On the Estimation of Technical and Allocative Efficiency in a Panel Stochastic Production Frontier System Model: Some New Formulations and Generalizations

Subal C. Kumbhakar* and Mike G. Tsionas†

Abstract

In this paper we propose some alternative formulations and estimation of technical and allocative inefficiency in the presence of some exogenous variables in the context of a panel stochastic frontier model which includes time-invariant firm effects (heterogeneity) along with time-varying technical inefficiency and random noise. These exogenous variables are used to explain technical and allocative inefficiency as well as firm heterogeneity. The presence of these exogenous variables allows us to relax some of the assumptions made in a recent paper by Lai and Kumbhakar (2019). These variables also allow to add flexibility in estimating the model parameters as well as both technical and allocative inefficiency and costs therefrom. More specifically, the incidental parameters problem associated with firm heterogeneity in the production function as well in the first-order conditions of cost minimization can be avoided by parameterizing them in terms of the exogenous variables. We propose and implement model comparison based on Bayes factors and marginal likelihood.

Keywords: Productivity and Competitiveness; Endogeneity; Technical Efficiency; Allocative Efficiency; Incidental Parameters Problem.

Acknowledgments: We wish to thank the Editor Robert Dyson and three anonymous reviewers for their useful comments on an earlier version of the paper.

*State University of New York, Binghamton, NY 13902, USA, email: kkar@binghamton.edu
†Lancaster University Management School, LA1 4YX, U.K, email: m.tsionas@lancaster.ac.uk

1 Introduction

Stochastic frontier models going back to Aigner, Lovell and Schmidt (1977), and Meeusen and van den Broeck (1977) and its earlier extensions to accommodate endogeneity of inputs (Schmidt and Lovell 1979) assumed all the error components to be independently distributed with some specific distributions. The main advantage of assuming distributions on the error components is that on can get observation-specific estimates of technical and allocative inefficiency and costs therefrom. Some of the distributions assumptions can be relaxed if there are exogenous variables that can explain these inefficiencies. This might be especially useful in panel data models when one wants to control for firm heterogeneity in both the production function and the first-order conditions of cost minimization. These exogenous variables can also be used to explain technical inefficiency either in a pure parametric form or in a stochastic set-up.

In a recent paper Lai and Kumbhakar (2019) consider a state-of-the-art panel stochastic production frontier model and estimate it using a system that includes both technical and allocative inefficiency. The system consists of a Cobb-Douglas production function with fixed effects along with the first order conditions (FOCs) from cost minimization, and is a panel generalization of the cross-sectional model in Schmidt and Lovell (1979). By using the FOCs the model accounts for the endogeneity of inputs which is a problem in production function estimation dating back to Marschak and Andrews (1944). Allocative inefficiency is modeled as non-fulfillment of the FOCs which also include fixed effects. Lai and Kumbhakar (2019) propose to take care of the incidental parameters problem associated with the fixed effects in the production function as well as in the FOCs using the half-panel jackknife estimator of Dhaene and Jochmans (2015).

In terms of similarities with Lai and Kumbhakar (2018), our common departure point is that estimating the primal system (consisting of the production function and the associated FOCs for cost minimization) requires a) taking into account the econometric endogeneity of inputs, and b) modeling the fixed firm effects in both the production function and the FOCs. Compared to Lai and Kumbhakar (2018) our paper is different in a number of aspects:

In this paper, we propose alternative, more general specifications to the modeling of fixed effects assuming that some exogenous variables (z) are available to explain firm heterogeneity and inefficiency. In particular, we break new ground relative to Lai and Kumbhakar (2018) along the following lines:

i) Instead of assuming that firm effects (heterogeneity) are fixed, we assume that they are members of the wider class of Mundlak's (1961) model. For this, we specify firm heterogeneity as a parametric function of the z variables along with a random error added in it. This enables us to test whether firm heterogeneity is fixed (determined parametrically by the z variables) or random.

ii) We model technical inefficiency using both a distributional assumption (lognormality) and a parametric specification (using certain exogenous variables, say z) introduced by Paul and Shankar (2018) and further generalized in Tsionas and Mamatzakis (2019) where it is shown that the inefficiency function may depend on endogenous variables as well and estimation can be performed via the generalized method of moments technique.

iii) Mundlak's approach is examined in its full version as well as in its more traditional "stripped down" version in which firm effects are linear functions of the average values of available predetermined variables. Mundlak's device is an important way of modeling fixed effects as it nests both fixed and random effect specifications. Typically, the Mundlak device makes firm effects linear functions of the average values of certain exogenous variables (z) and adds an error term. In fact, this is a testable assumption, something that appears to be novel in the literature. In fact, we propose a generalized Mundlak specification which does not depend on the assumption that average values of zs are the only determinants of firm effects.

iv) We do not assume that firm effects and error terms in the first-order conditions of cost minimization, reflect allocative inefficiency. As a matter of fact, we allow for a general semi-parametric model that allows arbitrary deviations from these first-order conditions via artificial neural networks. The importance of additional terms coming from the semi-parametric model can be evaluated easily using Bayes factors.

v) Several alternative models are compared in terms of Bayes factors, focusing on fixed versus random effects, exact fulfillment of cost minimizing first-order conditions, parametric versus random inefficiency, as well as the importance of the semi-parametric modification of the production function and the first-order conditions.

vi) As we use a number of different models making different assumptions about the nature of firm effects, the parametrization of inefficiency and the nature of the first-order conditions it is important to select a particular model. In the context of Bayesian inference this can be done using Bayes factors or posterior model probabilities. However, once posterior model probabilities are available, sample distributions of important functions of interest (like returns to scale, inefficiency, cost of allocative inefficiency etc.) can be derived using Bayesian model averaging. Model averaging can be performed easily by weighting functions of interest arising from different models using the associated posterior model probabilities.

2 Models

2.1 General

Before proceeding, it might be useful to derive conditions for cost-minimization and illustrate the difference between the primal and dual problems. Suppose there is a single output y, and inputs $x \in \Re_+^J$ whose prices are $w \in \Re_+^J$. Production possibilities are given by the production function, $y = f(x)$, which satisfies the standard neoclassical properties. The cost minimization problem is:

$$C(w, y) = \min_{x \in \Re_+^J} : \sum_{j=1}^J w_j x_j,$$
$$\text{subject to } y = f(x).$$

In this problem w and y are predetermined and the optimal solution, $C(w, y)$ is the cost function which provides the minimum cost to produce output y. Using the Lagrange function, $\mathcal{L} = \sum_{j=1}^J w_j x_j + \lambda \{y - f(x)\}$, the first order conditions are as follows:

$$w_j = \lambda f_j(x) \, \forall j = 1, \ldots, J,$$
$$y = f(x),$$

where $f_j(x) = \frac{\partial f(x)}{\partial x_j} \, \forall j = 1, \ldots, J$. We can eliminate the Lagrange multiplier λ as follows:

$$\frac{w_j}{w_1} = \frac{f_j(x)}{f_1(x)} \,\forall j = 2, \ldots, J,$$
$$y = f(x).$$

There are J equations for the J unknown input demands which are, say, $x_j^*(w, y) \,\forall j = 1, \ldots, J$. Standard manipulations yield

$$\frac{w_j x_j}{w_1 x_1} = \frac{f_j(x) x_j / y}{f_1(x) x_1 / y} \,\forall j = 2, \ldots, J,$$

where $\mathcal{E}_j(x) \equiv f_j(x) x_j / y$, is the elasticity of the production function with respect to input x_j. Therefore, the cost-minimization conditions can be written as:

$$\ln x_j - \ln x_1 = \ln w_1 - \ln w_j + \ln \mathcal{E}_j(x) - \ln \mathcal{E}_1(x),$$
$$y = f(x).$$

This is the "primal" version of the system which consists of the production function and the $J - 1$ first-order conditions for cost minimization. In the "dual" version of the problem one can specify a functional form for the cost function $C(w, y)$ and then apply Shephard's lemma to obtain the following duality result:

$$\frac{\partial C(w, y)}{\partial w_j} = x_j^*(w, y) \,\forall j = 1, \ldots, J.$$

By simple algebraic operations these equations can be written as follows:

$$\frac{\partial \ln C(w, y)}{\partial \ln w_j} = s_j^*(w, y) \,\forall j = 1, \ldots, J,$$

where $s_j^*(w, y) \equiv \frac{w_j x_j^*(w,y)}{\sum_{j'=1}^{J} w_{j'} x_{j'}^*(w,y)} \,\forall j = 1, \ldots, J$, represent the cost shares of each input.

Therefore, there are two ways to estimate the technology. In the dual approach one would have to estimate the cost function along with its associated $J - 1$ share equations (as by definition cost shares sum to unity). In the primal approach one would have to estimate the production function along with the $J - 1$ first-order conditions. In theory, the dual approach can be used to obtain both technical and allocative inefficiency in a way that respects the so-called Greene's problem but the resulting system of equations allowing for allocative distortions is highly nonlinear. From this perspective, estimating the primal system is easier, no matter how complicated the production function may be.

2.2 New models

We start with the formulation used in Lai and Kumbhakar (2019) with a slight change in notations:

$$\ln y_{it} = \beta_0 + \sum_{j=1}^{J} \beta_j \ln x_{jit} + \tau_{1i} + v_{1it} - u_{it}, i = 1, ..., n, t = 1, ..., T, \tag{1}$$

where y_{it} is output for firm i and date t, x_{jit} is the jth input for firm i and date t , τ_{1i} is firm heterogeneity representing time-invariant managerial skill, v_{1it} is a two-sided error term, u_{it} is a non-negative error component that represents technical inefficiency in production, and $\beta_j > 0, j = 1, ..., J$. From the FOC of cost minimization (where inputs are endogeneous choice variables and output is predetermined) we obtain:

$$\frac{\partial \ln y_{it}/\partial \ln x_{jit}}{\partial \ln y_{it}/\partial \ln x_{1it}} = \frac{w_{jit}x_{jit}}{w_{1it}x_{1it}} = \frac{\beta_j}{\beta_1}e^{\zeta_{jit}}, 2 = 1, ..., J. \tag{2}$$

These conditions can be expressed as follows:

$$\ln x_{1it} - \ln x_{jit} = \ln(w_{jit}/w_{1it}) + (\ln \beta_1 - \ln \beta_j) + \tau_{ji} + v_{jit}, j = 2, ..., J. \tag{3}$$

where $\zeta_{jit} = \tau_{ji} + v_{jit}$ is a two-sided error component that represents deviations from the exact fulfillment of FOC and, therefore, represents allocative inefficiency.[1] Specifically, v_{jit} is a two-sided error component and τ_{ji} is a fixed effect (specific for each input and firm).

This system of equations (which is nonlinear in the parameters) can be estimated using the method of maximum likelihood (ML) after noticing that the Jacobian of transformation from $\boldsymbol{v}_{it} = [v_{1it}, ..., v_{Jit}]'$ to the endogenous variables $\boldsymbol{X}_{it} = [\ln x_{1it}, ..., \ln x_{Jit}]'$ is simply $\sum_{j=1}^{J} \beta_j$. Since there is a large number of fixed effects $\boldsymbol{\tau}_i = [\tau_{1i}, ..., \tau_{Ji}]'$ whose number grows with the number of firms (n) estimation by ML yields consistent estimators when both n and T grow to infinity but not for fixed T as n tends to infinity (Neyman and Scott, 1948, and Schmidt, 1988[2]).

We can write (1) and (3) as follows:

$$\boldsymbol{F}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\beta}) = \boldsymbol{\tau}_i \mathbf{1}_J + \boldsymbol{v}_{it} - u_{it}\boldsymbol{\iota}, i = 1, ..., n, t = 1, ..., T, \tag{4}$$

where $\boldsymbol{Z}_{it} \in \Re^{d_Z}$ is a vector of exogenous variables, $\beta = [\beta_1, ..., \beta_J]'$, $\boldsymbol{\iota} = [1, 0, ..., 0]'$, $\mathbf{1}_J = [1, 1, ..., 1]'$. One can either keep β_0 as a parameter but impose the identifiability restriction: $\sum_{i=1}^{n} \tau_{1i} = 0$, or drop β_0 and leave $\sum_{i=1}^{n} \tau_{1i}$ unrestricted (that is, we set $\beta_0 = 0$). We use the first option without loss of generality. Moreover

$$\boldsymbol{F}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\beta}) = \begin{bmatrix} \ln y_{it} - \beta_0 - \sum_{j=1}^{J} \beta_j \ln x_{jit} \\ \ln x_{1it} - \ln x_{2it} - \ln(w_{2it}/w_{1it}) - (\ln \beta_1 - \ln \beta_2) \\ \vdots \\ \ln x_{1it} - \ln x_{Jit} - \ln(w_{Jit}/w_{1it}) - (\ln \beta_1 - \ln \beta_J) \end{bmatrix}, i = 1, ..., n, t = 1, ..., T. \tag{5}$$

[1] Greene's problem arises only if we consider the cost function and associated share equations. In this paper, we use the production function and associated first-order conditions for cost minimization. Therefore, the terms that account for allocative inefficiency in the first-order conditions do not appear in the production function itself. However, the computation of allocative inefficiency becomes involved in this case, an issue that we take up in equations (7)-(10).

[2] The incidental parameters problem has not been resolved as of yet although a number of contributions have been made. For details see Arellano and Hahn (2006), and Lancaster (2002). For the Bayesian approach see Liseo (2005). The Bayesian approach to the incidental parameters problem is still developing. The discussion boils down to considering different kinds of asymptotics, viz. $n \to \infty$ with fixed T, $T \to \infty$ with fixed n, or $n, T \to \infty$ possibly at different rates. When $n, T \to \infty$ the incidental parameters problem does not appear as estimators are consistent although, of course, they may be biased in finite samples particularly in nonlinear or dynamic models.

The above model can be estimated (as in Lai and Kumbhakar (2019)) using the distributional assumptions:

$$v_{it} \sim iid\, \mathcal{N}_J(\mathbf{0}, \mathbf{\Sigma}),\; u_{it} \sim iid\, \mathcal{N}_+(0, \sigma_u^2), i = 1, ..., n, t = 1, ..., T, \tag{6}$$

where $\mathcal{N}_+(0, \sigma_u^2)$ denotes the half-normal distribution and $\mathcal{N}_J(\mathbf{0}, \mathbf{\Sigma})$ denotes the J-variate normal distribution with mean a zero vector, and covariance matrix $\mathbf{\Sigma}$. To refer to this model later we call it **Model 0**. If $\boldsymbol{\tau_i}$ are parameters, the residuals from the production function in (1), i.e., $v_{1it} - u_{it}$, can be used to estimate technical inefficiency from the Jondrow et al. (1982) formula. For the system defined in (1) and (3) it is straightforward to derive the input demand functions (x_{jit}) with and without ζ_{jit} and derive the algebraic formula for the cost of allocative inefficiency from

$$\sum_j w_{jit} x_{jit}|_{\zeta_{jit} \neq 0} - \sum_j w_{jit} x_{jit}|_{\zeta_{jit} = 0} \equiv C(\boldsymbol{w_{it}}, y_{it})|_{\zeta_{jit} \neq 0} - C(\boldsymbol{w_{it}}, y_{it})|_{\zeta_{jit} = 0}. \tag{7}$$

Following Schmidt and Lovell (1979) and Lai and Kumbhakar (2019), we compute cost of allocative inefficiency from

$$CAI_{it} = \ln C(\boldsymbol{w_{it}}, y_{it})|_{\zeta_{jit} \neq 0} - \ln C(\boldsymbol{w_{it}}, y_{it})|_{\zeta_{jit} = 0}, \tag{8}$$

which when multiplied by 100 can be interpreted as the percentage increase in cost due to allocative inefficiency. For the system defined in (1) and (3)is given by the expression

$$CAI_{it} = (E_{it} - \ln r)\,, \tag{9}$$

where

$$E_{it} = \frac{1}{r} \sum_{j=2}^{J} \beta_j \zeta_{jit} + \ln\left(\beta_1 + \sum_{j=2}^{J} \beta_j \exp(-\zeta_{jit})\right), \tag{10}$$

and $r = \sum_{j=1}^{J} \beta_j$. That is, cost is increased by $100\,(E_{it} - \ln r)$ percent due to allocative inefficiency. If there is no input allocative inefficiency (i.e., $\zeta_{jit} = 0$ for all i, j, t), then $E_{it} - \ln r = 0$. Similarly, cost of technical inefficiency is $\frac{1}{r} u_{it}$, i.e., cost is increased by $(\frac{1}{r} u_{it})100$ percent due to technical inefficiency, u_{it}. It is worth emphasizing that to compute CAI it is necessary to solve for x_{jit} using the system in ((1)) and ((3)). This requirement restricts the use of flexible functional form for the production function in ((1)).

3 Model I: Parameterized technical inefficiency

An alternative to assuming u_{it} to be random and distributed half-normally is to specify it as a parametric function of the Z variables. Following Paul and Shankar (2018) and Tsionas and Mamatzakis (2019) in Model I we fully parameterize technical efficiency as[3]

$$e^{-u_{it}} = \Phi(\boldsymbol{Z}'_{it}\boldsymbol{\gamma}) \Rightarrow u_{it} = -\ln \Phi(\boldsymbol{Z}'_{it}\boldsymbol{\gamma}), \tag{11}$$

where $\boldsymbol{Z}_{it}$ is the vector of exogenous variables explaining inefficiency, $\boldsymbol{\gamma}$ is the corresponding parameter vector, and $\Phi(.)$ denotes a cumulative distribution function that takes values in $(0,1]$. One can use any cumulative distribution function, for example, the standard normal cumulative distribution function $\Phi(z)$ which is: $\Phi(z) = \int_{-\infty}^{z}(2\pi)^{-1/2}e^{-\xi^2/2}d\xi$.

To continue with this approach, we set $\boldsymbol{F}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\beta}) - \ln \Phi(\boldsymbol{Z}'_{it}\boldsymbol{\gamma}) \equiv \boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\beta}, \boldsymbol{\gamma})$, and write the system as

$$\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = \boldsymbol{\tau}_i \mathbf{1}_J + \boldsymbol{v}_{it}, i = 1, ..., n, t = 1, ..., T, \tag{12}$$

where

$$\boldsymbol{v}_{it} \sim iid\, \mathcal{N}_J(\mathbf{0}, \boldsymbol{\Sigma}), i = 1, \ldots, n, t = 1, ..., T. \tag{13}$$

The firm heterogeneity vector $\boldsymbol{\tau}_i$ creates the incidental parameters problem when the elements of $\boldsymbol{\tau}_i$ are treated as parameters because its dimension increases with n. One way to deal with the problem is to use either the first difference or the within transformation to the model in (12) as in Lai and Kumbhakar (2019). Another alternative is to parameterize explicitly the correlation between the firm effects and the predetermined variables (as suggested by Mundlak (1978) and Chamberlain (1980)):

$$\begin{aligned} \tau_{1i} &= \textstyle\sum_{t=1}^{T} \boldsymbol{Z}'_{it}\boldsymbol{a}_{1t} + \omega_{1i}, \\ \tau_{2i} &= \textstyle\sum_{t=1}^{T} \boldsymbol{Z}'_{it}\boldsymbol{a}_{2t} + \omega_{2i}, \\ &\vdots \\ \tau_{Ji} &= \textstyle\sum_{t=1}^{T} \boldsymbol{Z}'_{it}\boldsymbol{a}_{Jt} + \omega_{Ji}, \end{aligned} \tag{14}$$

where $\boldsymbol{a}_{jt}$ is a vector of parameters $(j = 1, ..., J, t = 1, ..., T)$, and $\boldsymbol{\omega}_i = [\omega_{1i}, ..., \omega_{Ji}]', i = 1, ..., n$ is a vector of random effects or error terms uncorrelated with the $\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}$ variables. We will call this the **General Mundlak Formulation** to model firm effects.[4]

To reduce the number of $\boldsymbol{a}_{jt}$ parameters, we consider a parsimonious formulation and assume that firm effects depend on the mean (over time) values of the predetermined variables, i.e.,

$$\tau_{ji} = \overline{\boldsymbol{Z}}'_i \boldsymbol{\alpha}_j + \omega_{ji}, j = 1, ..., J, i = 1, ..., n, \tag{15}$$

where $\overline{\boldsymbol{Z}}_i = T^{-1}\sum_{t=1}^{T} \boldsymbol{Z}_{it}, i = 1, ..., n$. We call this **Means-based Mundlak Formulation**.

[3]As $u_{it} \geq 0$, technical efficiency $e^{-u_{it}} \in (0,1]$. We wish to thank an anonymous reviewer for raising the point that exogenous variables can, in fact, appear in this equation as well as in the production function and no modifications are needed in terms of estimation. Another point raised by the reviewer concerns the use of the Cobb-Douglas production function. The advantage of the Cobb-Douglas is that it is self-dual in the sense that the cost function corresponding to a Cobb-Douglas production function is also Cobb-Douglas and its parameters can be recovered. Clearly, alternative functional forms can be clearly estimated along with the system of their first-order conditions using all the new devices proposed in this paper (Mundlak generalized specification, parametric specification of inefficiency, neural networks in first-order conditions to allow for semi—parametric formulation, etc.) For other functional forms like the translog, we do not have the cost function in closed form and, therefore, we cannot recover allocative inefficiency and its cost. This is a major research problem and we are currently working on it hoping to provide results in the not too distant future.

[4]Note that in Chamberlain (1980), Mundlak (1978), the firm effects are expressed as functions of the covariates. Since in our case covariates are inputs (x) which are endogenous, we specify firm-effects as functions of exogenous variables.

$$\boldsymbol{\tau}_i = \sum_{t=1}^{T} (\mathbf{I}_J \otimes \boldsymbol{Z}'_{it})\boldsymbol{a}_t + \boldsymbol{\omega}_i, i = 1, ..., n, \tag{16}$$

where $\mathbf{I}_J$ denotes the $J \times J$ identity matrix, and $\boldsymbol{a}_t = [\boldsymbol{a}'_{1t}, \boldsymbol{a}'_{2t}, ..., \boldsymbol{a}'_{Jt}]', t = 1, ..., T$ is a parameter vector whose dimensionality is Jd_Z. In total, there are TJd_Z such parameters. For fixed T and large n, this approach is desirable as, usually, we have a large number of firms and relatively few time periods. Denoting $\boldsymbol{Z}_i = [\mathbf{I}_J \otimes \boldsymbol{Z}'_{it}, t = 1, ..., T]$, a $J \times d_Z JT$ matrix, and $\boldsymbol{a} = [\boldsymbol{a}'_1, \boldsymbol{a}'_2, ..., \boldsymbol{a}_T]'$, a $d_Z JT \times 1$ vector, we have:

$$\boldsymbol{\tau}_i = \boldsymbol{Z}_i \boldsymbol{a} + \boldsymbol{\omega}_i, i = 1, ..., n. \tag{17}$$

Therefore, the system consists of (12), (13), (17) along with the following distributional assumption:

$$\boldsymbol{\omega}_i \sim iid\, \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}), i = 1, ..., n. \tag{18}$$

The advantages of the Mundlak device are many. *First*, it allows explicitly for correlation between a vector of firm effects and the input variables via the predetermined variables. *Second*, when $\boldsymbol{a} = \mathbf{0}_{d_Z JT}$, the model reduces to a standard random effects model where the random effects are the $\boldsymbol{\omega}_i$s. *Third*, when $\boldsymbol{\Omega} = \mathbf{O}_{J\times J}$, the model reduces to a fixed effects model in the sense that the $\boldsymbol{\tau}_i$s are parametric functions of $\boldsymbol{Z}_i$, i.e. $\boldsymbol{\tau}_i = \boldsymbol{Z}_i \boldsymbol{a}$. Thus this fixed effects model is not the same as the one in which the $\boldsymbol{\tau_i}$ are parameters.

Denoting $\boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{\gamma}']' \in \Re^d$ where d stands for the dimensionality of the parameter vector, the likelihood function is:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{\Omega}; \mathcal{D}) = \textstyle\int_{\Re^J} (2\pi)^{-nT/2} |\boldsymbol{\Sigma}|^{-nT/2} \prod_{i=1}^{n} \prod_{t=1}^{T} \exp\left\{-\tfrac{1}{2}\left[\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\theta}) - \boldsymbol{\tau}_i\right]' \boldsymbol{\Sigma}^{-1} \left[\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\theta}) - \boldsymbol{\tau}_i\right]\right\} \cdot \\ (2\pi)^{-n/2} |\boldsymbol{\Omega}|^{-n/2} \textstyle\prod_{i=1}^{n} \exp\left\{-\tfrac{1}{2}\left(\boldsymbol{\tau}_i - \boldsymbol{Z}_i \boldsymbol{a}\right)' \boldsymbol{\Omega}^{-1} \left(\boldsymbol{\tau}_i - \boldsymbol{Z}_i \boldsymbol{a}\right)\right\} \mathrm{d}\boldsymbol{\tau}_i.\end{aligned} \tag{19}$$

where $\mathcal{D} = \{y_{it}, X_{it}, \boldsymbol{Z}_{it}, i = 1, ..., n, t = 1, ..., T\}$ denotes the entire set of data. In a Bayesian framework, suppose we have a prior of the form

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{\Omega}) \propto p(\boldsymbol{\Sigma}, \boldsymbol{\Omega}) \cdot p(\boldsymbol{\theta}, \boldsymbol{a} | \boldsymbol{\Sigma}, \boldsymbol{\Omega}) \propto |\boldsymbol{\Sigma}|^{-(J+1)/2} \cdot |\boldsymbol{\Omega}|^{-(J+1)/2} \cdot p(\boldsymbol{\theta}, \boldsymbol{a}) \propto |\boldsymbol{\Sigma}|^{-(J+1)/2} \cdot |\boldsymbol{\Omega}|^{-(J+1)/2} \cdot p(\boldsymbol{a}). \tag{20}$$

The prior for the covariance matrices, $p(\boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(J+1)/2}$ and $p(\boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{-(J+1)/2}$ is the standard "non-informative" prior (Zellner, 1971, p. 24, formula 8.9) and $p(\boldsymbol{\theta}, \boldsymbol{a}) \propto$ const., viz. a flat prior for all parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$. For the Mundlak parameters, $\boldsymbol{a}$, we assume a prior of the form:

$$\boldsymbol{a} \sim \mathcal{N}_{TJd_Z}\left(\mathbf{0}, h^2 \mathbf{I}\right), \tag{21}$$

where the parameter $h = 10^3$ so that the prior is relatively uninformative. We use this "regularization prior" because the number of parameters in $\boldsymbol{a}$ is potentially large. The prior is rather diffuse for all practical purposes. For example, in our application, which is the same as in Lai and Kumbhakar (2019) we have $n = 192$, $T = 12$, $J = 3$ and $d_Z = 4$ if relative input prices, a time trend, and output are used as the predetermined variables. This leads to $TJd_Z = 108$ parameters. Although this number of parameters may seem exceedingly large, in fact, (14) is much more general compared to (15). Note that (15) involves only $d_Z J = 9$ parameters so the question of model fit versus parsimony remains to be resolved on empirical grounds.

We write the prior for all the parameters as:

$$p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{\Omega}) \propto |\boldsymbol{\Sigma}|^{-(J+1)/2} |\boldsymbol{\Omega}|^{-(J+1)/2} \cdot \exp\left(-\tfrac{1}{2h^2}\boldsymbol{a}'\boldsymbol{a}\right). \tag{22}$$

Using Bayes' theorem, the posterior is:

$$\begin{gathered} p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{\Omega}|\mathcal{D}) \propto \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{\Omega}; \mathcal{D}) p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{\Omega}) \propto \\ \int_{\Re^J} |\boldsymbol{\Sigma}|^{-nT/2-(J+1)/2} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^n \sum_{t=1}^T \left[\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\theta}) - \boldsymbol{\tau}_i\right]' \boldsymbol{\Sigma}^{-1} \left[\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\theta}) - \boldsymbol{\tau}_i\right]\right\} \times \\ |\boldsymbol{\Omega}|^{-n/2-(J+1)/2} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^n \left(\boldsymbol{\tau}_i - \boldsymbol{Z}_i\boldsymbol{a}\right)' \boldsymbol{\Omega}^{-1} \left(\boldsymbol{\tau}_i - \boldsymbol{Z}_i\boldsymbol{a}\right) - \tfrac{1}{2h^2}\boldsymbol{a}'\boldsymbol{a}\right\} \mathrm{d}\boldsymbol{\tau}_i. \end{gathered} \tag{23}$$

Alternatively, we can consider the posterior augmented with the firm effects to avoid the multivariate integration:

$$\begin{gathered} p(\boldsymbol{\theta}, \boldsymbol{\Sigma}, \boldsymbol{a}, \boldsymbol{\Omega}, \boldsymbol{\tau}|\mathcal{D}) \propto |\boldsymbol{\Sigma}|^{-nT/2-(J+1)/2} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^n \sum_{t=1}^T \left[\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\theta}) - \boldsymbol{\tau}_i\right]' \boldsymbol{\Sigma}^{-1} \left[\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\theta}) - \boldsymbol{\tau}_i\right]\right\} \times \\ |\boldsymbol{\Omega}|^{-n/2-(J+1)/2} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^n \left(\boldsymbol{\tau}_i - \boldsymbol{Z}_i\boldsymbol{a}\right)' \boldsymbol{\Omega}^{-1} \left(\boldsymbol{\tau}_i - \boldsymbol{Z}_i\boldsymbol{a}\right) - \tfrac{1}{2h^2}\boldsymbol{a}'\boldsymbol{a}\right\}, \end{gathered} \tag{24}$$

where $\boldsymbol{\tau} = [\boldsymbol{\tau}'_1, ..., \boldsymbol{\tau}'_n]'$. The posterior in (24) can be analyzed using numerical techniques organized around Markov Chain Monte Carlo (MCMC), and especially the Gibbs sampler with data augmentation. The Gibbs sampler provides a sequence of draws $\left\{\boldsymbol{\theta}^{(s)}, \boldsymbol{\Sigma}^{(s)}, \boldsymbol{a}^{(s)}, \boldsymbol{\Omega}^{(s)}, \boldsymbol{\tau}^{(s)}, s = 1, ..., S\right\}$ (in general, not $i.i.d$) which converges in distribution to the posterior non-normalized density provided by (24).

Once the parameters are estimated, we can compute costs of technical and allocative inefficiency using the same procedure discussed earlier for Model 0. Since technical inefficiency is fully parameterized, it can be estimated from (11) without using the Jondrow et al. (1982) procedure. However, there is no change in the computation of CAI.

4 Model II: Lognormal distribution for technical inefficiency

In Model II we replace the parametric assumption on u_{it} in (11) with a distributional assumption, viz.,

$$\ln u_{it} \sim \mathcal{N}(\boldsymbol{Z}'_{it}\boldsymbol{\delta}, \sigma_u^2), i = 1, ..., n, t = 1, ..., T, \tag{25}$$

i.e., u_{it} is assumed to follow a lognornal distribution where $\boldsymbol{\delta}$ is the vector of parameters associated with the exogenous variables $\boldsymbol{Z}_{it}$. The modeling of τ_i is performed along the lines of Mundlak discussed before.

This distribution more convenient relative to the truncated normal specification both in terms of estimation and interpretation of the $\boldsymbol{\delta}$ parameters. In the truncated normal model the $\boldsymbol{Z_{it}}$ variables appear in a complicated way in the mean function of u_{it} and it is cumbersome to derive marginal effects of the $\boldsymbol{Z_{it}}$ variables or interpret the $\boldsymbol{\delta}$ parameters. In (25) the interpretation is straightforward. For example, if Z_{kit} is in log, $\partial E[\ln(u_{it})]/\partial Z_{kit} = \delta_k$, that is the δ_k parameters are elasticities of the environmental variables on mean inefficiency.

From the distributional assumption in (25) we have:

$$p(u_{it}|\boldsymbol{Z}_{it}, \boldsymbol{\delta}, \sigma_u^2) = \frac{1}{u_{it}} \left(2\pi\sigma_u^2\right)^{-1/2} \exp\left\{-\tfrac{1}{2\sigma_u^2}\left(\ln u_{it} - \boldsymbol{Z}_{it}'\boldsymbol{\delta}\right)^2\right\}, i = 1, ..., n, t = 1, ..., T. \quad (26)$$

Since marginal likelihoods and Bayes factors can be computed, we can compare u being a random versus deterministic. Comparing (7) and (22) is important as, in principle, it allows to determine whether technical inefficiency is random. In turn, this allows to compare different approaches to efficiency as in (7) and (22).

Technical details on estimation of this model are presented in the Technical Appendix. Once the parameters are estimated, we can use the procedure under Model 0 to estimate technical and allocative inefficiency and costs therefrom.

5 Model III: Generalization of Allocative Inefficiency

Following Schmidt and Lovell (1979), Lai and Kumbhakar (2019) modeled allocative inefficiency as deviations from the *exact* fulfillment of the FOCs of cost minimization. This departure is captured by a random variable which might have a non-zero mean (for each input pair) in a cross-sectional model (as in Schmidt and Lovell (1979)) and a firm-specific mean for each input pair in a panel data (as in Lai and Kumbhakar (2019)). Note that the deviations from the exact FOCs is likely to depend on input prices (Farrell (1957)). This can be easily seen from the Farrell type figure explaining costs of technical and allocative inefficiency as in Figure 4.1 in Kumbhakar and Lovell (2000), p. 160 and Figure 8.2 in Kumbhakar, Wang and Horncastle (2015). However, this important feature is not incorporated in any of the stochastic frontier model that estimated allocative inefficiency and cost therefrom. To allow allocative inefficiency to depend on input prices and possibly on some environmental variables ($\boldsymbol{ZA}_{it}$), which might be different from $\boldsymbol{Z}_{it}$ – the ones that are used to explain technical inefficiency, we write the FOC in (3) in a more general form as follows:

$$\ln x_{1it} - \ln x_{jit} = \ln(w_{jit}/w_{1it}) + (\ln\beta_1 - \ln\beta_j) + f_j\left(\{\ln(w_{jit}/w_{1it}), j = 2, ..., J\}, \boldsymbol{ZA}_{it}\right) + \tau_{ji} + v_{jit}, j = 2, ..., J, \quad (27)$$

where $f_j(.,.)$ is a function of log relative input prices and $\boldsymbol{ZA}_{it}$. Note that $\boldsymbol{ZA}_{it}$ might be different from $\boldsymbol{Z}_{it}$, for at least some elements if not all. Suppose $\boldsymbol{w}_{it} = [\ln(w_{2it}/w_{1it}), ..., \ln(w_{Jit}/w_{1it})]'$ is the $(J-1) \times 1$ vector of log relative input prices so that

$$\ln x_{1it} - \ln x_{jit} = \ln(w_{jit}/w_{1it}) + (\ln\beta_1 - \ln\beta_j) + f_j\left(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it}\right) + \tau_{ji} + v_{jit}, j = 2, ..., J. \quad (28)$$

The functions $f_j\left(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it}\right), j = 2, ..., J$ allow for deviations of the FOCs from those of cost minimization in a systematic way via input prices as noted in Farrell (1957). Of course, τ_{ji}, v_{jit} also represent deviations from the FOCs.

The purpose of introducing additional features into allocative inefficiency via $f_j(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it}), j = 2, ..., J$, is to relate deviations from cost minimization explicitly to input prices and some other ZA variables. The τ_{ji} components might be related to unobserved input- and firm-specific factors such as managerial skill in allocating different inputs. Finally, the v_{jit} component is purely random and is outside the control of firms. Thus we allow for deviations in the form of functions $f_j(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it}), j = 2, ..., J$, in order to examine whether difference in input prices can explain allocative inefficiency beyond those that can be attributed to random errors or time-invariant managerial skills. If this is the case, then we expect $f_j(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it}) = 0 \, \forall j = 2, ..., J$, empirically. If we denote $\zeta^e_{jit} = f_j(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it}) + \tau_{ji} + v_{jit}$, then $\zeta^e_{jit} = \zeta_{jit} + f_j(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it})$. That is, allocative inefficiency in Lai-Kumbhakar becomes a special case of our extended model. If $f_j(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it}) = 0 \, \forall j = 2, ..., J$ then there is no systematic deviation explained by exogenous variables other than, possibly, τ_{ji} and the error terms v_{jit}. Using ζ^e_{jit} as allocative inefficiency for the input pair $(j, 1)$, we can define CAI in our extended model as before, i.e., $CAI^e_{it} = (E^e_{it} - \ln r)$ where $E^e_{it} = \frac{1}{r}\sum_{j=2}^J \beta_j \zeta^e_{jit} + \ln\left(\beta_1 + \sum_{j=2}^J \beta_j \exp(-\zeta^e_{jit})\right)$ and $r = \sum_{j=1}^J \beta_j$.

An important feature of the extension in Model III is that we do not assume any functional form of $f_j(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it})$. We model this function semi-parametrically using **artificial neural networks** (**ANN**):

$$f_j(\boldsymbol{w}_{it}, \boldsymbol{ZA}_{it}) \equiv f_j(\boldsymbol{W}_{it}) = \sum_{g=1}^{G_j} \delta_{1jg} \varphi\left(\boldsymbol{W}'_{it}\boldsymbol{\delta}_{jg}\right), j = 2, ..., J, \tag{29}$$

where $\boldsymbol{W}_{it} \equiv [\boldsymbol{w}'_{it}, \boldsymbol{ZA}'_{it}]'$, δ_{1j} and $\boldsymbol{\delta}_{jg}$ are parameters, and $\varphi(.)$ is an activation function, for which we assume:

$$\varphi(\xi) = \frac{1}{1 + e^{-\xi}}, \, \xi \in \Re. \tag{30}$$

For identification purposes, we assume $\delta_{1j1} < \delta_{1j2} < ... < \delta_{1jG}$. It is well known that an ANN can approximate any functional form provided the number of nodes G_j increases. For simplicity we assume $G_j = G \in \{1, 2, ...\}$ and we select the value of G using a data-based procedure (viz. Bayes factors). The introduction of (29) does not introduce any new statistical problems other than we have to specify a prior for $\boldsymbol{\delta} = [\delta_{1jg}, \boldsymbol{\delta}'_{2g}, ..., \boldsymbol{\delta}'_{Jg}, g = 1, ..., G]'$. Specifically, we assume the same prior as in (21), viz.:

$$\boldsymbol{\delta} \sim \mathcal{N}\left(\mathbf{0}, h^2\mathbf{I}\right), \tag{31}$$

where $h = 10^3$ which can be dominated easily by the data. There are different ways in which (29) can be zero. One possibility is that $\delta_{1jg} = 0 \, \forall j, g$. Another possibility is $\boldsymbol{\delta}_{jg} = \mathbf{0} \, \forall j, g$ irrespective of the values of δ_{1jg}.

We do not favor one specification over the others *a priori* and prefer to perform Bayesian model comparison in the system of production function, FOC from cost minimization, with and without (29). Instead, we examine whether (29) provides a better fit to the data relative to the "cost" of having more parameters. We call this the **ANN specification**. Estimation of the models is presented in the Appendix.

Based on the specifications of Models I-III, we end up with several different models which are described in Table 1. Model 1a (Model 1b) use the parametric specification of inefficiency, the full Mundlak specification of $\boldsymbol{\tau}$ and without

Table 1: Models

	Mundlak full specification for $\boldsymbol{\tau}$ as in (14)	Mundlak means- specification for $\boldsymbol{\tau}$ as in (15)
Without ANN for the FOC of cost minimization		
Parametric inefficiency as in (11)	Model 1a	Model 3a
Lognormal as in (25)	Model 2a	Model 4a
Including ANN for the FOC of cost minimization as in (29)		
Parametric inefficiency as in (11)	Model 1b	Model 3b
Lognormal inefficiency as in (25)	Model 2b	Model 4b

(with) the deterministic semi-parametric functions (ANN) explaining allocative inefficiency. Similarly, Model 3a (Model 3b) use the parametric specification of inefficiency, the restricted Mundlak specification of $\boldsymbol{\tau}$ and without (with) the semi-parametric functions (ANN) explaining allocative inefficiency. Model 2a (Model 2b) use the lognormal specification of inefficiency, the full Mundlak specification of $\boldsymbol{\tau}$ and without (with) the semi-parametric functions (ANN) explaining allocative inefficiency. Finally, Model 4a (Model 4b) use the lognormal specification of inefficiency, the restricted Mundlak specification of $\boldsymbol{\tau}$ and without (with) the semi-parametric functions (ANN) explaining allocative inefficiency. The difference between type "a" and type "b" models is in the inclusion of semi-parametric functions explaining allocative inefficiency in terms of the environmental variables.

6 Empirical application

We use the same data as in Lai and Kumbhakar (2019) which was also used by Rungsuriyawiboon and Stefanou (2008). We have panel data on $n = 82$ U.S. electric power generation plants during $1986 - 1997$ ($T = 12$). The three inputs are labor and maintenance, fuel, and capital. Output is net steam electric power generation in megawatt-hours, defined as power produced using fossil-fuel fired boilers to generate steam for turbine generators. We added a time trend variable the production function to capture technical change. In the absence of environmental variables in the data, we use input prices, output, and time trend as the Z variables. The ZA variables are those in Z without the input prices. MCMC is implemented using 150,000 draws discarding the first 50,000 to mitigate possible start up effects.

All models are defined in Table 1. Therefore, the models are as follows:

- Model 1a: Parametric inefficiency and full (generalized) Mundlak device for fixed effects.
- Model 2a: Lognormal inefficiency and full (generalized) Mundlak device for fixed effects.[5]
- Model 1b: ANN semi-parametric component in firm effects, parametric inefficiency and full (generalized) Mundlak device for fixed effects.

[5]Mundlak (1978) specified the fixed firm effects as a functions of the temporal averages of the regressors. In eq (14) we consider the firm effects (fixed and/or random) as functions of some Z variables observed over the entire time periods plus unobserved errors. We call it generalized Mundlak formulation because each of the firm effects is assumed to depend on a set of Z variables which are not the mean values of the regressors. In the simpler version of it (eq 15), the firm effects are assumed to be related to (time) mean of the Z variables. These formulations provide generalization of both fixed effects and random effects. Note that even if the firm effects are random, we allow correlations between the firm effects and the regressors (X variables) because X and Z variables are correlated.

- Model 2b: ANN semi-parametric component in firm effects, Lognormal inefficiency and full (generalized) Mundlak device for fixed effects.
- Model 3a: Parametric inefficiency and restricted Mundlak device for fixed effects. Model 4a: Lognormal inefficiency and restricted Mundlak device for fixed effects.
- Model 3b: ANN semi-parametric component in firm effects, Parametric inefficiency and restricted Mundlak device for fixed effects.
- Model 4a: Lognormal inefficiency and restricted Mundlak device for fixed effects.
- Model 4b: ANN semi-parametric component in firm effects, Lognormal inefficiency and restricted Mundlak device for fixed effects.

Therefore the models differ in several respects: The nature of firm effects, the nature of inefficiency and whether we should allow for a semi-parametric ANN formulation in the FOCs of cost minimization.

Since we are interested in extensive model comparisons, the Bayesian framework is the appropriate vehicle for it. The Bayes factor in favor of a model (say I) against another model (say II) is $BF_{I:II} = \frac{p_I(\mathcal{D})}{p_{II}(\mathcal{D})}$, if $BF_{I:II} > 1$ where $\mathcal{D}$ denotes the data and $p_\iota(\mathcal{D}) = \int p_\iota(\mathcal{D}, \boldsymbol{\lambda})\text{d}\,\boldsymbol{\lambda} = \int \mathcal{L}_\iota(\boldsymbol{\lambda}; \mathcal{D})p_\iota(\boldsymbol{\lambda})\text{d}\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ represents all parameters and latent variables (like firm effects and inefficiency) in model $\iota \in \{I, II\}$. Therefore $p_\iota(\mathcal{D})$ represents the integrating constant of the posterior distribution, that is: $p(\boldsymbol{\lambda}|\mathcal{D}) = \frac{\mathcal{L}_\iota(\boldsymbol{\lambda};\mathcal{D})p_\iota(\boldsymbol{\lambda})}{\int \mathcal{L}_\iota(\boldsymbol{\lambda};\mathcal{D})p_\iota(\boldsymbol{\lambda})d\boldsymbol{\lambda}} = \frac{\mathcal{L}_\iota(\boldsymbol{\lambda};\mathcal{D})p_\iota(\boldsymbol{\lambda})}{p_\iota(\mathcal{D})}$, and $\iota \in \{I, II\}$. More generally, the posterior probability of a model $\iota \in \mathcal{I} \subset \mathbb{Z}$ (where $\mathbb{Z}$ denotes the set of integers) is defined as: $P_\iota(\mathcal{D}) = \frac{p_\iota(\mathcal{D})}{\sum_{i\in\mathcal{I}} p_i(\mathcal{D})}$, $\iota \in \mathcal{I}$. Clearly, $\sum_{\iota\in\mathcal{I}} P_\iota(\mathcal{D}) = 1$.

To approximate the marginal or integrated likelihood (also known as “evidence”), we note that

$$p_\iota(\mathcal{D}) = \frac{\mathcal{L}_\iota(\boldsymbol{\lambda}; \mathcal{D})p_\iota(\boldsymbol{\lambda})}{p(\boldsymbol{\lambda}|\mathcal{D})} \; \forall \iota \in \mathcal{I} \subset \mathbb{Z}, \tag{32}$$

which holds for all values of $\boldsymbol{\lambda}$. Therefore, we can use any point, say the posterior mean $\overline{\boldsymbol{\lambda}} = E(\boldsymbol{\lambda}|\mathcal{D})$ to obtain:

$$p_\iota(\mathcal{D}) = \frac{\mathcal{L}_\iota(\overline{\boldsymbol{\lambda}}; \mathcal{D})p_\iota(\overline{\boldsymbol{\lambda}})}{p(\overline{\boldsymbol{\lambda}}|\mathcal{D})} \; \forall \iota \in \mathcal{I} \subset \mathbb{Z}. \tag{33}$$

The only unknown quantity in this expression is the denominator, $p(\overline{\boldsymbol{\lambda}}|\mathcal{D})$. However, if we use a Laplace approximation we obtain: $p(\overline{\boldsymbol{\lambda}}|\mathcal{D}) \simeq (2\pi)^{-dim(\boldsymbol{\lambda})/2}|\boldsymbol{V}|^{-1/2}$, where $dim(.)$ denotes the dimensionality of a vector, and $\boldsymbol{V}$ is the covariance matrix of $\boldsymbol{\lambda}$ which we can approximate closely using the available MCMC draws, viz. $\boldsymbol{V} = S^{-1}\sum_{s=1}^{S}\left(\boldsymbol{\lambda}^{(s)} - \overline{\boldsymbol{\lambda}}\right)\left(\boldsymbol{\lambda}^{(s)} - \overline{\boldsymbol{\lambda}}\right)'$, and $\overline{\boldsymbol{\lambda}} = S^{-1}\sum_{s=1}^{S}\boldsymbol{\lambda}^{(s)}$. In turn, the log of (33) can be approximated easily based on information from MCMC.

Bayes factors and posterior model probabilities strongly favored an ANN with $G = 1$. For example, the Bayes factor in favor of $G = 1$ and against $G = 2$ turned out to be 125.32, 144.12, 212.14 and, nearly, 512 for models 1b through 4b, respectively. Therefore, in Table 2, for these specifications we proceed under the assumption that $G = 1$. For G_0 (the number of nodes of ANN in the production function) we have $G = 1$ for the ANN model. Performing model comparison only for these models, and $G \in \{1, ..., 5\}$, the posterior probability of ANN ($G = 3$) was nearly 97% so, in the interest

Table 2: Bayes factors and posterior model probabilities for different models

	Bayes factor	posterior model probability
	No ANN in FOCs	
1a	1.000	0.0317
2a	0.032	0.001
3a	0.085	0.0027
4a	0.014	0.0004
	ANN in FOCs	
1b	**24.812**	**0.7875**
2b	3.717	0.118
3b	1.130	0.0359
4b	0.718	0.0228

Notes: The Bayes factor in favor of a model (say I) and against another model (say II) is $BF_{I:II} = \frac{p_I(\mathcal{D})}{p_{II}(\mathcal{D})}$, where $\mathcal{D}$ denotes the data and $p_\iota(\mathcal{D}) = \int p_\iota(\mathcal{D}, \boldsymbol{\lambda}) d\, \boldsymbol{\lambda} = \int \mathcal{L}_\iota(\boldsymbol{\lambda}; \mathcal{D}) p_\iota(\boldsymbol{\lambda}) d\boldsymbol{\lambda}$, where $\boldsymbol{\lambda}$ represents all parameters and latent variables (like individual effects and inefficiency) in model $\iota \in \{I, II\}$. Therefore $p_\iota(\mathcal{D})$ represents the integrating constant of the posterior distribution, that is: $p(\boldsymbol{\lambda}|\mathcal{D}) = \frac{\mathcal{L}_\iota(\boldsymbol{\lambda};\mathcal{D}) p_\iota(\boldsymbol{\lambda})}{\int \mathcal{L}_\iota(\boldsymbol{\lambda};\mathcal{D}) p_\iota(\boldsymbol{\lambda}) d\boldsymbol{\lambda}} = \frac{\mathcal{L}_\iota(\boldsymbol{\lambda};\mathcal{D}) p_\iota(\boldsymbol{\lambda})}{p_\iota(\mathcal{D})}$, and $\iota \in \{I, II\}$. The posterior probability of a model $\iota \in \mathcal{I} \subset \mathbb{Z}$ is defined as: $P_\iota(\mathcal{D}) = \frac{p_\iota(\mathcal{D})}{\sum_{i \in \mathcal{I}} p_i(\mathcal{D})}$, $\forall \iota \in \mathcal{I} \subset \mathbb{Z}$.

Table 3: Bayes factors for other model features

Model	against fixed effects	against Means-based Mundlak in (15)	against lognormal specification as in (26)
1a	1.000	342.81	77.62
2a	3,122.55	685.32	101.43
3a	2,891.43	302.12	71.10
4a	1,440.12	244.21	34.71
1b	717.12	5,522.32	5,832.33
2b	34,021.12	33,072.67	31,515
3b	**35,617.87**	3,516.13	7,771.10
4b	32,381,18	3,155,71	3,212.32

Notes: The Table provides Bayes factors against fixed effects, specification (15) and specification (26). We normalize so that the Bayes factor is equal to one for Model 1a. We use $G = G_0 = 1$.

of brevity (viz. not having to discuss and average over a large number of different models per cases a, b, c, and d, we proceed on the assumption that the values of G and G_0 are as given above. From Table 2, it turns out that the best model is 1b whose Bayes factor relative to model 1a is 24.812 and its posterior model probability is 0.7875. This implies that we do need ANN in FOCs, along with parametric inefficiency as specified in (11) along with a full Mundlak specification for the fixed effects. As full Mundlak specifications are rarely, if ever, used this means that they should be given some attention in future empirical work. The fact that the posterior probability of model 1b is much greater than the other posterior model probabilities in Table 2, implies that the assumptions of lognormality for inefficiecy and the exclusion of ANNs from the FOCs are quite detrimental in terms of fit of the alternative models.

To compare parametric versus lognormal (stochastic) technical inefficiency, viz. models “a” versus “b”, that is models 1a, 2a versus models 1b, 2b, Bayes factors (or posterior odds when the prior odds ratio is equal to one) are reported in Table 2. The Bayes factor in favor of stochastic inefficiency (model 2b versus 1b) is $\frac{34{,}021.12}{3{,}122.55} = 10.90$, approximately. The evidence against simple fixed effects and against the Means-based Mundlak specification is overwhelming in all models in Table 2. Models 1a-4a show moderate to strong evidence against of stochastic inefficiency, while models 1b-4b show overwhelming evidence in favor of it. Therefore, there is ample evidence against the parametric inefficiency model.

The posterior model probability for Model 1b is approximately 79% so this model clearly stands out in the light of the data. However, all other models have jointly a posterior model probability of nearly 21% so we cannot ignore them and there is scope for Bayesian Model Averaging (BMA)[6]. In practical terms, this implies that the cost minimization FOC need to allow for semi-parametric formulations including relative prices and environmental variables.

In Table 3, we take up a "grand comparison" between all models (1a to 4a and 1b to 4b). The comparison is performed in three aspects: First, in terms of the nature of fixed effects. Second, in terms of the Mundlak specification and, third, against the lognormal distributional specification. The Bayes factors reported in Table 3 are in favor of a particular "column" model (1a-4a and 1b-4b) and against "row" specifications such as fixed effects, the Mundlak means-specification, and lognormality. From the evidence in Table 3, the model with the highest Bayes factor (relative to model 1a) is model 3b (the Bayes factor is nearly 35,617.

Posterior means and posterior standard deviations of the most important functions of interest are summarized in Table 4. In terms of technical efficiency, the results vary widely. For example, technical efficiency in model 2b has a posterior mean of 89.13% (posterior standard deviation 0.69%). Allocative inefficiency is 4.58% (posterior standard deviation 3.27%) for the labor-fuel input pair, and 4.06% (3.61%) for the capital-fuel pair for model 1b, whereas they are 2.22% (1.54%) and 2.91% (2.93%), respectively for model 2b. The large posterior standard deviations suggest that there is considerable heterogeneity across different firms. The posterior standard deviations are much smaller for other models (for the "b" type). In terms of the cost of allocative inefficiency, model 2b suggests that it is 14.97% (posterior standard deviation 2.21%). All other models imply much higher costs -close to 24% for models of the "a" type. For model 1b, cost of allocative inefficiency is much lower. The different implications are, clearly, attributed to the different assumptions and specifications underlying each model. The same results are true for fixed effects in the production function or the cost-minimizing fist-order conditions. Relative to model 2b, costs of allocative inefficiency are much higher in Lai and Kumbhakar (2019, Table 4) and they average almost 30% with standard deviation 12.35%. Our estimates are about half of the mean cost of allocative inefficient in Lai and Kumbhakar (2019). In their study, technical efficiency averages 75.12% (standard deviation 15.03%). In our case, it averages 89.13% (posterior standard deviation 0.69%) and, therefore, almost 15 percentage higher compared to Lai and Kumbhakar (2019) with a much small posterior standard deviation.

In terms of fixed effects in the first-order cost-minimizing equations, model 2b suggests that the posterior mean is 0.338 (posterior standard deviation 0.0232) in the labor-fuel equation, and 0.0171 (posterior standard deviation 0.0113) in the capital-fuel equation, suggesting that more labor and capital is used than necessary. In contrast, these estimates are 0.012 (standard deviation 0.449) and -3.76 (standard deviation 0.47) in Lai and Kumbhakar (2019, Table 4) suggesting a slight overuse of labor and a quite large under-utilization of capital, relative to fuel. Part of the reason is, of course, that in our model 2b, we use an ANN formulation which is quite flexible and also that the full Mundlak specification for the treatment of firm effects is much more general than standard fixed effects formulations. Marginal posterior distributions of fixed effects with and without ANN are reported in Figure 1 (left and right panel, respectively). In Figure 2 we report posterior distributions of technical efficiency for all models.

Evidence on the posterior densities of allocative inefficiency in Figure 3 suggests that, after using Bayesian Model

[6] Of course all other models, *individually* have negligible posterior probability so they are strongly rejected by the data, although we do take them into account in BMA, as collectively they account for 26% in terms of posterior model probability

Table 4: Empirical results

	1a	2a	3a	4a	1b	2b	3b	4b
	Technical Efficiency							
Posterior means	0.9085	0.8351	0.7911	0.7433	0.9484	0.8913	0.8069	0.8263
Posterior s.d.	(0.0137)	(0.0109)	(0.0219)	(0.0084)	(0.0116)	(0.0069)	(0.0068)	(0.0125)
	Fixed Effects (Production Function)							
Posterior means	0	0	0	0	0	0	0	0
Posterior s.d.	(0.0181)	(0.0218)	(0.0082)	(0.0144)	(0.0342)	(0.0229)	(0.0055)	(0.0180)
	Fixed Effects (Labor-fuel Equation)							
Posterior means	0.0388	0.021	0.0095	0.0228	0.0438	0.0338	0.0096	0.0232
Posterior s.d.	(0.0369)	(0.0206)	(0.0047)	(0.0192)	(0.0345)	(0.0232)	(0.0067)	(0.0145)
	Fixed Effects (Capital-fuel Equation)							
Posterior means	0.024	0.0427	0.0091	0.0307	0.0222	0.0171	0.0063	0.0219
Posterior s.d.	(0.0146)	(0.0313)	(0.0065)	(0.0223)	(0.0179)	(0.0113)	(0.0077)	(0.0159)
	Allocative Inefficiency (Labor-fuel Equation)							
Posterior means	0.0447	0.0316	0.0074	0.0331	0.0458	0.022	0.0081	0.0198
Posterior s.d.	(0.0298)	(0.0217)	(0.0054)	(0.0325)	(0.0327)	(0.0154)	(0.0065)	(0.0224)
	Allocative Inefficiency (Capital-fuel Equation)							
Posterior means	0.0528	0.0335	0.0092	0.0267	0.0406	0.0291	0.01	0.0265
Posterior s.d.	(0.0351)	(0.0265)	(0.0087)	(0.0206)	(0.0361)	(0.0293)	(0.0068)	(0.0236)
	Cost of Allocative Inefficiency							
Posterior means	0.2428	0.2345	0.2389	0.2407	0.0958	0.1497	0.0663	0.1149
Posterior s.d.	(0.0316)	(0.0335)	(0.043)	(0.0286)	(0.0249)	(0.0221)	(0.0151)	(0.0200)

Notes: Reported in the Table are posterior means with posterior standard deviations (s.d) in parentheses. For definitions of the models, see Table 1. Returns to scale are defined as $RTS = \sum_{j=1}^{J} \left\{ \beta_j + \frac{\partial f_0(\boldsymbol{x}_{it}, \boldsymbol{z}_{it})}{\partial \ln x_{jit}} \right\}$. Output elasticities are defined as $\epsilon_{jit} = \beta_j + \frac{\partial f_0(\boldsymbol{x}_{it}, \boldsymbol{z}_{it})}{\partial \ln x_{jit}}$, $j = 1, ..., J$. We impose the restriction that elasticities should be non-negative via rejection sampling in MCMC.

Averaging, allocative inefficiencies are fairly large for both equations, ranging from –5% to over 15%.

As posterior model probabilities are available, we can perform Bayesian Model Averaging (BMA) as follows. For any function of interest $\psi(\mathcal{D}, \boldsymbol{\lambda})$ which takes the form $\psi_\iota(\mathcal{D}, \boldsymbol{\lambda})$, for a particular model $\iota \in \mathcal{I} \subset \mathbb{Z}$, the BMA version is:

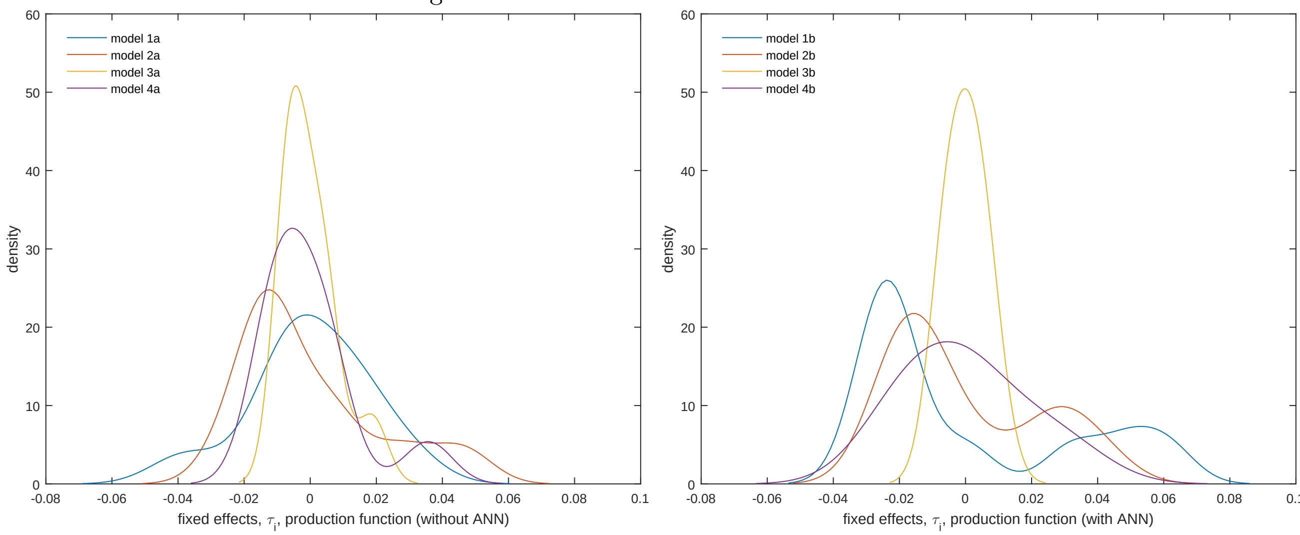
$$\psi_{BMA}(\mathcal{D}, \boldsymbol{\lambda}) = \sum_{\iota \in \mathcal{I}} P_\iota(\mathcal{D}) \psi_\iota(\mathcal{D}, \boldsymbol{\lambda}). \tag{34}$$

Functions of interest include firm effects, technical and allocative inefficiency, etc.

In Kumbhakar and Lai (2019, Table 3) the coefficients of log inputs in the Cobb-Douglas production function are 0.2219, 0.7795, and 0.0083. The last coefficient has standard error 0.102 suggesting that it is not significant. The means of elasticities are, respectively 0.35, 0.60 and 0.13 with small posterior standard deviations (0.014, 0.010, and 0.023, respectively). These estimates are, clearly, different compared to Lai and Kumbhakar (2019). As we use different models in this study this is, more or less, expected.

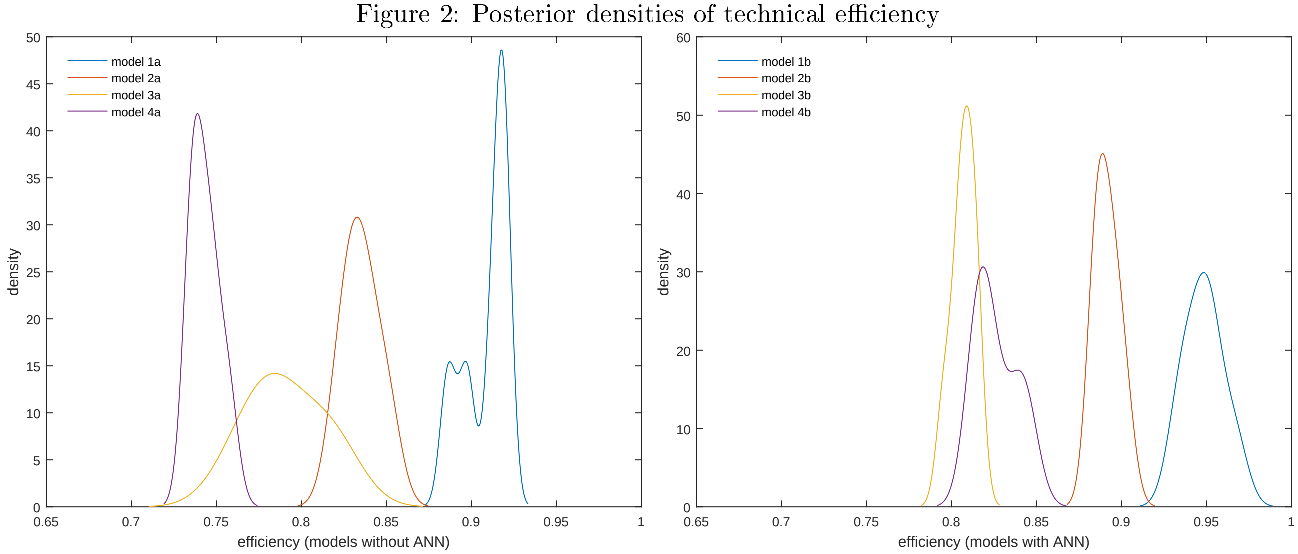
Clearly, there is an issue in that estimates of various functions of interest differ across models. The issue is less serious than it seems at first sight as, finally, we can either select the best model (using the Bayes factors reported in Tables 2 and 3) or perform BMA as we do in Figure 4. We should mention that the concern of different results from different

Figure 1: Posterior densities of firm effects



Notes: The Figures report the posterior distributions of fixed effects for all variants of Models 1 and 2. For each MCMC draw we save the fixed effects and then we average across the sample and parameter draws to account for parameter uncertainty. Since a constant appears in the production function we normalize, without loss of generality, the sum of fixed effects to zero for each equation in the system of production function and the first-order conditions of cost minimization. No such normalization is necessary in the first-order conditions for cost minimization.

Figure 2: Posterior densities of technical efficiency



Notes: The Figures report the posterior distributions of technical efficiency for all variants of Models 1 and 2. Technical efficiency is defined as $r_{it} = e^{-u_{it}}$. For each MCMC replication we save the draws for u_{it}, and then we average across the sample and parameter draws to account for parameter uncertainty.

Figure 3: Posterior distributions of allocative efficiency

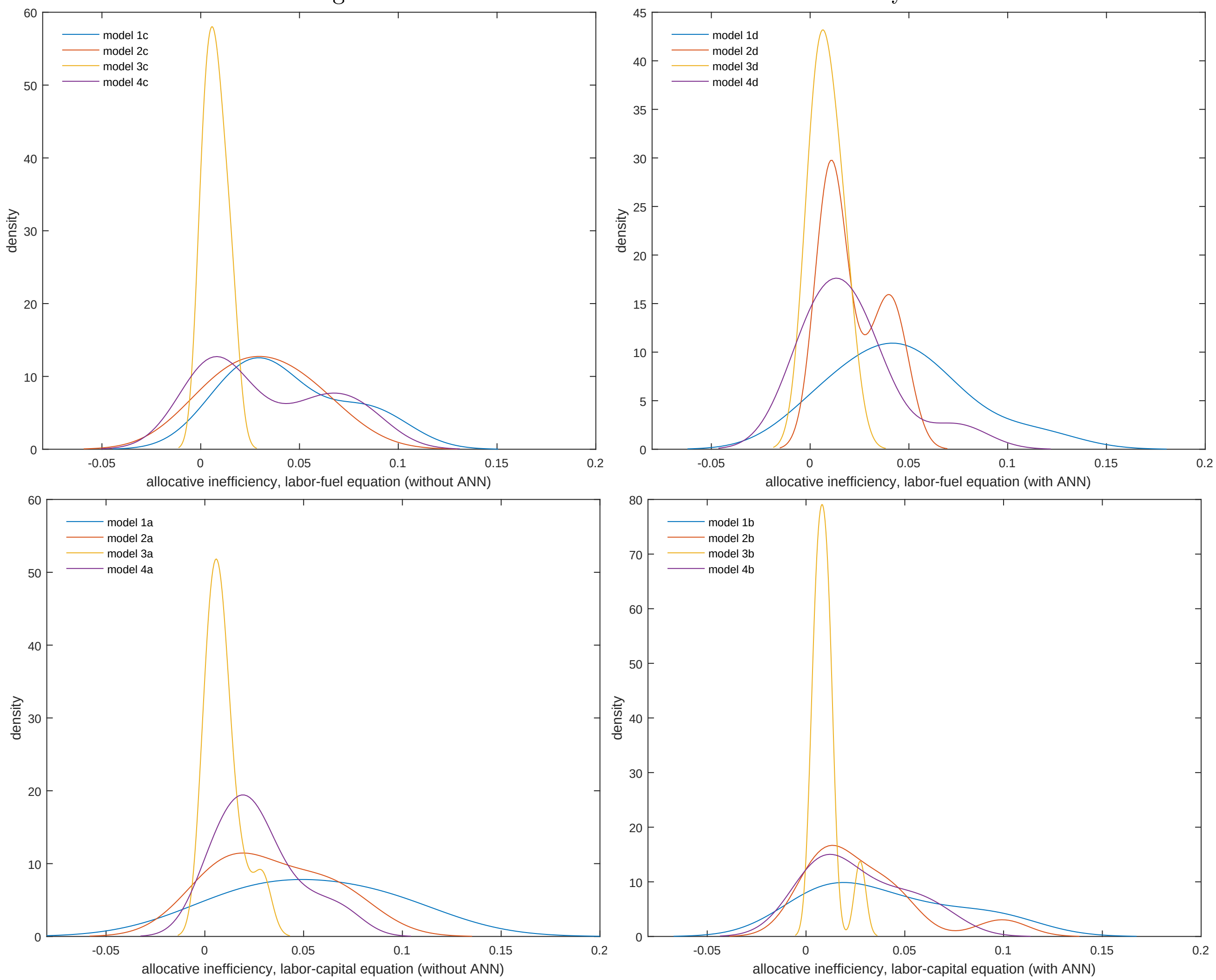
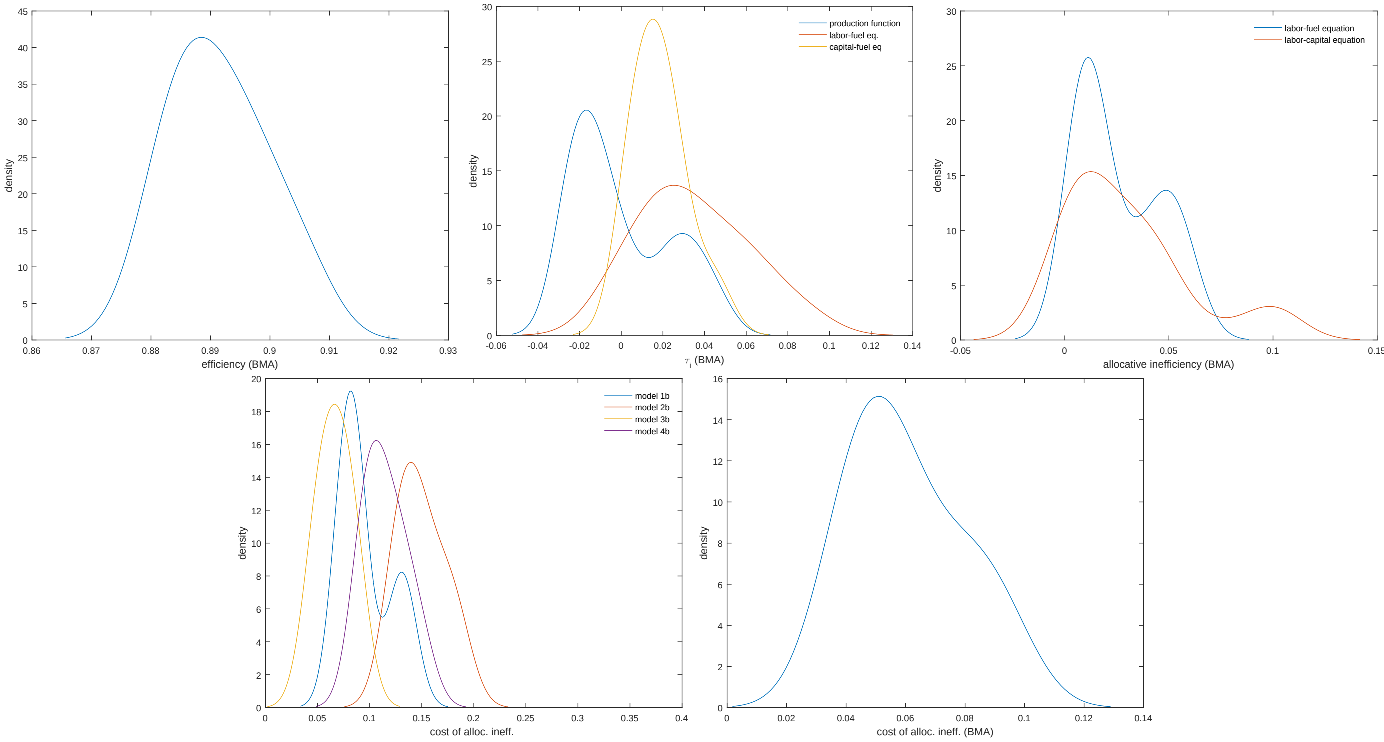


Figure 4: Bayesian Model Averaging (BMA) densities



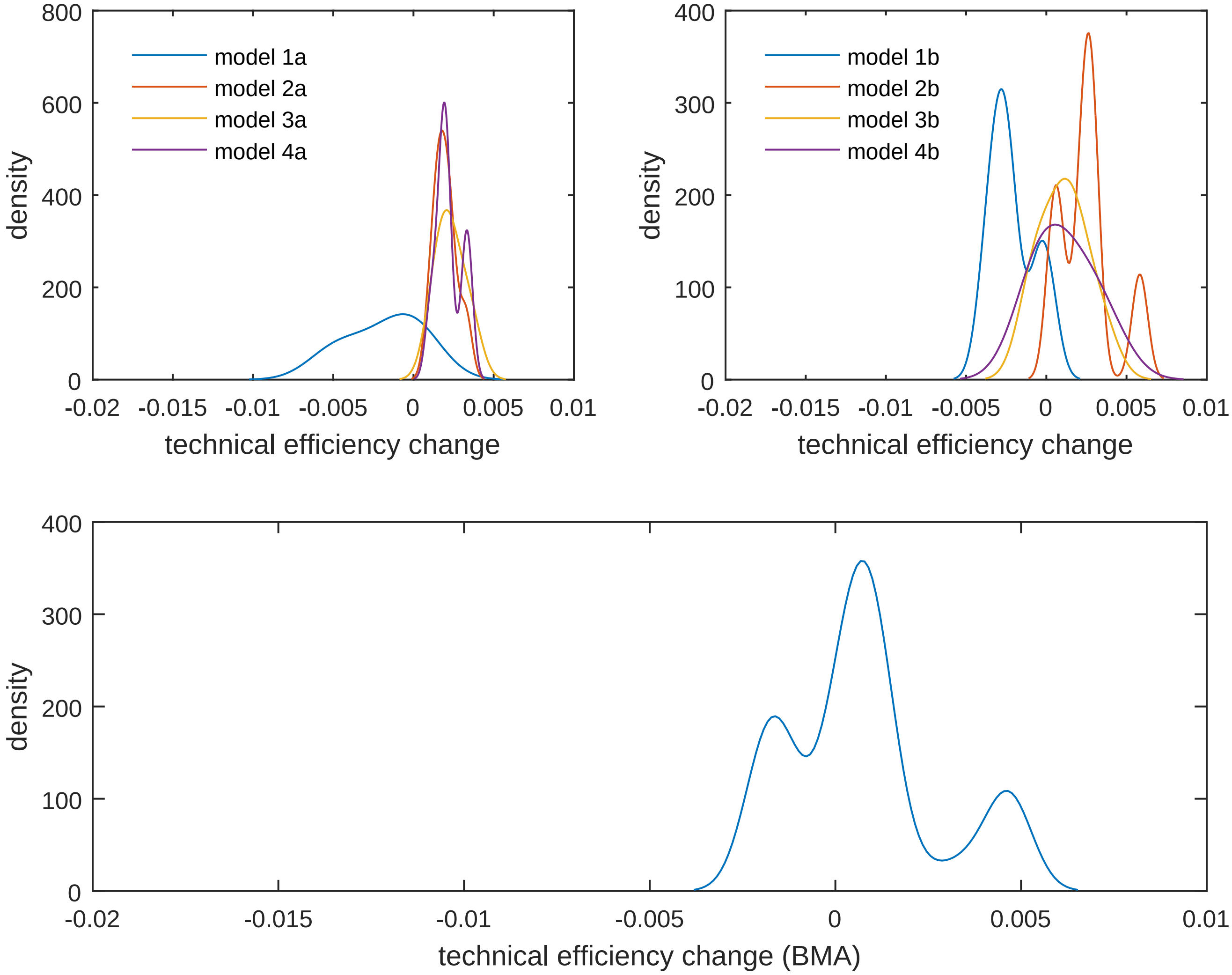
Notes: The Figures show the posterior distribution of technical efficiency, allocative efficiency, cost of allocative efficiency, and firm effects. BMA is based on posterior model probabilities reported in Table 2.

models is indeed valid but on the positive side, however, BMA takes into account not only parameter uncertainty but also uncertainty with respect to which model is most likely to be "true" or "closer to the truth" given the data. Of course, as different models involve different assumptions about many aspects (nature of firm effects, semi-parametric terms in FOCs, nature of inefficiency, etc.) they also tend to provide different results in terms of functions of interest. In a Bayesian inference framework, the most reasonable approach is to recognize the presence of model uncertainty and, therefore, use BMA as we do in Figure 4.

Posterior densities of technical efficiency change are reported in Figure 5. Models of the "a" variety imply little technical efficiency change (averaging 0.025). For models of the "b" type, efficiency change ranges from –0.5% to 0.5%, and the same is true. The BMA density reported in the lower panel of Figure 5 shows that efficiency change has been close to zero ranging from –0.5% to slightly over 0.5%. Therefore, despite the presence of considerable technical efficiency (close to 89% for model 2b in Table 4), it appears that management has done little to improve this score. In the BMA density of efficiency change, it is clear that it is bimodal with two modes near zero and 0.4%, respectively so there is some heterogeneity across firms in terms of improving efficiency, albeit by small amounts, in most models (see upper panel of Figure 5).

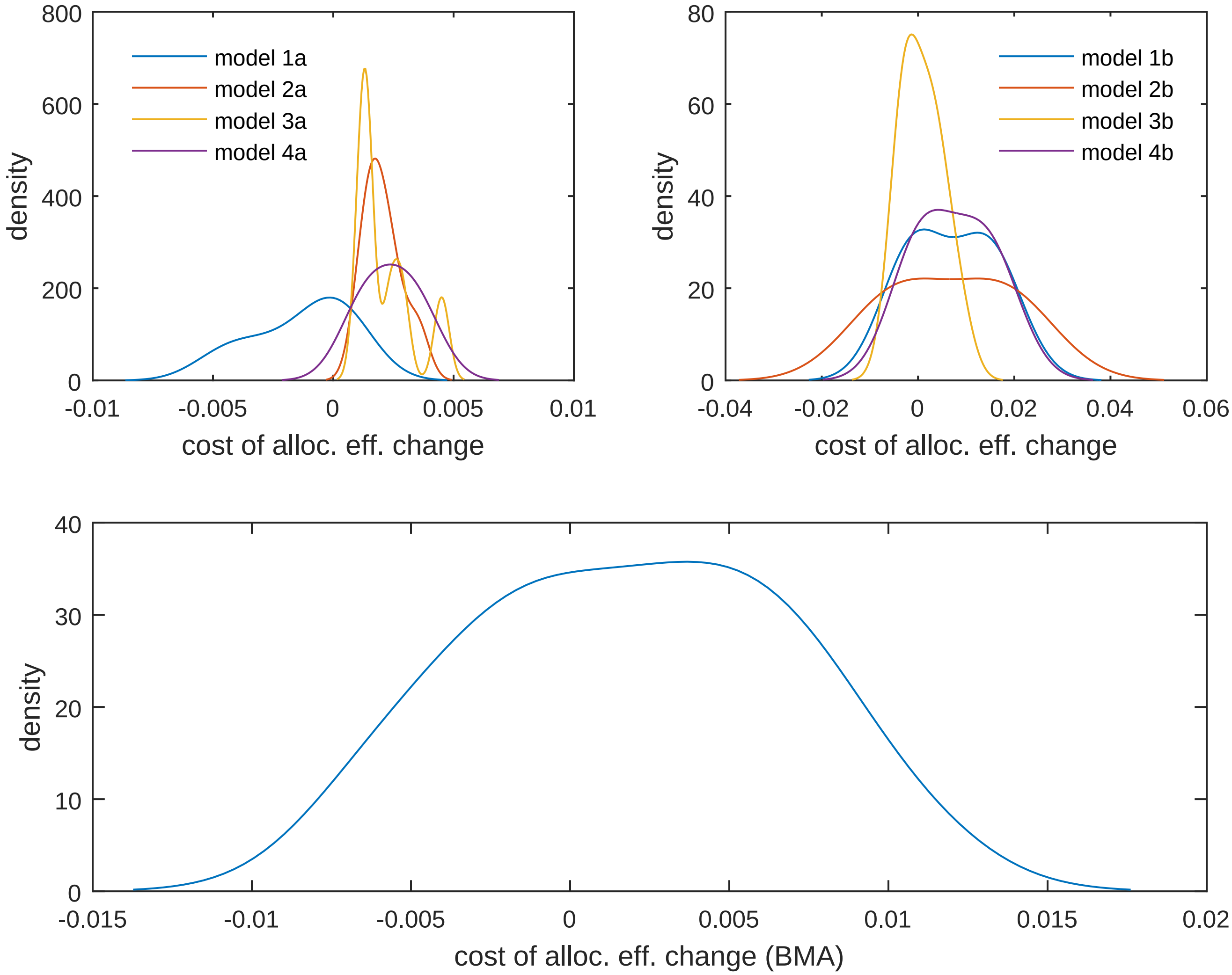
An important feature is change in cost of allocative efficiency provided in Figure 6. Although posterior densities differ widely across models for models of the "a" and "b" variety, this change in costs is rather small averaging, respectively, 0.5% and 1%. The BMA density is clearly bimodal with modes close to –0.002 and 0.005. The range of all these densities is confined to values close to zero suggesting that costs of allocative inefficiency change are rather small, despite the fact that allocative inefficiencies themselves are large (Figure 3). Specifically, under-utilization of capital does not seem to be

Figure 5: Posterior densities of technical efficiency change



Notes: The Figures show the posterior distribution of technical efficiency change, defined as $EC = \hat{u}_{it} - \hat{u}_{i,t-1}$ in (29), where $\hat{u}_{it}$ is estimated efficiency change. The measure is calculated using draws of u_{it} and then taking first differences. In turn, we average across all observations and MCMC draws to account for parameter uncertainty. BMA is based on posterior model probabilities reported in Table 2.

Figure 6: Posterior densities of cost of allocative inefficiency change



Notes: The Figures show the posterior distribution of the cost of allocative inefficiency change. In turn, we consider its first difference, $\triangle CAI_{it} = CAI_{it} - CAI_{i,t-1}$. In turn, we average across all observations and MCMC draws to account for parameter uncertainty. BMA is based on posterior model probabilities reported in Table 2.

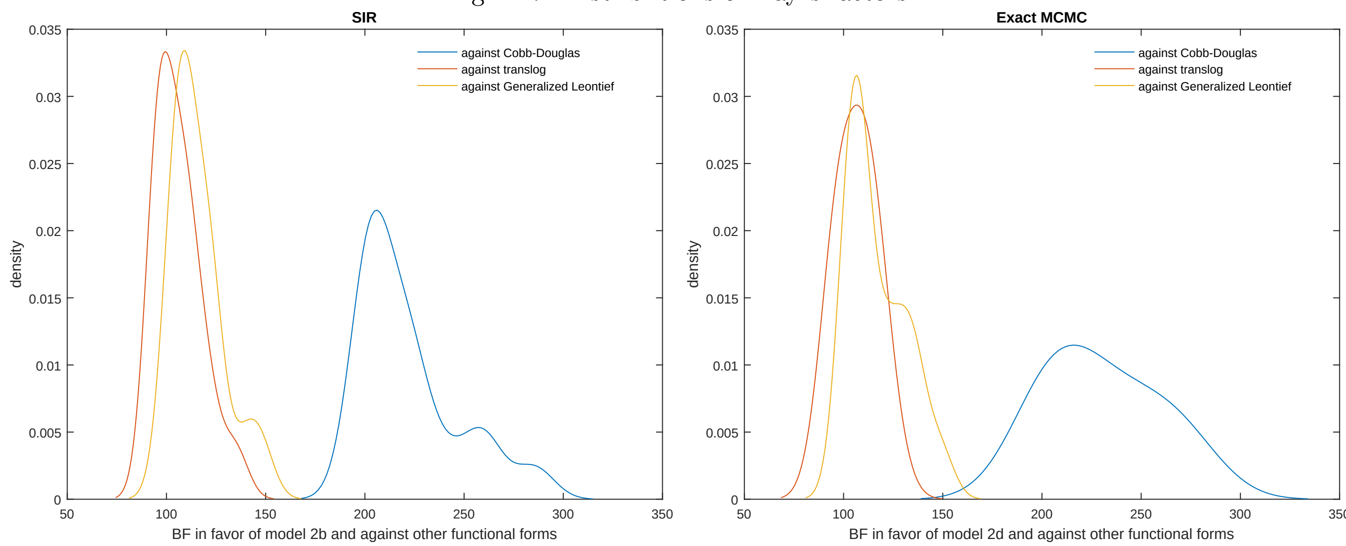
the case as in Lai and Kumbhakar (2019, Table 4).

For model comparison, the underlying production function can be more general than Cobb-Douglas. It is straightforward to estimate any production function along with the FOCs with allocative inefficiency embedded.[7] In (the left panel of) Figure 7, we report distributions of Bayes factors in favor of model 2b and against, Cobb-Douglas, translog and Generalized Leontief, all estimated along with the FOC for cost minimization and fixed effects ï¿œ la Lai and Kumbhakar (2019). Clearly, the new model performs better compared to the translog and generalized Leontief, meaning that, in the light of the data, the new model performs well in terms of fit and number of parameters. This is important, as the Cobb-Douglas production is not flexible.[8] Specifically, the posterior odds (with prior odds equal to 1) average over 50, but they extend from 30 to over 120. According to Kass and Raftery (1995) in order to have "strong" evidence against a model the Bayes factor should exceed 20. According to Jeffreys it should be greater than 31.6, which is far in the left tail

[7]The problem is in computing costs of allocative inefficiency which requires solving for input quantities with and without allocative inefficiency. This is non-trivial. See, for example, Kumbhakar and Wang (2006) who addressed this for a translog production function. No analytical expression for CAI can be derived for the translog case.

[8]In spite of this we used the extended CD specification (ANN) because it helped us in computation of costs of allocative inefficiency.

Figure 7: Distributions of Bayes factors



of our reported Bayes factors using either SIR or exact MCMC.

The distribution of these Bayes factors arises because we omit randomly b observations at a time where b is randomly selected from $\{1, 2, ..., 20\}$. We perform this for a total of 10,000 times. Posterior simulation is performed using sampling-importance-resampling (SIR, Rubin, 1987, 1988) to approximate the new posterior distribution. From the original MCMC sample (100,000 draws after burn-in) we consider 20,000 replications to implement SIR. In (the right panel of) Figure 7, we report the same distributions but this time we consider exact MCMC inference using 150,000 draws with a burn-in phase of 50,000. We do this in the interest of comparing SIR with exact MCMC. It turns out that SIR provides an accurate enough approximation, at least in terms of average Bayes factors and their spread across the sample.

Concluding remarks and further research

In this paper we addressed the issue of specification and estimation of technical inefficiency and costs therefrom using a system approach. In particular, we consider specifications to address the incidental parameters problem raised in Lai and Kumbhakar (2019) because of fixed firm effects in the production function as well as in the first-order conditions of cost minimization. To deal with the incidental parameters problem, we draw the insight of Mundlak (1978) and expressed firm effects as either functions some exogenous variables (full version) or the mean of the exogenous variables (restricted version). Extensive model comparison and Bayesian model averaging are implemented using Bayes factors, and associated posterior model probabilities.

In terms of future research, it might be important to examine the role of the Mundlak formulation in more general panel data models. The Mundlak formulation resolves the incidental parameters problem as it depends on a finite-dimensional vector of parameters, at least in its restricted version, while its more general version is expected to work better compared to alternatives. Intermediate cases between the full and restricted versions are worthwhile to examine, as they can yield significant reduction in the number of parameters. This is, however, a data - specific problem. The semi-parametric modification of the first-order conditions for cost minimization is quite important since it allows allocative inefficiency to depend on input prices in a flexible way, and it is in the line of Farrell (1957). However, use of all these extensions requires availability of some predetermined/exogenous variables (ideally variables other than input prices and output which are assumed to be exogenous in a cost minimizing framework).

Another important extension would be to examine the behavior of different estimators in the context of dynamic production systems, i.e., in the context of a dynamic panel data model with individual effects. Linear dynamic panel data models have been examined in the context of single equations using the Generalized Method of Moments (GMM) estimator. Systems, on the other hand, have received much less attention. Dynamic production models arise naturally in the context of intertemporal optimization models of the firm or in models with adjustment costs in the short-run. Another generalization of the models considered here is the presence of individual effects in the semi-parametric terms in the first-order conditions. This problem is significantly more difficult as additional individual effects are included in the nonlinear part of the model. Although there are numerically efficient techniques to deal with the problem (Greene, 2005a, b) they are still subject to the incidental parameters problem and an application of the Mundlak device is likely to be necessary. Endogeneity in stochastic frontier models has also been considered in Amsler, Prokhorov, and Schmidt (2016) where variants of 2SLS and LIML are considered. Such techniques may be profitably employed using the information provided by first-order conditions for cost minimization, at least in an "informal" way. This may help to provide different, and perhaps, simpler estimators to deal with both endogeneity and the incidental parameters problem.

Technical Appendix

To provide access to the augmented posterior in (24), we have the following posterior conditional distributions from which random drawings is obtained[9]:

$$\boldsymbol{\tau}_i|\cdot \sim \mathcal{N}(\hat{\boldsymbol{\tau}}_i,\ \boldsymbol{V}_i)\ \forall i = 1, ..., n, \tag{A.1}$$

where $\mathbf{F}_i = [\mathcal{F}(\boldsymbol{X}_{i1}, \boldsymbol{Z}_{i1}, \boldsymbol{\theta})', \ldots \mathcal{F}(\boldsymbol{X}_{iT}, \boldsymbol{Z}_{iT}, \boldsymbol{\theta})']'$, $\mathbf{W}_i = [\boldsymbol{Z}_1'\boldsymbol{a}, ..., \boldsymbol{Z}_T'\boldsymbol{a}]'$, $\mathcal{Y}_i = [\mathbf{F}_i', \mathbf{W}_i']'$,

$\hat{\boldsymbol{\tau}}_i = \left[\mathbf{I} \otimes \boldsymbol{\Sigma}^{-1} + h^2\mathbf{I}\right]^{-1} \left[(\mathbf{I} \otimes \boldsymbol{\Sigma}^{-1}) + h^2\mathbf{I}\right] \mathcal{Y}_i, i = 1, ..., n$, and $\boldsymbol{V}_i = \left[\mathbf{I} \otimes \boldsymbol{\Sigma}^{-1} + h^2\mathbf{I}\right]^{-1}\ \forall i = 1, ..., n.$

For covariance matrices, we have:

$$p(\boldsymbol{\Sigma}|\cdot) \propto |\boldsymbol{\Sigma}|^{-(nT+J+1)/2} \exp\left(-\tfrac{1}{2} tr \mathbb{A}_{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}\right), \tag{A.2}$$

$$p(\boldsymbol{\Omega}|\cdot) \propto |\boldsymbol{\Omega}|^{-(n+J+1)/2} \exp\left(-\tfrac{1}{2} tr \mathbb{A}_{\boldsymbol{\Omega}} \boldsymbol{\Omega}^{-1}\right), \tag{A.3}$$

where $\mathbb{A}_{\boldsymbol{\Sigma}} = \sum_{i=1}^{n}\sum_{t=1}^{T} \left(\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\theta}) - \boldsymbol{\tau}_i\right)\left(\boldsymbol{\mathcal{F}}(\boldsymbol{X}_{it}; \boldsymbol{Z}_{it}, \boldsymbol{\theta}) - \boldsymbol{\tau}_i\right)'$, and $\mathbb{A}_{\boldsymbol{\Omega}} = \sum_{i=1}^{n}\sum_{t=1}^{T} \left(\boldsymbol{\tau}_i - \boldsymbol{Z}_i\boldsymbol{a}\right)\left(\boldsymbol{\tau}_i - \boldsymbol{Z}_i\boldsymbol{a}\right)'$. These are in the form of an inverted Wishart distribution (Zellner, 1981, pp. 395–396, particularly B.53, and 8.15 in p. 227).

[9]The symbol "$|\cdot$" denotes conditioning on all other parameters and latent variables in the augmented posterior distribution.

For the parameters in the Mundlak specification we have:

$$\boldsymbol{a}|\cdot \sim \mathcal{N}\left(\hat{\boldsymbol{a}},\ \boldsymbol{V}_{\boldsymbol{a}}\right), \tag{A.4}$$

where $\hat{\boldsymbol{a}} = [\boldsymbol{Z}'(\mathbf{I}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{Z}+h^2\mathbf{I}]^{-1}\boldsymbol{Z}'(\mathbf{I}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{\tau}$, $\boldsymbol{V}_{\boldsymbol{a}} = [\boldsymbol{Z}'(\mathbf{I}\otimes\boldsymbol{\Omega}^{-1})\boldsymbol{Z}+h^2\mathbf{I}]^{-1}$, $\boldsymbol{Z} = diag(\boldsymbol{Z}_1, ..., \boldsymbol{Z}_n)$, and $\boldsymbol{\tau} = [\boldsymbol{\tau}'_1, ..., \boldsymbol{\tau}'_N]'$.

Drawing structural parameters of the production function and the inefficiency function, $\boldsymbol{\theta}$, is not straightforward as they enter nonlinearly in (12). In this case we use the Girolami and Calderhead (2011) Riemannian MCMC method which is based on first- and second-order derivatives of the log posterior. These derivatives are computed numerically.

When considering parameters in the lognormal distributional in (26), most steps are valid, except that $\boldsymbol{\theta}$ contains only the production function parameters, and we have to draw $u_{it}, i = 1, ..., n, t = 1, ..., T$, $\boldsymbol{\delta}$, as well as σ_u^2. Under a flat prior of the form: $p(\boldsymbol{\delta}, \sigma_u) \propto \sigma_u^{-1}$, we have the following posterior conditional distributions, viz.,

$$\boldsymbol{\delta}|\cdot \sim \mathcal{N}\left((\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\log\mathbf{u},\ \sigma_u^2(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\right), \tag{A.5}$$

where $\boldsymbol{Z} = [\boldsymbol{Z}_{it}, i = 1, .., n, t = 1, ..., T]$, $\mathbf{u} = [u_{it}, i = 1, ..., n, t = 1, ..., T]$,

$$\tfrac{(\log\mathbf{u}-\boldsymbol{Z}\boldsymbol{\delta})'(\log\mathbf{u}-\boldsymbol{Z}\boldsymbol{\delta})}{\sigma_u^2}|\cdot \sim \chi^2_{nT}. \tag{A.6}$$

Finally, to draw $\log\mathbf{u}$, we follow an optimal rejection technique. As candidate densities we use independent exponential densities whose parameters are selected so as to maximize the acceptance rate relative to its posterior conditional distributions. The required optimizations are always valid as the posterior conditional density of $\log u_{it}$ is log-concave.

For models including a semi-parametric modification, e.g., the models in Section 5 , we follow the same procedures with the Girolami and Calderhead (2011) Riemannian MCMC method for nonlinear parameters, including the parameters of the ANN functions.

The Girolami and Calderhead (2011) algorithm

Regarding the Girolami and Calderhead (2011, GC) algorithm to update draws for $\boldsymbol{\theta}$, we use local information about both the gradient and the Hessian of the log-posterior conditional of $\boldsymbol{\theta}$ at the existing draw. A Metropolis test is again used for accepting the candidate so generated but the GC algorithm moves considerably faster relative to our naive scheme previously described. It has been found that the GC algorithm performs much better than a standard Metropolis-Hastings algorithm, and autocorrelations are, more often than not, much smaller.

Suppose $\mathscr{L}(\boldsymbol{\theta}) = \log p(\boldsymbol{\theta}|\boldsymbol{X})$ is used to denote for simplicity the log posterior of $\boldsymbol{\theta}$. Moreover, define:

$$\boldsymbol{G}(\boldsymbol{\theta}) = \text{est.cov}\tfrac{\partial}{\partial\boldsymbol{\theta}}\log p(\boldsymbol{X}|\boldsymbol{\theta}), \tag{A.7}$$

which is the empirical counterpart of

$$\boldsymbol{G}_o(\boldsymbol{\theta}) = -\mathbb{E}_{\mathscr{Y}|\boldsymbol{\theta}}\tfrac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log p(\boldsymbol{X}|\boldsymbol{\theta}). \tag{A.8}$$

The Langevin diffusion is given by the stochastic differential equation:

$$d\boldsymbol{\theta}(t) = \tfrac{1}{2}\tilde{\nabla}_{\boldsymbol{\theta}}\mathscr{L}\{\boldsymbol{\theta}(t)\}\,dt + d\mathbf{B}(t), \tag{A.9}$$

where

$$\tilde{\nabla}_{\boldsymbol{\theta}}\mathscr{L}\{\boldsymbol{\theta}(t)\} = -\boldsymbol{G}^{-1}\{\boldsymbol{\theta}(t)\}\cdot\tilde{\nabla}_{\boldsymbol{\theta}}\mathscr{L}\{\boldsymbol{\theta}(t)\}, \tag{A.10}$$

is the so called "natural gradient" of the Riemann manifold generated by the log posterior. The elements of the Brownian motion are

$$\begin{aligned}\boldsymbol{G}^{-1}\{\boldsymbol{\theta}(t)\}\,d\mathbf{B}_i(t) =& |\boldsymbol{G}\{\boldsymbol{\theta}(t)\}|^{-1/2}\sum_{j=1}^{K_\beta}\tfrac{\partial}{\partial\boldsymbol{\theta}}\left[\boldsymbol{G}^{-1}\{\boldsymbol{\theta}(t)\}_{ij}\,|\boldsymbol{G}\{\boldsymbol{\theta}(t)\}|^{1/2}\right]dt \\ &+\left[\sqrt{\boldsymbol{G}\{\boldsymbol{\theta}(t)\}}d\boldsymbol{B}(t)\right]_i.\end{aligned} \tag{A.11}$$

The discrete form of the stochastic differential equation provides a proposal as follows:

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_i =& \boldsymbol{\theta}_i^o + \tfrac{\varepsilon^2}{2}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\nabla_{\boldsymbol{\theta}}\mathscr{L}(\boldsymbol{\theta}^o)\right\}_i - \varepsilon^2\sum_{j=1}^{K_\theta}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\frac{\partial\boldsymbol{G}(\boldsymbol{\theta}^o)}{\partial\boldsymbol{\theta}_j}\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\right\}_{ij} \\ &+\tfrac{\varepsilon^2}{2}\sum_{j=1}^{K_\theta}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\right\}_{ij}\operatorname{tr}\left\{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\frac{\partial\boldsymbol{G}(\boldsymbol{\theta}^o)}{\partial\boldsymbol{\theta}_j}\right\} + \left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)}\boldsymbol{\xi}^o\right\}_i \\ =& \boldsymbol{\mu}(\boldsymbol{\theta}^o,\varepsilon)_i + \left\{\varepsilon\sqrt{\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)}\boldsymbol{\xi}^o\right\}_i,\end{aligned} \tag{A.12}$$

where $\boldsymbol{\beta}^o$ is the current draw. We select ε so that, approximately, 20% of candidates are eventually accepted. The proposal density is:

$$\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^o \sim \mathcal{N}_{K_\theta}\left(\tilde{\boldsymbol{\theta}}, \varepsilon^2\boldsymbol{G}^{-1}(\boldsymbol{\theta}^o)\right). \tag{A.13}$$

Finally, convergence to the invariant distribution suggests using the Metropolis-Hastings probability:

$$\min\left\{1, \frac{p\left(\tilde{\boldsymbol{\theta}}|\cdot,\mathscr{Y}\right)q\left(\boldsymbol{\theta}^o|\tilde{\boldsymbol{\theta}}\right)}{p\left(\boldsymbol{\theta}^o|\cdot,\mathscr{Y}\right)q\left(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^o\right)}\right\}. \tag{A.14}$$

References

Amsler, C., A. Prokhorov, and P. Schmidt, (2016), "Endogeneity in stochastic frontier models", Journal of Econometrics, 190 (2), 280–288.

Arellano, M., and J. Hahn (2006): "Understanding Bias in Nonlinear Panel Models: Some Recent Developments," in Advances in Economics and Econometrics, Ninth World Congress, Chapter 12, pp. 381–409, ed. by R. Blundell, W.

Newey, and T. Persson, Cambridge University Press.

Dhaene, G., and K. Jochmans, (2015), “Split-panel Jackknife Estimation of Fixed-effect Models”, Review of Economic Studies, 82, 991–1030.

Farrell, M. J. (1957) The measurement of productive efficiency, J. R. Statistic. Soc. A, 120, 253-281.

Girolami, M., and B. Calderhead (2011), “Riemann manifold Langevin and Hamiltonian Monte Carlo methods”, Journal of the Royal Statistical Society Series B, 73 (2), 123–214.

Greene, W., (2005a), “Fixed and Random Effects in Stochastic Frontier Models”, Journal of Productivity Analysis, 23 (1), 7–32.

Greene, W., (2005b), “Reconsidering heterogeneity in panel data estimators of the stochastic frontier model”, Journal of Econometrics, 126 (2), 269–303.

Kass, R.E. and Raftery, A.E., (1995), “Bayes factors”, Journal of the American Statistical Association, 90 (430), 773–795.

Kumbhakar, S.C., and Wang, H.-J. (2006), Estimation of technical and allocative inefficiency: A primal system approach, Journal of Econometrics134 (2), 419–440.

Lai, H.-P., and S. C. Kumbhakar, (2019), “Technical and Allocative Efficiency in a Panel Stochastic Production Frontier System Model”, European Journal of Operational Research, 278(1), 255-265.

Lancaster, T., (2002), “Orthogonal Parameters and Panel Data”, Review of Economic Studies, 69, 647–666.

Liseo, B. (2005), “The Elimination of Nuisance Parameters”, in Handbook of Statistics Vol. 25. (D. K. Dey and C. R. Rao, eds.), 193–219. Elsevier, Amsterdam.

Marschak, J. and W. H. Andrews, Jr. (1944). “Random Simultaneous Equations and the Theory of Production”, Econometrica, 12 (3/4), 143–205.

Michaelides, P. G., A. T. Vouldis, and E. G. Tsionas, (2010), “Globally Flexible Functional Forms: The Neural Distance Function”, European Journal of Operational Research, 206 (2), 456–469.

Michaelides, P. G., E. G. Tsionas, A. T. Vouldis, and K. N. Konstantakis, (2015), “Global Approximation to Arbitrary Cost Functions: A Bayesian Approach with Application to US Banking”, European Journal of Operational Research, 241 (1), 148–160.

Mundlak, Y., (1961), Empirical Production Function Free of Management Bias, Journal of Farm Economics, 43, 44–56.

Mundlak, Y., (1978), On the Pooling of Time Series and Cross Section Data, Econometrica, 46, 69–85.

Neyman, J., and E. L. Scott, (1948), “Consistent Estimates Based on Partially Consistent Observations,” Econometrica, 16: 1–32.

Paul, S., S. Shankar (2018), “On Estimating Efficiency Effects in a Stochastic Frontier Model”, European Journal of Operational Research 271 (2), 769–774.

Rubin, D. B. (1987), “Comment on ’The calculation of posterior distributions by data augmentation’, by M. A. Tanner and W. H. Wong”, Journal of the American Statistical Association 82, 543–546.

Rubin, D. B. (1988), "Using the SIR Algorithm to Simulate Posterior Distributions", In Bayesian Statistics 3, ed. J. M. Bernardo, M. H. DeGroot , D. V. Lindley, and A. F. M. Smith, 395–402. Oxford: Oxford University Press.

Rungsuriyawiboon, S. and S. Stefanou, (2008), "The Dynamics of Efficiency and Productivity in U.S. Electric Utilities," Journal of Productivity Analysis, 30, 177–190.

Schmidt, P., (1988), "Estimation of a Fixed-Effect Cobb-Douglas System Using Panel Data", Journal of Econometrics, 37, 361–380.

Schmidt, P. and C. A. K. Lovell, (1979), "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers, Journal of Econometrics, 9: 343—366.

Tsionas, M. G., E. Mamatzakis (2019), "Further Results on Estimating Inefficiency Effects in Stochastic Frontier Models", European Journal of Operational Research, 275 (3), 1157–1164.

Zellner, A., "An Introduction to Bayesian Inference in Econometrics", Wiley, New York.