Constraints in models of production and cost via slack-based measures

October 7, 2021

Abstract

In this paper we propose the use of stochastic frontier models to impose theoretical regularity constraints (like monotonicity and concavity) on flexible functional forms. These constraints take the form of inequalities involving the data and the parameters of the model. We address a major concern when statistically endogenous variables are present in these inequalities. We present results with and without endogeneity in the inequality constraints. In the system case (e.g. cost-share equations) or more generally, in production function-first order conditions case, we detect an econometric problem which we solve successfully. We provide an empirical application to U.S. electric power generation plants during $1986 - 1997$, previously used by several authors.

1   Introduction

In many areas of applied economics operations research and applied economics , equations or systems of equations are often estimated that must satisfy certain theoretical constraints either globally or locally (that is at a specific point of approximation). Other times the equations must satisfy certain motononicity or other constraints at each observation. Although globally flexible functional forms that satisfy the constraints (globally), often practitioners use flexible forms that cannot satisfy these restrictions using only parametric restrictions. Instead, the constraints also involve the data. Suppose, for example, we have a translog production function of the form:

$$y = \beta_1 + \beta_2 x_1 + \beta_3 x_2 + \tfrac{1}{2}\beta_4 x_1^2 + \tfrac{1}{2}\beta_5 x_2^2 + \beta_6 x_1 x_2, \tag{1}$$

where y, x_1, x_2 denote the logs of output, capital and labor. The input elasticities must be positive so we have the constraints:

$$\begin{aligned} \beta_2 + \beta_4 x_1 + \beta_6 x_2 &\geqslant 0, \\ \beta_3 + \beta_6 x_1 + \beta_5 x_2 &\geqslant 0. \end{aligned} \tag{2}$$

For problems related to the use of the translog see, for example, O'Donnell (2018, p. 286, footnote 11). We expand on this point below.

Imposing constraints has given rise to a significant literature including O'Donnell, Rambaldi and Doran (2001) and O'Donnell and Coelli (2005). McCausland (2008) uses orthogonal polynomials while other authors proposed the use of neural networks (Vouldis, Michaelides and Tsionas, 2010). Diewert and Wales (1987) spell out the conditions that must be satisfied while Gallant and Golub (1984), and Lau (1978) represent earlier attempts. Ivaldi, Ladoux, Ossard and Simioni (1996) compare different functional forms in a concise way.

The dominant approach seems to be the one adopted by O'Donnell, Rambaldi and Doran (2001) and O'Donnell and Coelli (2005) who impose the constraints by assuming a different technology for each firm. Terrell (1996) is more in line with Geweke (1986) while Wolff, Heckelei, and Mittelhammer (2010) present new "local" approaches. In this paper we retain the original problem: Given an equation or a system of equations of the traditional form, is it possible to use standard Markov Chain Monte Carlo (MCMC) methods to perform Bayesian inference subject to many data– and parameter–specific inequality constraints? This problem is fundamentally different from Geweke (1991) since we have, more often than not, a number of constraints exceeding the number of equations, and the inequality constraints must be imposed exactly without regard for their posterior probability. Surely, the constraints must not be imposed at all data points when they depend on the data as, for example, a translog cost or production function would reduce to a Cobb-Douglas which is not second-order flexible.

It turns out that there are two approaches to solve the problem. In the first approach that we call "naive", all inequality constraints are converted to equalities using a surplus formulation. The surpluses are modeled within

the stochastic frontier approach. This is, essentially, the approach in Huang and Huang (2019) which has been proposed independently of this paper. In the second approach, we take up a major problem with this model; viz. the fact that *endogenous variables may appear in the inequality constraints*. To the best of our knowledge, the problem (and a potential solution) has not been realized before. As we mentioned, the first approach has been proposed independently by Huang and Huang (2019) where the surpluses are assumed to follow independent half-normal distributions. The problems with this approach are: *First*, the surpluses cannot be independent as violation of some constraints (say monotonicity) are known to have effects on other constraints (like curvature). *Second*, endogeneity is not taken into account although endogenous variables appear in the frontier equations that impose the constraints. In turn, specialized methods can be used. *Third*, it is well known that imposition of theoretical constraints (which is necessary in any functional form to account for the information provided by neoclassical production theory) affects estimates of technical inefficiency, so the *surpluses should be correlated with the one-sided error term in the production or cost function.*

Moreover, we apply the new techniques to the translog as it is used widely. If researchers want to use globally flexible functional forms that satisfy monotonicity and curvature via only parametric restrictions (e.g. Koop et al., 1997 and the Generalized McFadden functional form), they can certainly do it and the methods in this paper are not necessary. However, as the translog is quite popular, we use it here as our benchmark case. Another case of flexible functional forms where the constraints also depend on the data has been analyzed in Tsionas (2016).

2   Equations and constraints linear in the parameters

2.1   General

Let us consider the simplest case of an equation which is linear in the parameters:

$$y_i = g(z_i)' \beta + v_i - u_i, \quad i = 1, ..., n, \tag{3}$$

where z_i is an $m \times 1$ vector of basic covariates, β is a $k \times 1$ vector of parameters, $g : \mathbb{R}^m \to \mathbb{R}^k$ is a vector function, and u_i is a nonnegative error component representing technical inefficiency. The translog or polynomials in certain variables z_i, are leading examples. Apparently we can write:

$$y_i = x_i' \beta + v_i - u_i, \tag{4}$$

where $x_i = g(z_i)$ is $k \times 1$. Suppose we require the function $g(z)' \beta$ to be monotonic, and without loss of generality assume that all first order partial derivatives must be nonnegative, that is $Dg(z)\beta \geq \mathbf{0}_{(m \times 1)}$. Since no parameters are involved in g, it is clear that the (transposed) Jacobian $Dg(z) \equiv X_0$, which is $m \times k$, is a simple function of z and, therefore, a simple function of X.

Suppose that for the ith observation, we have $X_{i0} = \begin{bmatrix} 0 & 1 & 0 & x_{i1} & 0 & x_{i2} \\ 0 & 0 & 1 & 0 & x_{i2} & x_{i1} \end{bmatrix}$, $(i = 1, \ldots, n)$.

Bayesian inference in the linear model subject to a few inequality constraints has been considered by Geweke using both importance sampling (Geweke, 1986) and Gibbs sampling (Geweke, 1996). Here we follow a different approach, independently proposed by Huang and Huang (2019).

Suppose we write the constraints as follows:

$$\mathbf{0}_{(m \times 1)} = X'_{i0}\beta + v_{i0} - \widetilde{u}_{i0}, i = 1, ..., n, \tag{5}$$

where v_{i0} is an $m \times 1$ two-sided error term and $\widetilde{u}_{i0}$ is an $m \times 1$ nonnegative error component. Here, v_{i0} represents noise in the inequality constraints and inequality constraints themselves are represented by $\widetilde{u}_{i0}$. So, (5) imposes the constraints $X'_{i0}\beta \geq \mathbf{0}_{m \times 1}$ up to measurement errors (v_{i0}). Moreover, $\widetilde{u}_{i0}$ represents slacks in the constraints.

If we write the equations together we have:

$$
\begin{aligned}
y_i &= x'_i\beta + v_i - u_i, i = 1, ..., n, \\
\mathbf{0}_{(m \times 1)} &= X_{i0}\beta + v_{i0} - \widetilde{u}_{i0}, i = 1, ..., n.
\end{aligned}
\tag{6}
$$

We are now ready to specify our distributional assumptions on the error components:

$$\mathbf{v}_i = [v_i, v'_{i0}]' \sim N_{m+1}(\mathbf{0}, \Sigma), \ \mathbf{u}_i = [u_i, \widetilde{u}'_{i0}]' \sim N^+_{m+1}(\mathbf{0}, \Phi), \ i = 1, \ldots, n. \tag{7}$$

In this specification the two-sided and one-sided error components are correlated across equations. This specification, unlike the one in Huang and Huang (2019), has certain advantages. First, the error terms v_i and v_{i0} are allowed to be correlated, as the imposition of theoretical constraints affects parameter estimates in the first equation of (6). Second, the one-sided error terms u_i and $\widetilde{u}_{i0}$ are allowed to be correlated, since the extent of violation of certain constraints is very likely to affect the degree of violation of other constraints.

2.2 Simplified example

For ease of presentation and to establish the techniques, we assume $\Sigma = \begin{bmatrix} \sigma^2 & \\ & \omega^2 I_m \end{bmatrix}$, and $\Phi = \begin{bmatrix} 0 & \\ & \phi^2 I_m \end{bmatrix}$.
In this case, we have no technical inefficiency (viz. $u_i = 0$) and the scale parameter of surpluses, $\widetilde{u}_i$, have the same scale parameter (ϕ). Clearly, the scales could have been different (Huang and Huang, 2019) but most importantly Φ *should allow for correlations between the violations of different constraints*. Here, we focus on the simplest possible case to provide the background of the new approach. *For a fixed value of ω, which controls the degree of satisfaction*

of the constraints, the posterior distribution of this model is given by:

$$p\left(\beta,\sigma,\phi,u|y,X,\omega\right)\propto$$
$$\sigma^{-(n+1)}\phi^{-(n+1)}\exp\left[-\frac{(y-X\beta)'(y-X\beta)}{2\sigma^2}-\frac{(u-X_0\beta)'(u-X_0\beta)}{2\omega^2}-\frac{u'u}{2\phi^2}\right]\cdot p\left(\beta,\sigma,\phi\right),\tag{8}$$

where

$$X_{0\;(nm\times k)}=\begin{bmatrix}X_{10}\\X_{20}\\\dots\\X_{n0}\end{bmatrix}.\tag{9}$$

For the prior, we have:

$$p\left(\beta,\sigma,\phi\right)\propto p(\beta|\sigma,\phi)p(\sigma|\phi)p(\phi)\propto p(\beta)p(\sigma)p(\phi).\tag{10}$$

The only parameters of interest are the elements of β, and possibly σ, but not ϕ, which like u, are artificial parameters introduced to facilitate Bayesian inference. Alternatively, u represents prior parameters with a prior given by (5) in which ϕ and ω are parameters. The user may have high relative prior precision with respect to the degree of satisfaction of the constraints (so parameter ω can be set in advance) but in other respects the user does not particularly care about how far in the acceptable region are the parameters. Of course, if this is not the case it can always be controlled via choice of an informative prior of β.

Suppose for simplicity that $p\left(\beta,\sigma,\phi\right)\propto\sigma^{-1}\phi^{-1}$. Then we can use the Gibbs sampler based on the following standard full posterior conditional distributions:

$$\beta|\sigma,\phi,u,y,X\sim N_k\left(\bar{\beta},\bar{V}\right),\tag{11}$$

where $\bar{\beta}=\left(\omega^2X'X+\sigma^2X_0'X_0\right)^{-1}\left(\omega^2X'y+\sigma^2X_0'u\right)$, and $\bar{V}=\sigma^2\omega^2\left(\omega^2X'X+\sigma^2X'X\right)^{-1}$,

$$\frac{(y-X\beta)'(y-X\beta)}{\sigma^2}|\beta,\phi,u,y,X\sim\chi_n^2,\tag{12}$$

$$\frac{u'u}{\phi^2}|\sigma,u,y,X\sim\chi_n^2,\tag{13}$$

and finally:

$$u_i|\beta,\sigma,\phi,y,X\sim N_+\left(\frac{\phi^2}{\phi^2+\omega^2}X_0\beta,\;\frac{\phi^2\omega^2}{\phi^2+\omega^2}1_{nm}\right),\;i=1,\dots,n,\tag{14}$$

where 1_{nm} is a vector of ones whose dimensionality is $nm\times1$. For details on the derivations, see Tsionas (2000). Generating random draws from these full posterior conditional distributions is straightforward. The last distribution

is quite standard in Bayesian analysis of the normal – half normal stochastic model. So far we have assumed that ω can be set in advance. This is, of course, a possibility. If the user does not feel comfortable about this choice then one can use the following prior:

$$\frac{\overline{q}}{\omega^2} \sim \chi^2_{\overline{n}}, \tag{15}$$

where $\overline{n}, \overline{q} \geq 0$ are prior parameters. In this case, the posterior conditional of ω is:

$$\frac{\overline{q} + (u - X_0\beta)'(u - X_0\beta)}{\omega^2} | \beta, \sigma, \phi, u, y, X \sim \chi^2_{n+\overline{n}}. \tag{16}$$

The interpretation of (15) is that from a fictitious sample of size $\overline{n}$ the average ω^2 would be close to $\frac{(u-X_0\beta)'(u-X_0\beta)}{\overline{n}}$. There is nothing wrong with setting these parameters so that the prior is extremely informative if that is necessary; for example $\overline{n} = n$, and $\overline{q} = 0.001$. The interpretation, in this case, is that we need the errors v_0 to be quite small so that the constraints satisfy "exactly" the inequality constraints. Of course, this is related to Theil's mixed estimator.

2.3   An artificial example

Following Parmeter, Sun, Henderson and Kumbhakar (2009) we have the following model:

$$y = 10 + 3x + x^2 - 3x^3 + x^4 + v,$$

where the x's are generated as uniform in the interval $[0, 2.5]$, and $v \sim N\left(0, \ 0.1^2\right)$. We have $n = 100$ observations and the x's are sorted. The constraint is that we need this function to be non-decreasing that is $3 + 2x - 9x^2 + 4x^3 \geqslant 0$ at all observed data points.

In this case $x_i = \left[1, x, x^2, x^3, x^4\right]$, and $x_{0i} = \left[0, 1, 2x, -9x^2, 4x^3\right]'$. So the model is $y_i = x_i'\beta + v_i$ subject to the constraints $x_{0i}'\beta \geqslant 0$. For this example we set $\overline{n} = 50$ and $Q = 0.001$. Gibbs sampling is implemented using 15,000 passes the first 5,000 of which are discarded to mitigate possible start up effects. *The Least Squares (LS) fit has 22 violations of the constraints, while the Bayes fit has none.* The Bayes fit is computed as follows. For each draw $\beta^{(s)}$ we compute $f_i^{(s)} = X_1\beta^{(s)}$. After the burn–in period our estimate of the fit is $\hat{f}_i = S^{-1} \sum_{s=1}^{S} x_{i1}'\beta^{(s)}$ which is equivalent to $\hat{f}_i = x_i'\hat{\beta}$ where $\hat{\beta} = E(\beta|y, X)$ is the posterior expectation of β that can be approximated arbitrarily well (since it is simulation – consistent) by $\hat{\beta} = S^{-1} \sum_{s=1}^{S} \beta^{(s)}$. The same is true for the derivative. These computations involve only a standard Gibbs sampling scheme and the trivial computation of the posterior mean of β. In Figure 1 we present the original data points, the LS fit, the Bayes fit and the constrained least squares (LS) fit as it is more appropriate to compare restricted LS with the Bayesian estimates.

Even in this case one may argue that the selection of parameters results in satisfaction of the constraints but

Figure 1: LS and Bayes fit in an artificial example
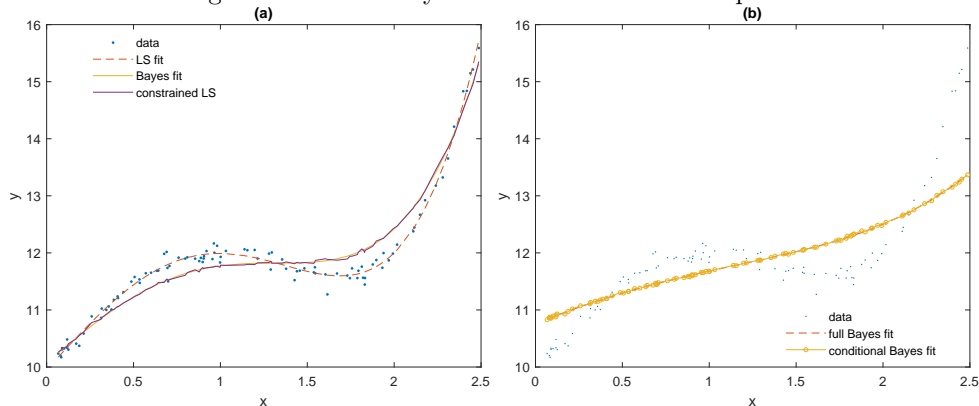
it does not guarantee the best fit. This criticism is not totally unfounded. For example, it would be possible to select these parameters so as to pass a line with positive slope through the points that would, apparently, satisfy all the constraints. Therefore, it may be necessary to device a mechanism by which ω is truly adjusted to the data so as to guarantee the best possible fit and also satisfying the constraints. Therefore, we search directly over the minimum value of ω that guarantees satisfaction of all constraints. It turns out that this value is 1.076 (when we use 15,000 Gibbs passes and the first 5,000 are discarded). It turned out that this problem requires a fine grid of values (of the order 10^{-4}) in the relevant range which is determined empirically by trial-and-error.

Despite the effort it does not appear that the results are any better compared to standard Bayes analysis when ω is assigned a prior. The results are presented in panel (b) of Figure 1. By "full Bayes fit" we mean the fit when ω is drawn from its full conditional distribution. By "conditional Bayes fit" we mean the fit when a detailed search is made over ω to determine the optimal value $\omega^* = 1.076$ that constitutes the value for which all constraints are satisfied along with fully Bayesian solutions for β, σ, ϕ, and u.

3   The general formulation

3.1   Posterior

In the general case of technical inefficiency and correlated error components we can write the posterior of the model in (6) as follows:

$$p(\beta, \Sigma, \Phi, \mathbf{u}|y, X) \propto$$
$$|\Sigma|^{-n/2}|\Phi|^{-n/2} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}(\psi_i - X_i\beta + \mathbf{u}_i)'\Sigma^{-1}(\psi_i - X_i\beta + \mathbf{u}_i) - \tfrac{1}{2}\sum_{i=1}^{n}\mathbf{u}_i'\Phi^{-1}\mathbf{u}_i\right\} \cdot \tag{17}$$
$$p(\beta, \Sigma, \Phi),$$

where $p(\beta, \Sigma, \Phi)$ denotes the prior, $\psi_i = \begin{bmatrix} y_i \\ \mathbf{0}_{(m \times 1)} \end{bmatrix}$, and $X_i = \begin{bmatrix} x_i' \\ X_{io}' \end{bmatrix}$. As we mentioned above, $\mathbf{u}_i = \begin{bmatrix} u_i \\ \widetilde{u}_i \quad {}_{(m \times 1)} \end{bmatrix}$, for all $i = 1, \ldots, n$. Our prior is a reference flat prior:[1]

$$p(\beta, \Sigma, \Phi) \propto |\Sigma|^{-(m+2)/2} |\Phi|^{-(m+2)/2}. \tag{18}$$

Therefore, the posterior becomes:

$$p(\beta, \Sigma, \Phi, \mathbf{u}|y, X) \propto \tag{19}$$
$$|\Sigma|^{-(n+m+2)/2} |\Phi|^{-(n+m+2)/2} \exp\left\{ -\tfrac{1}{2} \sum_{i=1}^{n} (\psi_i - X_i\beta + \mathbf{u}_i)' \Sigma^{-1} (\psi_i - X_i\beta + \mathbf{u}_i) - \tfrac{1}{2} \sum_{i=1}^{n} \mathbf{u}_i' \Phi^{-1} \mathbf{u}_i \right\}.$$

In this formulation, Φ is a general (positive-definite) covariance matrix which allows for the fact that violations of different constraints may be related in an unknown way. The posterior can be analyzed easily using MCMC as shown in part A.1 of the Technical Appendix.

3.2 Economic applications

3.2.1 Systems of equations

Many important systems of equations like the translog can be written in the form:[2]

$$
\begin{aligned}
y_{t1} &= x_t' \beta_{J(1)} + v_{t1}, \\
y_{t2} &= x_{t0}' \beta_{J(2)} + v_{t2}, \\
&\;\;\vdots \\
y_{tM} &= x_{t0}' \beta_{J(M)} + v_{tM},
\end{aligned}
\tag{20}
$$

where $\beta_{J(m)}$ represents a particular selection of elements of vector β with the indices in $J(m)$, $m = 1, ..., M$. We agree that $\beta_{J(1)} = \beta$ so that $J(1) = \{1, 2, ..., d\}$. Suppose, for example, that we have a translog cost function with

---

[1] For a positive definite matrix Δ whose dimension is $d \times d$ the reference prior is: $p(\Delta) \propto |\Delta|^{-(d+1)/2}$. In our case, we have $d = m + 1$.

[2] Parmeter, Sun, Henderson and Kumbhakar (2009) and Kumbhakar, Park, Simar and Tsionas (2007).

K inputs and N outputs[3]:

$$\ln C = \alpha_0 + \sum_{k=1}^{K} \alpha_k \ln p_k + \sum_{n=1}^{N} \beta_n \ln y_n + \tfrac{1}{2} \sum_{k=1}^{K} \sum_{l=1}^{K} \alpha_{kl} \ln p_k \ln p_l +$$
$$\tfrac{1}{2} \sum_{n=1}^{N} \sum_{s=1}^{N} \beta_{ns} \ln y_n \ln y_s + \sum_{k=1}^{K} \sum_{n=1}^{N} \gamma_{kn} \ln p_k \ln y_n, \qquad (21)$$

where C is cost. We assume linear homogeneity with respect to prices which can be imposed directly by dividing C and all prices by p_K. The share equations are:

$$S_k \equiv \frac{w_k e^{x_{kit}}}{\sum_{k'=1}^{K} w_{k'} e^{x_{k'it}}} = \alpha_k + \sum_{l=1}^{K} \alpha_{kl} \ln p_l + \sum_{n=1}^{N} \gamma_{kn} \ln y_n, \quad k = 1, \ldots, K-1. \qquad (22)$$

Clearly, x_t consists of the regressors in the cost function, whose dimensionality is $d = 1 + K + N + \frac{K(K+1)}{2} + \frac{N(N+1)}{2} + KN$.

Also $x_{t0} = \begin{bmatrix} 1 \\ \ln p_t \\ \ln y_t \end{bmatrix}$ whose dimensionality is $L = K + N + 1$. Moreover $\beta_{J(m)} = A_m \beta$ where A_m is an $L \times d$ selection matrix (consisting of ones and zeros), for all $m = 2, \ldots, M$, and $A_1 = I_{(d \times d)}$. Defining $x_{t0}' A_m = x_{tm}'$ for $m = 1, \ldots, M$, we can write the full system in the form:

$$y_{t1} = x_{t1}'\beta + v_{t1}$$
$$y_{t2} = x_{t2}'\beta + v_{t2}$$
$$\vdots \qquad\qquad (23)$$
$$y_{tM} = x_{tM}'\beta + v_{tM},$$

where $M = K + 1$. The complete system is $Y_t = X_t \beta + v_t$ or

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{v}, \qquad (24)$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix}, \; Y_t = \begin{bmatrix} y_{t1} \\ \vdots \\ y_{tM} \end{bmatrix}, \; X_t = \begin{bmatrix} x_{t1}' \\ \cdots \\ x_{tM}' \end{bmatrix}, \; \mathbf{X} = \begin{bmatrix} X_1 \\ \cdots \\ X_T \end{bmatrix}.$$

The output cost elasticities are:

$$\frac{\partial \ln C}{\partial \ln y_n} = \beta_n + \sum_{k=1}^{K} \gamma_{kn} \ln p_k + \sum_{s=1}^{N} \beta_{ns} \ln y_s = x_{t0}' \beta_{I(n)} = x_{t0}' D_n \beta, \qquad (25)$$

---

[3]Unfortunately, whether inputs are strongly disposable or not, cost functions simply cannot be translog functions; see, for example, footnote 11 on p. 286 in O'Donnell (2018). Here, we use the translog simply as an approximation to the true but unknown functional form.

where D_n is an $L \times d$ selection matrix, and $I(n)$ represents the proper set of indices. Therefore,we have:

$\frac{\partial \ln C}{\partial \ln y_n} = z'_{tn}\beta$, where $z'_{tn} = x'_{t0}D_n$, for $n = 1, \ldots, N$.

Monotonicity with respect to prices and outputs implies the following restrictions:

$$x'_{tm}\beta \geqslant 0, \quad m = 2, \ldots, K, \tag{26}$$

$$z'_{tn}\beta \geqslant 0, \quad n = 1, \ldots, N, \tag{27}$$

for all $t = 1, \ldots, T$. In total, we have $r = T(K + N - 1)$ monotonicity restrictions that we can represent in the form:

$$\mathbf{0}_{r \times 1} = \mathbf{W}\beta + \xi + \mathbf{u}, \tag{28}$$

where $\mathbf{W}$ is $r \times d$ and $W'_t = [(x'_{tm}, m = 2, \ldots, K), (z'_{tn}, n = 1, \ldots, N)]$ is the tth row of $\mathbf{W}$. We assume $\xi \sim N_r(\mathbf{0}, \omega^2\mathbf{I})$, and $\mathbf{u} \sim N_r^+(\mathbf{0}, \sigma_u^2\mathbf{I})$. Further, we assume: $\mathbf{v} \sim N(\mathbf{0}, \Sigma \otimes \mathbf{I}_T)$. Therefore, the complete system along with monotonicity constraints is:

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{0}_{r \times 1} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{W} \end{bmatrix} \beta + \begin{bmatrix} \mathbf{v} \\ \xi \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{MT \times 1} \\ \mathbf{u} \end{bmatrix}. \tag{29}$$

Moreover, we assume $\Sigma = \sigma^2\mathbf{I}$.

The conditional posterior distributions required to implement Gibbs sampling are presented in part A.2 of the Technical Appendix.

3.2.2 Concavity

Diewert and Wales (1987) showed that concavity of the translog cost function requiring negative semidefinitiness of $\nabla_{pp}C(p, y)$ amounts to negative semidefiniteness of $\mathbf{M}_t = \mathbf{A} - diag(\mathbf{S}_t) + \mathbf{S}_t\mathbf{S}'_t$, where $\mathbf{S}_t$ is the vector of shares, and $\mathbf{A}$ is the $K \times K$ matrix $[\alpha_{kl}]$. This matrix (after recalling $M = K + 1$) is

$$\mathbf{M}_t = \mathbf{A} - \begin{bmatrix} x'_{t2}\beta & & & \\ & x'_{t3}\beta & & \\ & & \ddots & \\ & & & x'_{tM}\beta \end{bmatrix} + \begin{bmatrix} x'_{t2}\beta x'_{t2}\beta & x'_{t2}\beta x'_{t3}\beta & \cdots & x'_{t2}\beta x'_{tM}\beta \\ x'_{t2}\beta x'_{t3}\beta & x'_{t3}\beta x'_{t3}\beta & \cdots & x'_{t3}\beta x'_{tM3}\beta \\ & & \cdots\cdots\cdots & \\ x'_{t2}\beta x'_{tM}\beta & x'_{t3}\beta x'_{tM}\beta & \cdots & x'_{tM}\beta x'_{tM}\beta \end{bmatrix}. \tag{30}$$

Suppose the eigenvalues of $\mathbf{M}_t$ are $\lambda'_t(\beta) = [\lambda_{t1}(\beta), \lambda_{t2}(\beta), \ldots, \lambda_{tK}(\beta)]$. Suppose $\mathbf{\Lambda}(\beta)$ is the $T \times K$ matrix whose rows are $\lambda'_t(\beta)$, $t = 1, \ldots, T$. The concavity restrictions can be expressed in the form:

$$-\boldsymbol{\Lambda}\left(\beta\right) = \zeta + \mathbf{w}, \quad \zeta \sim N_{TK}\left(\mathbf{0}, \ \Omega^2 \mathbf{I}\right), \ \mathbf{w} \sim N_{TK}^+\left(\mathbf{0}, \ \sigma_w^2 \mathbf{I}\right), \tag{31}$$

where Ω^2 and σ_w^2 are parameters. If we set Ω^2 to a small number, the meaning of this expression is that all eigenvalues of the M_t matrix are nonnegative. In practice, we can treat Ω as a parameter to examine systematically the extent of violation of the constraint(s). Moreover, it is straightforward to have different Ω parameters for different constraints or treat Ω as a general covariance matrix.

4 Endogeneity issues

In the case of the cost function where input prices and outputs are taken as predetermined or in the case of (4) where the covariates are weakly exogenous, the Bayesian techniques we have described, can be applied easily. However, *there are many instances in which the covariates or explanatory variables are endogenously determined.* An extremely important case is when (4) represents a production function. Under the assumption of cost minimization, inputs are decision variables (and, therefore, endogenous) while output (the left-hand-side variable) is predetermined. Moreover, economic exogeneity and econometric exogeneity are different things. If econometric exogeneity is rejected this does not mean that the economic assumptions are wtong. Measurement errors, for example, would be a typical example. Lai and Kumbhakar (2019) consider the case of a Cobb-Douglas production function along with the first-order conditions for cost minimization. To summarize the approach of Lai and Kumbhakar (2019), we have the following Cobb-Douglas production function:

$$y_{it} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kit} + v_{it} - u_{it}, i = 1, \ldots, n, t = 1, \ldots, T, \tag{32}$$

where y_{it} is log output for firm i and date t, x_{kit} is the log of kth input for firm i and date t , v_{it} is a two-sided error term, u_{it} is a non-negative error component that represents technical inefficiency in production, and $\beta_k > 0, k = 1, ..., K$. Suppose also input prices are w_{kit}. From the first-order conditions of cost minimization (where inputs are endogenous choice variables and output is predetermined) we obtain:

$$\frac{\partial y_{it}/\partial x_{kit}}{\partial y_{it}/\partial x_{1it}} = \frac{w_{kit}x_{kit}}{w_{1it}x_{1it}} = \frac{\beta_k}{\beta_1}e^{v_{kit}}, 2 = 1, ..., K. \tag{33}$$

These conditions can be expressed as follows:

$$x_{1it} - x_{kit} = \ln(w_{kit}/w_{1it}) + (\ln\beta_1 - \ln\beta_j) + v_{kit}, k = 2, \ldots, K. \tag{34}$$

The constraints are only on the parameters β_k $(k = 1, \ldots, K)$ so this is not a very interesting example. However, if we generalize (32) to the translog case, we have:

$$y_{it} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kit} + \tfrac{1}{2} \sum_{k=1}^{K} \sum_{k'=1}^{K} \beta_{kk'} x_{kit} x_{k'it} + v_{it} - u_{it}, i = 1, \ldots, n, t = 1, \ldots, T. \tag{35}$$

Suppose all parameters are collected into the vector β whose dimension is $d = 1 + K + \frac{K(K+1)}{2}$, after imposing symmetry, viz. $\beta_{kk'} = \beta_{k'k}$, $k, k' = 1, \ldots, K$. The first-order conditions for cost minimization are as follows:

$$\frac{\partial y_{it}/\partial x_{kit}}{\partial y_{it}/\partial x_{1it}} = \frac{w_{kit} x_{kit}}{w_{1it} x_{1it}} = \frac{\beta_k + \sum_{k'=1}^{K} \beta_{kk'} \ln x_{k',it}}{\beta_1 + \sum_{k'=1}^{K} \beta_{1k'} \ln x_{k',it}} e^{v_{kit}}, k = 2, \ldots, K. \tag{36}$$

These equations can be rewritten as follows:

$$x_{kit} - x_{1it} =$$
$$\ln \left(\beta_k + \sum_{k'=1}^{K} \beta_{kk'} x_{k',it} \right) - \ln \left(\beta_1 + \sum_{k'=1}^{K} \beta_{1k'} x_{k',it} \right) - \ln(w_{kit}/w_{1it}) + v_{kit}, k = 2, \ldots, K. \tag{37}$$

Moreover, it is convenient to rewrite (35) in the form:

$$y_{it} = \psi(x_{it})\beta + v_{it} - u_{it}, i = 1, \ldots, n, t = 1, \ldots, T, \tag{38}$$

where $\psi(x_{it}) = [1, x_{1it}, \ldots, x_{Kit}, \tfrac{1}{2}x_{1it}^2, \ldots, \tfrac{1}{2}x_{Kit}^2, x_{1it}x_{2it}, \ldots, x_{(K-1)it}x_{Kit}]'$ are the nonlinear terms in the translog functional form.

From (35) and (37) we have a system of K equations in K endogenous variables.

The economic restrictions are as follows. First we have monotonicity:

$$\beta_k + \sum_{k'=1}^{K} \beta_{kk'} x_{k',it} \geq 0, k = 1, \ldots, K. \tag{39}$$

This imposes the following set of restrictions:

$$0 = \beta_k + \sum_{k'=1}^{K} \beta_{kk'} x_{k',it} + \tilde{v}_{k,it} + \tilde{u}_{k,it}, k = 1, \ldots, K. \tag{40}$$

Following Diewert and Wales (1978, p. 58), given the monotonicity restrictions, we need the matrix $B =$

$[\beta_{kk'}, k, k' = 1, ..., K]$ to be negative semi-definite. Therefore, it is necessary and sufficient that the eigenvalues of B, say $\Lambda(\beta) = [\lambda_1(\beta), \ldots, \lambda_K(\beta)]'$ are all non-positive. This imposes the following set of nonlinear restrictions:

$$-\text{vec}\boldsymbol{\Lambda}(\beta)_{(TK \times 1)} = v^o_{(TK \times 1)} + u^o_{(TK \times 1)}. \tag{41}$$

From (40) and (41) we have $2K$ additional equations so in total we have K endogenous variables but $3K$ stochastic equations. From the econometric point of view, this is, clearly, *a major problem, as we lack $2K$ endogenous variables to complete the system in (35), (37), (40), and (41).* Let us write the entire system as follows:

$$\begin{bmatrix} -y_{it} \\ x_{2it} - x_{1it} \\ \vdots \\ x_{Kit} - x_{1it} \\ \mathbf{0}_K \\ \mathbf{0}_K \end{bmatrix} = \begin{bmatrix} -\psi(x_{it})\beta \\ g_2(x_{it}; \beta) \\ \vdots \\ g_K(x_{it}; \beta) \\ \mathbf{m}(x_{it}; \beta) \\ \mathbf{s}(x_{it}; \beta) \end{bmatrix} + \begin{bmatrix} v_{it,1} \\ v_{it,2} \\ \vdots \\ v_{it,K} \\ \tilde{v}_{it,(K \times 1)} \\ v^o_{it,(K \times 1)} \end{bmatrix} + \begin{bmatrix} u_{it,1} \\ u_{it,2} \\ \vdots \\ u_{it,K} \\ \tilde{u}_{it,(K \times 1)} \\ \tilde{u}_{it,(K \times 1)} \end{bmatrix} \tag{42}$$

where $g_k(x_{it}; \beta) = \ln\left(\beta_k + \sum_{k'=1}^K \beta_{kk'} \ln x_{k',it}\right) - \ln\left(\beta_1 + \sum_{k'=1}^K \beta_{1k'} \ln x_{k',it}\right) - \ln(w_{kit}/w_{1it}), k = 2, ..., K, \mathbf{m}(x_{it}; \beta) = [m_1(x_{it}; \beta), ..., m_K(x_{it}; \beta)]', m_k(x_{it}; \beta) = \beta_k + \sum_{k'=1}^K \beta_{kk'} \ln x_{k',it}, k = 1, ..., K, \mathbf{s}(x_{it}; \beta) = [s_1(x_{it}; \beta), ..., s_K(x_{it}; \beta)]', s_k(x_{it}; \beta) = \lambda_k(x_{it}; \beta), k = 1, ..., K$. Let us write the system in (42) compactly as follows:

$$\mathbf{Y}_{it} = \mathbf{f}(x_{it}; \beta) + \mathbf{v}_{it} + \mathbf{u}_{it}. \tag{43}$$

Although we have $3K$ equations there are only K endogenous variables. To provide $2K$ additional equations, it seems that the only possibility is to assume that $\mathbf{U}_{it} \equiv [\tilde{u}'_{it}, \breve{u}'_{it}]'$ are, in fact, *endogenous* variables. *This provides, indeed, the missing $2K$ additional equations.* To accomplish this, we depart from the assumption that $\mathbf{U}_{it}$ is a (vector) random variable, and, instead, we make use of the following device originally proposed by Paul and Shankar (2018) and further developed by Tsionas and Mamatzakis (2019):

$$\mathbf{U}_{it} = \begin{bmatrix} U_{it,1} \\ \vdots \\ U_{it,2K} \end{bmatrix} = \begin{bmatrix} -\ln \Phi(x'_{it}\gamma_1) \\ \vdots \\ -\ln \Phi(x'_{it}\gamma_K) \end{bmatrix} \equiv \mathbf{h}(x_{it}; \gamma), \tag{44}$$

where $\Phi(\cdot)$ is any distribution function (for example, the standard normal), and $\gamma = [\gamma'_1, \ldots, \gamma'_K]' \in \mathbb{R}^{2K}$ is a vector of parameters. The idea of Paul and Shankar (2018) is that efficiency, $r = e^{-u}$ is in the interval $(0, 1]$ and, therefore,

r can be modeled using any distribution function. In turn, we have:

$$
\begin{bmatrix} -y_{it} \\ \Delta x_{it} \\ \widetilde{u}_{it} \end{bmatrix} = \begin{bmatrix} -\psi(x_{it})\beta \\ \mathbf{g}(x_{it};\beta) \\ \mathbf{h}(x_{it};\beta,\gamma) \end{bmatrix} + \begin{bmatrix} v_{it,1} \\ \widetilde{v}_{it} \\ \mathring{v}_{it,(2K\times 1)} \end{bmatrix} + \begin{bmatrix} u_{it,1} \\ \widetilde{u}_{it} \\ \mathbf{0}_{(2K\times 1)} \end{bmatrix},
\tag{45}
$$

where $\mathring{v}_{it} = [\tilde{v}'_{it}, \check{v}'_{it}]'$, $\mathbf{V}_{it} = [v_{it,1}, \ldots, v_{it,K}, \mathring{v}'_{it}]'$, $\mathbf{U}_{it} = [u_{it,1}, \ldots, u_{it,K}]'$, $\triangle x_{it} = [x_{2it} - x_{1it}, \ldots, x_{Kit} - x_{1it}]'$, $\mathbf{g}(x_{it};\beta) = [g_2(x_{it};\beta), \ldots, g_K(x_{it};\beta)]'$, $\widetilde{v}_{it} = [v_{it,2}, \ldots, v_{it,K}]'$, $\widetilde{u}_{it} = [u_{it,2}, \ldots, u_{it,K}]'$. MCMC for this model is detailed in the Technical Appendix (sections A.4 and A.5).

5 Empirical application

We use the same data as in Lai and Kumbhakar (2019) which have been used before 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. We use a production frontier approach. Output is net steam electric power generation in megawatt-hours. Input prices are available and a time trend is included in both the production function and the predetermined variables. MCMC is implemented using 150,000 draws discarding the first 50,000 to mitigate possible start up effects. Since we have 984 observations we impose the constraints in (40) and (41) at randomly chosen points. The reason is that imposing the constraints in (40) and (41) at all points, compromises the flexibility of the translog and reduces it to the Cobb-Douglas production function, which is, clearly, very restrictive. We select the points by using the following methodology. Suppose we impose the constraints at the means of the data and P other points $(P = 1, \ldots, \overline{P})$ where $\overline{P} < nT$. The points are randomly chosen and we set $\overline{P} = 500$ which is, roughly, half the number of available observations. P itself is randomly chosen, uniformly distributed in $\{1, 2, ..., \overline{P}\}$. We repeat the process 10,000 times and we compute the marginal likelihood of the model. The marginal likelihood is defined as:

$$
\mathcal{M}(\mathscr{D}) = \int p(\beta, \gamma, \sigma_u, \mathbf{u}_1; \mathscr{D}) d\beta \, d\gamma \, d\sigma_u \, d\mathbf{u}_1.
\tag{46}
$$

The integral is not available analytically but can be computed numerically using the methodology of Perrakis, Ntzoufras and Tsionas (2015). In turn, we select the value of P as well as the particular points at which the constraints are imposed, by maximizing the value of $\mathcal{M}(\mathscr{D})$. For each P, we average across all datasets with this number of points, and we present the normalized log marginal likelihood in Figure A.1.

Marginal posterior densities of input elasticities are reported in Figure 1. Without imposing the theoretical constraints, we have a number of violations in fuel and capital elasticities and even labor (where there is a distinct mode around zero). After imposing the constraints, the marginal posteriors are much more concentrated around their mean or median, showing that *imposition of theoretical constraints improves the accuracy of statistical inference*
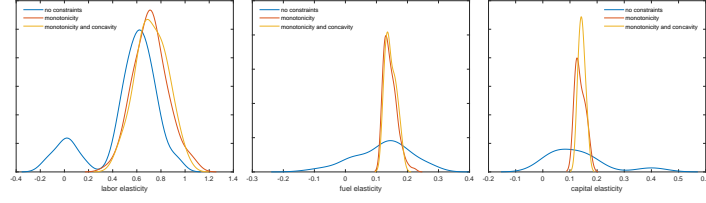
for these elasticities.

Marginal posterior densities of aspects of the model are reported in Figure 2. Technical efficiency is defined as $r_{it} = e^{-u_{it}}$ where $u_{it} \geq 0$ represents technical inefficiency. Technical change (TC) is defined as the derivative of the log production function with respect to time, viz. $TC_{it} = \frac{\partial E(y_{it})}{\partial t}$. Efficiency change ($EC_{it}$) is $EC_{it} = \frac{u_{it} - u_{i,t-1}}{u_{i,t-1}}$. Productivity growth ($PG_{it}$) is $PG_{it} = TC_{it} + EC_{it} + SCE_{it}$, where SCE_{it} is the scale effect (Kumbhakar et al., 2015, equation 11.8). Under monotonicity and / or concavity, technical efficiency averages 85% and ranges from 78% to 93%. Without imposition of theoretical constraints technical efficiency is considerable lower, averaging 78% and ranging from 74% to 84%. Therefore, *imposing the constraints is quite informative for efficiency* and delivers results that are different compared to an unrestricted translog production function. Technical change averages 1% and ranges from -3% to 5% *per annum*. Efficiency change is much more pronounced when monotonicity and / or concavity restrictions are imposed. Without the restrictions, it averages 1% and ranges from -1% to 3.5%. With the restrictions in place, it averages 3.2% and ranges from -1% to 6%. In turn, productivity growth (the sum of technical change, efficiency change and scale effect) averages 4.2% relative to only 2% in the translog model without the constraints.

In relation to (13), let us define $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \boldsymbol{\sigma}'_1 \\ \boldsymbol{\sigma}_1 & \boldsymbol{\Sigma}^* \end{bmatrix}$, where σ_{11} is the variance of $v_{it,1}$, $\boldsymbol{\sigma}_1$ represents the vector of covariances between $v_{it,1}$ and $\widetilde{v}_{it}$, and $\boldsymbol{\Sigma}^*$ is the covariance matrix of $\widetilde{v}_{it}$. To examine whether the artificial error terms $\widetilde{v}_{it}$ are of quantitative importance, we can use the measure $|\boldsymbol{\Sigma}^*|/\sigma_{11}$. This measure provides the (generalized) variability of $\widetilde{v}_{it}$ in terms of the variance of $v_{it,1}$, viz. the stochastic error in the production function.

Aspects of the posterior distribution of the model are reported in Figures 2 and 3.

The marginal posterior densities of measure $|\boldsymbol{\Sigma}^*|/\sigma_{11}$ are reported in the upper left panel of Figure 4. The (generalized) variance of two-sided error terms in the constraints is only 3.5% relative to the variance of production function error term, implying that the one-sided error terms account for most of the variability of the equations corresponding to the restrictions. In the upper right panel of Figure 5, we report the marginal posterior density of $\lambda = \frac{\sigma_u^2}{\sigma_{11}}$ which represents the signal-to-noise ratio in frontier models. This ratio averages 1.3 (ranging from 0.5 to 2.7) without the constraints and 2.5 (ranging from 1.5 to 3.7) when the constraints are imposed. This suggests that the *imposition of theoretical constraints allows for more precise inferences in the stochastic frontier model*. To allow for the presence of the constraints, it is more appropriate to define the signal-to-noise ratio as $\lambda^* = \frac{\sigma_u^2}{|\boldsymbol{\Sigma}|}$. The marginal posterior density of λ^* is reported in the bottom panel of Figure 4. Evidently, the new measure is lower compared to λ but it is still considerably larger than the ratio without imposition of the theoretical constraints.

Figure 3: Marginal posterior densities of input elasticities

Notes: Input elasticities are defined as $\frac{\partial E(y_{it})}{\partial x_{it}}$. in relation to (35).
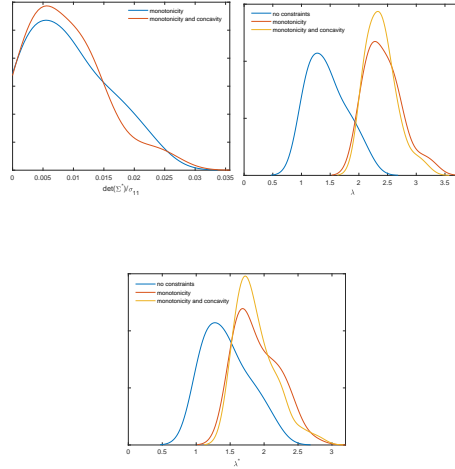
Figure 4: Marginal posterior densities of aspects of the model
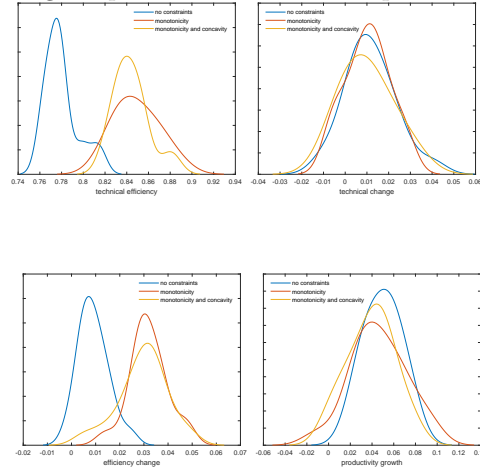
Notes: Technical efficiency is defined as $r_{it} = e^{-u_{it}}$ where $u_{it} \geq 0$ represents technical inefficiency. Technical change (TC) is defined as the derivative of the log production function with respect to time. Efficiency change (EC) is $EC = \frac{u_{it} - u_{i,t-1}}{u_{i,t-1}}$. Productivity growth ($PG$) is $PG = TC + EC$.

Figure 5: Marginal posterior densities of $\det(\boldsymbol{\Sigma}^*)/\sigma_{11}$ and λ^*

Notes: In relation to (13), let us define $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \boldsymbol{\sigma}'_1 \\ \boldsymbol{\sigma}_1 & \boldsymbol{\Sigma}^* \end{bmatrix}$, where σ_{11} is the variance of $v_{it,1}$, $\boldsymbol{\sigma}_1$ represents the vector of covariances between $v_{it,1}$ and $\widetilde{v}_{it}$, and $\boldsymbol{\Sigma}^*$ is the covariance matrix of $\widetilde{v}_{it}$. To examine whether the artificial error terms $\widetilde{v}_{it}$ are of quantitative importance, we can use the measure $|\boldsymbol{\Sigma}^*|/\sigma_{11}$. This measure provides the (generalized) variability of $\widetilde{v}_{it}$ in terms of the variance of $v_{it,1}$, viz. the stochastic error in the production function. To allow for the presence of the constraints, it is more appropriate to define the signal-to-noise ratio as $\lambda^* = \frac{\sigma_u^2}{|\boldsymbol{\Sigma}|}$.

Posterior moments are presented in Table 1.

Another important issue is whether posterior predictive densities of efficiency estimates are more informative relative to unconstrained estimates. From Table in O'Donnell, Shumway, and Ball (1999) who used Metropolis-Hastings to impose the constraints, unconstrained maximum likelihood estimates and Bayes estimates that impose concavity, sometimes yield higher efficiency estimates and sometimes they yield lower efficiency estimates. Standard errors without concavity and standard errors with concavity are, more often than not, lower in the second case but there are some exceptions. On the other hand, the results of O'Donnell and Coelli (2005) suggest that imposing monotonicity and curvature yields more precise estimates (Table 3 and Figures 2–9).

In Figure 5, we present posterior predictive densities of efficiency for nine randomly selected observations.

In line with O'Donnell and Coelli (2005) or O'Donnell, Shumway, and Ball (1999) we find that, more often than not, the posterior predictive efficiency densities are more concentrated around their modal values. The posterior predictive efficiency densities of other farms behave in the same way, and results are available on request. A related issue is whether imposition of monotonicity and curvature result in stochastic dominance over the model without these restrictions. From the evidence in Figure 6, where we report normalized cumulative distribution functions (cdf) we have stochastic dominance of the model with monotonicity and curvature only for farms 6 and 9. Therefore, as a rule, imposition of the restrictions does not, necessarily, imply stochastic dominance mostly because the average posterior predictive efficiency estimates change as well.

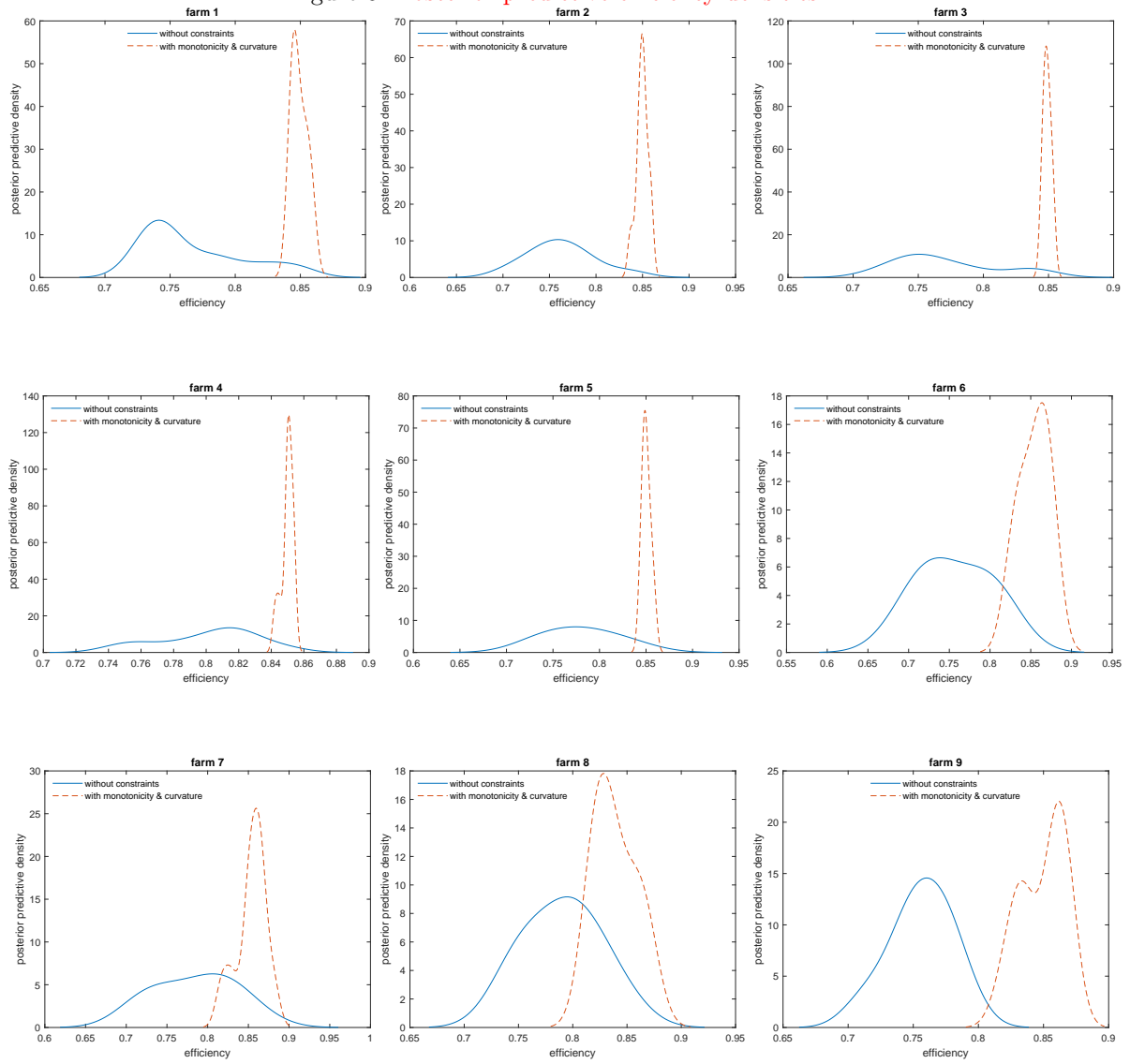Figure 6: Posterior predictive efficiency densities

Table 1: Posterior moments and functions of interest

	unconstrained	monotonicity	monotonicity & curvature
labor elasticity	0.527	0.712	0.723
	(0.261)	(0.139)	(0.132)
fuel elasticity	0.120	0.145	0.146
	(0.090)	(0.021)	(0.018)
capital elasticity	0.127	0.138	0.144
	(0.112)	(0.018)	(0.012)
technical efficiency	0.779	0.849	0.844
	(0.016)	(0.020)	(0.017)
technical change	0.012	0.010	0.009
	(0.010)	(0.010)	(0.011)
efficiency change	0.09	0.032	0.030
	(0.06)	(0.008)	(0.010)
scale effect	0.029	0.0037	-0.0004
	(0.006)	(0.016)	(0.010)
productivity growth	0.049	0.046	0.039
	(0.017)	(0.025)	(0.021)
λ	1.446	2.427	2.389
	(0.326)	(0.314)	(0.258)
λ^*	1.260	1.898	1.876
	(0.199)	(0.301)	(0.292)

Notes: Reported are posterior means with posterior standard deviations in parentheses.

Concluding remarks

An issue of great practical importance is the imposition of theoretical inequality constraints on cost or production functions. These constraints can be handled efficiently using a novel formulation that converts inequality constraints to equalities using surpluses which are treated in the context of stochastic frontier analysis. The idea has been developed independently by Huang and Huang (2019). However, the authors did not deal with the case of cost-share systems (in which more problems arise and need to be addressed) neither they allowed for correlation between violations of monotonicity and curvature which is quite likely in practice. There are two problems that are successfully resolved in this paper. First, the constraints are not independent as it is known, for example, that imposing monotonicity leads to fewer violations of concavity. Second, when explanatory variables are endogenous, special endogeneity problems arise which cannot be solved easily. In turn, proposed are special techniques to address this issue, and they are shown to perform well in an empirical application.

TECHNICAL APPENDIX

Figure 7: Posterior predictive cumulative distribution functions
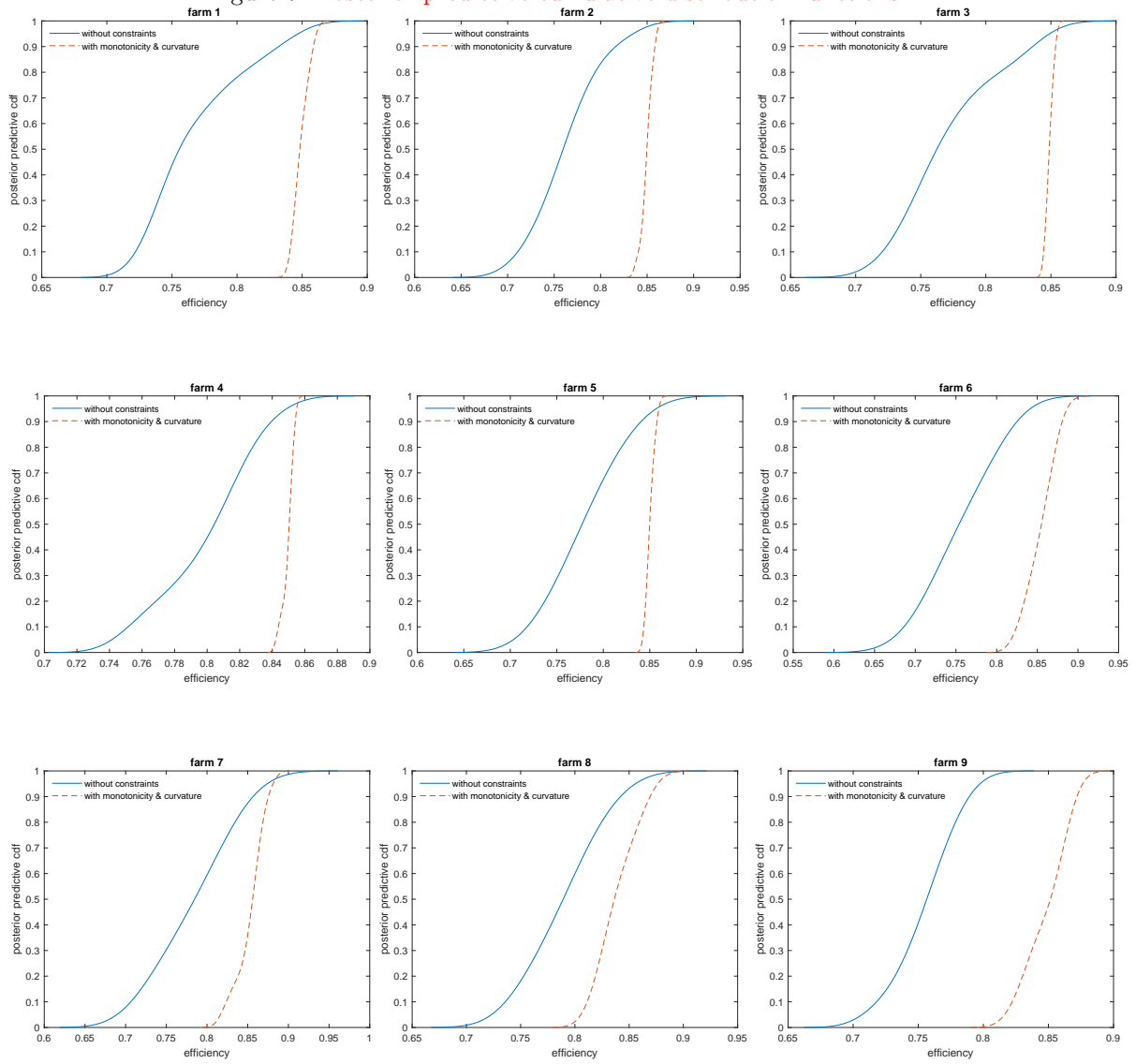
5.1 MCMC associated with (19)

Conditional on all other parameters the posterior density of $\mathbf{u}_i$ is as follows:

$$\mathbf{u}_i|\beta, \Sigma, \Phi, y, X \sim N_{m+1}^+ \left(\hat{\mathbf{u}}_i, \mathbf{V} \right), \ i = 1, \ldots, n, \tag{1}$$

where $\hat{\mathbf{u}}_i = - \left(\Sigma^{-1} + \Phi^{-1} \right)^{-1} \left(\psi_i - X_i\beta \right), \ i = 1, \ldots, n,$ and $\mathbf{V} = \left(\Sigma^{-1} + \Phi^{-1} \right)^{-1}$.

We remind that in the main part we have assumed $\Sigma = \begin{bmatrix} \sigma^2 & \\ & \omega^2 I_m \end{bmatrix}$, and $\Phi = \begin{bmatrix} 0 & \\ & \phi^2 I_m \end{bmatrix}$. Here, we allow

for general covariance matrices with the following prior:

$$p(\Sigma, \Phi) \propto |\Sigma|^{-(n+m+1)/2} |\Phi|^{-(n+m+1)/2}. \tag{2}$$

We adopt general covariance matrices because of the following reasons. It is always possible to fix Σ in advance

in the form $\Sigma = \begin{bmatrix} \sigma^2 & \\ & \omega^2 I_m \end{bmatrix}$, where σ^2 remains an unknown parameter but ω is fixed to a small value (say

0.001). In this case, the posterior conditional distribution of σ^2 is:

$$\frac{(y + \mathbf{u} - X\beta)'(y + \mathbf{u} - X\beta)}{\sigma^2} |\beta, \omega, \mathbf{u}, y, X \sim \chi_n^2. \tag{3}$$

We do not recommend this practice for three reasons. First, assuming independent two-sided errors is too restrictive as imposition of one constraint is not independent of imposing other constraints. Second, assuming independent one-sided errors is too restrictive as well. Third, a common value of ϕ is too restrictive as different constraints may require different surplus. However, with regard to the third point, it is not difficult to replace $\phi^2 I_m$ with an $m \times m$ diagonal matrix whose diagonal elements are $\phi_1, \ldots, \phi_m$. Moreover, it is possible to treat ϕ as a general covariance matrix (denoted Φ), as we do here.

The posterior conditional distributions are in the well-known inverted-Wishart family:

$$p(\Sigma|\beta, \Phi, \mathbf{u}, y, X) \propto |\Sigma|^{-(n+m+1)/2} \exp \left(-\tfrac{1}{2} tr \mathbf{A}_\Sigma \Sigma^{-1} \right),$$
$$p(\Phi|\beta, \Sigma, \mathbf{u}, y, X) \propto |\Phi|^{-(n+m+1)/2} \exp \left(-\tfrac{1}{2} tr \mathbf{A}_\Phi \Phi^{-1} \right), \tag{4}$$

where $\mathbf{A}_\Sigma = \sum_{i=1}^n (\psi_i - X_i\beta + \mathbf{u}_i)(\psi_i - X_i\beta + \mathbf{u}_i)'$, and $\mathbf{A}_\Omega = \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i'$. Finally, the posterior conditional distribution of β is a multivariate normal:

$$\beta|\Sigma, \Omega, \mathbf{u}, y, X \sim N \left(\hat{\beta}, \mathbf{V}_\beta \right), \tag{5}$$

where $\hat{\beta} = (\widetilde{X}'\Sigma^{-1}\widetilde{X})^{-1}\widetilde{X}'\Sigma^{-1}(y + \mathbf{u})$, $\mathbf{V}_\beta = (\widetilde{X}'\Sigma^{-1}\widetilde{X})^{-1}$, and $\widetilde{X} = \begin{bmatrix} X \\ X_0 \end{bmatrix}$.

All conditionals are in standard forms which facilitate random number generation for implementation of MCMC.

5.2 MCMC associated with (29).

We remind that we assume $\Sigma = \begin{bmatrix} \sigma^2 & \\ & \omega^2 I_m \end{bmatrix}$, $\xi \sim N(\mathbf{0}, \omega^2 \mathbf{I})$, and $\mathbf{u} \sim N_r^+(\mathbf{0}, \sigma_u^2 I)$. In this case, we have no technical inefficiency (viz. $u_i = 0$) and the scale parameter of surpluses, $\widetilde{u}_i$, have the same scale parameter.

It is not difficult to show that

$$\beta|\cdot, \mathbf{Y}, \mathbf{X} \sim N_d\left(\hat{\beta}, \ \mathbf{V}\right), \tag{6}$$

where $|\cdot, \mathbf{Y}, \mathbf{X}$ denotes conditioning on all other parameters, latent variables, and the data, and

$$\hat{\beta} = \left[\omega^2 \mathbf{X}'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T\right)\mathbf{X} + \sigma^2 \mathbf{W}'\mathbf{W}\right]^{-1}\left[\omega^2 \mathbf{X}'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T\right)\mathbf{Y} - \mathbf{W}'\mathbf{u}\right],$$
$$\mathbf{V} = \sigma^2 \omega^2 \left[\omega^2 \mathbf{X}'\left(\mathbf{\Sigma}^{-1} \otimes \mathbf{I}_T\right)\mathbf{X} + \sigma^2 \mathbf{W}'\mathbf{W}\right]^{-1}. \tag{7}$$

Also:

$$\mathbf{u}|\cdot, \mathbf{Y}, \mathbf{X} \sim N_r^+\left(\hat{\mathbf{u}}, \ \sigma_*^2 \mathbf{I}_r\right), \tag{8}$$

where $\hat{\mathbf{u}} = -\frac{\sigma_u^2}{\sigma_u^2 + \omega^2}\mathbf{W}\beta$, and $\sigma_*^2 = \frac{\sigma_u^2 \omega^2}{\sigma_u^2 + \omega^2}$. Draws from the conditional distributions of σ_u^2 and ω^2 (given $\beta, \mathbf{u}, \mathbf{\Sigma}, \mathbf{Y}, \mathbf{X}$) can be obtained as

$$\frac{\mathbf{u}'\mathbf{u} + \bar{S}_u}{\sigma_u^2}|\cdot, \mathbf{Y}, \mathbf{X} \sim \chi_{r+\bar{r}_u}^2, \quad \frac{(\mathbf{W}\beta + \mathbf{u})'(\mathbf{W}\beta + \mathbf{u}) + \bar{S}_\omega}{\omega^2}|\cdot \sim \chi^2\left(r + \bar{r}_\omega\right), \tag{9}$$

where $\bar{S}_\omega$, $\bar{S}_u$, $\bar{r}_\omega$ and $\bar{r}_u$ represent prior parameters. The priors are such that

$\frac{\bar{S}_\omega}{\omega^2} \sim \chi^2\left(\bar{r}_\omega\right)$, and independently $\frac{\bar{S}_u}{\sigma_u^2} \sim \chi^2\left(\bar{r}_u\right)$, $\frac{\bar{S}}{\sigma^2} \sim \chi_{\bar{r}}^2$.

The values of the hyperparameters are $\bar{r}_u = \bar{r}_\omega = \bar{r} = 1$, and $\bar{S}_u = \bar{S}_\omega = \bar{S} = 10^{-6}$.

The conditional posterior distribution of σ^2 is

$$\frac{\bar{S} + \sum\limits_{t=1}^{T}(Y_t - X_t\beta)'(Y_t - X_t\beta)}{\sigma^2}|\cdot, \mathbf{Y}, \mathbf{X} \sim \chi_{nT+r}^2. \tag{10}$$

.

22

5.3 MCMC associated with (45).

We write the system in (45) as follows:

$$\mathbf{Y}_{it} = \mathbf{F}(x_{it}; \beta, \gamma) + \mathbf{V}_{it} + \begin{bmatrix} \mathbf{U}_{it} \\ \mathbf{0}_{(2K \times 1)} \end{bmatrix}, \tag{11}$$

We need the Jacobian of transformation from $\mathbf{V}_{it}$ to the endogenous variables $[x'_{it}, \widetilde{u}'_{it}]'$. The required Jacobian is:

$$J_{it}(\beta) = \sum_{k=1}^{K} \left(\beta_k + \sum_{k'=1}^{K} \beta_{kk'} x_{k',it} \right) = 1, \tag{12}$$

because $\sum_{k=1}^{K} \beta_k = 1$ and $\sum_{k=1}^{K} \sum_{k'=1}^{K} \beta_{kk'} = 0$ so that cost shares sum to unity. These constraints can be easily implemented in advance.

We are now able to make the following distributional assumptions:

$$\mathbf{V}_{it} \sim iid\ N_K(\mathbf{0}, \mathbf{\Sigma}),\ u_{it,1} \sim iid\ N_+(\mathbf{0}, \sigma_u^2),\ i = 1, \ldots, n,\ t = 1, \ldots, T. \tag{13}$$

Under these assumptions, the likelihood function of the model is as follows:

$$\begin{aligned} \mathcal{L}(\beta, \gamma, \mathbf{\Sigma}, \mathbf{\Omega}; \mathscr{D}) &\propto \int_{\mathbb{R}_+^{nT}} |\mathbf{\Sigma}|^{-nT/2} \sigma_u^{-nT/2} \cdot \exp\left(-\frac{\sum_{i=1}^{n} \sum_{t=1}^{T} u_{it,1}^2}{2\sigma_u^2} \right) \cdot \\ &\quad \exp\left\{ -\tfrac{1}{2} \left[\mathbf{Y}_{it} - \mathbf{F}(x_{it}; \beta, \gamma) \right] \mathbf{\Sigma}^{-1} \left[\mathbf{Y}_{it} - \mathbf{F}(x_{it}; \beta, \gamma) \right] \right\} d\mathbf{u}_1. \end{aligned} \tag{14}$$

where $\mathscr{D}$ denotes the data, $\mathbf{u}_1 = [u_{it}, i = 1, \ldots, n, t = 1, \ldots, T]$. Our prior for the parameters is given by:

$$p(\beta, \gamma, \mathbf{\Sigma}, \sigma_u) \propto |\mathbf{\Sigma}|^{-(d_\Sigma + 1)/2} \cdot \sigma_u^{-1}, \tag{15}$$

where $d_\Sigma = 3K$ is the dimensionality of Σ.

By Bayes' theorem, the posterior is $p(\beta, \gamma, \mathbf{\Sigma}, \sigma_u | \mathscr{D}) \propto \mathcal{L}(\beta, \gamma, \mathbf{\Sigma}, \sigma_u; \mathscr{D}) \cdot p(\beta, \gamma, \mathbf{\Sigma}, \sigma_u)$. To access the posterior, we consider the augmented (by $\mathbf{u}_1$) posterior which is:

$$\begin{aligned} p(\beta, \gamma, \mathbf{\Sigma}, \sigma_u, \mathbf{u}_1; \mathscr{D}) &\propto |\mathbf{\Sigma}|^{-(nT+3K+1)/2} \sigma_u^{-(nT+1)/2} \cdot \exp\left(-\frac{\sum_{i=1}^{n} \sum_{t=1}^{T} u_{it,1}^2}{2\sigma_u^2} \right) \cdot \\ &\quad \exp\left\{ -\tfrac{1}{2} \left[\mathbf{Y}_{it} - \mathbf{F}(x_{it}; \beta, \gamma) \right]' \mathbf{\Sigma}^{-1} \left[\mathbf{Y}_{it} - \mathbf{F}(x_{it}; \beta, \gamma) \right] \right\}. \end{aligned} \tag{16}$$

We use MCMC to obtain draws from $p(\beta, \gamma, \mathbf{\Sigma}, \sigma_u, \mathbf{u}_1; \mathscr{D})$. It is possible to integrate out $\mathbf{\Sigma}$ analytically from (16) using properties of the inverted Wishart distribution (Zellner, 1971, p. 229, formula (8.24) and p. 243, formula

(8.86)) and we obtain:

$$p(\beta, \gamma, \sigma_u, \mathbf{u}_1; \mathscr{D}) \propto \sigma_u^{-(nT+1)/2} \cdot \exp\left(-\frac{\sum_{i=1}^n \sum_{t=1}^T u_{it,1}^2}{2\sigma_u^2}\right) \cdot$$
$$|\sum_{i=1}^n \sum_{t=1}^T \left[\mathbf{Y}_{it} - \mathbf{F}(x_{it}; \beta, \gamma)\right]\left[\mathbf{Y}_{it} - \mathbf{F}(x_{it}; \beta, \gamma)\right]'|^{-nT/2}. \tag{17}$$

First, we draw σ_u from its posterior conditional:

$$\frac{\sum_{i=1}^n \sum_{t=1}^T u_{it,1}^2}{\sigma_u^2} | \beta, \gamma, \mathbf{u}_1, \mathscr{D} \sim \chi_{nT}^2. \tag{18}$$

To update β and γ we use a Girolami and Calderhead (2011) algorithm which uses first- and second-order derivative information from the log of (17).

From Figure (A.1) we see that the log marginal likelihood is largest when we use, approximately, 50 points to impose the monotonicity and curvature restrictions.

Figure 1: Normalized log marginal likelihood
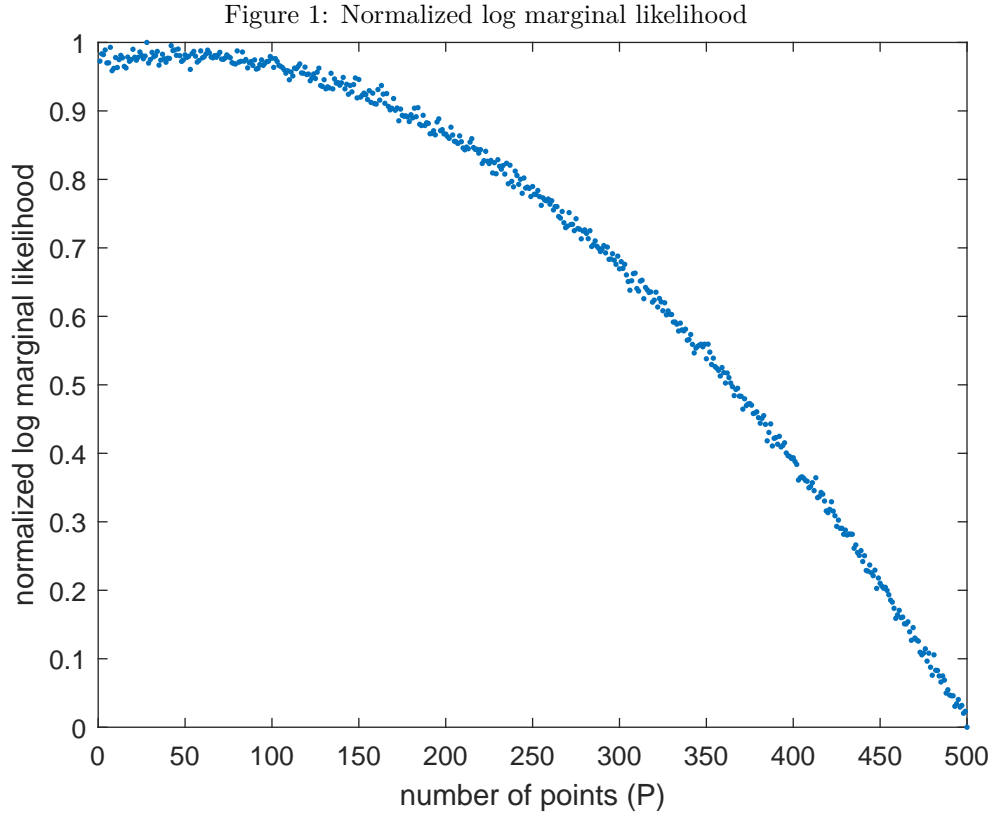
Notes: The log marginal likelihood is normalized in the interval $[0,1]$. We select the value of P as well as the particular points at which the constraints are imposed, by maximizing the value of $\mathcal{M}(\mathscr{D})$. For each P, we average across all data sets with this number of points. The constraints include monotonicity as well as concavity.

5.4 Comparison with other techniques to impose the restrictions

Other techniques to impose the constraints include i) to use LS in the production function and examine points at which the restrictions do not hold. ii) One can in turn use the accept-reject algorithm used by Terrell (1996) and the Metropolis-Hastings (M-H) algorithm used by Griffiths et al. (2000) and O'Donnell and Coelli (2005). Terrell (1996) used an accept-reject Gibbs sanpling algorithm to impose monotonicity and concavity constraints on the parameters of a cost function (through parametric restrictions only). A problem is that it may be necessary to generate an extremely large number of candidate MCMC draws before finding one that is acceptable. O'Donnell and Coelli (2005) simulate from the constrained posterior using a random-walk Metropolis-Hastings algorithm. The same criticism that applies to Terrell's (1996) approach possibly holds for the O'Donnell and Coelli (2005) approach as the acceptance rate could be too high or too low and it requires some tuning. Moreover, its performance can potentially be rather poor due to autocorrelation in MCMC.

Here, we consider both Terrell's (1996) and ODonnell and Coelli's (2005) approach. First, we consider auto-correlation functions (acf's) for these techniques compared to ours in Figure (A2). Relative numerical efficiency (RNE, Geweke, 1992) for the various techniques is reported in Figure (A.3). If i.i.d sampling from the posterior were feasible, then RNE would be equal to one. As we use MCMC, we have, of course, to settle for lower values, provided they are not small in order to be sure that we are exploring the posterior in a thorough manner.

Autocorrelation functions drop to zero more rapidly compared to Terrell (1996) and O'Donnell and Coelli (2005). We use the same number of MCMC draws and the same length of burn-in.

In Figure (A.4) we report the density of percentage differences between Terrell (1996) and O'Donnell (2005) for production function parameters (panel (a)), technical inefficiency (panel (b)), returns to scale (panel (c)), and productivity growth in panel (d). The rank correlations were from 0.30 to 0.60.

These differences between our method and Terrell (1996) or O'Donnell and Coelli (2005) can be explained by their higher acf's which do not help convergence. As we see from Figure (A.5) Geweke's (1992) convergence diagnostic (GCD, (which is distributed as standard normal, asymptotically in the number of draws) fails to support convergence for Terrell's (1996) or O'Donnell and Coelli's (2005) approach (as it is greater than 2 in absolute value, more often than not). Results for technical inefficiency (panel (b)), and scale parameters (panel (c)) are similar. Reported in panel (c) is GCD for the latent variables in our constraints. From GCD we can see that MCMC has converged.
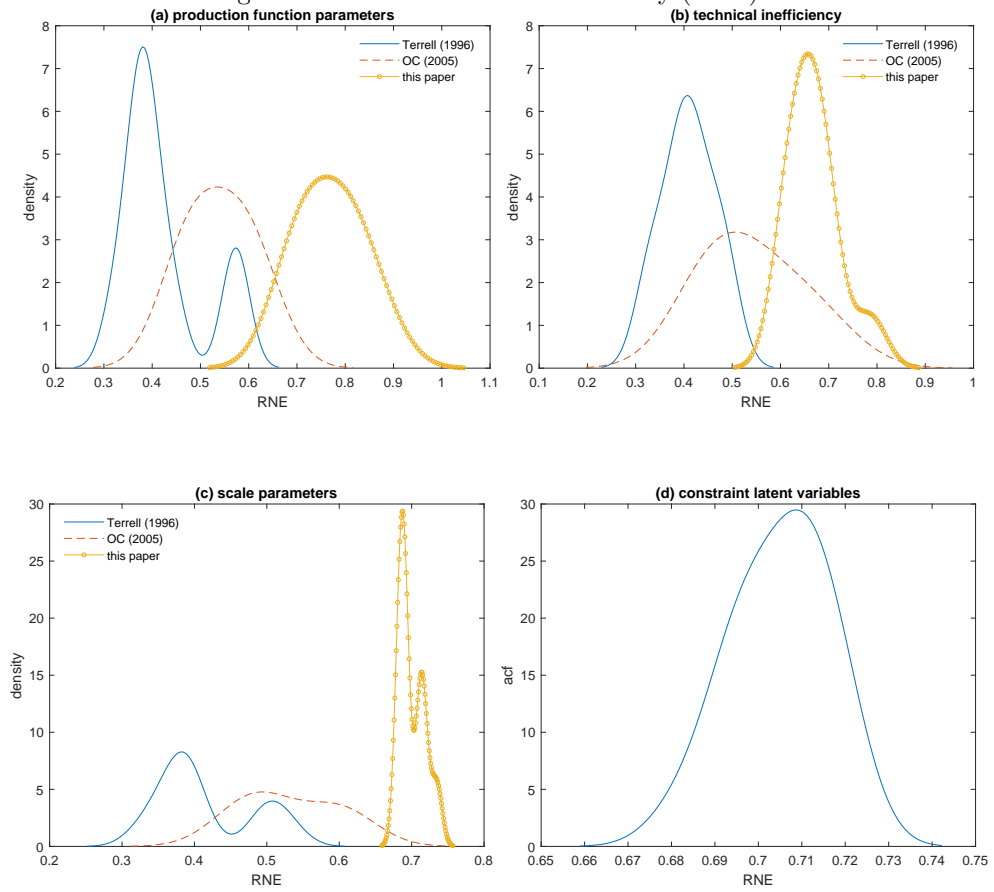
Compliance with Ethical Standards:
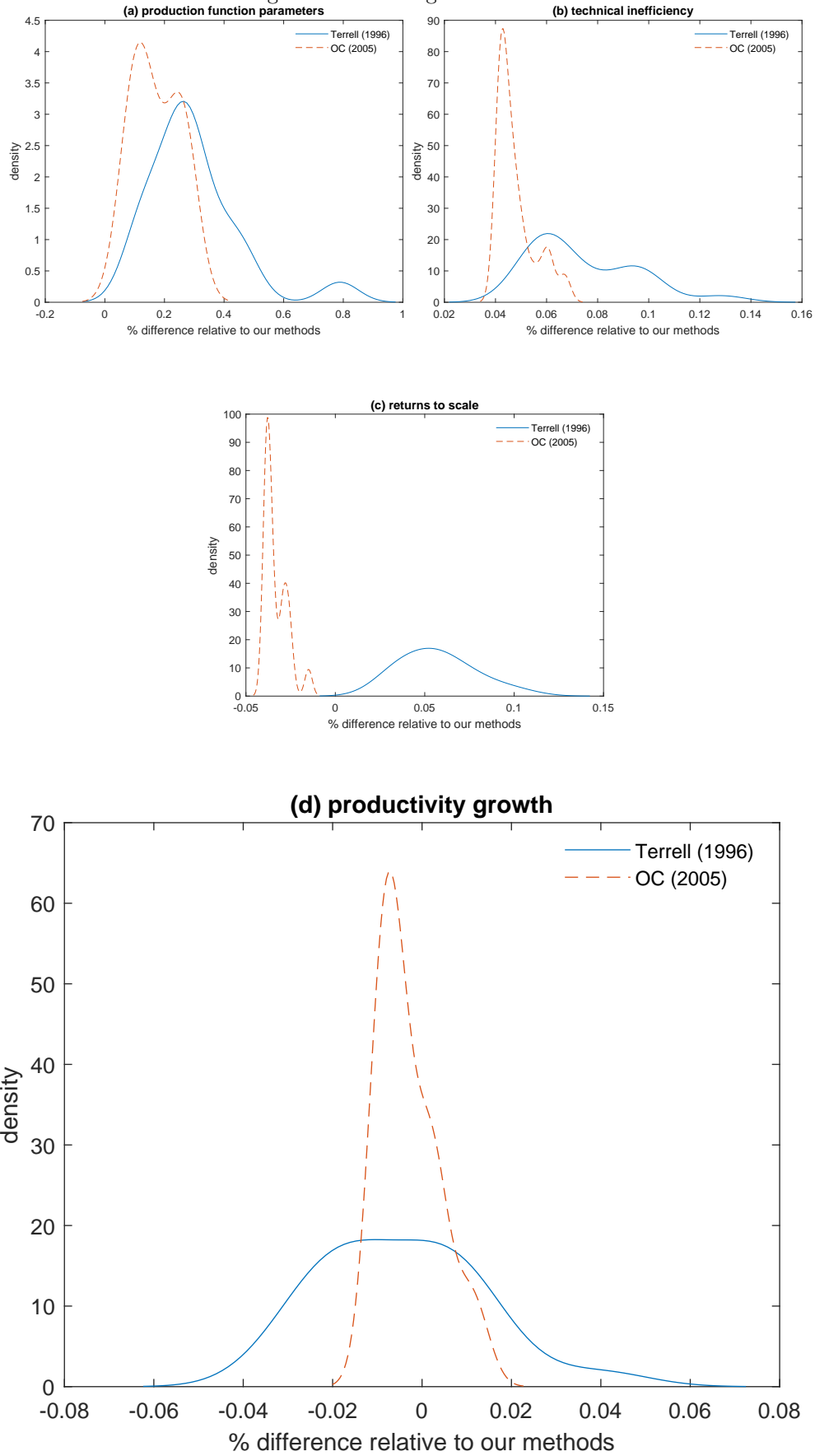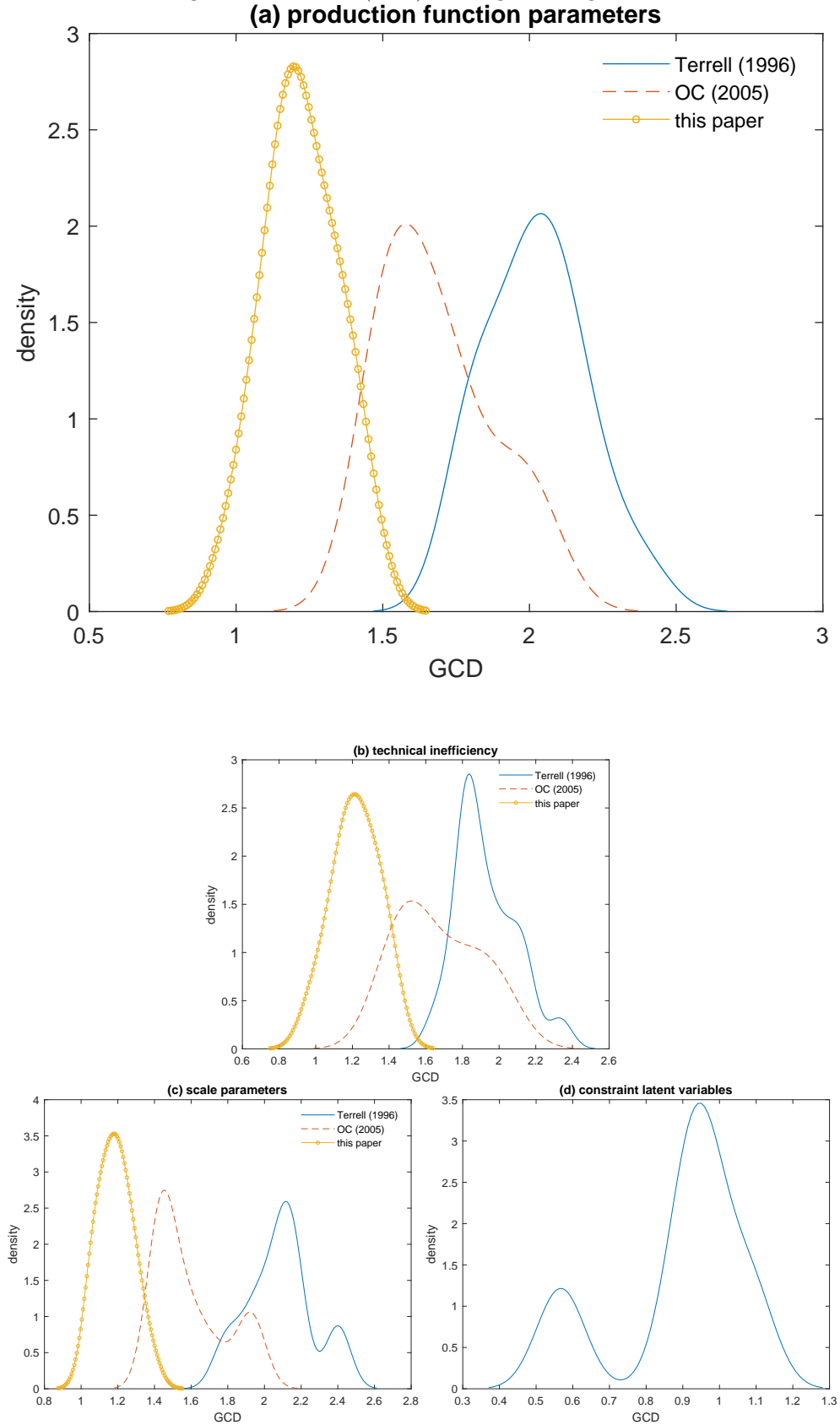
Figure 2: Autocorrelation functions (acf)
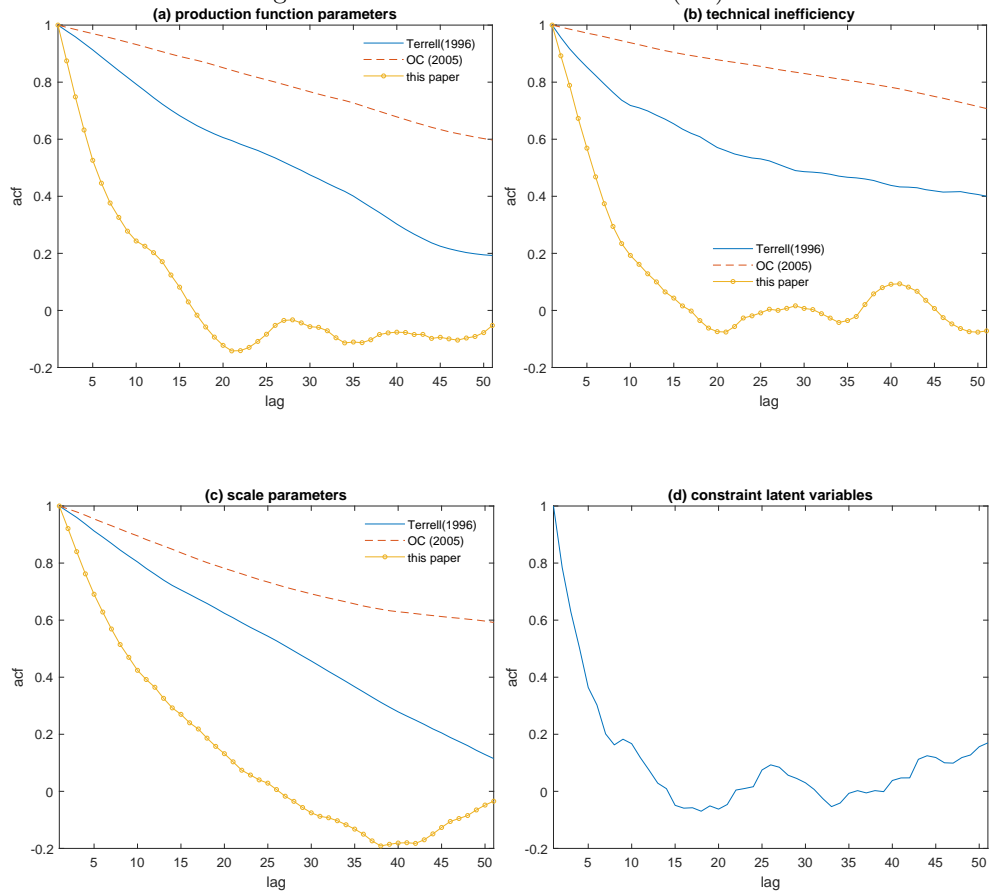
Notes: OC (2005) is O'Donnell and Coelli (2005). For each set of parameters (in panels (a), (c), and (d)) we report the median acf across all parameters.

Figure 3: Relative Numerical Efficiency (RNE)

Notes: OC (2005) is O'Donnell and Coelli (2005). For each set of parameters (in panels (a), (c), and (d) we report the median acf across all parameters.

Figure 4: Percentage differences

(a) production function parameters

(b) technical inefficiency

(c) returns to scale

(d) productivity growth

Notes: OC (2005) is O'Donnell and Coelli (2005). For each set of parameters (in panels (a), (c), and (d) we report the median acf across all parameters.

Figure 5: Geweke's (1992) convergence diagnostic

References

Diewert, W. E., and T. J. Wales (1987). Flexible Functional Forms and Global Curvature Conditions. Econometrica 55, 43–68. Gallant, A. R. and G. H. Golub (1984). Imposing Curvature Restrictions on Flexible Functional Forms. Journal of Econometrics, 26, 295–322. Geweke, J. (1986). Exact inference in the inequality constrained normal linear regression model, Journal of Applied Econometrics 1 (1), 127–141. Geweke, J. (1991). Efficient simulation from the multivariate normal and Student-t distributions subject to linear constraints and the evaluation of constraint probabilities, Computer Science and Statistics. Proceedings of the 23rd Symposium on the Interface. Seattle Washington. Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Clarendon Press, Oxford, UK, 169–193. Geweke, J. (1996). Bayesian inference for linear models subject to linear inequality constraints, Federal Reserve Bank of Minneapolis WP 552 [1995], and also in: A. Zellner and J. S. Lee (eds), Modeling and Prediction: Honouring Seymour Geisser, Springer, New York. 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. Huang, T. H., and Z. Huang (2019). Imposing regularity conditions to measure Banks' productivity changes in Taiwan using a stochastic approach, unpublished manuscript. Koop, G., J. Osiewalski, and M. F. J. Steel (1997). Bayesian efficiency analysis through individual effects: Hospital cost frontiers. Journal of Econometrics 76 (1–2), 77–105. Kumbhakar, S.C., B.U. Park, L. Simar and M. G. Tsionas (2007). Nonparametric stochastic frontiers: A local maximum likelihood approach, Journal of Econometrics 137 (1), 1–27. Kumbhakar, S. C., H-J Wang, A. P. Horncastle (2015). A practitioner's guide to stochastic frontier analysis using Stata. Cambridge University Press, New York. 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. Lau, L.J. (1978). Testing and imposing monotonicity, convexity, and quasi-convexity constraints. In: Production Economics: A Dual Approach to Theory and Applications (volume 1), eds. Fuss, M., and D. McFadden, Amsterdam: North-Holland, 1978, pp. 409–453. Ivaldi, M., N. Ladoux, H. Ossard and M. Simioni (1996). Comparing Fourier and translog specifications of multiproduct technology: Evidence from an incomplete panel of French farmers. Journal of Applied Econometrics, 11 (6), 649–667. McCausland, W. J. (2008). On Bayesian analysis and computation for functions with monotonicity and curvature restrictions, Journal of Econometrics 142 (1), 484–507. O'Donnell,

C. (2018). Productivity and efficiency analysis: An economic approach to measuring and explaining managerial performance, Springer. O'Donnell, C.J., and T. Coelli (2005). A Bayesian approach to imposing curvature on distance functions. Journal of Econometrics 126 (5), 493–523. O'Donnell, C.J., A.N. Rambaldi and H.E. Doran (2001). Estimating economic relationships subject to firm- and time-varying equality and inequality constraints, Journal of Applied Econometrics, 16 (6), 709–726. Parmeter, C.F., K. Sun, D. J. Henderson and S. C. Kumbhakar (2009). Regression and inference under smoothness restrictions, no date, presentation at the 44th Annual CAE. See also paper with same title, dated October 14, 2009, working paper. Paul, S., and S. Shankar (2018). On estimating efficiency effects in a stochastic frontier model. European Journal of Operational Research 271 (2), 769–774. Perrakis, K., I. Ntzoufras, and M. G. Tsionas (2015). On the use of marginal posteriors in marginal likelihood estimation via importance sampling. Computational Statistics & Data Analysis, 77 (C), 54–69. Rungsuriyawiboon, S., and S. Stefanou, (2008), The dynamics of efficiency and productivity in u.s. electric utilities. Journal of Productivity Analysis, 30, 177–190. Terrell, D. (1996). Imposing monotonicity and concavity restrictions in flexible functional forms, Journal of Applied Econometrics 11, 179–194. Tsionas, M. G. (2000). Full Likelihood Inference in Normal-Gamma Stochastic Frontier Models. Journal of Productivity Analysis 13, 183–205. Tsionas, M. G. (2016). "When, Where, and How" of efficiency estimation: Improved procedures for stochastic frontier modeling. Journal of the American Statistical Association 112 (519), 948–965. Tsionas, M. G., and E. Mamatzakis (2019). Further results on estimating inefficiency effects in stochastic frontier models. European Journal of Operational Research, 275 (3), 1157–1164. Wolff, H., T. Heckelei, T., and R. C. Mittelhammer (2010). Imposing curvature and monotonicity on flexible functional forms: an efficient regional approach, Computational Economics 36 (4), 309–339. Vouldis, A. T., P. G. Michaelides, and M. G. Tsionas (2010). Estimating semi-parametric output distance functions with neural-based reduced form equations using LIML, Economic Modelling 27 (3), 697–704. Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, Wiley, New York.