University of Dundee

Semi-Parametric Analysis of Efficiency and Productivity using Gaussian Processes

Emvalomatis, Gregory

Published in:
Econometrics Journal

DOI:
[10.1093/ectj/utz013](https://doi.org/10.1093/ectj/utz013)

Publication date:
2020

Document Version
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):
Emvalomatis, G. (2020). Semi-Parametric Analysis of Efficiency and Productivity using Gaussian Processes. *Econometrics Journal*, 23(1), 48-67. <https://doi.org/10.1093/ectj/utz013>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Semi-Parametric Analysis of Efficiency and Productivity using Gaussian Processes

GRIGORIOS EMVALOMATIS[†]

[†]*Economic Studies, University of Dundee, Perth Road, Dundee, DD1 4HN, UK.*

E-mail: g.emvalomatis@dundee.ac.uk

Received: January 2019

Summary This paper proposes a fully Bayesian semi-parametric method for efficiency and productivity analysis based on Gaussian processes. The proposed technique frees the researcher from having to specify a functional form for the production frontier and is shown in simulated data to perform as well as flexible parametric models when correct distributional assumptions are imposed on the inefficiency component of the error term and slightly better when incorrect assumptions are made. The technique is applied to a panel dataset of US electric utilities, where Total Factor Productivity growth is estimated and decomposed using both parametric and semi-parametric techniques.

Keywords: *Gaussian-process regression, stochastic frontier, TFP decomposition.*

1. INTRODUCTION

Soon after the first formal definition of technical efficiency (Koopmans, 1951) and the development of a quantitative measure of it (Debreu, 1951), the first empirical application of the concept was presented by Farrell (1957). Farrell's approach of expressing each Decision Making Unit's (DMU) observed inputs and outputs as a linear combination of the inputs and outputs of other DMUs heavily influenced the development of Data Envelopment Analysis (DEA) (Charnes et al., 1978, and Banker et al., 1984) and Free Disposal Hull (FDH) methods (Deprins et al., 1984). In his discussion of Farrell's paper, Winsten (1957) noted that, under certain assumptions, the production frontier can be estimated by shifting upwards the predicted line from a standard linear regression model, thus laying the foundations for parametric efficiency measurement, which eventually led to the development of the stochastic frontier approach (Aigner et al., 1977, and Meeusen and van den Broeck, 1977). For the two decades following the introduction of DEA and stochastic frontier models, the two approaches developed in parallel, with very little interaction.

The major advantage of non-parametric techniques, such as DEA or FDH, is that they have minimal specification requirements: deciding what are the inputs in a production process and what the outputs is enough to proceed with application. By constructing a piecewise linear frontier and measuring DMU efficiency against it, FDH and DEA rely on the 'mild' monotonicity and, most frequently, curvature assumptions about the frontier of the production possibilities set. On the contrary, with its firm foundations on statistical concepts, stochastic frontier analysis estimates the parameters of a pre-specified frontier, while, in the simplest cases, an assumption on the distribution of inefficiency is also necessary. However, rarely exists a convincing argument to justify the choice of one distribution over another and this has led to a proliferation of alternative specifications. On the other hand, these statistical foundations enable stochastic frontier models to

handle noise in the data naturally. The original non-parametric techniques were heavily criticized for their complete disregard of noise. Furthermore, if the objective of a study is to examine the effects of environmental factors (conditions which cannot be affected by a DMU) on inefficiency, the need to impose a distributional assumption on inefficiency turns from a nuisance to a useful tool. Due to the absence of a concrete specification of the data-generating process, simple non-parametric techniques cannot provide information on environmental influences in a consistent way. Finally, when examining the evolution of Total Factor Productivity (TFP) and its components within a sector or economy, the stochastic frontier approach originally had an advantage because, due to its ability to estimate output elasticities and other marginal effects, it could build directly on existing concepts, such as the Solow residual.

Many attempts have been made to alleviate the shortcomings of non-parametric methods. The first issue to be tackled was handling noise in the data. Various approaches have been developed, with the most prominent ones being based on bootstrapping (Simar and Wilson, 1998) and the notion of partial frontiers (Cazals et al., 2002). Simar and Wilson (2007) provide an extension of the bootstrap methods that can account for environmental influences in DEA models and Bădin et al. (2012) present methods with the same objective, but based on the approach introduced by Cazals et al. (2002). Regarding TFP measurement and decomposition, Färe et al. (1994) show how DEA can be used to construct and decompose a Malmquist productivity index and Simar and Wilson (1999) extend the method to allow for noise in the data.

Considerably less effort has been put from the parametric branch of the efficiency and productivity analysis literature to bridge the gap between parametric and non-parametric methods. Two distinct issues had to be addressed: the specification of the production frontier and the requirement for a distributional assumption on inefficiency. Regarding the first issue, Fan et al. (1996) propose an approach based on kernel-density estimation of the frontier, while maintaining the need for a distributional assumption. Kumbhakar et al. (2007) use a local maximum-likelihood approach to avoid specifying the frontier and impose a relatively mild assumption on the distribution of inefficiency. Regarding distributional assumptions, Park and Simar (1994) treat the distribution of inefficiency as unknown, but their approach can only be applied when panel data are available. Griffin and Steel (2004) use Dirichlet processes to flexibly model inefficiency in a Bayesian context, while also maintaining the parametric specification of the frontier.

This article proposes a new method of relaxing the functional assumptions imposed on stochastic frontier models by using an extension to Gaussian-process regression. In terms of its objective, the paper is similar in spirit to Fan et al. (1996) and Kumbhakar et al. (2007), but uses a fully-Bayesian approach. This is not done to stir controversy on whether Bayesian or frequentist methods should be used in general or simply to provide an alternative to existing frequentist techniques. Rather, the practical implication of having a Bayesian counterpart to the semi-parametric stochastic frontier is that all results available in the Bayesian stochastic frontier literature can be used directly. Due to its reliance on data augmentation (Tanner and Wong, 1987), the Bayesian approach allows imposing elaborate structures on inefficiency, such as autoregressive or persistent/transient inefficiency, or even combining the method presented here with the one proposed by Griffin and Steel (2004) to avoid the specification of a distribution for inefficiency altogether. Additionally, when panel data are available, accounting for unobserved heterogeneity is straightforward in a Bayesian semi-parametric frontier. Finally, due to the derivatives of the frontier estimated by Gaussian-process regression following Gaus-

sian processes themselves, calculation and decomposition of TFP growth is simple, while accounting for uncertainty effortlessly.

Gaussian-process regression has its origins in geostatistics, where it was introduced as a spatial interpolation method. Based on a Gaussian process with a, possibly multi-dimensional, continuous index set, Gaussian-process regression fits a curve through the observed data without imposing strong restrictions on its shape and can be used to predict the values of the output(s) of a process for arbitrary values of the input(s). As such, the technique can be viewed as a generalization of Bayesian linear regression with infinitely many basis functions, while posterior consistency properties (Choi and Schervish, 2007 and van der Vaart and van Zanten, 2009) make it appealing in settings with many observations. Due to its flexibility, the technique has found wide applications, primarily in the field of machine learning. Although Gaussian processes themselves have long been an integral part of time-series econometrics, applications of Gaussian-process regression have only recently appeared in the econometrics literature. For example, Kasy (2018) uses Gaussian-process regression to model the relationship between healthcare expenditures and coinsurance levels and Ludkovski et al. (2018) to model mortality rates as a function of the individual's age and calendar year. Ruseckaite et al. (2018) use multivariate Gaussian-process regression in a mixture-amount model to capture the dependence of the mixture parameters on the amounts themselves. In a time-series context, Nesreen et al. (2010) compare multiple machine-learning techniques in terms of their ability to forecast the values of 1045 time series and find Gaussian-process regression to be among the two best-performing techniques according to various criteria, while Yeo et al. (2015) conclude that Gaussian-process regression outperforms ARMA and GARCH models in forecasting wine prices.

The following section formally defines the concepts of technical efficiency and Gaussian process and illustrates how the two can be combined in a semi-parametric model designed to measure DMU-level efficiency. The estimation approach is presented in the subsequent section, where the curve that a Gaussian-process regression would fit through the data is now altered to envelop the data, save for the effect of statistical noise. Section 4 discusses possible limitations of the technique along with remedies, as well as extensions to panel datasets and the measurement and decomposition of productivity growth. An illustration is provided in section 5 using simulated data and in the very simple context of a single-input production function. An empirical application using the US electric utilities dataset constructed by Rungsuriyawiboon and Stefanou (2007) is presented in section 6, where, apart from estimating efficiency scores and output elasticities with respect to inputs, TFP growth is also estimated and the results are compared to a parametric stochastic frontier model. Concluding comments are given in the final section.

2. MODELLING ASSUMPTIONS

Consider a production process in which a vector of inputs, $\mathbf{x} \in \mathbb{R}_+^K$, is transformed into a single output, $y \in \mathbb{R}_+$, according to the production function $g(\mathbf{x})$. The technical efficiency of a DMU that employs inputs $\tilde{\mathbf{x}}$ to produce output $\tilde{y}$ is defined as the ratio of observed output to maximum attainable output, given $\tilde{\mathbf{x}}$:

$$\widehat{\text{TE}} \equiv \frac{\tilde{y}}{g(\tilde{\mathbf{x}})} \Rightarrow \log \tilde{y} = \log g(\tilde{\mathbf{x}}) - \tilde{u} \quad (2.1)$$

where $\tilde{u} \equiv -\log \widetilde{\text{TE}}$. Because technical efficiency assumes values in the unit interval, by construction, $\tilde{u}$ can take only non-negative values.

Let $f(\mathbf{x}) \equiv \log g(\mathbf{x})$ and assume that it follows a Gaussian process, which, following Rasmussen and Williams (2006, p.13), is defined as “a collection of random variables, any finite number of which have a joint Gaussian distribution”. Let $m(\mathbf{x})$ denote the mean function of the Gaussian process before looking at any relevant data (in the prior) and let its covariance (kernel) function take the typical squared-exponential form: $k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp \{-\|\mathbf{x} - \mathbf{x}'\|^2 / (2\sigma_\ell^2)\}$.¹ In the context of a production function, the random variables behind the Gaussian process are the values of $f(\cdot)$ obtained for different values of the input vector, $\mathbf{x}$. The specification of the kernel function implies that the covariance of the values of this function evaluated at two input vectors, $\mathbf{x}$ and $\mathbf{x}'$, is a decreasing function of the Euclidean distance between these vectors. In other words, the closer two input vectors are to each other in $\mathbb{R}_+^K$, the more likely they are to produce similar amounts of output. σ_ℓ^2 in the kernel function defines the length-scale of the process, with larger values of this parameter making large differences between the variables obtained from the process less likely, for any given distance of the input vectors. Finally, σ_f^2 is the variance of $f(\cdot)$ for any given vector of inputs. That is, uncertainty with respect to maximum attainable output is already embedded in the Gaussian process and technical efficiency, as defined in equation (2.1), becomes a random variable itself.

Suppose now that observations, $\{\mathbf{x}_i, y_i\}_{i=1}^N$, are available for N DMUs. Using the definitions given above, we assume the following data-generating process:

$$\begin{aligned} \log y_i &= f(\mathbf{x}_i) + v_i - u_i \\ f &\sim \text{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \\ v_i &\sim \text{N}(0, \sigma_v^2) \\ u_i &\sim \text{N}^+(0, \sigma_u^2) \end{aligned} \tag{2.2}$$

where a normally-distributed error term, v_i , is added to the right-hand side of the equation that determines the value of log-output and u_i is treated as a random draw from a half-normal distribution with pre-truncation variance parameter σ_u^2 .² The last assumption is used here only because of its simplicity and alternative distributional assumptions, such as the exponential, truncated-normal, Gamma or Weibull, or even more elaborate structures, where the distribution of u_i is a function of determinants of inefficiency, are certainly feasible. The specification of the data-generating process is very similar to a standard stochastic frontier model, with the only difference being that the frontier itself is expressed as a Gaussian process instead of a parametric function of the inputs.

The primary objective in an application of a standard Gaussian-process regression is to infer the shape of $f(\cdot)$ from the data and, based on this, make predictions on the values of output for input vectors, not necessarily in the observed (training) data. On the contrary, the main objective in an efficiency-measurement exercise is to estimate the levels of inefficiency for each DMU and, possibly, the output elasticities with respect to the inputs. In both cases the parameters that enter the data-generating process need to be estimated prior to or along with estimating efficiency scores or the derivatives of $f(\cdot)$.

¹Because $m(\mathbf{x})$ is a prior, the mean function is usually set to zero for any value of $\mathbf{x}$. We will keep carrying this mean function along as this comes at minimal notational cost, but we will restrict it to zero in the empirical application.

²Given that, by its nature, $f(\cdot)$ already incorporates statistical noise, the v_i term is not necessary. However, it is common practice in standard Gaussian-process regression to include such an error term, whose purpose is to capture statistical noise without affecting the smoothness of the frontier.

3. ESTIMATION AND INFERENCE

Given a dataset of N observations on inputs and outputs, define $\mathbf{X}$ as the $N \times K$ matrix obtained from stacking by row the transposed input vectors and $\mathbf{y}$ as the $N \times 1$ vector that contains the logarithm of the corresponding output values. Let $\mathbf{m}$ be the $N \times 1$ vector of the values of the Gaussian process' mean function, evaluated at each input vector and $\mathbf{K}$ be the symmetric $N \times N$ matrix of the corresponding kernel function, evaluated at each combination of input vectors. Note that $\mathbf{K}$ is a function of the parameters that appear in the process' kernel function and $\mathbf{m}$ would depend on the values of any parameters appearing in the specification of the process' mean function. Finally, let $\mathbf{f}$ denote the $N \times 1$ random vector consisting of the random variables that are defined by evaluating $f(\cdot)$ at the N input vectors and $\mathbf{u}$ be the $N \times 1$ random vector that contains the corresponding inefficiency terms. Neither $\mathbf{f}$ nor $\mathbf{u}$ are observed, but they can be treated as latent data during estimation. The complete-data likelihood can be written as:

$$\begin{aligned} p(\mathbf{y}, \mathbf{f}, \mathbf{u} | \mathbf{X}) &= p(\mathbf{y} | \mathbf{f}, \mathbf{u}) \times p(\mathbf{f} | \mathbf{X}) \times p(\mathbf{u}) \\ &= \frac{|\sigma_v^2 \mathbf{I}_N|^{-1/2}}{(2\pi)^{N/2}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{f} + \mathbf{u})^\top (\sigma_v^2 \mathbf{I}_N)^{-1} (\mathbf{y} - \mathbf{f} + \mathbf{u}) \right\} \\ &\times \frac{|\mathbf{K}|^{-1/2}}{(2\pi)^{N/2}} \exp \left\{ -\frac{1}{2} (\mathbf{f} - \mathbf{m})^\top \mathbf{K}^{-1} (\mathbf{f} - \mathbf{m}) \right\} \\ &\times \prod_{i=1}^N \left(\frac{2}{\pi \sigma_u^2} \right)^{1/2} \exp \left\{ -\frac{u_i^2}{2\sigma_u^2} \right\} \end{aligned} \quad (3.3)$$

where the first line after the second equality symbol is due to the normality and independence (across observations) assumption on v_i , the second line is due to $f(\cdot)$ following a Gaussian process and, therefore, $\mathbf{f}$ following a multivariate-normal distribution, and the last line is due to the u_i s being independent and each of them following a half-normal distribution. Implicitly, all densities in the first line of this expression are conditional on the values of relevant parameters.

Because $\mathbf{f}$ appears in the complete-data likelihood only inside multivariate-normal densities, it can be integrated-out analytically, leading to:

$$\begin{aligned} p(\mathbf{y}, \mathbf{u} | \mathbf{X}) &= p(\mathbf{y} | \mathbf{X}, \mathbf{u}) \times p(\mathbf{u}) \\ &= \frac{|\mathbf{K} + \sigma_v^2 \mathbf{I}_N|^{-1/2}}{(2\pi)^{N/2}} \exp \left\{ -\frac{(\mathbf{y} - \mathbf{m} + \mathbf{u})^\top (\mathbf{K} + \sigma_v^2 \mathbf{I}_N)^{-1} (\mathbf{y} - \mathbf{m} + \mathbf{u})}{2} \right\} \\ &\times \prod_{i=1}^N \left(\frac{2}{\pi \sigma_u^2} \right)^{1/2} \exp \left\{ -\frac{u_i^2}{2\sigma_u^2} \right\} \end{aligned} \quad (3.4)$$

In a second step, $\mathbf{u}$ can be integrated-out from the last expression, leading to an observed-data likelihood function, which in principle, can be maximized numerically with respect to the parameters (σ_f^2 , σ_ℓ^2 , σ_v^2 , σ_u^2 and any parameters that appear in the specification of process' mean function). The observed-data likelihood function, however, involves an N -dimensional truncated-normal distribution and the amount of computations involved in its approximation is prohibitive, even for moderate values of N .³ Instead we use a

³The Geweke-Hajivassiliou-Keane simulator (Geweke, 1991, Hajivassiliou et al., 1996, and Keane, 1994)

Bayesian approach in which the unobserved u_i s are integrated from the likelihood by simulation using data-augmentation (Tanner and Wong, 1987).

To keep the upcoming expressions simple, we re-parameterize the model in terms of precision parameters ($\tau_f \equiv 1/\sigma_f^2$, $\tau_\ell \equiv 1/\sigma_\ell^2$ and $\tau_v \equiv 1/\sigma_v^2$) and collect their logarithms in a vector $\boldsymbol{\theta}$, along with any parameters that appear in the specification of the Gaussian process' prior mean function. Independent Gamma priors are imposed on the precision parameters, which result in log-Gamma priors for their logarithms:

$$p(\log \tau_j) = \frac{b_j^{a_j}}{\Gamma(a_j)} \exp \{a_j \log \tau_j - b_j e^{\log \tau_j}\} \quad (3.5)$$

where j is a placeholder for f , ℓ and v and the a s and b s are, respectively, parameter-specific shape and rate hyper-parameters for the original Gamma priors. τ_v is a typical precision parameter for a statistical-noise term and vague or informative priors can be placed on it. On the contrary, the hyper-parameters for τ_f and τ_ℓ may have a larger impact on the results, depending on the range of the variables (y and $\mathbf{x}$). If τ_f is restricted in the prior to be very large relative to the values of y , then the fitted frontier will be pooled too much towards the prior mean function, $m(\mathbf{x})$. Similarly, if τ_ℓ is restricted to be very large relative to the values of $\mathbf{x}$, $k(\cdot, \cdot)$ will tend to zero fast as the distance between two data points increases in $\mathbb{R}^K$ and, thus, the fitted frontier will tend to disregard information from nearby points and will, as a result, follow the data very closely. The opposite will happen if τ_ℓ is restricted in the prior to be too small. On the other hand, if vague priors are used for both τ_f and τ_ℓ , small values for one parameter are likely to be compensated by large values for the other and numerical stability issues may be encountered.

Regarding the inefficiency component of the error term, a Gamma prior is imposed on $1/\sigma_u^2$ with shape parameter, a_u , equal to seven and rate parameter, b_u , equal to $1/2$. These values of the hyper-parameters for $1/\sigma_u^2$ result in a prior median efficiency of approximately 83%, but can be adjusted accordingly to reflect other prior median efficiencies in specific applications. If the Gaussian process' prior mean function contains any parameters, priors for these should also be specified at this stage and would typically take the form of normal distributions.

Implementing a Gibbs sampler with data augmentation requires sampling from the full or complete conditionals of the parameters, $\boldsymbol{\theta}$ and σ_u^2 , as well as of the latent data, $\mathbf{f}$ and $\mathbf{u}$. The full conditional of $\boldsymbol{\theta}$ can be obtained either conditionally on the values of $\mathbf{f}$ from the expression in (3.3) or marginally with respect to $\mathbf{f}$ from the expression in (3.4). Very few simplifications can be made in either case. Using the complete-data likelihood in (3.3) we obtain:

$$\begin{aligned} \pi(\boldsymbol{\theta}|\mathbf{f}, \cdot) &\propto |\sigma_v^2 \mathbf{I}_N|^{-1/2} |\mathbf{K}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{f} + \mathbf{u})^\top (\sigma_v^2 \mathbf{I}_N)^{-1} (\mathbf{y} - \mathbf{f} + \mathbf{u}) \right\} \\ &\times \exp \left\{ -\frac{1}{2} (\mathbf{f} - \mathbf{m})^\top \mathbf{K}^{-1} (\mathbf{f} - \mathbf{m}) \right\} \\ &\times \exp \{a_f \log \tau_f - b_f e^{\log \tau_f} + a_\ell \log \tau_\ell - b_\ell e^{\log \tau_\ell} + a_v \log \tau_v - b_v e^{\log \tau_v}\} \end{aligned} \quad (3.6)$$

could provide such an approximation, but, apart from being very demanding computationally, in applications with a large N , numerical stability issues would make the approach impractical.

while the marginal likelihood in (3.4) leads to:

$$\pi(\boldsymbol{\theta}|\cdot) \propto |\mathbf{K} + \sigma_v^2 \mathbf{I}_N|^{-1/2} \exp \left\{ -\frac{(\mathbf{y} - \mathbf{m} + \mathbf{u})^\top (\mathbf{K} + \sigma_v^2 \mathbf{I}_N)^{-1} (\mathbf{y} - \mathbf{m} + \mathbf{u})}{2} \right\} \\ \times \exp \left\{ a_f \log \tau_f - b_f e^{\log \tau_f} + a_\ell \log \tau_\ell - b_\ell e^{\log \tau_\ell} + a_v \log \tau_v - b_v e^{\log \tau_v} \right\} \quad (3.7)$$

Inversion of $\mathbf{K}$, as required in the former expression, can quickly become numerically unstable as the number of observations increases. On the contrary, the latter expression requires inversion of $\mathbf{K} + \sigma_v^2 \mathbf{I}_N$, with the second summand bounding the eigenvalues of the matrix to be inverted away from zero. For this reason we opt using the second expression for the full conditional of $\boldsymbol{\theta}$, which results in a partially-collapsed Gibbs sampler.

The full conditional of $\boldsymbol{\theta}$ does not belong to any known parametric family of distributions. Nevertheless, Metropolis-Hastings updates are feasible. We use a random-walk proposal for a move from the current value of $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$:

$$q(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \frac{|\mathcal{T}\mathbf{C}|^{-1/2}}{(2\pi)^{3/2}} \exp \left\{ -\frac{(\boldsymbol{\theta}^* - \boldsymbol{\theta})^\top (\mathcal{T}\mathbf{C})^{-1} (\boldsymbol{\theta}^* - \boldsymbol{\theta})}{2} \right\} \quad (3.8)$$

where $\mathcal{T}$ is a tuning parameter, adjusted such that approximately 30%-40% of the proposed moves are accepted and $\mathbf{C}$ is an estimate of the variance matrix of $\boldsymbol{\theta}$, obtained during the adaptation phase of the Gibbs sampler. The logarithm of the Metropolis-Hastings ratio for a move from $\boldsymbol{\theta}$ to $\boldsymbol{\theta}^*$ is:

$$\log \text{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \log \pi(\boldsymbol{\theta}^*|\cdot) - \log q(\boldsymbol{\theta}, \boldsymbol{\theta}^*) - \log \pi(\boldsymbol{\theta}|\cdot) + \log q(\boldsymbol{\theta}^*, \boldsymbol{\theta}) \\ = -\frac{1}{2} (\log |\mathbf{K}_{\boldsymbol{\theta}^*} + \sigma_v^{2*} \mathbf{I}_N| - \log |\mathbf{K}_{\boldsymbol{\theta}} + \sigma_v^2 \mathbf{I}_N|) \\ -\frac{1}{2} (\mathbf{y} - \mathbf{m} + \mathbf{u})^\top \left((\mathbf{K}_{\boldsymbol{\theta}^*} + \sigma_v^{2*} \mathbf{I}_N)^{-1} - (\mathbf{K}_{\boldsymbol{\theta}} + \sigma_v^2 \mathbf{I}_N)^{-1} \right) (\mathbf{y} - \mathbf{m} + \mathbf{u}) \\ + a_f (\log \tau_f^* - \log \tau_f) - b_f (e^{\log \tau_f^*} - e^{\log \tau_f}) \\ + a_\ell (\log \tau_\ell^* - \log \tau_\ell) - b_\ell (e^{\log \tau_\ell^*} - e^{\log \tau_\ell}) \\ + a_v (\log \tau_v^* - \log \tau_v) - b_v (e^{\log \tau_v^*} - e^{\log \tau_v}) \quad (3.9)$$

where $\mathbf{K}_{\boldsymbol{\theta}}$ and $\mathbf{K}_{\boldsymbol{\theta}^*}$ are the matrices obtained by evaluating the kernel function, $k(\cdot, \cdot)$, at each combination of the observed input points using, respectively, the current and the proposed values of $\boldsymbol{\theta}$.

As in a parametric stochastic frontier model and using the complete-data likelihood function in (3.3), it is easy to show that the complete conditional of $1/\sigma_u^2$ is Gamma with shape and rate parameters $\frac{N}{2} + a_u$ and $\frac{1}{2} \sum_i u_i^2 + b_u$, respectively, and the complete conditional of each u_i is normal with mean $-\frac{(y_i - f_i)\sigma_u^2}{\sigma_v^2 + \sigma_u^2}$ and variance $\frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}$, truncated from below at zero.⁴ DMU-specific estimates of efficiency scores can be obtained by storing the values of u_i from every iteration of the Gibbs sampler and, given the relationship between u_i and technical efficiency ($u = -\log \text{TE}$), summarizing (across iterations) the exponential of minus these values.

⁴ f_i in the expression for the mean of u_i is the current draw from the posterior of $\mathbf{f}$ for observation i and the term $y_i - f_i$ can be viewed as corresponding to the residual of a parametric stochastic frontier.

The complete conditional of $\mathbf{f}$, derived using the complete-data likelihood in (3.3), is normal with mean:

$$\tilde{\mathbf{m}} = \mathbf{m} + \mathbf{K} (\mathbf{K} + \sigma_v^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{m} + \mathbf{u}) \quad (3.10)$$

and variance:

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{K} (\mathbf{K} + \sigma_v^2 \mathbf{I})^{-1} \mathbf{K} \quad (3.11)$$

These expressions are similar to typical Gaussian-process regression, with the only difference being that $\mathbf{u}$ is now added to the last term in the expression for $\tilde{\mathbf{m}}$. This complete conditional is the density of $\mathbf{f}$ after having observed $\mathbf{y}$ and conditional on $\mathbf{u}$ and is of particular importance, as it conveys information about the shape of the frontier function, $f(\cdot)$, and its derivatives. In fact, the frontier function follows a Gaussian process in the posterior, with mean and covariance functions:

$$\tilde{m}(\mathbf{x}_*) = m(\mathbf{x}_*) + K(\mathbf{X}, \mathbf{x}_*)^\top (\mathbf{K} + \sigma_v^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{m} + \mathbf{u}) \quad (3.12)$$

and:

$$\tilde{k}(\mathbf{x}_*, \mathbf{x}'_*) = k(\mathbf{x}_*, \mathbf{x}'_*) - K(\mathbf{X}, \mathbf{x}_*)^\top (\mathbf{K} + \sigma_v^2 \mathbf{I})^{-1} K(\mathbf{X}, \mathbf{x}'_*) \quad (3.13)$$

respectively, where $\mathbf{x}_*$ and $\mathbf{x}'_*$ are any two input vectors, not necessarily in the observed data, and $K(\mathbf{X}, \mathbf{x}_*)$ is an $N \times 1$ vector obtained by evaluating the Gaussian process' prior kernel function, $k(\cdot, \cdot)$, at each observed input vector, $\mathbf{x}_i$, and $\mathbf{x}_*$.⁵ The gradient of $\tilde{m}(\cdot)$ at $\mathbf{x}_*$ is:

$$\frac{\partial \tilde{m}(\mathbf{x}_*)}{\partial \mathbf{x}_*} = \frac{\partial m(\mathbf{x}_*)}{\partial \mathbf{x}_*} + \frac{1}{\sigma_\ell^2} \tilde{\mathbf{X}}^\top \left(K(\mathbf{X}, \mathbf{x}_*)^\top \odot (\mathbf{K} + \sigma_v^2 \mathbf{I})^{-1} (\mathbf{y} - \mathbf{m} + \mathbf{u}) \right) \quad (3.14)$$

where $\tilde{\mathbf{X}} \equiv [\mathbf{x}_1 - \mathbf{x}_* \quad \mathbf{x}_2 - \mathbf{x}_* \cdots \quad \mathbf{x}_N - \mathbf{x}_*]^\top$ and ' $\odot$ ' denotes the Hadamard product. The posterior mean and kernel functions, as well as the gradient of $\tilde{m}(\cdot)$ depend on the values of $\boldsymbol{\theta}$ and $\mathbf{u}$. These unobservables, however, can be marginalized by evaluating the gradient in every iteration of the Gibbs sampler and then summarizing these values across iterations. In this way, any uncertainty with respect to the values of random variables or parameters is transmitted to the estimates of the frontier and its gradient.

The Gibbs sampler, as described above, is given in Algorithm 1. This algorithm is broken into blocks related to different unobserved quantities. By far, the most computationally intensive steps are the ones that appear in blocks 1 and 2, as they involve inversion or decomposition of $N \times N$ matrices. The Metropolis-Hastings ratio in block 1 requires inversion of $\mathbf{K}_{\boldsymbol{\theta}^*} + \sigma_v^{2*} \mathbf{I}$ and calculation of the logarithm of its determinant. These calculations can be combined by evaluating the Cholesky decomposition of $\mathbf{K}_{\boldsymbol{\theta}^*} + \sigma_v^{2*} \mathbf{I}$ and then using this to invert the matrix, as well as to calculate its determinant as the product of the squares of the diagonal elements of the resulting triangular matrix. Block 2 involves sampling from a multivariate-normal density with variance matrix $\mathbf{K}_{\boldsymbol{\theta}} - \mathbf{K}_{\boldsymbol{\theta}} (\mathbf{K}_{\boldsymbol{\theta}} + \sigma_v^2 \mathbf{I})^{-1} \mathbf{K}_{\boldsymbol{\theta}}$, which implicitly requires the Cholesky decomposition of this matrix. Given that the computational complexity of a Cholesky decomposition is

⁵This result can be derived by noting that:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{m} - \mathbf{u} \\ \mathbf{m} \end{bmatrix}, \begin{bmatrix} \mathbf{K} + \sigma_v^2 \mathbf{I} & K(\mathbf{X}, \mathbf{X}_*) \\ K(\mathbf{X}, \mathbf{X}_*)^\top & K(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right)$$

where $\mathbf{f}_*$ is a random vector obtained from the Gaussian process for arbitrary input points, $\mathbf{X}_*$. Using the properties of the normal distribution we obtain the probability density function of $\mathbf{f}_* | \mathbf{y}$ and since this is true for any input points, $\mathbf{X}_*$, $f(\cdot)$ follows a Gaussian process in the posterior.

Algorithm 1 Gibbs Sampler and Gradient of the Posterior Mean Function

```

set initial values for  $\boldsymbol{\theta}$ ,  $\sigma_u^2$  and  $\mathbf{u}$ 
calculate  $\mathbf{m}_{\boldsymbol{\theta}}$  and  $\mathbf{K}_{\boldsymbol{\theta}}$  using these initial values
for  $g = 1 : G$  do
  # Block 1: sampling for  $\boldsymbol{\theta}$ 
  draw  $\boldsymbol{\theta}^*$  from the proposal,  $q(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ 
  calculate  $\mathbf{m}_{\boldsymbol{\theta}^*}$  and  $\mathbf{K}_{\boldsymbol{\theta}^*}$  using the proposed value  $\boldsymbol{\theta}^*$ 
  accept the proposed move with probability  $\alpha = \min\{1, \exp\{\log \text{MH}(\boldsymbol{\theta}, \boldsymbol{\theta}^*)\}\}$ 
  (log MH  $(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$  is given in equation (3.9))

  # Block 2: sampling for  $\mathbf{f}$ 
  draw  $\mathbf{f}$  from  $N(\tilde{\mathbf{m}}_{\boldsymbol{\theta}}, \tilde{\mathbf{K}}_{\boldsymbol{\theta}})$ 
  (the expressions for  $\tilde{\mathbf{m}}_{\boldsymbol{\theta}}$  and  $\tilde{\mathbf{K}}_{\boldsymbol{\theta}}$  are given in equations (3.10) and (3.11))

  # Block 3: inefficiency term
  draw  $1/\sigma_u^2$  from Gamma  $(\frac{N}{2} + a_v, \frac{1}{2} \sum_i u_i^2 + b_v)$ 
  for  $i = 1 : N$  do
    draw  $u_i$  from  $N^+\left(-\frac{(y_i - f_i)\sigma_u^2}{\sigma_v^2 + \sigma_u^2}, \frac{\sigma_v^2 \sigma_u^2}{\sigma_v^2 + \sigma_u^2}\right)$ 
  end for

  # Block 4: gradient of  $\tilde{m}(\cdot)$ 
  evaluate  $\frac{\partial \tilde{m}(\mathbf{x}_*)}{\partial \mathbf{x}_*}$  at any required point,  $\mathbf{x}_*$ , given the current values of  $\boldsymbol{\theta}$  and  $\mathbf{u}$ 
  (the expression for the gradient of the posterior mean is given in equation (3.14))

  # Block 5: storing values
  store the values of  $\boldsymbol{\theta}$ ,  $\sigma_u^2$ ,  $\mathbf{u}$ ,  $\frac{\partial \tilde{m}(\mathbf{x}_*)}{\partial \mathbf{x}_*}$  and possibly  $\mathbf{f}$  from the current iteration
end for
drop the first few draws to avoid dependence of the results on initial values (burn-in)
and summarize the remaining draws from the posterior

```

$\mathcal{O}(N^3)$, computation time may become prohibitive for N greater than a few thousands. Block 3 is particular to the distribution imposed on the u_i s and this is the only part of the algorithm that needs to be modified when different distributional assumptions are employed. No more decompositions or inversions are needed in the calculation of the gradient of the posterior mean function and the computations in block 4 are minimal, given that the part of $K(\mathbf{X}, \mathbf{x}_*)$ that does not depend on $\boldsymbol{\theta}$ can be calculated once and reused in every iteration of the Gibbs sampler.

4. ALTERNATIVE SPECIFICATIONS AND EXTENSIONS

An implication of the Gaussian process' kernel function taking the squared-exponential form is that, given the values of the parameters, the covariance of two values on the frontier depends solely on the Euclidean distance between the two corresponding input vectors. Thus, the units of measurement of the inputs may have an impact on the results. An easy way to avoid this issue is to standardize the input data prior to estimation. In this way, differences in input vectors will have a similar effect on the parameters and

other unknown quantities, irrespective of the direction in which these differences appear. A more elaborate approach would be to allow for different length-scale parameters (σ_ℓ^2 s) in the kernel function for each input. This would lead to a kernel function of the form:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp \left\{ -\frac{(\mathbf{x} - \mathbf{x}')^\top \mathbf{\Lambda}^{-1} (\mathbf{x} - \mathbf{x}')}{2} \right\} \quad (4.15)$$

where $\mathbf{\Lambda}$ is a $K \times K$ diagonal matrix, whose diagonal stores the input-specific σ_ℓ^2 s. This kernel function, however, has the potential to over-parameterize the problem, especially when K is large, as gaps in the data are more likely to appear in such applications. A less extreme approach would be to group inputs according to a criterion, for example quasi-fixed and variable, and restrict σ_ℓ^2 s to be common to inputs in the same group.

A major criticism that parametric efficiency analysis methods receive is that they require a distributional assumption to be imposed on the inefficiency component of the error term. Although the distribution used in applications rarely has any major impact on the ranking of DMUs according to their efficiency scores (Ruggiero, 1999, and Kumbhakar and Lovell, 2000, p.90), it may affect the absolute levels of efficiency when examined either at the DMU level (Baccouche and Kouki, 2003) or at the sample-mean level (Greene, 1990). This criticism extends to the methods described in this paper, as u_i is still treated as a random variable with support on $\mathbb{R}_+$. Alternative distributions can be imposed on u_i and the only place where Algorithm 1 needs to be altered is in block 3. Furthermore, given the resemblance of the Gaussian-process approach to a parametric stochastic frontier, results already available in the literature can be used directly.⁶

Accounting for additive unobserved heterogeneity in the case of panel data is also straightforward. Assuming that group-specific effects are normally distributed and uncorrelated with the inputs, $\alpha_i \sim N(0, \sigma_\alpha^2)$, these can be added to the right-hand side of the equation that determines log-output and, subsequently, integrated-out from the likelihood via data augmentation. Such an approach would require subtracting the latent group effects from the observed $\mathbf{y}$ in every place in Algorithm 1 and adding a block where σ_α^2 and each α_i are sampled from their respective complete conditionals.

Finally, given that the derivatives of the Gaussian process' posterior mean function are easy to evaluate, when time-series or panel data are available the technique can be used to measure and decompose TFP growth. Defining TFP growth as growth rate in output net of growth in inputs and assuming that DMUs are allocatively efficient results in:

$$\widehat{\text{TFP}} = \frac{\varepsilon - 1}{\varepsilon} \sum_{k=1}^K \varepsilon_k \cdot \hat{x}_k + \frac{\partial f}{\partial t} - \frac{\partial u}{\partial t} \quad (4.16)$$

where $\varepsilon_k \equiv \frac{\partial f}{\partial \log x_k}$, $\varepsilon \equiv \sum_{k=1}^K \varepsilon_k$ and $\hat{x}_k$ denotes the growth rate in input k . The first summand in this expression is the scale effect, the second the technical change effect and the last the efficiency change effect. Notice that the logarithm of the production frontier, f , now includes a time trend as an additional explanatory variable. With panel data the growth rate in the Malmquist productivity index for DMU i between periods $t-1$ and t can be evaluated by taking the simple arithmetic mean of elasticities and other

⁶See, for example, van den Broeck et al. (1994) for exponentially-, Erlang- and truncated-normally-distributed u_i s, Tsionas (2000, 2007) for Gamma- and Weibull-distributed u_i s, and Griffin and Steel (2004) for a non-parametric treatment of u_i .

derivatives across two adjacent periods (Orea, 2002):

$$\begin{aligned} \widehat{\text{TFP}}_{it} = & \frac{1}{2} \sum_{k=1}^K \left(\frac{\varepsilon_{i,t-1}-1}{\varepsilon_{i,t-1}} \varepsilon_{k,i,t-1} + \frac{\varepsilon_{it}-1}{\varepsilon_{it}} \varepsilon_{k,it} \right) \cdot \hat{x}_{k,it} \\ & + \frac{1}{2} \left(\left. \frac{\partial f}{\partial t} \right|_{i,t-1} + \left. \frac{\partial f}{\partial t} \right|_{it} \right) - \Delta u_{it} \end{aligned} \quad (4.17)$$

As before, uncertainty with respect to the values of the u_{it} s and of the derivatives can be accounted for by evaluating the TFP-growth components in every iteration of the Gibbs sampler and then averaging.

5. ILLUSTRATION USING SIMULATED DATA

This section provides a simple illustration of the technique in the case of a single-input production function and using simulated data. 300 values for the input, x , are drawn from a log-normal distribution with location parameter equal to 2 and scale parameter equal to 0.1. The corresponding u_i s are drawn from a half-normal distribution with pre-truncation variance (σ_u^2) equal to 0.09. The values for the dependent variable are generated as $y_i = (6x_i + 5x_i^2 - 0.2x_i^3) e^{v_i - u_i}$, where the v_i s are draws from a normal distribution with mean zero and variance (σ_v^2) equal to 0.04.

Using these data, the production frontier and the efficiency scores were estimated using a Gaussian process and a parametric stochastic frontier. For the Gaussian process $\log y$ is treated as the dependent variable and x as the input variable, although using $\log x$ instead of x produces very similar results. The parametric stochastic frontier is specified as quadratic in $\log x$, along with a constant term and is estimated using Bayesian methods and the same priors for the inefficiency and noise components of the error term as in the semi-parametric frontier. Of course, the parametric model is misspecified, but by being quadratic in $\log$ -input, it provides a second-order approximation to the true frontier. Prior to estimation with the parametric model the input variable was normalized by its geometric mean and, thus, the point of expansion of the Taylor series is the geometric mean of the data.

Panel (a) in Figure 1 presents the true frontier and its estimates from the Gaussian process and the parametric frontier. As expected, both techniques approximate the true frontier better in the region around the mean of the data. For the parametric model this is because the geometric mean is the point around which the Taylor series is expanded and, therefore, it can mimic the curvature of the true frontier well. By being a semi-parametric procedure, the Gaussian process exploits mostly local information from the data and, therefore, performs better in regions where data density is high. Indeed, for large values of x , where data are sparser, the Gaussian-process frontier tends to follow the data quite closely, rather than approximating the true frontier. On the contrary, the parametric frontier's shape is determined by all data points and, as such, it cannot capture the changing curvature of the true frontier when data are sparse. Panel (b) in Figure 1 plots the estimated efficiency scores, obtained as the posterior expected value of each e^{-u_i} , from the semi-parametric and the fully parametric frontiers against the true efficiency scores. From this plot it appears that the two techniques perform equally well in estimating DMU-specific efficiency.

Panels (c) and (d) in Figure 1 present graphs corresponding to panels (a) and (b), but for the case where a wrong distributional assumption is imposed on the inefficiency

Figure 1. Semi-parametric and fully parametric stochastic frontiers in the single-input case.

component of the error term. In particular, the same data on x and y are used as before, where each u_i is generated as a draw from a half-normal distribution, but both the semi-parametric and parametric models are estimated assuming that the u_i s follow an exponential distribution with rate parameter λ . The performance of the models deteriorates visibly when considering both the approximation to the true frontier and the estimates of the efficiency scores, although the semi-parametric model produces slightly more accurate efficiency scores.

A numerical description of these results appears in Table 1. The upper part of this table presents the true values of the common parameters in the data-generating process (σ_v^2 and σ_u^2), as well as the average output elasticity over data points, along with estimates from the semi-parametric and parametric specifications and using both alternative assumptions on the distribution of u_i . The differences in the posterior means of the quantities of interest between the two techniques are negligible, whether one considers the results from the correctly and incorrectly specified models. The performance of the semi-parametric and parametric specifications in terms of prediction accuracy of the efficiency scores is assessed using the Mean Absolute Percentage Error (MAPE), Mean

Table 1. Estimates of common parameters and prediction accuracy of efficiency scores.

	TRUE	Correct assumption on the distribution of u_i		Wrong assumption on the distribution of u_i	
		GP Frontier	Parametric Frontier	GP Frontier	Parametric Frontier
mean ε	1.399	1.412	1.413	1.412	1.411
σ_v^2	0.040	0.041	0.043	0.050	0.053
σ_u^2	0.090	0.079	0.077	–	–
λ	–	–	–	7.733	8.616

Prediction accuracy of technical efficiency scores

	GP Frontier	Parametric Frontier	GP Frontier	Parametric Frontier
MAPE	12.772	12.838	17.199	18.013
MAD	9.008	9.011	11.429	11.886
MSD	1.221	1.217	2.078	2.268

Note: MAPE = $\frac{100}{N} \sum_i \left| \frac{A_i - P_i}{A_i} \right|$, MAD = $\frac{100}{N} \sum_i |A_i - P_i|$ and MSD = $\frac{100}{N} \sum_i (A_i - P_i)^2$, where A_i is the true technical efficiency for observation i and P_i is its estimate from the respective model.

Absolute Deviation (MAD) and Mean Squared Deviation (MSD). When the correct distributional assumption is imposed on u_i the two techniques perform equally well in terms of prediction accuracy. However, when a wrong assumption is made the Gaussian-process frontier slightly outperforms the parametric frontier.

The results presented in this section were obtained after setting the values of the shape hyper-parameters for $\tau_f \equiv 1/\sigma_f^2$ and $\tau_\ell \equiv 1/\sigma_\ell^2$ to one and the respective rate hyper-parameters to 0.1. Thus, the prior expected values of these two precision parameters are equal to 10 and their prior variances equal to 100. These are rather vague priors, given the range of the dependent and independent variables. Results obtained using different sets of hyper-parameters are available in the supplementary appendix associated with this paper. The main findings discussed here persist under different hyper-parameter values, with a notable exception: when both τ_f and τ_ℓ are forcefully restricted to be close to 10 in the prior, the fitted frontier tends to follow the data too closely wherever data density is low. Thus, a value for τ_ℓ close to 10 is, in the context of the data, large enough to make the estimated frontier alternate curvature. This is because an observation that is more than one unit of x away from its neighbors can hardly use any information on the location of the frontier nearby (the value of $k(\cdot, \cdot)$ with $\tau_f = \tau_\ell = 10$ and $|x - x'|^2 = 1$ is as low as 0.067, which is extremely low for the range of the dependent variable).

6. EMPIRICAL APPLICATION

This section presents an empirical application of the technique developed in section 3 and the extensions discussed in section 4 to a panel dataset of US electric utilities. This dataset was constructed and first used by Rungsuriyawiboon and Stefanou (2007). It contains information on 81 investor-owned utilities, which use fossil-fuel fired boilers to

produce electricity, each one of them observed annually between 1986 and 1997. Output is measured in megawatt hours of electricity and inputs are aggregated into three categories: (a) quantity of fuel, measured as fuel cost divided by a price index for fuel, which was specifically constructed for this dataset, (b) labour and maintenance, defined as costs attributed to labour and maintenance, divided by a price index, again, specifically constructed for this dataset, and (c) capital stock, measured in the base year at replacement cost and adjusted for new investment, depreciation and retirement in subsequent years.

Fully parametric and semi-parametric stochastic frontier models are estimated. The independent variables in the Gaussian-process frontier are the logarithms of the three inputs and a time trend variable. To avoid dependence of the results on the units of measurement, all independent variables are standardized such that their arithmetic mean is equal to zero and their variance equal to one. The dependent variable is the logarithm of output, which for numerical stability purposes is transformed prior to estimation by subtracting its arithmetic mean from each observation. This last transformation has no implications for the interpretation of the results. The parametric frontier is specified as translog in the three inputs and, additionally, includes a time trend, the square of the time trend and interaction terms between the logarithms of the inputs and the time trend variable. Prior to estimation of the parametric model, the input and output variables are normalized by their geometric means and the time trend variable transformed such that its arithmetic mean is equal to zero. These transformations make the parameter estimates associated with first-order terms directly interpretable as output elasticities, evaluated at the geometric mean of the data. In the parametric model, these elasticities coincide to the ones obtained by first evaluating the elasticities at each data point and then taking the arithmetic mean across observations.

Following Emvalomatis (2012), an autoregressive structure is imposed on the firm-specific efficiency scores:

$$\begin{aligned} s_{it} &= \delta + \rho s_{i,t-1} + \xi_{it}, & \xi_{it} &\sim N\left(0, \sigma_\xi^2\right) \\ s_{i0} &= \frac{\delta}{1-\rho} + \xi_{i0}, & \xi_{i0} &\sim N\left(0, \frac{\sigma_\xi^2}{1-\rho^2}\right) \end{aligned} \quad (6.18)$$

where $s_{it} \equiv \log \frac{TE_{it}}{1-TE_{it}}$ is a hidden state variable. This one-to-one relationship between s_{it} and TE_{it} implies that $\log TE_{it} = s_{it} - \log(1 + e^{s_{it}})$ and the observed equation of the stochastic frontier model becomes:

$$y_{it} = \alpha_i + f(\mathbf{x}_{it}) + v_{it} + \log TE_{it} \quad (6.19)$$

where, by assumption, $v_{it} \sim N(0, \sigma_v^2)$ and $\alpha_i \sim N(0, \sigma_\alpha^2)$ is a firm-specific random effect. The stochastic dependence of efficiency scores over time is justified from an adjustment-cost point of view (Ahn and Sickles, 2000, Tsionas, 2006, and Emvalomatis, 2012) and, as argued by Skevas et al. (2018), it provides a balance between imposing strong assumptions on the evolution of efficiency over time as, for example, in the Battese and Coelli (1992) model and no dependence at all. This aspect becomes particularly relevant when estimating and decomposing TFP growth.

The estimates of the common parameters in the semi-parametric and parametric specifications appear in Table 2.⁷ The two specifications produce very similar results, with

⁷The results presented here are obtained from a Gibbs sampler that ran for 220,000 iterations. The first 20,000 draws were discarded (burn-in) and, to reduce the degree of autocorrelation in the draws without using excessive amounts of machine memory, only one in ten from the remaining 200,000 were stored.

Table 2. Estimates of common parameters/elasticities.

	Gaussian-Process Specification			Translog Specification		
	mean	st.dev.	90% interval	mean	st.dev.	90% interval
ε_K	0.1639	0.0284	[0.117, 0.210]	0.2132	0.0287	[0.165, 0.260]
ε_L	0.0533	0.0159	[0.027, 0.079]	0.0552	0.0195	[0.022, 0.086]
ε_F	0.5346	0.0200	[0.502, 0.568]	0.5267	0.0217	[0.491, 0.562]
ε_t	0.0218	0.0015	[0.019, 0.024]	0.0169	0.0017	[0.014, 0.020]
σ_α^2	0.0776	0.0198	[0.051, 0.114]	0.0662	0.0182	[0.042, 0.100]
σ_v^2	0.0011	0.0002	[0.001, 0.001]	0.0009	0.0002	[0.001, 0.001]
δ	0.1210	0.0342	[0.071, 0.182]	0.1113	0.0311	[0.064, 0.166]
ρ	0.9376	0.0151	[0.911, 0.960]	0.9250	0.0167	[0.896, 0.951]
σ_ξ^2	0.1616	0.0309	[0.116, 0.216]	0.1288	0.0223	[0.093, 0.167]

overlapping 90% credible intervals for all parameters/elasticities. In both specifications firms operate, on average, at the decreasing returns to scale part of the production function, while the semi-elasticity with respect to time implies average productivity increase due to technical change at rates of 2.18% and 1.69% in the two specifications. Average technical efficiency across both the time and firm dimensions of the panel is 83% for the semi-parametric specification and 78.3% for the parametric.

Apart from estimation of the parameters/output elasticities, TFP growth between each successive year and for each firm is calculated and decomposed using the formula in (4.17). The evolution of TFP and its components is presented in Figure 2, after taking the average across firms and normalizing the productivity index to unity in the first year for which data are available. The two specifications produce almost identical results for TFP growth and imply an average annual growth rate of approximately 1.79% and 1.63%, respectively. Although the patterns of TFP evolution are very similar in the two specifications, the contribution of each component to TFP growth reveals some important differences. In particular, the technical change component in the parametric specification is very smooth, by construction. On the contrary, the semi-parametric model implies a slowdown in technical progress in the early 1990's, followed by an acceleration later on. Two events that occurred during this period may explain this finding: the first Gulf war, that took place in 1991, led to an increase in oil prices, and the 1992 Energy Policy Act led to deregulation of the electricity production and distribution industry in the US. With respect to the increased oil prices, although a production frontier is estimated here, which should theoretically not depend on input prices, the quantity of fuel is obtained by dividing fuel costs by a price index which may be unable to capture sharp changes in the prices of some categories of fuel. More importantly, the 1992 Energy Policy Act forced electricity producers to deliver power to third parties at cost-based prices and

The shape hyper-parameters for τ_f and τ_ℓ are set to one and the rate hyper-parameters to 0.1. Results obtained using different values for these hyper-parameters are available in the supplementary appendix. Largely because the independent variables were standardized prior to estimation of the semi-parametric model, they have a limited range and, thus, the choice of hyper-parameter values has negligible impact on the results, except in cases where these values severely restrict the range of τ_f and τ_ℓ .

Figure 2. TFP evolution and decomposition in US electric utilities using semi-parametric and fully parametric stochastic frontiers.

removed incentives which, for decades, led to over-investment in the sector. It is, therefore, reasonable that in anticipation of these policy changes, electric utilities refrained from investing in non-essential equipment or premature capital replacement, thus leading to a slowdown in technical progress.

Due to the parametric specification's inability to model the technical change component in a flexible way, any discrepancies between the "true" and estimated technical change effects are left to be captured by the efficiency change and, to a lesser extent, the scale effects. Indeed, for the period during which the cumulative technical change effect obtained from the parametric model falls behind that from the semi-parametric, the cumulative efficiency change effect compensates for the difference. This compensation continues later on, but in the opposite direction, and the cumulative efficiency change effect from the two specifications is almost the same in the last period covered by the data. Overall, although estimates of TFP differ only slightly, the two specifications attribute changes in TFP to different components.

7. CONCLUSIONS

Although the distinction between parametric and non-parametric techniques of efficiency and productivity analysis tends to eclipse, at least in terms of the information that the techniques can extract from the data, efforts to bridge the gap overwhelmingly come from the non-parametric branch of the relevant literature. The few papers that relax functional specification assumptions in stochastic frontier models use frequentist approaches, which are usually specific to other parts of the specification of the model. This paper proposes a fully Bayesian approach to semi-parametric stochastic frontier analysis based on Gaussian-process regression. This approach requires no specification of a functional form of the production frontier, but maintains the need for a distributional assumption on inefficiency. However, due to the ability of Bayesian techniques to incorporate additional components in the specification of the model, the Gaussian-process approach can be combined with any valid distributional assumption on inefficiency or even with non-parametric modelling of the inefficiency component of the error term, in the spirit of Griffin and Steel (2004). Furthermore, unobserved heterogeneity in the case of panel data can be handled easily, while the availability of results on the distribution of the derivatives of the frontier make calculation and decomposition of TFP growth a straightforward task.

The technique developed here is found to perform equally well or slightly better than a flexible parametric specification of a stochastic frontier model using simulated data in the very simple context of a single-input production process and under correct and incorrect distributional assumptions on inefficiency. The technique is also applied to a panel dataset of US electric utilities, in a specification that allows for additive firm effects and autoregressive inefficiency. Output elasticities and other parameters obtained from the Gaussian-process frontier are very similar to those that come from a translog frontier. The evolution of TFP from the two alternative specifications is also very similar, but the semi-parametric model allows for a more flexible pattern in the technical progress component and, as a result, does not overload the efficiency change and scale components of TFP growth.

Gaussian process regression has a long history in the fields of geostatistics and machine learning and results already available in the literature can be applied directly to efficiency and productivity analysis to build more elaborate models. For example, noise in the inputs can be handled as in McHutchon and Rasmussen (2011), while monotonicity and curvature constraints can be imposed on the production frontier using the approaches proposed by Riihimäki and Vehtari (2010) and Wang and Berger (2016). As relative input prices, along with the assumption that the DMUs' objective is to minimize cost, provide information on the slope of the frontier in the input space, this information can be incorporated in the estimation process using the result that a Gaussian process and its derivatives jointly follow another Gaussian process (Rasmussen and Williams, 2006, p.191).

REFERENCES

- Ahn, S. C. and R. C. Sickles (2000). Estimation of long-run inefficiency levels: a dynamic frontier approach. *Econometric Reviews*, 19, 461–92.
- Aigner, D., C. A. K. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production functions. *Journal of Econometrics*, 6, 21–37.
- Baccouche, R. and M. Kouki (2003). Stochastic production frontier and technical inefficiency.

- ciency: A sensitivity analysis. *Econometric Reviews*, 22, 79–91.
- Banker, R. D., A. Charnes, and W. W. Cooper (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30, 1078–92.
- Battese, G. E. and T. J. Coelli (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis*, 3, 153–69.
- Bădin, L., C. Daraio, and L. Simar (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, 223, 818–33.
- Cazals, C., J.-P. Florens, and L. Simar (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106, 1–25.
- Charnes, A., W. W. Cooper, and E. Rhodes (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429–44.
- Choi, T. and M. J. Schervish (2007). On posterior consistency in nonparametric regression problems. *Journal of Multivariate Analysis*, 98, 1969–87.
- Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19, 273–92.
- Deprins, D., L. Simar, and H. Tulkens (1984). Measuring labor efficiency in post offices. In M. Marchand and P. Pestieau and H. Tulkens (Eds.), *The Performance of Public Enterprises Concepts and Measurements*, 247–63. Amsterdam: Elsevier.
- Emvalomatis, G. (2012). Adjustment and unobserved heterogeneity in dynamic stochastic frontier models. *Journal of Productivity Analysis*, 37, 7–16.
- Fan, Y., Q. Li, and A. Weersink (1996). Semiparametric estimation of stochastic production frontier models. *Journal of Business and Economic Statistics*, 14, 460–68.
- Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84, 66–83.
- Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A* 120, 253–81.
- Geweke, J. (1991). Efficient simulation from the multivariate normal and student-*t* distributions subject to linear constraints and the evaluation of constraint probabilities. In E. M. Keramidas and S. M. Kaufman (Eds.), *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, 571–78. Fairfax Station, VA: Interface Foundation of North America.
- Greene, W. H. (1990). A Gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46, 141–63.
- Griffin, J. E. and M. Steel (2004). Semiparametric Bayesian inference for stochastic frontier models. *Journal of Econometrics*, 123, 121–52.
- Hajivassiliou, V., D. McFadden, and P. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives. Theoretical and computational results. *Journal of Econometrics*, 72, 85–134.
- Kasy, M. (2018). Optimal taxation and insurance using machine learning – Sufficient statistics and beyond. *Journal of Public Economics*, 167, 205–19.
- Keane, M. P. (1994). A computationally practical simulation estimator for panel data. *Econometrica*, 62, 95–116.
- Koopmans, T. C. (1951). Analysis of production as an efficient combination of activities. In T. C. Koopmans (Ed.), *Activity analysis of production and allocation*, 33–97. New York, NY: John Wiley & Sons.

- Kumbhakar, S. C. and C. A. K. Lovell (2000). *Stochastic Frontier Analysis*. Cambridge: Cambridge University Press.
- Kumbhakar, S. C., B. U. Park, L. Simar, and E. G. Tsionas (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach. *Journal of Econometrics*, 137, 1–27.
- Ludkovski, M., J. Risk, and H. Zail (2018). Gaussian process models for mortality rates and improvement factors. *ASTIN Bulletin*, 48, 1307–47.
- McHutchon, A. and C. E. Rasmussen (2011). Gaussian process training with input noise. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems 24*, 1341–49. Curran Associates, Inc.
- Meeusen, W. and J. van den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18, 435–44.
- Nesreen K. A., A. F. Atiya, N. El Gayar and H. El-Shishiny (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29, 594–621.
- Orea, L. (2002). Parametric decomposition of a generalized Malmquist productivity index. *Journal of Productivity Analysis*, 18, 5–22.
- Park, B. U. and L. Simar (1994). Efficient semiparametric estimation in a stochastic frontier model. *Journal of the American Statistical Association*, 89, 929–36.
- Rasmussen, C. E. and C. K. I. Williams (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press.
- Riihimäki, J. and A. Vehtari (2010). Gaussian processes with monotonicity information. In Y. W. Teh and M. Titterton (Eds.), *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Volume 9 of *Proceedings of Machine Learning Research*, Chia Laguna Resort, Sardinia, Italy, 645–52. PMLR.
- Ruggiero, J. (1999). Efficiency estimation and error decomposition in the stochastic frontier model: A Monte Carlo analysis. *European Journal of Operational Research*, 115, 555–63.
- Rungsuriyawiboon, S. and S. E. Stefanou (2007). Dynamic efficiency estimation: An application to U.S. electric utilities. *Journal of Business and Economic Statistics*, 25, 226–38.
- Ruseckaite, A., D. Fok, and P. Goos (2018). Flexible mixture-amount models using multivariate Gaussian processes. *Journal of Business & Economic Statistics*, *Forthcoming*.
- Simar, L. and P. W. Wilson (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science*, 44, 49–61.
- Simar, L. and P. W. Wilson (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research*, 115, 459–71.
- Simar, L. and P. W. Wilson (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136, 31–64.
- Skevas, I., G. Emvalomatis, and B. Brümmer (2018). Productivity growth measurement and decomposition under a dynamic inefficiency specification: The case of German dairy farms. *European Journal of Operational Research*, 271, 250–61.
- Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82, 528–40.
- Tsionas, E. G. (2000). Full likelihood inference in normal-gamma stochastic frontier models. *Journal of Productivity Analysis*, 13, 183–205.

- Tsionas, E. G. (2006). Inference in dynamic stochastic frontier models. *Journal of Applied Econometrics*, 21, 669–76.
- Tsionas, E. G. (2007). Efficiency measurement with the Weibull stochastic frontier. *Oxford Bulletin of Economics and Statistics*, 69, 693–706.
- van den Broeck, J., G. Koop, J. Osiewalski, and M. F. Steel (1994). Stochastic frontier models: A Bayesian perspective. *Journal of Econometrics*, 61, 273–303.
- van der Vaart, A. W. and J. H. van Zanten (2009). Adaptive Bayesian estimation using a Gaussian random field with inverse Gamma bandwidth. *Annals of Statistics*, 37, 2655–75.
- Wang, X. and J. O. Berger (2016). Estimating shape constrained functions using Gaussian processes. *SIAM/ASA Journal on Uncertainty Quantification*, 4, 1–25.
- Winsten, C. B. (1957). Discussion on Mr. Farrell’s paper. *Journal of the Royal Statistical Society, Series A* 120, 282–84.
- Yeo, M., T. Fletcher, and J. Shawe-Taylor (2015). Machine learning in fine wine price prediction. *Journal of Wine Economics*, 10, 151–72.