The Estimation and Decomposition of Tourism Productivity: Into a More Robust Estimation using Bayesian Techniques

Introduction

"Productivity isn't everything but in the long run it's almost everything" (Jones, 2007, p. 1).

Despite being a high priority on the World Tourism Organization (UNWTO) research agenda, the productivity analysis of the tourism industry has not received much attention in the tourism literature. There is a continuous effort at most tourism destinations to strengthen the productivity of their tourism industry (Cvelbar et al. 2015). As stated by Assaf and Dwyer (2013), with the tourism industry often perceived as a low productivity industry, productivity analysis is "crucial to evaluating tourism sustainability and reshaping tourism activities. There is a direct link between productivity and profitability, as when productivity increases, the tourism industry's competitiveness in labour, capital and real estate markets also increase".

Often misleading is the definition of productivity in the tourism industry. The various league tables providing productivity indicators of the tourism industry "neither takes explicit account of productivity in tourism" (Blake, 2006, p.1100). Productivity is a complex phenomenon and involves several components; hence using simple metrics to reflect the overall tourism productivity can be misleading for policy implications (Peypoch et al., 2012). Over the last decade, there has been an increasing focus on analysing the performance of the tourism industry using the concept of "technical efficiency" (Peypoch and Solonandrasana, 2006; Peypoch et al., 2012; Assaf and Josiassen, 2012). However, while technical efficiency is a comprehensive measure of performance, it is only one component of productivity- productivity growth is not driven by technical efficiency alone, but by other factors such as "innovation" and "output growth" (Coelli et al. 2005). The concept of "competitiveness "should not also be used to reflect the productivity of the tourism industry (Assaf and Josiassen, 2012) - productivity is a major driver of "competitiveness", and not "competitiveness" itself (Cvelbar et al. 2015).

In their recent paper, Assaf and Dwyer (2013) emphasized that the highly popular "Travel & Tourism Competeviness Index" published by the World Economic Forum and widely used by tourism destinations should not be used as an index of productivity- it does not rank destinations based on their tourism productivity (Cvelbar et al. 2015). There is clearly a need to complement such index with a robust productivity index that takes into consideration the unique multiple input and output characteristics of the tourism industry (Assaf and Josiassen, 2012). The Malmquist productivity index, for example, recently used in tourism to measure tourism productivity (Barros, 2005, Gracolici et al. 2007; Peypoch, 2008) is an important step in the right direction-It is a comprehensive index that takes into account multiple inputs and outputs in the measurement of tourism productivity, and can be decomposed into measures of efficiency growth and technical growth.

Motivated by the above, the aim of this paper is to extend the current literature on tourism productivity, addressing several important gaps in the literature. For the first time, we introduce a highly advanced total factor productivity index that allows for a rich decomposition of the sources of productivity growth in the tourism industry. We focus on different decompositions such as output growth, input growth, efficiency growth and technical growth, with each of these components providing an important source of policy implication for the tourism industry[1]. As our index is highly complicated we use the Bayesian approach based on Sequential Monte Carlo / Particle Filtering (SMC/PF) to perform the computations.

Importantly our index also introduces four important innovations to the tourism literature. First, we account for heterogeneity between multiple tourism destinations, something that has been completely ignored in related studies. As it is well known that considerable heterogeneity exists between tourism destinations, a failure to account for this can result in biased conclusions (Assaf and Tsionas, 2015). Second, our index accounts for potential endogeneity in inputs using a reduced form equation that also takes into account the fact that productivity and inputs cannot be independent of each other. Third, we develop our index at the macro and not at the micro level as is the case with most studies in the tourism literature. As stated by Assaf and Dwyer (2013) "for productivity measures to be even more useful and relevant to public policy and regulation, they need to relate to the overall tourism industry, and not just to particular sectors of the industry". Fourth and finally, we focus on cross-country comparisons; our aim is to provide each destination with a more accurate assessment of the international standing of their tourism industry. This differentiates the study from most existing tourism benchmarking studies in the literature that are limited to one single destination.

The paper will proceed as follow. The next section provides a background of productivity and highlights some of the competing methods. Section 3 reviews the current literature on tourism productivity and highlights some of the existing gaps. Section 4 presents the model. Section 5…

Benchmarking and Productivity

Interest in productivity has revived in econometrics through the work of Olley and Pakes (1996) and Levinsohn and Petrin (2003). Across many industries, productivity remains one of most comprehensive and reliable benchmark (Coelli et al. 2005). While in tourism, studies have benchmarked tourism destinations with respect to several performance indicators such as customer satisfaction (Milman and Pizam, 1995), competitiveness (Kozak and Rimmington, 1999), and market share (Dwyer and Kim, 2003), the use of productivity remains largely limited. For tourism policy makers "all these issues are important, but the problem is that they lead to subjectivity in selecting the true benchmarking parameters" (Assaf and Dwyer, 2013).

A more obvious and established benchmark is productivity (Jones, 2007). Usually measured based on multiple inputs and outputs, productivity provides a more comprehensive benchmark and reduces the subjectivity in comparing between different industry leaders (Peypoch et al. 2012). To define productivity, we start with a production function of this form:

$$Y_{it} = \lambda_{it} f\left(X_{it}\right) \qquad (1)$$

where Y_{it} refers to the output, X_{it} is a vector of inputs, and λ refers to "how much output a given input is able to produce from a certain amount of inputs, given the technological level" (Gatto et al, 2011). The total factor productivity index (TFP) at a time period "*t*" is the ratio of produced output and total inputs used:

$$TFP_{it} \equiv \lambda_{it} = \frac{Y_{it}}{f\left(X_{it}\right)} \qquad (2)$$

As simple as it looks, the estimation of productivity in (2) is not that straightforward- particularly when there are multiple and outputs where finding the appropriate weights becomes challenging.

[1] Other indices including the Malmquist do not allow for such decomposition.

There is an array of methodologies, and the distinction between them is not just in terms of whether they use a deterministic vs. a parametric approach, but also in terms of whether they adapt a micro (i.e. firm) vs. a macro level approach (industry/country, etc.).

The early literature on the measurement of aggregate productivity growth started with "the Solow growth theory (1957)[2], in which the pattern of productivity growth essentially mirrors that of the so-called technologies progress (i.e. Solow residual)" (Gatto et al, 2011). Such approach is also known as "growth accounting", and despite the limitations, is still a very popular methodology. Recent extension of this method also includes the "level accounting" decomposition (Caselli, 2005), which has the advantage of providing not only growth measures but also estimates of productivity levels, and the so called "growth regressions" where productivity is not estimated as a residual (like "growth accounting"), and is not dependent on a specific functional form (Islam, 1995). This method has also the advantage of not requiring data on physical capital, which in most cases, are usually characterized by high measurement errors[3].

In tourism and other related industries. frontier methods have been the most popular for measuring both aggregate and firm level productivity. In contrast to the non-frontier methods they provide two unique advantage. First, they do not assume producers to be always using their full existing technology. When technical inefficiency is present, which is often the case, productivity is also affected resulting in a productivity change over time (unless technical efficiency is constant over time). Frontier methods have also a high flexible capability in disentangling the source of productivity change into technical efficiency changes and technological change. Technological change results from shift or the frontier of best practices over time, where an upward shift reflects a sign of innovation or technological progress, while efficiency growth reflects a growing ability of firms to improve their production with a given set of inputs. Both these measures have important implications for productivity improvement in the tourism industry. For instance, if the main source of productivity decline is negative change in technology, this would suggest that policies should be more directed towards investing in technological innovation.

To fully understand the difference between frontier and non-frontier methods, one can rewrite equation 1 with relative to the frontier function $Y_{it}^{*} = \lambda_{it}^{*} f^{*}(X_{it})$:

$$\frac{Y_{it}}{Y_{it}^{*}} = \lambda_{it} \frac{f(X_{it})}{f^{*}(X_{it})} \qquad (3)$$

where the difference between the observed output and the frontier in (3) is due to either a lack of ability to improve outputs given a input and the technology λ_{it}, or due to a lack of technical efficiency with respect to the frontier $\frac{f(X_{it})}{f^{*}(X_{it})}$.

The two popular methodologies for estimating productivity using the frontier methodology are the Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA). As both these methods are now well established in the tourism literature we do not reiterate their technical details here. Generally, DEA is a non-parametric frontier which envelops the input/output combination of the data in order to obtain the closest approximation possible of the best-practice frontier, and from here estimates the measures of productivity change, technological change and efficiency change. The main strengths of the method and which overcome several

---

[2] For more details, see Solow (1957).

[3] For a more detailed review of these methods refer to Gatto et al. (2011).

limitations of the growth accounting literature are that it: 1- does not require functional form for the technology, 2- does not impose any assumption on the market structure and does not make the hypothesis that markets are perfect (Gatto et al. 2001).

SFA also has the same strengths as DEA but has an additional advantage of accommodating for random error that is beyond the control of a firm. Several approaches for estimating productivity using SFA have also been proposed in the literature (Kumbhakar, 2000; Orea, 2002). In this paper, we build on this SFA literature and provide some important extensions that are of high relevance to the tourism industry. In contexts like ours, where the sample involves comparing between international and heterogeneous tourism destinations, using DEA may be even more sensitive to measurement error. With SFA being an econometric approach, one can also impose more advanced assumptions on the model, and produce more robust productivity estimates.

Current Gaps in the Literature

Before discussing the current gaps in the literature, it is important to emphasize that this review focuses mainly on the productivity studies within the tourism literature. While there are many DEA and SFA studies in tourism, these are mainly "efficiency" studies and not productivity studies[4]. To clearly highlight the current gaps, Table 1 lists the existing studies based on several criteria, including the methodology and the sample used, the extent of productivity decomposition, as well as assumptions made on the model.

Several trends can be identified from Table 1:

- First, most studies use the non-parametric DEA approach to estimate productivity, adopting well established indices such the Malmquist and Luenberger productivity indices[5]. None of these studies, however, use the SFA approach, which in contexts like comparing between international destinations, where the data is usually plagued by measurement error, has a clear advantage.
- Second, most existing studies focus on one tourism destination, or multiple destinations within one specific geographic region. Only one study has compared between international tourism destinations.
- Third, none of these studies impose a dynamic assumption on the model to allow for both short-run and long-run productivity estimates. For policy implications both these measures become important as while a destination might be performing well in the short-run, its long-run estimate may show otherwise.
- Fourth, with the exception of one study, none of the existing studies has accounted for heterogeneity in modelling tourism productivity. It would 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 doesn't exist. It is also a matter of economic and natural resources available in each country (the "feasible" touristic products mix) that prevents to consider "standard" the output of the production function. One can think at the different facilities demanded by business or leisure

[4] Assaf and Josiassen (2015) recently indicated that there are now more than 60 studies using SFA and DEA in the tourism literature. For a detailed review of these studies, we refer the reader to their paper.
[5] For more details on the Malmquist and Luenberger productivity indices refer to Fare et al. (2008).

tourists, but also at the differences between leisure tourism in Russia and in Guatemala (as an example). This heterogeneity could leads to a strong bias in the estimation of the frontier's parameters, in the productivity estimates and hence in the ranking proposed. The fact that the recent literature uses one destination (or different regions of a same destination) is not also enough to convince that there isn't a problem of heterogeneity in the production technology and - consequently - a bias in the evaluation of destination' performance.

- Fifth, most of the existing studies have focused only on two types of productivity decomposition: efficiency change and technological change. We believe that providing a richer decomposition can help better identify the sources of productivity growth in the tourism industry.
- Sixth, and finally, none of these studies has used the Bayesian approach. In complicated models like ours where impose a dynamic assumption on the model and account for heterogeneity, the Bayesian approach provides higher flexibility than traditional estimation methods such as the Maximum Likelihood.

The present paper aims to address all the above limitations. We introduce a unique productivity index that takes up the idea that productivity is a dynamic process, and develop appropriate methods of estimation in the context of multiple-input, multiple-output production which is, typically, the characteristic of the tourism industry. An input distance function is used to describe the technology. We address the problem of unobserved heterogeneity, that is not simple and cannot be captured using fixed-effects formulations. Full heterogeneity requires that the parameters of the input distance function change across individuals and over time. This shift of the frontier also generates growth that is different from productivity growth and can be identified. Our model also addresses the classical endogeneity problem in inputs by posing a reduced form for inputs, taking the assumption that inputs are not necessarily uncorrelated with productivity or the random error term in the input distance function. For the time in this area, use the Bayesian approach to perform the computation, using highly advanced Sequential Monte Carlo / Particle Filtering (SMC/PF) techniques.

We provide a rich decomposition of the providing index, deriving measures such as input growth, output growth, efficiency growth, frontier growth and productivity growth, where each of these components provides an important source of policy implication for the tourism industry. We also derive short-term and long terms measure of these estimates. Our sample is unique in that we compare between more than 100 international tourism destinations, providing hence better complement to other international statistical releases published by organizations such the United Nations World Tourism Organization (UNWTO) or the World Travel and Tourism Council (WTTC).

The Model

As stated, we develop our model here using the frontier approach[6]. Suppose $X \in \mathbb{R}^K$ is a vector of inputs, $Y \in \mathbb{R}^M$ is a vector of outputs and $Z \in \mathbb{R}^{d_z}$ is a vector of contextual or environmental variables. Our starting point is an input-oriented distance function of the form

$$D(X, Y, Z; \beta) = 1, \tag{1}$$

where $\beta \in \mathbb{R}^p$ is a vector of parameters. After imposing homogeneity of degree one with respect to input and using lower-case letters to denote logs and $x_1 = \log X_1$, $x_2 = \log\left(\frac{X_2}{X_1}\right), \ldots, x_K = \log\left(\frac{X_K}{X_1}\right)$ we have:

$$\begin{gathered} x_1 = f\left(x_2, \ldots, x_K, y_1, \ldots, y_M, z_1, \ldots, z_{d_z}; \beta\right) + v_1 - u \equiv \\ f\left(x_{(-1)}, y, z; \beta\right) + v_1 - u, \end{gathered} \tag{2}$$

where v_1 is a usual econometric error term, and $u \geq 0$ represents technical inefficiency (in the form of radial input over-utilization). If we denote $w = \left[x'_{(-1)}, y', z'\right]' \in \mathbb{R}^{d_w}$ so that the distance function takes the form:

$$x_1 = f\left(w\right) + v_1 + u, \tag{3}$$

We can use the translog functional form:

$$\begin{gathered} f(w) = a_o + a'w + \tfrac{1}{2} w'\Gamma w = \\ a_o + \sum_{j=1}^{d_w} a_j w_j + \tfrac{1}{2} \sum_{j=1}^{d_w} \sum_{h=1}^{d_w} \gamma_{jh} w_j w_h. \end{gathered} \tag{4}$$

The parameter vector β consists of the parameters in the above expression, viz. a_o, a and Γ. From these expressions we have the inputs distance function in the form:

$$\begin{aligned} x_1 &= a_o + a'w + \tfrac{1}{2} w'\Gamma w + v_1 + u \Rightarrow \\ x_1 &= a_o + \sum_{j=1}^{d_w} a_j w_j + \tfrac{1}{2} \sum_{j=1}^{d_w} \sum_{h=1}^{d_w} \gamma_{jh} w_j w_h + v_1 + u. \end{aligned} \tag{5}$$

Assuming the availability of panel data we can write the equation as follows:

---

[6] As the frontier approach is now well established in the tourism literature, we do not provide so much background details here. For a rich overview refer to Assaf and Josiassen (2016).

$$x_{1,it} = a_{o,it} + a'w_{it} + \tfrac{1}{2}w_{it}'\Gamma w_{it} + v_{1,it} + u_{it} \Rightarrow$$
$$x_{1,it} = a_{o,it} + \sum_{j=1}^{d_w} a_j w_{j,it} + \tfrac{1}{2}\sum_{j=1}^{d_w}\sum_{h=1}^{d_w} \gamma_{jh} w_{j,it} w_{h,it} + v_{1,it} + u_{it}, \quad (6)$$

where $a_{o,it}$ captures firm and time effects, $i = 1,...,n,\ t = 1,...,T.$. Our interest here focuses on productivity growth (PG) which is equal to technical change (TC) plus efficiency change (EC) where TC is, typically, measured by including time effects or a trend in (6).

Two problems arise and we can solve them at the same time. First, $x_{(-1),it}$ in (6) is endogenously determined. Second, productivity cannot be independent of the inputs used. To proceed we use the following reduced form equation for $x_{(-1),it}$ in (7):

$$x_{(-1),it} = \Pi z_{it} + V_{(-1),it}, \quad (7)$$

where Π is a $\left(K-1\right)\times d_z$ matrix of reduced form coefficients and $V_{(-1),it}$ is a $K \times 1$ error term. We assume the following error structure:

$$\left[v_{1,it}, V_{(-1),it}\right]' \sim N_K\left(0, \Sigma\right). \quad (8)$$

In this way endogeneity of $x_{(-1),it}$ is taken into account.

Another problem that empirical researchers often face, and which is of high importance when comparing between tourism destinations, is unobserved heterogeneity. The challenge is that unobserved heterogeneity cannot be captured using fixed-effects formulation as in (6). For this reason we assume that the parameters of the frontier are country-specific and time- specific, as follows:

$$x_{1,it} = a_{o,it} + a_{it}'w_{it} + \tfrac{1}{2}w_{it}'\Gamma_{it} w_{it} + v_{1,it} + u_{it} + \omega_{it} \Rightarrow$$
$$a_{o,it} + \sum_{j=1}^{d_w} a_{j,it} w_{j,it} + \tfrac{1}{2}\sum_{j=1}^{d_w}\sum_{h=1}^{d_w} \gamma_{jh,it} w_{j,it} w_{h,it} + v_{1,it} + u_{it} + \omega_{it}. \quad (9)$$

Let us denote

$$\beta_{it} = \left[a_{o,it}, a_{it}', vec\left(\Gamma_{it}\right)\right]' \in \mathbb{R}^{d_\beta}. \quad (10)$$

To model unobserved heterogeneity, we use a dynamic stochastic time-varying parameters framework:

$$\beta_{it} = b_i + A_i\beta_{i,t-1} + \Lambda z_{it} + e_{it},\ \ e_{it} \sim N_{d_\beta}\left(0, \Omega\right),\ \ i = 1,...,n, t = 1,...,T, \quad (11)$$

where b_i is a $d_\beta \times 1$ vector, A is a $d_\beta \times d_\beta$ matrix, Ω is the $d_\beta \times d_\beta$ covariance matrix of the error term and Λ is a $d_\beta \times d_z$ matrix of coefficients. In (11) we allow for stochastic time-varying parameters of the distance function where the dynamics of the parameter vector are country-specific through b_i and A_i. The model is quite general but we need shrinkage prior in order to estimate the parameters with accuracy. Our hierarchical prior for this model is:

$$\begin{aligned} &b_i \sim N_{d_\beta}\left(\overline{b},\ \overline{\Sigma}_b\right),\ i = 1,\ldots,n, \\ &a_i \equiv vec(A_i) \sim N_{d_\beta^2}\left(\overline{a},\ \overline{\Sigma}_A\right),\ i = 1,\ldots,n. \end{aligned} \tag{12}$$

The prior covariance matrices $\overline{\Sigma}_b$ and $\overline{\Sigma}_A$ control the degree of shrinkage. Our prior is that the z_{it} s are adequate in modeling the evolution of parameters so we would like to have $A_i = O_{d_\beta^2}$, viz. the zero matrix. Therefore, we set all elements of $\overline{a}$ equal to zero depending and we set $\overline{\Sigma}_A = h_A^2 I_{d_\beta^2}$ where h_A is a shrinkage parameter. We set $h_A = 10$ as we do not wish to place much confidence in our prior belief about A_i being a zero matrix. We also try a model without instruments in which case our prior for A_i is that it is an identity matrix, viz. $I_{d_\beta^2}$.

For vector b_i we do not have much prior information so we set $\overline{b} = 0$, $\overline{\Sigma}_b = h_b^2 I_{d_\beta}$ and we set, again, $h_b = 10$ as we do not wish to place much confidence in our prior belief.

When the number of exogenous variables is large or when the basic exogenous variables are few but we have to consider squares and cross-products, we need some way of controlling the proliferation of parameters. In this study, we adopt the procedure of Bayesian Compressed Regression (BCR) of Guhaniyogi and Dunson (2015). Specifically we replace the model in (9) and (11) by the following:

$$x_{(-1),it} = \Pi \tilde{z}_{it} + V_{(-1),it}, \tag{9b}$$

$$\beta_{it} = b_i + A_i \beta_{i,t-1} + \Lambda \tilde{z}_{it} + e_{it},\ e_{it} \sim N_{d_\beta}\left(0,\ \Omega\right),\ i = 1,\ldots,n, t = 1,\ldots,T, \tag{11b}$$

where

$$\tilde{z}_{it} = \Psi z_{it}, \tag{13}$$

is an $r \times 1$ vector of compressed variables resulting from z_{it} through the application of a linear transformation using the $r \times d_z$ matrix Ψ. Here, r is the rank (dimensionality) of the compressed regressors. Guhaniyogi and Dunson (2015) avoid estimation of $\Psi = \left[\Psi_{ij}\right]$ altogether by drawing its elements randomly as follows:

$$
\begin{aligned}
&P\left(\Psi_{ij}=\psi\right)=\psi^{2},\\
&P\left(\Psi_{ij}=0\right)=2\psi\left(1-\psi\right),\\
&P\left(\Psi_{ij}=-\psi\right)=\left(1-\psi\right)^{2},
\end{aligned}
\tag{14}
$$

where ψ is a parameter randomly drawn from (0.1, 1] –the lower bound is 0.1 instead of 0 for numerical stability. We search over different random draws, the rank r and the parameter ψ for $R=10^6$ times. We select the appropriate matrix Ψ, the rank r and the parameter ψ by maximizing the marginal likelihood of the model which is a natural byproduct of our Sequential Monte Carlo / Particle-Filtering techniques.

As will be described later, in our application we have 22 z's including a time trend. Taking squares of non-categorical variables and their interactions we have almost 230 exogenous variables that cannot possibly be used in conjunction with (9) and (13). In our empirical application we find that $r=12$ so we have huge dimensionality reduction in effect. The optimal parameter ψ turned out to be 0.31 based on maximizing the marginal likelihood of the model.

In this study we use Sequential Monte Carlo / Particle Filtering (SMC/PF) to perform the computations, see Technical Appendix. We use 10^6 particles per iteration for 15,000 iterations the first 5,000 of which are discarded to mitigate possible start-up effects. Our results remained the same when we used an additional 10,000 iterations with 10^7 particles per iteration. Convergence was monitored using the standard diagnostics (Geweke, 1992) and obtain within the first 5,000 iterations we discard. To ensure convergence further we use random initial conditions from the prior and run 100 different SMC chains for the baseline prior. We impose monotonicity and concavity restrictions in the translog functional form using rejection sampling. Specifically, we first impose these restrictions at the means of the data say $\overline{\mathbf{d}}$. Then we impose the same conditions at $\overline{\mathbf{d}}\pm h\mathbf{s}$ for $h=0.1,\ 0.2,\ 0.3,\ldots,\overline{h}$. At $\overline{h}=1.5$ led to acceptance of the constraints at almost every observed point.

Decomposition

Suppose we have a distance function

$$
D(x,y;\beta)=u+\omega+v, \tag{15}
$$

a la OP and LP, where β is parameters, u is inefficiency and w is productivity all in log terms, v is the error term. We have:

$$
\sum_{k=1}^{K}\frac{\partial D}{\partial x_k}\frac{\partial x_k}{\partial t}+\sum_{m=1}^{M}\frac{\partial D}{\partial y_m}\frac{\partial y_m}{\partial t}+\sum_{j=1}^{p}\frac{\partial D}{\partial \beta_j}\frac{\partial \beta_j}{\partial t}=\frac{\partial u}{\partial t}+\frac{\partial \omega}{\partial t}+tfpg \tag{16}
$$

where $tfpg$ is a "modified Solow residual" or TFP growth. Also $\frac{\partial D}{\partial x_i} = e_{x_i}$ is an input elasticity, $\frac{\partial D}{\partial y_j} = e_{y_j}$ is an output elasticity, $\frac{\partial x_i}{\partial t} = \dot{x}_i$, $\frac{\partial y_i}{\partial t} = \dot{y}_i$, $\frac{\partial u}{\partial t} = \dot{u}$ and $\frac{\partial \omega}{\partial t} = \dot{\omega}$ are relative rates of change as we already have logs. So we end up with

$$tfpg = \sum_{k=1}^{K} e_{x_k} \dot{x}_k + \sum_{m=1}^{M} e_{y_m} \dot{y}_m + \sum_{j=1}^{p} e_{\beta_j} \dot{\beta}_j, \tag{17}$$

where, as usually, we omit efficiency change. However, we do not omit the last term in (17) which is:

$$\sum_{k=1}^{K} \frac{\partial D}{\partial \beta_k} \frac{\partial \beta_k}{\partial t} = \sum_{k=1}^{K} \frac{\partial D}{\partial \beta_k} \frac{\partial \beta_k}{\partial t} \left(\frac{1}{\beta_k} \right) \beta_k = \sum_{k=1}^{K} e_{\beta_k} \dot{\beta}_k. \tag{18}$$

Growth can be decomposed to the following components in (17). The first component corresponds to change in the inputs. The second component corresponds to change in the outputs. The third component corresponds to a change in the frontier due to parameter changes.

To summarize, first, we have an input change component:

$$IC_{it} = \sum_{k=1}^{K} e_{x_k} \dot{x}_k \tag{19}$$

where β_{it}^{x} contains the appropriate elements of β_{it} from (11).

Second, an output change component:

$$OC_{it} = \sum_{m=1}^{M} e_{y_m} \dot{y}_m \tag{20}$$

where β_{it}^{y} contains the appropriate elements of β_{it} from (11).

Third, a frontier change component:

$$FC_{it} = \sum_{j=1}^{p} e_{\beta_j} \dot{\beta}_j. \tag{21}$$

Of course, apart from these, we have an efficiency change component:

$$EC_{it} = \frac{u_{i,t+1} - u_{it}}{u_{it}} \simeq \Delta \log u_{i,t+1}. \tag{22}$$

Data

To estimate the frontier model in (2) we need first to define the inputs (x_{it}), outputs (y_{it}), as well as the vector of environmental variables (z_{it}). Following the majority of studies in the literature (Barros et al. 2011; Cracolici et al. 2007), we select the following inputs: the number employees working in the tourism industry, capital investments made on the tourism industry, and number of rooms in accommodation properties[7], and outputs: the number of international tourism arrivals, receipts per capita from domestic tourism, receipts from international tourism, and average length of stay of international tourists. For the environmental variables z_{it}, we follow closely Assaf and Tsionas (2015) and define twenty one variables that we think can affect the production process in the tourism industry. These include variables reflecting the "infrastructure", "human resource and natural and environmental quality" of a tourism destination. The study by Assaf and Tsionas (2015) has shown that all the variables play "a critical role in attracting tourism outputs and hence ignoring them represent an important shortcoming that might bias the benchmarking outcomes"[8]. We provide in Table 2 a more detailed breakdown of these quality variables, as well as descriptive statistics of all variables included in the model. We used several sources to collect our data including the United Nations World Tourism Organization, Euromonitor database, tourism satellite accounts of some countries, as well as Eurostat database. Most of the quality variables were collected from the World Economic Forum, Executive Opinion Survey. The final sample included 101 tourism destinations over 4 years of data (2008-2012).

---

[7] Tourism accommodation represents any facility that regularly (or occasionally) provides overnight accommodation for tourists.

[8] We thank the authors for making the data available.

Table 1. Review of Productivity Studies in the Tourism Literature

Study	Methodology	Sample	Heterogeneity	Dynamic Structure
Barros (2005)	Malmquist DEA Index	42 Portuguese Hotels	No	No
Cracolici et al. (2001)	Malmquist DEA Index	103 Italian Regions	No	No
Peypoch and Solonandrasana (2008)	Luenberger DEA productivity index	10 French Hotels	No	No
Barros et al. (2008)	Luenberger DEA productivity index	15 Portuguese Hotels	No	No
Assaf and Dwyer (2013)	Metafrontier DEA approach	97 International Tourism Destinations	Yes	No
Goncalves (2013)	Luenberger DEA productivity index	64 French Ski Resorts	No	No
Barros and Alves (2007)	Malmquist DEA Index	42 Portuguese Hotels	No	No
Peypoch and Sbai (2011)	Luenberger DEA productivity index	15 Moroccan Hotels	No	No
Chen and Soo (2007)	Stochastic Cost Frontier	47 Taiwanese Hotels	No	No
Sun et al. (2015)	Malmquist DEA Index	31 Chinese Provinces	No	No

Barros, C. P., & Alves, F. P. (2004). Productivity in the tourism industry. *International Advances in Economic Research*, *10*(3), 215-225.

Chen, C. F., & Soo, K. T. (2007). Cost structure and productivity growth of the Taiwanese international tourist hotels. *Tourism Management*, *28*(6), 1400-1407.

Sun, J., Zhang, J., Zhang, J., Ma, J., & Zhang, Y. (2015). Total Factor Productivity Assessment of Tourism Industry: Evidence from China. *Asia Pacific Journal of Tourism Research*, *20*(3), 280-294

Table 2. Descriptive Statistics of All Model Variables

Variable	Mean	Std. Dev.	Min	Max
Inputs				
Number of rooms in accommodation properties	194.77	540.61	1.00	4932.80
Capital spending on tourism	6.17	18.35	0.02	164.47
Number of employees working in the tourism industry	2113.86	7391.92	13.30	63779.20
Outputs				
Number of international tourism arrivals	8563.77	13588.25	21.40	83168.50
Receipts per capita from domestic tourism	17437.45	69882.41	0.90	681780.00
Receipts from international tourism	8701.64	17235.30	32.00	160289.00
Average length of stay of international tourists	7.20	5.65	1.10	35.10
Infrastructure Quality Indicators **(1-7 scale)**				
Quality of road infrastructure	3.51	1.07	1.63	6.35
Quality of air infrastructure	4.09	1.20	2.05	6.74
Quality of ICT infrastructure	3.50	1.28	1.28	6.04
Human Resource Quality Indicators **(1-7 scale)**				
Quality of education	3.88	0.96	1.97	6.24
The existence of specialized research and training	4.27	0.86	2.43	6.47
Quality of labour training	4.09	0.76	2.22	5.95
Availability of qualified labour	5.23	0.51	2.69	6.39
The flexibility of recruiting foreign employees	4.45	0.73	2.21	6.18
Attitude of population toward international tourists	6.26	0.42	4.09	6.90
Natural & Environmental Quality Indicators				
No. of world heritage natural sites	4.21	1.03	2.18	6.63
Quality of the natural environment	3.97	1.05	2.25	6.38
Total known species	4.61	0.84	2.06	6.38
No. of world heritage cultural sites	2.18	0.95	1.00	7.00
Access to improved sanitation.	1.60	0.99	1.00	7.00
Access to improved drinking water	4.51	0.94	2.08	6.75
Stringency of environmental regulation	2.37	1.22	1.00	7.00
Enforcement of environmental regulation	1.65	0.90	1.00	7.00
Sustainability of T&T industry development	5.78	1.56	1.00	7.00
Threatened species	6.10	1.33	1.00	7.00
Environmental treaty ratification	4.93	1.22	1.00	7.00
Carbon dioxide emission	5.36	1.25	1.00	7.00

TECHNICAL APPENDIX

Particle filtering

The particle filter methodology can be applied to state space models of the general form:

$$y_T \sim p(y_t \mid x_t),\ s_t \sim p(s_t \mid s_{t-1}), \tag{A.1}$$

where s_t is a state variable. For general introductions see Gordon (1997), Gordon et al. (1993), Doucet et al (2001) and Ristic et al. (2004).

Given the data Y_t the posterior distribution $p(s_t \mid Y_t)$ can be approximated by a set of (auxiliary) particles $\left\{s_t^{(i)}, i=1,...,.N\right\}$ with probability weights $\left\{w_t^{(i)}, i=1,...,N\right\}$ where $\sum_{i=1}^{N} w_t^{(i)} = 1$. The predictive density can be approximated by:

$$p(s_{t+1} \mid Y_t) = \int p(s_{t+1} \mid s_t) p(s_t \mid Y_t) ds_t \simeq \sum_{i=1}^{N} p(s_{t+1} \mid s_t^{(i)}) w_t^{(i)}, \tag{A.2}$$

and the final approximation for the filtering density is:

$$p(s_{t+1} \mid Y_t) \propto p(y_{t+1} \mid s_{t+1}) p(s_{t+1} \mid Y_t) \simeq p(y_{t+1} \mid s_{t+1}) \sum_{i=1}^{N} p(s_{t+1} \mid s_t^{(i)}) w_t^{(i)}. \tag{A.3}$$

The basic mechanism of particle filtering rests on propagating $\left\{s_t^{(i)}, w_t^{(i)}, i=1,..,N\right\}$ to the next step, viz. $\left\{s_{t+1}^{(i)}, w_{t+1}^{(i)}, i=1,..,N\right\}$ but this often suffers from the weight degeneracy problem. If parameters $\theta \in \Theta \in \Re^k$ are available, as is often the case, we follow Liu and West (2001) parameter learning takes place via a mixture of multivariate normals:

$$p(\theta \mid Y_t) \simeq \sum_{i=1}^{N} w_t^{(i)} N(\theta \mid a\theta_t^{(i)} + (1-a)\bar{\theta}_t, b^2 V_t), \tag{A.4}$$

where $\bar{\theta}_t = \sum_{i=1}^{N} w_t^{(i)} \theta_t^{(i)}$, and $V_t = \sum_{i=1}^{N} w_t^{(i)} (\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$. The constants a and b are related to shrinkage and are determined via a discount factor $\delta \in (0,1)$ as $a = (1-b^2)^{1/2}$ and $b^2 = 1 - [(3\delta - 1)/2\delta]^2$. See also Casarin and Marin (2007).

Andrieu and Roberts (2009), Flury and Shephard (2011) and Pitt et al. (2012) provide the Particle Metropolis-Hastimgs (PMCMC) technique which uses an unbiased estimator of the likelihood function $\hat{p}_N(Y \mid \theta)$ as $p(Y \mid \theta)$ is often not available in closed form.

Given the current state of the parameter $\theta^{(j)}$ and the current estimate of the likelihood, say $L^j = \hat{p}_N(Y \mid \theta^{(j)})$, a candidate θ^c is drawn from $q(\theta^c \mid \theta^{(j)})$ yielding $L^c = \hat{p}_N(Y \mid \theta^c)$. Then, we set $\theta^{(j+1)} = \theta^c$ with the Metropolis - Hastings probability:

$$A = \min\left\{1, \frac{p(\theta^c)L^c}{p(\theta^{(j)})L^j}\frac{q(\theta^{(j)} \mid \theta^c)}{q(\theta^c \mid \theta^{(j)})}\right\}, \tag{A.5}$$

otherwise we repeat the current draws: $\left\{\theta^{(j+1)}, L^{j+1}\right\} = \left\{\theta^{(j)}, L^j\right\}$.

Hall, Pitt and Kohn (2014) propose an auxiliary particle filter which rests upon the idea that adaptive particle filtering (Pitt et al., 2012) used within PMCMC requires far fewer particles that the standard particle filtering algorithm to approximate $p(Y \mid \theta)$. From Pitt and Shephard (1999) we know that auxiliary particle filtering can be implemented easily once we can evaluate the state transition density $p(s_t \mid s_{t-1})$. When this is not possible, Hall, Pitt and Kohn (2014) present a new approach when, for instance, $s_t = g(s_{t-1}, u_t)$ for a certain disturbance. In this case we have:

$$p(y_t \mid s_{t-1}) = \int p(y_t \mid s_t) p(s_t \mid s_{t-1}) ds_t, \tag{A.6}$$

$$p(u_t \mid s_{t-1}; y_t) = p(y_t \mid s_{t-1}, u_t) p(u_t \mid s_{t-1}) / p(y_t \mid s_{t-1}). \tag{A.7}$$

If one can evaluate $p(y_t \mid s_{t-1})$ and simulate from $p(u_t \mid s_{t-1}; y_t)$ the filter would be fully adaptable (Pitt and Shephard, 1999). One can use a Gaussian approximation for the first-stage proposal $g(y_t \mid s_{t-1})$ by matching the first two moments of $p(y_t \mid s_{t-1})$. So in some way we find that the approximating density $p(y_t \mid s_{t-1}) = N\left(E(y_t \mid s_{t-1}), V(y_t \mid s_{t-1})\right)$. In the second stage, we know that $p(u_t \mid y_t, s_{t-1}) \propto p(y_t \mid s_{t-1}, u_t) p(u_t)$. For $p(u_t \mid y_t, s_{t-1})$ we know it is multimodal so suppose it has M modes are $\hat{u}_t^m$, for $m = 1, \ldots, M$. For each mode we can use a Laplace approximation. Let $l(u_t) = log\left[p(y_t \mid s_{t-1}, u_t) p(u_t)\right]$. From the Laplace approximation we obtain:

$$l(u_t) \simeq l(\hat{u}_t^m) + \tfrac{1}{2}(u_t - \hat{u}_t^m)' \nabla^2 l(\hat{u}_t^m)(u_t - \hat{u}_t^m). \tag{A.8}$$

Then we can construct a mixture approximation:

$$g(u_t \mid x_t, s_{t-1}) = \sum_{m=1}^{M} \lambda_m (2\pi)^{-d/2} \mid \Sigma_m \mid^{-1/2} \exp\left\{\tfrac{1}{2}(u_t - \hat{u}_t^m)' \Sigma_m^{-1}(u_t - \hat{u}_t^m\right\}, \tag{A.9}$$

where $\Sigma_m = -\left[\nabla^2 l(\hat{u}_t^m)\right]^{-1}$ and $\lambda_m \propto \exp\left\{l(u_t^m)\right\}$ with $\sum_{m=1}^{M} = 1$. This is done for each particle s_t^i. This is known as the Auxiliary Disturbance Particle Filter (ADPF).

An alternative is the independent particle filter (IPF) of Lin et al. (2005). The IPF forms a proposal for s_t directly from the measurement density $p(y_t \mid s_t)$ although Hall, Pitt and Kohn (2014) are quite right in pointing out that the state equation can be very informative.

In the standard particle filter of Gordon et al. (1993) particles are simulated through the state density $p(s_t^i \mid s_{t-1}^i)$ and they are re-sampled with weights determined by the measurement density evaluated at the resulting particle, viz. $p(y_t \mid s_t^i)$.

The ADPF is simple to construct and rests upon the following steps:

For $t = 0,..,T-1$ given samples $s_t^k \sim p(s_t \mid Y_{1:t})$ with mass π_t^k for $k = 1,...,N$.

1) For $k = 1,..,N$ compute $\omega_{t|t+1}^k = g(y_{t+1} \mid s_t^k)\pi_t^k$, $\pi_{t|t+1}^k = \omega_{t|t+1}^k / \sum_{i=1}^N \omega_{t|t+1}^i$.

2) For $k = 1,..,N$ draw $\tilde{s}_t^k \sim \sum_{i=1}^N \pi_{t|t+1}^i \delta_{s_t}^i (ds_t)$.

3) For $k = 1,..,N$ draw $u_{t+1}^k \sim g(u_{t+1} \mid \tilde{s}_t^k, y_{t+1})$ and set $s_{t+1}^k = h(s_t^k; u_{t+1}^k)$.

4) For $k = 1,..,N$ compute

$$\omega_{t+1}^k = \frac{p(y_{t+1} \mid s_{t+1}^k) p(u_{t+1}^k)}{g(y_{t+1} \mid s_t^k) g(u_{t+1}^k \mid \tilde{s}_t^k, y_{t+1})}, \pi_{t+1}^k = \frac{\omega_{t+1}^k}{\sum_{i=1}^N \omega_{t+1}^i}. \quad \text{(A.10)}$$

It should be mentioned that the estimate of likelihood from ADPF is:

$$p(Y_{1:T}) = \prod_{t=1}^T \left(\sum_{i=1}^N \omega_{t-1|t}^i \right) \left(N^{-1} \sum_{i=1}^N \omega_t^i \right). \quad \text{(A.11)}$$

Particle Metropolis adjusted Langevin filters

Nemeth, Sherlock and Fearnhead (2014) provide a particle version of a Metropolis adjusted Langevin algorithm (MALA). In Sequential Monte Carlo we are interested in approximating $p(s_t \mid Y_{1:t}, \theta)$. Given that:

$$p(s_t \mid Y_{1:t}, \theta) \propto g(y_t \mid x_t, \theta) \int f(s_t \mid s_{t-1}, \theta) p(s_{t-1} \mid y_{1:t-1}, \theta) ds_{t-1}, \quad \text{(A.12)}$$

where $p(s_{t-1} \mid y_{1:t-1}, \theta)$ is the posterior as of time $t-1$. If at time $t-1$ we have a set set of particles $\{s_{t-1}^i, i = 1,..,N\}$ and weights $\{w_{t-1}^i, i = 1,...N\}$ which form a discrete approximation for $p(s_{t-1} \mid y_{1:t-1}, \theta)$ then we have the approximation:

$$\hat{p}(s_{t-1} \mid y_{1:t-1}, \theta) \propto \sum_{i=1}^N w_{t-1}^i f(s_t \mid s_{t-1}^i, \theta). \quad \text{(A.13)}$$

See Andrieu et al. (2010) and Cappe at al. (2005) for reviews. From (A.13) Fernhead (2007) makes the important observation that the joint probability of sampling particle s_{t-1}^{i} and state s_t is:

$$\omega_t = \frac{w_{t-1}^{i} g(y_t \mid s_t, \theta) f(s \mid s_{t-1}^{i}, \theta)}{\xi_t^{i} q(s_t \mid s_{t-1}^{i}, y_t, \theta)}, \tag{A.14}$$

where $q(s_t \mid s_{t-1}^{i}, y_t, \theta)$ is a density function amenable to simulation and

$$\xi_t^{i} q(s_t \mid s_{t-1}^{i}, y_t, \theta) \simeq c g(y_t \mid s_t, \theta) f(s_t \mid s_{t-1}^{i}, \theta), \tag{A.15}$$

and c is the normalizing constant in (A.12).

In the MALA algorithm of Roberts and Rosenthal (1998)[9] we form a proposal:

$$\theta^{c} = \theta^{(s)} + \lambda z + \tfrac{\lambda^2}{2} \nabla \log p(\theta^{(s)} \mid Y_{1:T}), \tag{A.16}$$

where $z \sim N(0, I)$ which should result in larger jumps and better mixing properties, plus lower autocorrelations for a certain scale parameter λ. Acceptance probabilities are:

$$a(\theta^{c} \mid \theta^{(s)}) = \min\left\{1, \frac{p(Y_{1:T} \mid \theta^{c}) q(\theta^{(s)} \mid \theta^{c})}{p(Y_{1:T} \mid \theta^{(s)}) q(\theta^{c} \mid \theta^{(s)})}\right\}. \tag{A.17}$$

Using particle filtering it is possible to create an approximation of the score vector using Fisher's identity:

$$\nabla \log p(Y_{1:T} \mid \theta) = E\left[\nabla \log p(s_{1:T}, Y_{1:T} \mid \theta) \mid Y_{1:T}, \theta\right], \tag{A.18}$$

which corresponds to the expectation of:

$$\nabla \log p(s_{1:T}, Y_{1:T} \mid \theta) = \nabla \log p(\mid s_{1:T-1}, Y_{1:T-1} \mid \theta) + \nabla \log g(y_T \mid s_T, \theta) + \nabla \log f(s_T \mid s \mid_{T-1}, \theta),$$

over the path $s_{1:T}$. The particle approximation to the score vector results from replacing $p(s_{1:T} \mid Y_{1:T}, \theta)$ with a particle approximation $\hat{p}(s_{1:T} \mid Y_{1:T}, \theta)$. With particle i at time t-1 we can associate a value $\alpha_{t-1}^{i} = \nabla \log p(s_{1:t-1}^{i}, Y_{1:t-1} \mid \theta)$ which can be updated recursively. As we sample κ_i in the APF (the index of particle at time $t-1$ that is propagated to produce the i th particle at time t) we have the update:

$$\alpha_t^{i} = a_{t-1}^{\kappa_i} + \nabla \log g(y_t \mid s_t^{i}, \theta) + \nabla \log f(s_t^{i} \mid s_{t-1}^{i}, \theta). \tag{A.19}$$

---

[9] The benefit of MALA over Random-Walk-Metropolis arises when the number of parameters n is large. This happens because the scaling parameter λ is $O(n^{-1/2})$ for Random-Walk-Metropolis but it is $O(n^{-1/6})$ for MALA, see Roberts et al. (1997) and Roberts and Rosenthal (1998)

To avoid problems with increasing variance of the score estimate $\nabla \log p(Y_{1:t} \mid \theta)$ we can use the approximation:

$$\alpha_{t-1}^{i} \sim N(m_{t-1}^{i}, V_{t-1}). \tag{A.20}$$

The mean is obtained by shrinking α_{t-1}^{i} towards the mean of α_{t-1} as follows:

$$m_{t-1}^{i} = \delta \alpha_{t-1}^{i} + (1-\delta)\sum_{i=1}^{N} w_{t-1}^{i} \alpha_{t-1}^{i}, \tag{A.21}$$

where $\delta \in (0,1)$ is a shrinkage parameter. Using Rao-Blackwellization one can avoid sampling α_{t}^{i} and instead use the following recursion for the means:

$$m_{t}^{i} = \delta m_{t-1}^{\kappa_i} + (1-\delta)\sum_{i=1}^{N} w_{t-1}^{i} m_{t-1}^{i} + \nabla \log g(y_t \mid s_t^{i}, \theta) + \nabla \log f(s_t^{i} \mid s_{t-1}^{\kappa_i}, \theta), \tag{A.22}$$

which yields the final score estimate:

$$\nabla \log \hat{p}(Y_{1:t} \mid \theta) = \sum_{i=1}^{N} w_t^{i} m_t^{i}. \tag{A.23}$$

As a rule of thumb Nemeth, Sherlock and Fearnhead (2014) suggest taking $\delta = 0.95$. Furthermore, they show the important result that the algorithm should be tuned to the asymptotically optimal acceptance rate of 15.47% and the number of particles must be selected so that the variance of the estimated log-posterior is about 3. Additionally, if measures are not taken to control the error in the variance of the score vector, there is no gain over a simple random walk proposal.

Of course, the marginal likelihood is:

$$p(Y_{1:T} \mid \theta) = p(y_1 \mid \theta)\prod_{t=2}^{T} p(y_t \mid Y_{1:t-1}, \theta), \tag{A.24}$$

where

$$p(y_t \mid Y_{1:t-1}, \theta) = \int g(y_t \mid s_t) \int f(s_t \mid s_{t-1}, \theta)\, p(s_{t-1} \mid Y_{1:T-1}, \theta)\, ds_{t-1} ds_t, \tag{A.25}$$

provides, in explicit form, the predictive likelihood.

References

Andrieu, C., and G.O. Roberts. (2009) The pseudo-marginal approach for efficient computation. Ann. Statist., 37, 697–725.

Andrieu, C., A. Doucet, and R. Holenstein (2010), "Particle Markov chain Monte Carlo methods," Journal of the Royal Statistical Society: Series B, 72, 269–342.

Cappι, O., E. Moulines, E., and T. Rydιn (2005). Inference in Hidden Markov Models. Springer, Berlin.

Casarin, R., J.-M. Marin (2007). Online data processing: Comparison of Bayesian regularized particle filters. University of Brescia, Department of Economics. Working Paper n. 0704.

Doucet, A., N. de Freitas, and N. Gordon (2001). Sequential Monte Carlo Methods in Practice. New York: Springer.

Fearnhead, P. (2007). Computational methods for complex stochastic systems: a review of some alternatives to MCMC. Statistics and Computing, 18(2):151-171.

Flury, T., and N. Shephard, (2011) Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. Econometric Theory 27, 933-956.

Gordon, N.J. (1997). A hybrid bootstrap filter for target tracking in clutter. IEEE Transactions on Aerospace and Electronic Systems 33: 353– 358.

Gordon, N.J., D.J. Salmond, and A.F.M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE-Proceedings-F 140: 107–113.

Hall, J., M.K. Pitt, and R. Kohn (2014). Bayesian inference for nonlinear structural time series models. Journal of Econometrics 179 (2), 99–111.

Liu, J., M. West (2001). Combined parameter and state estimation in simulation-based filtering. In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), Sequential Monte Carlo Methods in Practice. Springer-Verlag.

Nemeth, C., P. Fearnhead (2014). Particle Metropolis adjusted Langevin algorithms for state-space models. Pre-print arXiv:1402.0694v1.

Pitt, M.K., R.S. Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. J. Econom. 171(2), 134–151.

B. Ristic, S. Arulampalam, and N. Gordon (2004). Beyond Kalman Filters: Particle Filters for Applications. Norwood, MA: Artech House.