_{it} is a vector of inputs, β_i is a vector of parameters, v_{it} represents measurement error, and u_{it} is a

[3] Multiple outputs can be dealt with easily through the use of distance functions.

random error that measures the deviations (i.e. inefficiency) from the frontier technology. u_{it} follows an asymmetric distribution, such as half-normal, truncated normal, exponential, or gamma. The model can be estimated using a variety of estimation methods such as Maximum Likelihood (ML), Bayesian, or Generalized Method of Moments (GMM).

A main distinction between the SF model in (2) and the DEA model in (1) is that the SF model involves an additional component (v_{it}) to account for the random noise effect. This makes the efficiency results less sensitive to random error in the data. In contrast to DEA, it is likely that none of the firms will land exactly on the frontier, as the inefficiency term u_{it} is a continuous random variable. Again, the distance from the frontier reveals the level of inefficiency. However, in contrast to DEA, which defines a piecewise surface or "frontier" that envelops the observations in the data, SF adopts a functional form $f(x_{it})$ for the estimation of the frontier and does so econometrically. One criticism for DEA is that it treats all random error in the data as a source of inefficiency. In addition, it is more sensitive to small sample sizes.

Current status of the frontier literature in tourism

Recently, Assaf and Josiassen (2016) conducted a comprehensive review on the use of frontier models in tourism. They summarized the characteristics of most applications, covering both micro and macro studies. The authors indicated that while the literature is rich and rapidly growing, it still suffers from important weaknesses. Our goal here is not to conduct another review on this topic. Rather, we start with the main recommendations of Assaf and Josiassen (2016) and elaborate further on what could be done to improve the literature in this area.

Assaf and Josiassen (2016) indicated that there are many important methodological issues within the context of frontier analysis that have been ignored in the literature. The authors highlighted the need for more use of dynamic and heterogeneous frontier models. Such models provide clear advantages in the estimation of stochastic frontier models and may lead to more accurate efficiency estimates. Importantly, most frontier studies in tourism do not account for potential endogeneity in the estimation of stochastic frontier models. Other important methodological considerations such as the issue of accounting for bad outputs or distinguishing between technical and allocative efficiencies have not also been applied. In terms of estimation methods, the Bayesian approach introduces important advantages to the estimation of SF models, but has yet to receive enough attention in the tourism literature.

We will elaborate further on these issues. Our goal is to provide tourism researchers with more guidance on the available performance models and their advantages. In particular, we focus on methodological considerations which, despite their importance, have not received enough attention in the tourism literature. We provide recommendations for better performance estimation within both the SF and DEA contexts.

Stochastic frontier models

As highlighted in the review by Assaf and Josiassen (2016), most SF applications in tourism have focused on simple applications of the model. There are important assumptions and model extensions that have been overlooked. We focus on five important methodological considerations in this section. We believe that addressing these methodological considerations is essential when measuring tourism performance.

Endogeneity

It is surprising that most SF applications in tourism have not taken into account the issue of endogeneity (Assaf & Tsionas, 2018b). It is well known in economics that the analysis of a stochastic frontier with multiple outputs is hampered by the fact that many, if not all, inputs and outputs may be endogenous, under a variety of behavioral assumptions. Ignoring the endogeneity problem can seriously bias the performance results.

The fundamental problem of endogeneity can be illustrated as follows. Given a production function $Y = F(X_1, \ldots, X_K)e^{v+\omega}$ where v is an error term and ω is productivity, both known to the firm, but unknown to us, the first-order conditions for profit maximization are:

$$\frac{\partial F(X)}{\partial X_k} = \frac{W_k}{P} e^{-(v+\omega)},\ k = 1, \ldots K, \tag{3}$$

where P is product price and W_k is the price of the kth input. Since the errors v and ω appear both in the production function and the first-order conditions, it turns out that inputs cannot be independent of these errors. To ease exposition, suppose we have a Cobb-Douglas production function:

$$\log Y = \beta_o + \sum_{k=1}^{K} \beta_k \log X_k + v + \omega \Leftrightarrow y = \beta_o + \sum_{k=1}^{K} \beta_k x_k + v + \omega. \tag{4}$$

Estimation by Ordinary Least Squares (OLS) would lead to inconsistent parameter estimators since x_k's and the errors are correlated. In econometrics this is known as simultaneity bias, since the x_k's and y are simultaneously determined. The problem, of course, occurs even when there is no productivity (that is, ω = 0).

The endogeneity problem can be illustrated in a somewhat simpler framework if we consider the following system:

$$y = \beta x + v,$$
$$x = \Pi z + e. \tag{5}$$

Generally, v and e are correlated (and the amount of correlation is related to the severity of the endogeneity problem) while the instrument z can be weak depending on its correlation with x. For example, a simple IV estimator is:

$$\widehat{\beta}_{IV} = \frac{\sum zy}{\sum zx}. \tag{6}$$

When the instrument is weak, the denominator is close to zero and the (finite-sample or large-sample) distribution will have large variance and possibly fat tails. Moreover, depending on the weakness of instruments and/or degree of endogeneity, the likelihood can have non-elliptical contours and can be highly irregular. Additionally, some instruments can be invalid in the sense that they can be direct determinants of y and/or they are correlated with the error "v". A new trend has steadily developed in econometrics that uses LASSO prior to developing procedures for instrument selection in cases where the number of instruments can be much larger than the number of observations.

There are certain solutions to the endogeneity problem that tourism researchers can adopt:

i) One can use GMM if instruments are available; instruments can be input prices, lagged inputs, etc.
ii) One can use a full-information approach in which the production function and the system of first-order conditions are jointly estimated (see Assaf and Tsionas, 2018b).
iii) Methods that do not depend on instruments can be used when the validity of these instruments is questioned and/or they are weak. The most prominent of these methods is a copula approach in which the joint distribution of v and ε is specified along with the marginal distributions (in the context of regression, see Park and Gupta, 2012).
iv) Closely related to GMM is a limited information approach in which the inputs are related to a set of instruments z through a reduced form that can be estimated as a system along with the model in (4):

$$x = \Pi(z) + \varepsilon, \tag{7}$$

where Π(z) is a certain matrix functional form, and ε is an error term that can be correlated with v and/or ω in (4). Such an approach was recently used by Tsionas, Assaf, Gillen, and Mattila (2017). The reduced form expresses endogenous outputs as functions of exogenous inputs (the instruments) in a flexible way.

Dynamic formulation

Another major issue that has been largely ignored in the tourism literature is the importance of estimating SF in a dynamic framework (Assaf & Josiassen, 2016). While several tourism studies have estimated the SF model in a panel data framework (e.g. Barros, 2004; Barros, 2006; Pérez-Rodríguez & Acosta-González, 2007) and allowed the performance term in (2) to vary over time, most of these models were not formulated in a dynamic framework "thereby meaning that an inefficient firm is not allowed to correct its inefficiency from the past" (Desli, Ray, & Kumbhakar, 2003, p. 623). Given the highly dynamic characteristics of the tourism industry, we believe that a dynamic model would provide a more realistic representation of firms or destinations in the industry. The dynamic model can allow differentiation between short-term and long-term performance measures. In addition, one can also derive impulse response functions and persistence measures to trace out the dynamic effect of performance over time.

It is extremely surprising how little attention has been paid to dynamic SF models in the tourism industry. Traditionally, formal dynamic frontier models, using dynamic programming techniques, have been proposed by Rungsuriyawiboon and Stefanou (2007). Kumbhakar (1991) suggests a random-effects model with time effects; the time effects allow for some dynamics but not dynamics in inefficiency per se. There are many studies where inefficiency has been estimated as a systematic function of time (Battese & Coelli, 1992; Cornwell, Schmidt, & Sickles, 1990; Kumbhakar, 1990; Kumbhakar & Wang, 2005; Lee & Schmidt, 1993), but again these are not formulated in a dynamic framework.

Some of the earliest dynamic models were introduced by Ahn, Good, and Sickles (2000), Desli et al. (2003) and Wang (2007). For instance, Wang (2007) introduced the following dynamic formulation:

$$u_{it} = \theta_1 u_{i,t-1} + \ldots + \theta_p u_{i,t-p} + \xi_{it} \tag{8}$$

where u_{it} is the inefficiency term, and ξ_{it} is an error term.

Tsionas (2006) introduced a more advanced formulation that would also allow inefficiency to be a function of some covariates z_{it}

$$\log u_{it} = z_{it}'\gamma + \rho \log u_{i,t-1} + \xi_{it} \tag{9}$$

This model has been used in many studies (e.g. Assaf, Gillen, & Barros, 2012; Zhang, Xu, Feng, & Jiao, 2015) and is highly popular for hypothesis testing because it can also assess the role of determinants (or variables) that drive efficiency. Recently, Tsionas and Assaf (2014) have proposed an extension to the model in (9). They argue that for many purposes an AR(1) model such as in (9) "could be adequate but when it comes to a detailed analysis of performance dynamics, one must formally estimate and statistically compare different ARMA models in order to use the most appropriate model" (Tsionas & Assaf, 2014, p.24). They proposed the following new model, which follows a general ARMA (L_1, L_2) model:

$$u_{it} = \sum_{p=1}^{L_1} \alpha_{ip} u_{i,t-p} + \varepsilon_{it} + \sum_{q=1}^{L_2} \gamma_{iq} \varepsilon_{i,t-q},\ \varepsilon_{it} \sim N(0, \sigma_\varepsilon^2),\ u_{it} \geq 0,\ i = 1, \ldots, n,\ t = 1, \ldots, T. \tag{10}$$

The model in (10) allows measurement for both short-run and long-run efficiencies. It can also be used to examine the impact of shocks on performance. We refer the reader to Tsionas and Assaf (2014) for more details. A nice discussion of dynamic stochastic frontier models is also available in Khalaf and Saunders (2016). These authors introduced a new approach for bias correction in the estimation of dynamic stochastic frontier models.

Heterogeneity

Most SF models in tourism have also ignored the issue of heterogeneity (Assaf & Tsionas, 2018b). We believe this is an important drawback that should be addressed in future papers. Heterogeneity refers to a general and realistic situation where equations of interest are different for different units in a sample. As stated by Assaf and Tsionas (2018b, p.133), "it would [be] hard to believe that the technology used to produce "tourism" in different tourism destinations is the same. If it differs, the "frontier technology of best practices" simply does not exist. The output of the production function cannot be considered "standardized" because it is dependent on the economic and natural resources available in each country (the "feasible" touristic products mix)".

In their review of the literature, Assaf and Tsionas (2018b) indicated that only one study has considered heterogeneity in the estimation of stochastic frontier models in tourism (Barros, Dieke, & Santos, 2010). Even in panel data applications, where one can at least use a fixed-effects formulation to have some account for unobserved heterogeneity, most tourism studies (e.g. Barros & Matias, 2006; Hu, Chiu, Shieh, & Huang, 2010 and Pérez-Rodríguez & Acosta-González, 2007) have not adopted a fixed-effects formulation.

Different options can be used by tourism researchers to address the issue of heterogeneity. With panel data, a quite general model has the following form:

$$y_{it} = \alpha_{it} + x_{it}'\beta_{it} + v_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{11}$$

where x_{it} is a k-dimensional vector of explanatory variables, v_{it} is an error term and all parameters depend on both firms and time. Without further assumptions, the model clearly fits trivially every possible data set. Although the model is quite general, we often assume

$$\alpha_{it} = \alpha_i,\ \beta_{it} = \beta_i,\ i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{12}$$

i.e. parameters are fixed over time but can differ among cross-sectional units. Typically, the errors are assumed to arise from the same distribution with a mean of zero and common variance or variance that can differ among cross-sectional units. It is important to understand that the "fixed-effects" specification of the form:

$$y_{it} = \alpha_i + x_{it}'\beta + v_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T, \tag{13}$$

is popular because it can be estimated easily without using dummy variables for the intercept. For example, using least-squares between $y_{it} - \bar{y}_i$ and $x_{it} - \bar{x}_i$ provides consistent estimators for the slope coefficients. Then, the fixed effects can be backed out as: $\hat{\alpha}_i = \bar{y}_i - \beta'\bar{x}_i$. This is a well-known Least-Squares Dummy Variables (LSDV) procedure. Fixed effects allow for inclusion, at least in an approximate way, of unobserved variables that are assumed to be slowly-varying or time-invariant.

Within the stochastic frontier framework, Schmidt and Sickles (1984) suggested using a model like (13) to account for unobserved heterogeneity.

$$\hat{\alpha}_i^* = max_i\, \hat{\alpha}_i - \hat{\alpha}_i. \tag{14}$$

Schmidt and Sickles (1984) treated α_i in (13) as a firm-specific inefficiency term. However, this approach was later criticized. Greene (2005a, p.277) argues that "by interpreting the firm specific term as 'inefficiency,' any unmeasured time invariant cross firm heterogeneity must be assumed away. The use of deviations from the maximum does not remedy this problem—indeed, if the sample does contain such heterogeneity, the comparison approach compounds it". Another problem with this approach is that it assumes efficiency to be time-invariant, which for most panel data is an unreasonable assumption.

Greene (2005a) suggested instead using a true fixed formulation of the following form to account for unobserved heterogeneity:

$$y_{it} = \alpha_i + x_{it}'\beta + v_{it} - u_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T. \tag{15}$$

The difference between this model and (13) is that it "retains the distributional assumptions of the stochastic frontier model, allows for freely time varying inefficiency, and allows the heterogeneity term to be correlated with the included variables" (Greene, 2005b, p11). Hence, this model provides more reasonable performance estimates while controlling for heterogeneity. Although the likelihood of the model is nonlinear in the parameters, Greene's model allows for separation of fixed effects from technical efficiency. Greene also proposes a versatile numerical procedure to maximize the likelihood function, relying on computationally efficient Gauss-Newton iterations to obtain the fixed effects.

Other approaches to address heterogeneity in estimating performance have also been proposed in the literature. Tsionas (2002), for instance, took a totally different approach to address this issue. The author criticized the use of fixed-effect formulations because they require the estimation of large numbers of parameters. He proposed instead a random coefficient formulation of the form:

$$y_{it} = \alpha + x_{it}'\beta_i + v_{it} - u_{it}, \tag{16}$$

where each firm has its own technology with parameters β_i which reflects heterogeneity of firms in their technology. Similar to the model in (15), this model has the advantage of separating efficiency from technological differences across firms.

Tsionas and Kumbhakar (2014) have also recently considered a more advanced model where firm effects and time effects can be separated from short-run and long-run (persistent) efficiency. We can write their model as follows:

$$y_{it} = x_{it}'\beta + v_{it} - u_{it}^{(+)} - u_i^{(+)} - u_t^{(+)} + \lambda_i + \mu_t, \; i = 1, \; ..., n, \; t = 1, \; ..., T. \tag{17}$$

Here, $u_{it}{}^{(+)}$ is short-run inefficiency and follows a distribution with non-negative support. $u_i{}^{(+)}$ is persistent inefficiency for a given firm (and follows a distribution with non-negative support). $u_t{}^{(+)}$ is time-varying inefficiency common to all firms (and follows a distribution with non-negative support). λ_i, μ_t are firm and time effects.

Finally, a very flexible approach to measuring hotel performance was recently introduced by Assaf and Tsionas (2018a). These authors allowed for a vector of environmental variables to directly influence heterogeneity (called, for this reason, observed heterogeneity). Their model is written as:

$$y_{it} = x_{it}'\beta(z_{it}) + v_{it}, \; i = 1, \; ..., n, \; t = 1, \; ..., T, \tag{18}$$

where z_{it} is $p \times 1$ vector of environmental variables (hotel size, hotel classification, etc.). Therefore, in their formulation make β dependent on z_{it} to control for heterogeneity. In (18), β is also random itself and has a random error term that differs by firm. The authors have estimated this model using the Bayesian approach and showed how the heterogeneous stochastic frontier model represents a better fit for the hotel industry than a homogenous stochastic frontier model. For more specific details on the model refer to Assaf and Tsionas (2018a).

To sum up, we believe it is important to address heterogeneity when measuring tourism performance. Different approaches are available and it may be advisable for future papers to compare the various approaches in one paper, or at least test whether heterogeneity exist within the data. It is surprising that such an issue has not been considered so far. Estimating tourism performance without accounting for heterogeneity may strongly bias the performance results.

An important question relates to how one should give "priority" to endogeneity, dynamic formulations, heterogeneity etc. As we rarely deal with experimental data, all these problems may be simultaneously present so, two options are available. The first option is to estimate the different models and compare them in terms of fit using the Bayes factor (see Section 0). The Bayes factor depends on the availability of marginal likelihood, which, in general, is hard to compute. The second option is to focus on a general encompassing model, which allows for all problems simultaneously. Adopting this option, one would have to consider (7) for endogeneity, (9) or (10) for a dynamic specification, and (11) for heterogeneity. We can write down a general model of this form, as follows:

$y_{it} = \alpha + x_{it}'\beta_i + v_{it} - u_{it}$, $i = 1, .., n$, $t = 1, \dots, T$, as in (16),

$\log u_{it} = z_{it}'\gamma_i + \rho_i \log u_{i,\,t-1} + \xi_{it}$, as in (9),

$x_{it} = \Pi(z_{it}) + \varepsilon_{it}$, as in (7).

To account for heterogeneity in general terms, the $\Pi(z_{it})$ function should depend on firm-specific parameters, say η_i. In turn, one can assume $\delta_i = [\beta_i', \gamma_i', \rho_i, \eta_i]' \sim iidN(\bar{\delta}, \Sigma_\delta)$, $i = 1, \; ..., n$..

Bad outputs

Another area of importance that we feel tourism researchers should pay strong attention to is the measurement of tourism performance subject to bad outputs. So far, this issue has not been taken into account (Assaf & Cvelbar, 2015). Bad outputs abound in many empirical processes. For example, CO_2 and SO_2 are produced along with good outputs in industrial sectors. Tourism generates some bad outputs as well, some of which are directly related to pollution, but others are related to cultural problems (Archer, Cooper, & Ruhanen, 2005). Within the hotel context, complaints on customer service have been used as negative or bad outputs when measuring hotel performance (Assaf & Cvelbar, 2015). As argued by Atkinson and Tsionas (2016), a comprehensive and fair performance analysis should give credit to firms for reducing the level of bad outputs.

The modeling of bad outputs is necessary because, if ignored, estimates of the technology will necessarily be biased and, almost always, such bad outputs are always present in empirical applications. Moreover, as Atkinson and Dorfman (2005) argued, firms should be credited when they reduce their levels of bad outputs just as they are credited when they reduce conventional inputs for given output (this concept being also known as input-specific efficiency). However, what constitutes bad output(s) is context-specific and more attention should be paid to a careful delineation of a potential list of bad output(s). As we mention below, there is also clearly a need for distinguishing bad outputs from inputs as, econometrically, a distance function that describes the technology set cannot distinguish between the two, as they satisfy the same properties.

So far, most research on bad outputs in the performance literature has focused on the manufacturing and agriculture sectors. Assaf et al. (2017) have recently developed a new stochastic frontier model for modeling the production of bad (or undesirable) outputs in the service industry. What we know about bad outputs is that they are produced jointly with good outputs. So, given an input vector x, the technology is described by the set:

$$T = \{(y, b): y, b \text{ can be produced using inputs } x\} \tag{19}$$

where y is the vector of good outputs and b is the vector of bad outputs. This implies clearly that production of y and b takes place at the same time, so they are jointly produced. Some authors (Fernandez et al., 2002) have assumed that:

$$b = b(y), \tag{20}$$

so that production of bad outputs depends on the level of production of good outputs. In other words, bad outputs are a straight technical by-product of good outputs. However, while such an approach may work well in the manufacturing context (more production implies more CO_2 emissions), it does not necessarily equally apply to the tourism industry. For example, more tourists do not necessarily imply more cultural or environmental problems.

Some applied researchers have adopted a stochastic frontier model in the form:

$$F(x, y, b) = 0. \tag{21}$$

This is a general transformation function, which allows for joint production. However, if we adopt this paradigm, there is nothing to distinguish inputs (x) from bad outputs (b). In the case of a single output we could write: $y = f(x,b)$ but nothing would really distinguish inputs from bad outputs. Both to increase good outputs and to avoid reverse causality, one could estimate the technology using the input functions: $x = g(y,b)$ if y and b are thought to be exogenously determined. Examples of such practices can be found by modeling bad outputs (e.g. non-performing loans in the banking sector). In the DEA literature, there are also many studies that treat bad outputs as inputs (Hailu and Veeman, 2001; Scheel, 2001; Seiford and Zhu, 2002).

This approach does not deal explicitly with how bad outputs are generated. Murty and Russell (2002) propose the following model:

$$F(x, y, b) \le 0,\ b \ge g(x), \tag{22}$$

and the technology is, in reality, the intersection of two technology sets, described by the two inequalities above. Of course, bad outputs in tourism are not direct determinants of inputs. Using rooms, beds, and facilities does not generate bad outputs on its own. Adopting an "agnostic" formulation like $F(x,y,b) \le 0$ does not make much sense on prior grounds, without additional equations to describe how bad outputs are generated. The approach of Murty and Russell (2002) makes sense from the point of view of physics and engineering; inputs are transformed into outputs and, in the process, pollution is generated necessarily from the underlying physics of commodity production. So, an attempt to make b a function of y's is more or less misguided. The argument may be less convincing in contexts outside strictly engineering-based views of economics.

This state of affairs reveals that distinguishing bad outputs from inputs requires an additional set of equations, detailing how the bad outputs are produced. The model by Assaf et al. (2017) addresses this issue and can be used in contexts such as tourism destinations or hotels. Before we discuss their model, it is important to note that one implication of introducing bad outputs in production is that they are potentially endogenous. It is well known that estimating a transformation function of the form $F(x,y,b) = 0$ is econometrically challenging. The reason is that several variables may be endogenous, but we do not have additional equations to describe the endogeneity and complete the system with as many equations as endogenous variables. In the case of a single output, we have: $y = f(x,b) + v$, where v is an error term. However, it is well known that inputs will be correlated with this error term under a variety of assumptions. Bad outputs are also likely to be correlated with v, if they are jointly produced depending on the y's or the x's. For example, if $b = g(y)$, it is obvious that b and v are correlated. If, instead, $b = h(x)$,since inputs are correlated with v, the b's are also correlated with v. Suppose for simplicity that:

$$\begin{aligned} y &= \beta x + \alpha b + v_1, \\ x &= \gamma z + v_2, \\ b &= \delta y + v_3, \end{aligned} \tag{23}$$

where the first equation is the production function, the second equation posits that inputs are related to some exogenous variable z and bad outputs are related to good outputs. The Jacobian determinant of transformation from v's to the endogenous variables (y, x, b) is $|1 - \alpha\delta|$, which shows that we have simultaneity. System estimation cannot be avoided in this case and full information maximum likelihood would have to take the Jacobian term into account.

This illustrates that a production or distance function has to be completed with additional equations for valid inferences in a likelihood-based framework. We need two sets of additional equations. The first relates to the inputs and the second to bad outputs. So, the first set needs to focus on the first-order conditions from cost minimization or profit maximization. Which behavioral assumption is adopted depends on the context and on what the researcher believes to be most appropriate. The second set of equations relies on an explication of the nature of bad outputs and how or why they are produced[4].

Assaf et al. (2017), for instance, adopted the following stochastic frontier formulation:

$$-y_1 = f(xy^*) + v_0 + u_0 \tag{24}$$

where $y^* = [y_2 - y_1, \ldots, y_M - y_1]'$, denotes the vector of all outputs, u_0 is inefficiency. Outputs are distinguished from inputs, as bad outputs do not appear in the production function. They introduced two additional equations to the stochastic frontier model in (24), one for the bad outputs, and one for inputs.

$$\begin{aligned} b &= \Pi_{11}x + \Pi_{12}y + \Pi_{13}z + \Pi_{14}u_0 + v_1 + u_1, \\ x &= \Pi_{20}y + \Pi_{21}z + v_2, \end{aligned} \tag{25}$$

where the bad outputs depend on selected inputs, selected outputs, other control variables (z), as well as inefficiency. The second equation is related to the inputs as discussed above and can help address the endogeneity problem. For a discussion of these issues see Kumbhakar and Tsionas (2016).

[4] For a discussion of these issues see Kumbhakar and Tsionas (2016).

It is clear that much more is expected from future research. More innovative models to examine the role of bad outputs in tourism are needed. Models should be able to distinguish what makes bad outputs different from inputs in the tourism context. Ignoring this feature may produce misleading estimates of the frontier technology, particularly if such considerations are not endogenized or become part of a coherent behavioral framework.

Bayesian inference

In their review of the literature, Assaf and Tsionas (2015) emphasized the importance of using the Bayesian approach for measuring tourism performance. In a recent paper, Assaf et al. (2017) also highlighted the importance of the Bayesian approach and provided different instructions and codes on how to estimate various SF models using the Bayesian approach. So far, very few papers have used the Bayesian approach to measure tourism performance (Assaf & Agbola, 2014; Assaf & Tsionas, 2018a, Assaf & Tsionas, 2018a; Arbelo, Arbelo-Pérez, & Pérez-Gómez, 2018; Tsionas & Assaf, 2014; Assaf and Tsionas, 2015).

Across several related disciplines, the use of the Bayesian approach is rapidly growing (Zyphur and Oswald, 2015). One attractive feature of the method is that it delivers sample-specific results and does not depend on asymptotic properties. Often, Bayesian estimators have excellent small-sample properties. In measuring tourism performance, in particular, the Bayesian approach is equipped with numerical techniques that can handle complicated models of performance, whereas the sampling—theory approach is quite difficult to implement. It can be applied to complicated versions of SF models, such as the ones described above, including bad output models and unobserved latent variables which are frequently dynamic (autoregressive, for example). Often it is difficult to associate standard errors or other measures of uncertainty to quantities like efficiency, etc. The Bayesian approach automatically considers this uncertainty through Markov Chain Monte Carlo (MCMC) techniques.

In dynamic panel SF models, the Bayesian approach is also very flexible. For instance, one can implement dynamic panel SF models with firm-specific coefficients, where the use of sampling-theory estimators is highly challenging and is still under development. This is an important handicap, which limits the scope of sampling-theory estimators in dynamic panel data models. Bayesian procedures are more straightforward to apply in dynamic models and are easier to understand, as lagged dependent variables do not create new problems in terms of estimation for the Bayesian approach. Recent papers have also used the Bayesian approach to avoid pre-imposing a functional form on the SF model. For instance, the problem of parametric assumptions regarding the functional form has been addressed in Tsionas and Izzeldin (2018) and Tsionas and Mallick (2018). More details and implementation using Bayesian techniques, are provided in these two papers. The Bayesian approach can also be useful in the context of DEA, as we will discuss below.

Another useful feature of the Bayesian approach is that model comparison (as opposed to frequentist model selection) is straightforward. Suppose we have a posterior distribution with density $p(\theta|D) \propto L(\theta;D)p(\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^D$ is the parameter vector, and $L(\theta;D)$, $p(\theta)$ denote the likelihood function and prior respectively. The marginal likelihood or "evidence" is defined as:

$$M(D) = \int_{\Theta} L(\theta;D)p(\theta)d\theta, \tag{26}$$

and summarizes all evidence (prior and data-based) of a model. If we have different models, indexed by $m \in \{1, \ldots, M\}$, we can define the marginal likelihood for each model as:

$$M_m(D) = \int_{\Theta_m} L_m(\theta_m;D)p_m(\theta_m)d\theta_m,\ m = 1,\ \ldots, M. \tag{27}$$

In turn, model comparison can be based on the Bayes factor in favor of model (1), say, and against any other model as follows:

$$BF_{1:m}(D) = \frac{M_m(D)}{M_m(D)},\ m \in \{2,\ldots, M\}. \tag{28}$$

In fact, the posterior odds ratio is equal to the Bayes factor times the prior odds ratio in favor of model "1" and against model "m". If all prior model probabilities are equal (and, therefore, equal to $\frac{1}{M}$) then the posterior odds ratio is equal to the Bayes factor. If $BF_{1:m}(D)$ is much larger than one, then there is evidence in favor of model. In fact, we can compute posterior model probabilities, as follows:

$$P_m(D) = \frac{M_m(D)}{\sum_{m'=1}^{M} M_{m'}(D)},\ m = 1,\ \ldots, M, \tag{29}$$

when prior model probabilities are the same. Posterior model probabilities present the quantitative evidence in favor of any given model in a way which is fully consistent with posterior probabilities and probabilistic arguments. Of course, computation of marginal likelihoods is a non-trivial issue, and especially so in models that involve latent variables. One prominent example of latent variables is technical inefficiencies.

Data envelopment analysis

Now that we have provided some recommendations for more robust performance estimation in the stochastic frontier context, we turn our attention to DEA. In general, more studies have used DEA than SF to measure tourism performance (e.g. Sigala, Airey, Jones, & Lockwood, 2004; Barros, 2005; Barros et al., 2011; Corne, 2015; Aissa and Goaied, 2016; and Cuccia et al., 2017). The method is flexible, easy to use, and can effectively handle multiple inputs and outputs. The DEA literature on tourism is well established, covering a variety of basic and advanced DEA models such as Window DEA (Pulina, Detotto, & Paba, 2010), network DEA (Hsieh & Lin, 2010), multiple stage DEA (Huang, Ho, & Chiu, 2014), and slack-based context-dependent DEA (Cheng, Lu, & Chung, 2010).

In contrast to SF, however, DEA does not account for measurement error in the data. It is well known that "production relationships are often stochastic in nature. If we ignore this circumstance, the efficiency assessment results will be biased and give misleading conclusions" (Shang, Wang, & Hung, 2010, p.2506). Most tourism data, for instance, is highly subject to noise or measurement error. It is encouraging that recent tourism studies have adopted different approaches to provide some statistical inferences for DEA scores. Methods such as bootstrapping and stochastic DEA have, for instance, been recently used for that purpose (Assaf & Agbola, 2011; Barros et al., 2011; Benito, Solana, & López, 2014; Chaabouni, 2019; Olesen & Petersen, 2016; Pulina & Santoni, 2018; Salman Saleh, Assaf, & Son Nghiem, 2012; Sellers-Rubio & Casado-Díaz, 2018; Shang et al., 2010; Yin, Tsai, & Wu, 2015).

We still believe, however, that there are some important problems to be addressed. It is well known, for instance, that the bootstrapping approach has only asymptotic justification, so it is not based on finite-sample inference. As highlighted by Tsionas and Papadakis (2010, p.309), "it is not guaranteed that in finite samples the bootstrap will perform well or that the associated inferences will be precise. Clearly, both accuracy and precision depend on the size of the sample, as well as the specific data set". Within the context of DEA, Simar and Wilson (1998, 2000, and 2007) introduced the bootstrap as a device to perform statistical inferences for efficiency scores. The only problem is that bootstrap DEA estimators converge slowly (at a rate $2/(r + 1)$; where r is the number of inputs and outputs) so unless we have few inputs and outputs, the slow convergence can be a problem. Reduction of inputs and outputs is, of course, possible; as one may examine correlations between the inputs or between the outputs, replace some of them with linear combinations, drop some of them, etc. Having a more principled way to perform such tasks would be advantageous in terms of future research.

A proper use of the bootstrap technology also leads to consistent estimation in what is known as the second-stage regression, where performance scores are regressed on a set of environmental variables (Simar & Wilson, 2007). Such processes have been used extensively in tourism research (Barros et al., 2011; Benito et al., 2014; Yin et al., 2015). One issue that has been ignored in these papers, however, is the importance of the separability condition. Daraio, Simar, and Wilson (2018) showed that if a certain separability assumption is violated, both second-stage results and first-stage efficiency scores might be meaningless. The separability assumption implies that environmental variables affect only the distribution of efficiency and not the production possibility set. The assumption is clearly restrictive, but can be tested. Moreover, a significance level of 20%, 40%, or even higher should be used to be on the safe side when separability does hold. If *p*-values of the separability tests are close to zero this is no help, but if they are marginal (say 1% or 5%) this makes things easier for practitioners. It seems that the unconditional approach to efficiency should be abandoned altogether unless, of course, separability tests are supportive of the notion that environmental variables (z) only affect the distribution of efficiency and not production itself. Conditional efficiency on the other hand is based on the idea of examining DEA sets localized with respect to z. The conditional efficiency model has been introduced by Cazals et al. (2002) and was developed further by Daraio and Simar (2005).

The bootstrap approach can also be extended to the context of stochastic DEA. In stochastic DEA, one admits that the data may be noisy, and this is taken into account by requiring that the DEA constraints do not hold exactly, but only with a certain probability (say 95%). So far, only one study has used stochastic DEA in the tourism literature. This method can bring several advantages, especially if accompanied with an approach that would provide formal statistical inference. Tsionas and Papadakis (2010) emphasized that while bootstrapping can be used with stochastic DEA, it introduces several statistical and conceptual problems. For instance, when using stochastic DEA, "it is assumed that the statistical model follows a multivariate normal distribution with unknown mean and covariance matrix" (Tsionas and Papadakis, 2010, p.309). The bootstrap, on the other hand "would generate alternative data sets that could have been observed based on estimated parameters from the sample. However, such parameters are unknown so uncertainty with respect to parameters is not accounted for" (Tsionas and Papadakis, 2010, p.309). In addition, the bootstrap is based on alternative data sets that could have been generated, but were never actually observed.

An attractive alternative for tourism researchers is to use the Bayesian approach with stochastic DEA. In Tsionas and Papadakis (2010), the authors assume a joint data-generating-process (DGP) for the inputs and outputs (say D) depending on parameters θ. The idea of Bayesian Stochastic DEA is to draw from the posterior $p(\theta|D)$, solve the stochastic DEA problem to compute efficiency scores, and average across all draws to obtain efficiency distributions robust to parameter uncertainty. In the bootstrap-DEA approach, the idea is different as the data is resampled (in the proper manner) and DEA scores are computed for each bootstrap sample. In Bayesian stochastic DEA, uncertainty arises from the fact that θ is unknown and enters the stochastic DEA problem in a natural way, (i.e. the constraints depend explicitly on θ), which is not the case in DEA. This is why the authors mention that "there can be no Bayesian inference approach in the non-stochastic DEA context: since the DEA problem defines an operator T that does not depend on the parameter vector, θ, it follows that, conditional on the data, there is no variability [in efficiency scores] and, therefore, no statistical inference questions arise" (Tsionas and Papadakis, p. 311). As the data is given, a Bayesian investigator cannot use sampling variability to justify uncertainty in efficiency scores. This is true for the particular definition of θ, which, in this context, is related directly to the joint distribution of D. However, it is imaginable that there can be other DGPs in which the definition of θ permits statistical inference in a way that relates θ and D through the operator T. This area of research is open and any further developments would be welcome. For example, the statistical literature discusses that the bootstrap and Bayesian posteriors are closely related in certain problems. If this is also the case in DEA or stochastic DEA remains to be investigated. This problem is not only of theoretical interest, there may also be computational advantages in a Bayesian approach to the problem.

Finally, other research areas that we highlighted above (such as the issue of accounting for bad outputs) are also possible with DEA (Färe & Grosskopf, 2004; Hua & Bian, 2007), but still have not received enough attention in the tourism literature. DEA is also flexible in settings where inputs and outputs are not necessarily quantitative. For instance, many input data in tourism (i.e. management competence) involve ordinal data. Special DEA models have been developed for such cases (see Cook & Seiford, 2009; Cooper, Park, & Yu, 1999 and Zhu, 2003) and we encourage their use in tourism research. Other DEA models that would be of interest are those that account for uncontrollable and categorical variables (see Cook & Seiford, 2009).

Concluding remarks

Despite the rapid advances in the use of frontier methods for measuring tourism performance, the literature still lacks methodological rigor. The present study highlighted the current status of the literature and provided several methodological recommendations for future research. We distinguished between non-parametric and parametric methods and discussed the current gaps in the literature. In general, we believe that more attention should be paid to the following methodological issues: endogeneity, bad outputs, dynamic formulation, heterogeneity, Bayesian estimation, bootstrapping and stochastic estimation. We discussed and presented several models that can be used by tourism researchers to account for these issues. Not only do these models provide more robust performance estimation, but will also increase the reliability of hypothesis testing around tourism performance.

As the literature is currently well developed, for the most part at least, it is important to estimate frontier models that are more robust and carefully checked for some important issues such as endogeneity and heterogeneity. For applied researchers we recommend trying alternative specifications before proceeding with a static SF model. These specifications include i) heterogeneity, ii) dynamics, iii) endogeneity, and iv) all of these in a unifying model. Estimating a unifying model may be tedious but the upside is that testing the individual specifications becomes easy through, for example, the use of Bayes factors. The same problems are open in DEA as well. Endogeneity is not a concern there, although in models such as Simar and Wilson (2007) it will make a difference. Dynamic models are also available in DEA although explicit assumptions about inefficiency are not made. This is an advantage, as assumptions are not made. However, it is also a drawback in the sense that if inefficiency is autocorrelated, having more efficient estimators becomes more important in small samples.

Statement of contributions

What is the contribution to knowledge, theory, policy or practice offered by the paper?

This paper is submitted to the VSI on Tourism performance. We provide a full overview of the literature, highlight some of the main gaps, and suggest areas in which there is room for improvement. We focus specifically on frontier methods, differentiating between non-parametric and parametric methods. We elaborate on key methodological issues such as endogeneity, bad outputs, dynamic formulation, heterogeneity, Bayesian estimation, bootstrapping, and stochastic estimation. For each of these areas we discuss and introduce some recent methodological breakthroughs that have been largely ignored in the tourism literature.

How does the paper offer a social science perspective/approach?

The focus on performance provides meaningful implications from the perspective of strategic management and financial management. The topic certainly falls under the umbrella of social science and may encourage more use of innovative performance models in hospitality and tourism research.

References

Ahn, S. C., Good, D. H., & Sickles, R. C. (2000). Estimation of long-run inefficiency levels: A dynamic frontier approach. *Econometric Reviews, 19*(4), 461–492.

Aissa, S. B., & Goaied, M. (2016). Determinants of Tunisian hotel profitability: The role of managerial efficiency. *Tourism Management, 52*, 478–487.

Altin, M., Koseoglu, M. A., Yu, X., & Riasi, A. (2018). Performance measurement and management research in the hospitality and tourism industry. *International Journal of Contemporary Hospitality Management, 30*(2), 1172–1189.

Anderson, R. I., Fish, M., Xia, Y., & Michello, F. (1999). Measuring efficiency in the hotel industry: A stochastic frontier approach. *International Journal of Hospitality Management, 18*(1), 45–57.

Arbelo, A., Arbelo-Pérez, M., & Pérez-Gómez, P. (2018). Estimation of profit efficiency in the hotel industry using a Bayesian stochastic frontier model. *Cornell Hospitality Quarterly, 59*(4), 364–375.

Archer, B., Cooper, C., & Ruhanen, L. (2005). The positive and negative impacts of tourism. *Global Tourism, 3*, 79–102.

Assaf, A. G., & Agbola, F. W. (2011). Modelling the performance of Australian hotels: A DEA double bootstrap approach. *Tourism Economics, 17*(1), 73–89.

Assaf, A. G., & Agbola, F. W. (2014). Efficiency analysis of the Australian accommodation industry: A Bayesian output distance function. *Journal of Hospitality and Tourism Research, 38*(1), 116–132.

Assaf, A. G., & Cvelbar, L. K. (2015). Why negative outputs are often ignored: A comprehensive measure of hotel performance. *Tourism Economics, 21*(4), 761–773.

Assaf, A. G., Gillen, D., & Barros, C. (2012). Performance assessment of UK airports: Evidence from a Bayesian dynamic frontier model. *Transportation Research Part E: Logistics and Transportation Review, 48*(3), 603–615.

Assaf, A. G., & Josiassen, A. (2016). Frontier analysis: A state-of-the-art review and meta-analysis. *Journal of Travel Research, 55*(5), 612–627.

Assaf, A. G., Oh, H., & Tsionas, M. (2017). Bayesian approach for the measurement of tourism performance: A case of stochastic frontier models. *Journal of Travel Research, 56*(2), 172–186.

Assaf, A. G., & Tsionas, M. (2018a). Measuring hotel performance: Toward more rigorous evidence in both scope and methods. *Tourism Management, 69*, 69–87.

Assaf, A. G., & Tsionas, M. (2018b). The estimation and decomposition of tourism productivity. *Tourism Management, 65*, 131–142.

Atkinson, S. E., & Tsionas, M. G. (2016). Directional distance functions: Optimal endogenous directions. *Journal of Econometrics, 190*(2), 301–314.

Barros, C. P. (2004). A stochastic cost frontier in the Portuguese hotel industry. *Tourism Economics, 10*(2), 177–192.

Barros, C. P. (2006). Analysing the rate of technical change in the Portuguese hotel industry. *Tourism Economics, 12*(3), 325–346.

Barros, C. P., Botti, L., Peypoch, N., Robinot, E., Solonandrasana, B., & Assaf, A. (2011). Performance of French destinations: Tourism attraction perspectives. *Tourism*

Management, 32(1), 141–146.
Barros, C. P., Dieke, P. U., & Santos, C. M. (2010). Heterogeneous technical efficiency of hotels in Luanda, Angola. *Tourism Economics, 16*(1), 137–151.
Barros, C. P., & Matias, Á. (2006). Assessing the efficiency of travel agencies with a stochastic cost frontier: A Portuguese case study. *International Journal of Tourism Research, 8*(5), 367–379.
Barros, C. P. (2005). Measuring efficiency in the hotel sector. *Annals of Tourism Research, 32*(2), 456–477.
Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis, 3*, 153–169.
Benito, B., Solana, J., & López, P. (2014). Determinants of Spanish regions' tourism performance: A two-stage, double-bootstrap data envelopment analysis. *Tourism Economics, 20*(5), 987–1012.
Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of econometrics, 106*(1), 1–25.
Chaabouni, S. (2019). China's regional tourism efficiency: A two-stage double bootstrap data envelopment analysis. *Journal of destination marketing & management, 11*, 183–191.
Chen, C. M., Delmas, M. A., & Lieberman, M. B. (2015). Production frontier methodologies and efficiency as a performance measure in strategic management research. *Strategic Management Journal, 36*(1), 19–36.
Cheng, H., Lu, Y. C., & Chung, J. T. (2010). Improved slack-based context-dependent DEA–A study of international tourist hotels in Taiwan. *Expert Systems with Applications, 37*(9), 6452–6458.
Cook, W. D., & Seiford, L. M. (2009). Data envelopment analysis (DEA)–thirty years on. *European Journal of Operational Research, 192*(1), 1–17.
Cooper, W. W., Park, K. S., & Yu, G. (1999). IDEA and AR-IDEA: Models for dealing with imprecise data in DEA. *Management Science, 45*(4), 597–607.
Corne, A. (2015). Benchmarking and tourism efficiency in France. *Tourism Management, 51*, 91–95.
Cornwell, C., Schmidt, P., & Sickles, R. (1990). Production frontiers with cross-sectional and time-series variation in efficiency levels. *Journal of Econometrics, 46*, 185–200.
Cuccia, T., Guccio, C., & Rizzo, I. (2017). UNESCO sites and performance trend of Italian regional tourism destinations: A two-stage DEA window analysis with spatial interaction. *Tourism Economics, 23*(2), 316–342.
Daraio, C., & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of productivity analysis, 24*(1), 93–121.
Daraio, C., Simar, L., & Wilson, P. W. (2018). Central limit theorems for conditional efficiency measures and tests of the 'separability' condition in non-parametric, two-stage models of production. *The Econometrics Journal, 21*(2), 170–191.
De Geuser, F., Mooraj, S., & Oyon, D. (2009). Does the balanced scorecard add value? Empirical evidence on its effect on performance. *The European Accounting Review, 18*(1), 93–122.
Deng, Z., Gao, Y., Liang, B., & Morrison, A. M. (2018). Efficiency evaluation of hotel operations in Mainland China based on the superefficiency SBM model. *Tourism Economics* (in press).
Desli, E., Ray, S. C., & Kumbhakar, S. C. (2003). A dynamic stochastic frontier production model with time-varying efficiency. *Applied Economics Letters, 10*(10), 623–626.
Evans, N. (2005). Assessing the balanced scorecard as a management tool for hotels. *International Journal of Contemporary Hospitality Management, 17*(5), 376–390.
Färe, R., & Grosskopf, S. (2004). Modeling undesirable factors in efficiency evaluation: Comment. *European Journal of Operational Research, 157*(1), 242–245.
Fernandez, C., Koop, G., & Steel, M. F. J. (2002). Multiple-output production with undesirable outputs: An application to nitrogen surplus in agriculture. *Journal of the American Statistical Association, 97*(458), 432–442.
Greene, W. (2005a). Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics, 126*(2), 269–303.
Greene, W. (2005b). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis, 23*(1), 7–32.
Hailu, A., & Veeman, T. S. (2001). Non-parametric productivity analysis with undesirable outputs: An application to the Canadian pulp and paper industry. *American Journal of Agricultural Economics, 83*(3), 605–616.
Herrero-Prieto, L. C., & Gomez-Vega, M. (2017). Cultural resources as a factor in cultural tourism attraction: Technical efficiency estimation of regional destinations in Spain. *Tourism Economics, 23*(2), 260–280.
Hsieh, L. F., & Lin, L. H. (2010). A performance evaluation model for international tourist hotels in Taiwan—An application of the relational network DEA. *International Journal of Hospitality Management, 29*(1), 14–24.
Hu, J.-L., Chiu, C.-N., Shieh, H.-S., & Huang, C.-H. (2010). A stochastic cost efficiency analysis of international tourist hotels in Taiwan. *International Journal of Hospitality Management, 29*(1), 99–107.
Hua, Z., & Bian, Y. (2007). DEA with undesirable factors. *Modeling data irregularities and structural complexities in data envelopment analysis* (pp. 103–121). Boston, MA: Springer.
Huang, C. W., Ho, F. N., & Chiu, Y. H. (2014). Measurement of tourist hotels' productive efficiency, occupancy, and catering service effectiveness using a modified two-stage DEA model in Taiwan. *Omega, 48*, 49–59.
Kaplan, R. S., & Norton, D. P. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review, 74*(January–February), 75–85.
Khalaf, L., & Saunders, C. J. (2016). Dynamic technical efficiency. *Productivity and efficiency analysis* (pp. 99–107). Cham: Springer.
Kumbhakar, S. C. (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics, 46*(1–2), 201–211.
Kumbhakar, S. C. (1991). Estimation of technical inefficiency in panel data models with firm- and time-specific effects. *Economics Letters, 36*(1), 43–48.
Kumbhakar, S. C., & Tsionas, E. G. (2016). The good, the bad and the technology: Endogeneity in environmental production models. *Journal of Econometrics, 190*(2), 315–327.
Kumbhakar, S. C., & Wang, H.-J. (2005). Estimation of growth convergence using a stochastic production frontier approach. *Economics Letters, 88*, 300–305.
Lee, Y., & Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical efficiency. *The measurement of productive efficiency: Techniques and applications*. Oxford University Press.
Lipe, M. G., & Salterio, S. E. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review, 75*(3), 283–298.
Luo, X., & Homburg, C. (2008). Satisfaction, complaint, and the stock value gap. *Journal of Marketing, 72*(4), 29–43.
Mendieta-Peñalver, L. F., Perles-Ribes, J. F., Ramón-Rodríguez, A. B., & Such-Devesa, M. J. (2018). Is hotel efficiency necessary for tourism destination competitiveness? An integrated approach. *Tourism Economics, 24*(1), 3–26.
Murty, S., & Russell, R. (2002). *On modeling pollution generating technologies University of California-Riverside (no. 2002-14). Working paper series*.
Ohe, Y., & Peypoch, N. (2016). Efficiency analysis of Japanese Ryokans: A window DEA approach. *Tourism Economics, 22*(6), 1261–1273.
Olesen, O. B., & Petersen, N. C. (2016). Stochastic data envelopment analysis—A review. *European Journal of Operational Research, 251*(1), 2–21.
Park, S., & Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science, 31*(4), 567–586.
Pérez-Rodríguez, J. V., & Acosta-González, E. (2007). Cost efficiency of the lodging industry in the tourist destination of Gran Canaria (Spain). *Tourism Management, 28*(4), 993–1005.
Pulina, M., Detotto, C., & Paba, A. (2010). An investigation into the relationship between size and efficiency of the Italian hospitality sector: A window DEA approach. *European Journal of Operational Research, 204*(3), 613–620.
Pulina, M., & Santoni, V. (2018). A two-stage DEA approach to analyse the efficiency of the hospitality sector. *Tourism Economics, 24*(3), 352–365.
Rungsuriyawiboon, S., & Stefanou, S. E. (2007). Dynamic efficiency estimation: An application to U.S. electric utilities. *Journal of Business and Economic Statistics, 25*(2), 226–238.
Salman Saleh, A., Assaf, A. G., & Son Nghiem, H. (2012). Efficiency of the Malaysian hotel industry: A distance function approach. *Tourism Analysis, 17*(6), 721–732.
Scheel, H. (2001). Undesirable outputs in efficiency valuations. *European Journal of Operational Research, 132*(2), 400–410.
Schmidt, P., & Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business and Economic Statistics, 2*(4), 367–374.
Seiford, L. M., & Zhu, J. (2002). Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research, 142*(1), 16–20.

Sellers-Rubio, R., & Casado-Díaz, A. B. (2018). Analyzing hotel efficiency from a regional perspective: The role of environmental determinants. *International Journal of Hospitality Management, 75*, 75–85.
Shang, J. K., Wang, F. C., & Hung, W. T. (2010). A stochastic DEA study of hotel efficiency. *Applied Economics, 42*(19), 2505–2518.
Sigala, M., Airey, D., Jones, P., & Lockwood, A. (2004). ICT paradox lost? A stepwise DEA methodology to evaluate technology investments in tourism settings. *Journal of Travel Research, 43*(2), 180–192.
Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science, 44*(1), 49–61.
Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics, 27*(6), 779–802.
Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics, 136*(1), 31–64.
Tsionas, E. G. (2002). Stochastic frontier models with random coefficients. *Journal of Applied Econometrics, 17*(2), 127–147.
Tsionas, E. G., & Kumbhakar, S. C. (2014). Firm heterogeneity, persistent and transient technical inefficiency: A generalized true random-effects model. *Journal of Applied Econometrics, 29*(1), 110–132.
Tsionas, E., Assaf, A. G., Gillen, D., & Mattila, A. S. (2017). Modeling technical and service efficiency. *Transportation Research Part B: Methodological, 96*, 113–125.
Tsionas, E. G. (2006). Inference in dynamic stochastic frontier models. *Journal of Applied Econometrics, 21*, 669–676.
Tsionas, E. G., & Assaf, A. G. (2014). Short-run and long-run performance of international tourism: Evidence from Bayesian dynamic models. *Tourism Management, 42*, 22–36.
Tsionas, M. G., & Izzeldin, M. (2018). Smooth approximations to monotone concave functions in production analysis: An alternative to nonparametric concave least squares. *European Journal of Operational Research, 271*(3), 797–807.
Tsionas, E. G., & Papadakis, E. N. (2010). A Bayesian approach to statistical inference in stochastic DEA. *Omega, 38*(5), 309–314.
Tsionas, M. G., & Mallick, S. K. (2018). A Bayesian semiparametric approach to stochastic frontiers and productivity. *European Journal of Operational Research, 274*(1), 391–402.
Wang, C. (2007). A dynamic stochastic frontier production model with time-varying efficiency: Comment. *Applied Economics Letters, 14*(6), 415–417.
Yin, P., Tsai, H., & Wu, J. (2015). A hotel life cycle model based on bootstrap DEA efficiency: The case of international tourist hotels in Taipei. *International Journal of Contemporary Hospitality Management, 27*(5), 918–937.
Zhang, Q., Xu, Z., Feng, T., & Jiao, J. (2015). A dynamic stochastic frontier model to evaluate regional financial efficiency: Evidence from Chinese county-level panel data. *European Journal of Operational Research, 241*, 907–916.
Zhu, J. (2003). Imprecise data envelopment analysis (IDEA): A review and improvement with an application. *European Journal of Operational Research, 144*(3), 513–529.
Zyphur, M. J., & Oswald, F. L. (2015). Bayesian estimation and inference: A user's guide. *Journal of Management, 41*(2), 390–420.

A. George Assaf (PhD) is an Associate Professor at the Isenberg School of Management, University of Massachusetts-Amherst. His work has appeared in several tourism, management and economic journals, such as Tourism Management, Journal of Travel Research, Annals of Tourism Research, and Journal of Retailing.

Mike Tsionas (PhD) is a Professor of Econometrics at Lancaster University. He serves as an Associate Editor of the Journal of Productivity Analysis. His work has appeared in leading statistics and econometric journals such as the Journal of the American Statistical Association (JASA), Journal of Econometrics, and Journal of Applied Econometrics, among others.