Some Models for Stochastic Frontiers with Endogeneity

William E. Griffiths and Gholamreza Hajargasht

Department of Economics

University of Melbourne, Australia

May 9, 2015

Abstract

We consider mostly Bayesian estimation of stochastic frontier models where one-sided inefficiencies and/or the idiosyncratic error term are correlated with the regressors. We begin with a model where a Chamberlain-Mundlak device is used to relate a transformation of time-invariant effects to the regressors. This basic model is then extended in two directions: First an extra one-sided error term is added to allow for time-varying efficiencies. Second, a model with an equation for instrumental variables and a more general error covariance structure is introduced to accommodate correlations between both error terms and the regressors. An application of the first and second models to Philippines rice data is provided.

Keywords: Technical efficiency, Instrumental variables, Gibbs sampling,

JEL Classification: C11, D24, C23, C12

Corresponding Author

William Griffiths
Economics Department,
University of Melbourne
Vic 3010, Australia
Phone: +61 3 8344 3622
Fax: +61 3 8344 6899
Email: wegrif@unimelb.edu.au
Email for Hajargasht: har@unimelb.edu.au

1. Introduction

Studies of stochastic frontier models that allow for correlation between inefficiency effects and regressors are few and have been mainly done under a fixed effects framework in which a panel data model with a two-sided error term is estimated first, and the inefficiency effects are later estimated by subtracting the effects from their maximum (see e.g. Sickles 2005 and references cited therein). Given that stochastic frontier models are more commonly estimated based on a one-sided random effects assumption, it is useful to investigate estimation within a framework where the one-sided random effects are correlated with the regressors. Also of interest are methods for accommodating correlation between the idiosyncratic error term and the regressors. The purpose of this paper is to propose a relatively general approach to modelling of stochastic frontiers with endogeneity, where one-sided efficiency effects, and idiosyncratic error terms, can be correlated with the regressors. We show that by transforming the inefficiency term into a normally distributed random term and modelling endogeneity through the mean or covariance of the normal errors, a range of stochastic frontier models with endogeneity can be handled.

We first consider a panel stochastic frontier model in which correlations between the effects and the regressors are based on a generalisation of the correlated random effects model proposed by Mundlak (1978), extended by Chamberlain (1984), and described further by Wooldridge (2010). Inefficiency effects are assumed to be correlated with the regressors through the mean of a transformation of the inefficiency errors. The main focus is on a log transformation implying the inefficiency errors have a lognormal distribution whose first argument depends on the regressors. Pursuing Bayesian estimation of the model, we derive conditional posterior densities for the parameters and the inefficiency errors for use in a Gibbs sampler. We then extend the model in two directions. Following Colombi et al. (2012), we add a time-varying inefficiency error leading to a model with both time invariant (permanent) and time-varying (transient) inefficiency errors; endogeneity is assumed to occur through correlation between the regressors and the time-invariant error. Necessary changes to the previously specified conditional posteriors are described. The second extension is to a more general model where endogeneity can exist because both the inefficiency errors and the idiosyncratic errors are correlated with the regressors. So that estimation can proceed, a "reduced form" type equation with instrumental variables is added to the earlier model. Details of how to estimate the model using both maximum simulated likelihood and Bayesian methods are provided.

The paper is organised as follows. The basic Mundlak-type model where the mean of the transformed error is a function of the regressors is considered in Section 2. In Section 3 we extend this model to include both permanent and transient inefficiency errors. Specification and estimation of the model that makes provision for instrumental variables and accommodates endogeneity more generally are considered in Section 4. An application using Philippine rice data and the models from Sections 2 and 3 is provided in Section 5.

2. Modelling correlation with a Chamberlain-Mundlak device

In the first instance we consider the following random effects stochastic production frontier model with a time invariant inefficiency term

$$y_{it} = \mathbf{x}_{1,it}\boldsymbol{\beta} - u_i + v_{it} . \tag{2.1}$$

In equation (2.1), $i = 1,...,N$ indexes the firms and $t = 1,...,T$ indexes time, $\mathbf{x}_{1,it}$ is a row vector of functions of inputs (e.g., logs of inputs and squared logs of inputs), y_{it} represents the logarithm of output, $\mathbf{x}_{1,it}\boldsymbol{\beta}$ is the log of the frontier production function (e.g., translog), u_i is a non-negative random error which accounts for time-invariant inefficiency of firm *i*, and v_{it} is an idiosyncratic error assumed to be $i.i.d.\,N\left(0,\sigma^2\right)$. The model can also represent a stochastic cost frontier, with y_{it} being the logarithm of cost, by changing "$-u_i$" to "$+u_i$".

In view of recent developments in the stochastic frontier literature – see, for example, Parmeter and Kumbhakar (2014) – having a model with time-invariant inefficiencies can be considered too restrictive. However, we include this model in the first instance as a stepping stone to more realistic time-varying inefficiency models considered in Sections 3 and 4.

To model correlation between the inefficiency error u_i and some or all of the inputs we assume that there is a transformation of u_i, call it $H\left(u_i\right)$, that is normally distributed with a mean that depends on the firm averages of some of the inputs or functions of them. These functions of the inputs are collected in the vector $\mathbf{x}_{2,it}$ and their firm averages are given by $\bar{\mathbf{x}}_{2,i} = T^{-1}\sum_{t=1}^{T}\mathbf{x}_{2,it}$. The resulting endogeneity model for describing how the inefficiency error is correlated with the inputs is given by

$$H(u_i) = \bar{\mathbf{x}}_{2,i}\boldsymbol{\gamma} + e_i\,, \tag{2.2}$$

with $e_i \sim i.i.d.\,N\left(0,\lambda^2\right)$. The most convenient transformation in the sense that it leads to recognisable conditional posterior distributions for implementing Gibbs sampling is the logarithmic one, $H(u_i) = \ln(u_i)$, implying that u_i has a lognormal distribution. Other transformations [e.g., $(u_i^{\rho} - 1)/\rho$ for some values of ρ] are possible.[1]

Equation (2.2) is an extension of the model considered by Mundlak (1978) for a conventional random effects panel data model with correlated effects. Modelling of endogeneity in this way, and its extension by Chamberlain (1984), have been referred to as the Chamberlain-Mundlak device, a device which has proved to be very useful in the context of nonlinear panel data models with endogeneity. It has been applied to model endogeniety in probit, fractional response, Tobit, sample selection, count data, double hurdle, unbalanced panel models, and models with cluster sampling. See Wooldridge (2010) for a review and for references to these applications. Also, when $H(u_i) = \ln(u_i)$, equation (2.2) can be written as $u_i = \exp\left\{\bar{\mathbf{x}}_{2,i}\gamma\right\}u_i^*$ where $u_i^* = \exp(e_i)$, implying the model can also be viewed as a stochastic frontier model with scaling properties. Alvarez et al. (2006) have studied and argued in favour of the scaling property in the context of models with environmental variables.

2.1 Prior specification

For Bayesian estimation of the model in (2.1)-(2.2), we begin by specifying prior distributions, and then present the conditional posterior densities that can be used for Gibbs sampling. For $\boldsymbol{\beta}$, we use the noninformative prior $p(\boldsymbol{\beta}) \propto 1$; for the variance of v_{it}, we use $\sigma^{-2} \sim G\left(A_{\sigma}, B_{\sigma}\right)$ where $G(A_{\sigma}, B_{\sigma})$ denotes a gamma density with shape parameter A_{σ} and scale parameter B_{σ}; a truncated normal

[1] One can in fact assume any distribution for u_i (e.g., exponential), with its cdf denoted by $F(u_i)$, and use the transformation $H(u_i) = \lambda\Phi^{-1}\left(F(u_i)\right) + \bar{\mathbf{x}}_{2,i}\boldsymbol{\gamma}$, but the posterior density must then include extra parameters from $F(u_i)$.

distribution denoted by $\boldsymbol{\gamma} \sim TN\left(\underline{\boldsymbol{\gamma}}, \mathbf{V}_\gamma; \mathbf{L}, \mathbf{U}\right)$ is used for $\boldsymbol{\gamma}$. The truncated normal parameters $\underline{\boldsymbol{\gamma}}$ and $\mathbf{V}_\gamma$ are what would be the prior mean vector and covariance matrix for $\boldsymbol{\gamma}$ if there were no truncation; $\mathbf{L}$ and $\mathbf{U}$ are vectors containing the lower and upper truncation points for each of the elements in $\boldsymbol{\gamma}$. For λ two alternative priors were considered: a gamma prior on λ^{-2} and a truncated uniform prior on λ, written as $\lambda^{-2} \sim G\left(A_\lambda, B_\lambda\right)$ and $\lambda \sim U\left(a_\lambda, b_\lambda\right)$, respectively.

The choice of priors for $\boldsymbol{\beta}$ and σ^{-2} is standard. For $\boldsymbol{\gamma}$ and λ, we experimented with several alternative priors, considering in each case their implications for (1) MCMC convergence, and (2) the marginal prior distributions of the inefficiency errors and their efficiencies, defined as $r_i = \exp\left(-u_i\right)$. Truncating a normal prior for $\boldsymbol{\gamma}$ to values that lead to reasonable efficiency values led to more precise estimates and improved MCMC convergence. A gamma prior for λ^{-2} is in line with most traditional priors specified for variance parameters, while use of a uniform prior for standard deviations in hierarchical models (which bear some similarity to our model) has been advocated by Gelman (2006). We defer discussion on the setting of values for the prior parameters to the application in Section 5.

2.2 Conditional posterior densities

To use Gibbs sampling for estimation we begin by considering the conditional posterior densities when $H(u_i) = \ln(u_i)$ and the prior $\lambda^{-2} \sim G\left(A_\lambda, B_\lambda\right)$ is used. Define $\mathbf{u}' = \left(u_1, u_2, \ldots, u_N\right)$; let $\mathbf{X}$ be a matrix with NT rows and typical row $\mathbf{x}_{1,it}$ and $\mathbf{X}_2$ be a matrix with N rows and typical row $\bar{\mathbf{x}}_{2,i}$. The joint posterior kernel for $\Theta = \left(\beta, \sigma^{-2}, \gamma, \lambda^{-2}, \mathbf{u}\right)$ is

$$p\left(\Theta \mid \mathbf{y}, \mathbf{X}, \mathbf{X}_2\right) \propto p\left(\mathbf{y} \mid \mathbf{X}, \beta, \sigma^{-2}, \mathbf{u}\right) p\left(\mathbf{u} \mid \mathbf{X}_2, \gamma, \lambda^{-2}\right) p(\beta)\, p\left(\sigma^{-2}\right) p(\gamma)\, p\left(\lambda^{-2}\right)$$

$$\propto \left(\sigma^{-2}\right)^{NT/2+A_\sigma-1} \exp\left\{-\frac{\sigma^{-2}}{2}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} + u_i\right)^2 + 2B_\sigma\right]\right\}\left(\lambda^{-2}\right)^{N/2+A_\lambda-1}\left[\prod_{i=1}^{N} u_i^{-1}\right] \qquad (2.3)$$

$$\exp\left\{-\left[\frac{\lambda^{-2}}{2}\sum_{i=1}^{N}\left(\ln u_i - \bar{\mathbf{x}}_{2,i}\boldsymbol{\gamma}\right)^2 + 2B_\lambda\right]\right\}\exp\left\{-\frac{1}{2}\left(\gamma - \underline{\gamma}\right)' \mathbf{V}_\gamma^{-1}\left(\gamma - \underline{\gamma}\right)\right\}\left[\prod_{s=1}^{S} I\left(L_s \le \gamma_s \le U_s\right)\right]$$

where $I\left(L_s \le \gamma_s \le U_s\right)$ is an indicator function, L_s, U_s and γ_s are elements of $\mathbf{L}, \mathbf{U}$ and $\boldsymbol{\gamma}$, respectively, and S is the dimension of $\boldsymbol{\gamma}$. If we use the uniform prior $\lambda \sim U\left(a_\lambda, b_\lambda\right)$, then the joint posterior density can be obtained from (2.3) by setting $A_\lambda = 1, B_\lambda = 0$, and including the indicator function $I\left(a_\lambda \le \lambda \le b_\lambda\right)$. From equation (2.3), and using $\mathbf{D} = \{\mathbf{y}, \mathbf{X}, \mathbf{X}_2\}$ to denote the available data, the following conditional posterior densities can be derived:

$$\left(\boldsymbol{\beta} \mid \boldsymbol{\Theta}_{-\beta}, \mathbf{D}\right) \sim N\left\{\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{y} + \mathbf{u} \otimes \mathbf{i}_T),\ \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right\}, \tag{2.4}$$

$$\left(\sigma^{-2} \mid \boldsymbol{\Theta}_{-\sigma^{-2}}, \mathbf{D}\right) \sim G\left(A_\sigma + NT/2,\ B_\sigma + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} + u_i\right)^2\right), \tag{2.5}$$

$$\left(\boldsymbol{\gamma} \mid \boldsymbol{\Theta}_{-\gamma}, \mathbf{D}\right) \sim TN\left\{\left(\lambda^{-2}\mathbf{X}_2'\mathbf{X}_2 + \mathbf{V}_\gamma^{-1}\right)^{-1}\left(\lambda^{-2}\mathbf{X}_2' \ln \mathbf{u} + \mathbf{V}_\gamma^{-1}\underline{\boldsymbol{\gamma}}\right),\ \left(\lambda^{-2}\mathbf{X}_2'\mathbf{X}_2 + \mathbf{V}_\gamma^{-1}\right)^{-1}\right\}\left[\prod_{s=1}^{S} I\left(L_s \le \gamma_s \le U_s\right)\right] \tag{2.6}$$

$$\left(\lambda^{-2} \mid \boldsymbol{\Theta}_{-\lambda^{-2}}, \mathbf{D}\right) \sim G\left\{\frac{N}{2} + A_\lambda,\ \frac{1}{2}\sum_{i=1}^{N}\left(\ln u_i - \overline{\mathbf{x}}_{2,i}\boldsymbol{\gamma}\right)^2 + B_\lambda\right\}, \tag{2.7}$$

$$p\left(u_i \mid \boldsymbol{\Theta}_{-u_i}, \mathbf{D}\right) \propto \exp\left\{\left(\lambda^{-2}\overline{\mathbf{x}}_{2,i}\boldsymbol{\gamma} - 1\right)\ln u_i - \lambda^{-2}\left(\ln u_i\right)^2/2 - \left(\sigma^{-2}/2\right)\left[Tu_i^2 + 2u_i\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta}\right)\right]\right\}. \tag{2.8}$$

When the prior $\lambda \sim U\left(a_\lambda, b_\lambda\right)$ is employed, the conditional posterior density for λ^{-2} becomes the truncated gamma density

$$\left(\lambda^{-2} \mid \boldsymbol{\Theta}_{-\lambda^{-2}}, \mathbf{D}\right) \sim TG\left\{\frac{N}{2} + 1,\ \frac{1}{2}\sum_{i=1}^{N}\left(\ln u_i - \overline{\mathbf{x}}_{2,i}\boldsymbol{\gamma}\right)^2\right\} I\left(a_\lambda \le \lambda \le b_\lambda\right). \tag{2.9}$$

All the densities in (2.4)-(2.8) are recognised densities which are straightforward to draw from, with the exception of $p\left(u_i \mid \boldsymbol{\Theta}_{-u_i}, \mathbf{D}\right)$. Depending on available software, it might also be less straightforward to draw from the truncated gamma density in (2.9). Since these exceptions are univariate distributions, a convenient method for drawing from them is the slice sampler of Neal (2003). If we do not set $H(u_i) = \ln(u_i)$, but instead use the general expression $H(u_i)$, then $\ln(u_i)$ is replaced by $H(u_i)$ in (2.6) and (2.7), and (2.8) becomes

$$p\left(u_i \mid \boldsymbol{\Theta}_{-u_i}, \mathbf{D}\right) \propto \frac{dH(u_i)}{du_i}\exp\left\{\lambda^{-2}\overline{\mathbf{x}}_{2,i}\boldsymbol{\gamma}\,H(u_i) - \lambda^{-2}\left(H(u_i)\right)^2/2 - \left(\sigma^{-2}/2\right)\left[Tu_i^2 + 2u_i\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta}\right)\right]\right\}. \tag{2.10}$$

3. Extension to a time-varying inefficiency model

A deficiency of the model considered in the previous section, and one that is likely to be particularly critical if the number of time periods is large, is the time invariance of the firm inefficiencies. One way to remedy this deficiency is to specify the inefficiency error as u_{it}, allowing it to vary freely over both firms and time. In this case we can specify $H(u_{it}) = x_{2,it}\gamma + e_{it}$ and derive a corresponding set of conditional posterior densities. We do so in Section 4, but for a more general model that also accommodates other forms of endogeneity. Another way to allow for time varying inefficiency is to add an extra one-sided random term η_{it} in the spirit of the generalised random effects model of Colombi et al. (2012), Kumbhakar and Tsionas (2014) and Filippini and Greene (2014). That is,

$$\begin{aligned} y_{it} &= \mathbf{x}_{1,it}\boldsymbol{\beta} - u_i - \eta_{it} + v_{it} \\ H(u_i) &= \bar{\mathbf{x}}_{2,i}\boldsymbol{\gamma} + e_i \end{aligned} . \qquad (3.1)$$

Here, u_i represents permanent and η_{it} transient inefficiencies. Ignoring the different components can lead to misleading estimates of inefficiency[2]. We assume only the permanent inefficiencies are correlated with the regressors, a situation likely to hold if the u_i represent systematic inefficiencies attributable to long-term input use and the short-run inefficiencies are less predictable and not within the control of the firm.

In addition to the assumptions of the previous section, we assume the η_{it} are i.i.d. and follow the exponential distribution $p(\eta_{it} \mid \delta) = \delta \exp(-\delta\eta_{it})$, and that η_{it}, u_i and v_{it} are uncorrelated. Using the same priors as before, the gamma prior for λ^{-2}, a gamma prior $G(A_\delta, B_\delta)$ for δ, and the transformation $H(u_i) = \ln u_i$, we can derive the following conditional posterior densities for the elements in $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \sigma^{-2}, \gamma, \lambda^{-2}, \delta, \mathbf{u}, \boldsymbol{\eta})$:

$$(\boldsymbol{\beta} \mid \boldsymbol{\Theta}_{-\beta}, \mathbf{D}) \sim N\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}(\mathbf{y} + \mathbf{u} \otimes \mathbf{i}_T + \boldsymbol{\eta}),\ \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}\right\}, \qquad (3.2)$$

[2] In the application in Section 5, we find that inclusion of η_{it} leads to a much larger estimate of total inefficiency.

$$\left(\sigma^{-2} \mid \boldsymbol{\Theta}_{-\sigma^{-2}}, \mathbf{D}\right) \sim G\left(A_\sigma + NT/2,\ B_\sigma + \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} + u_i + \eta_{it}\right)^2\right), \tag{3.3}$$

$$\left(\eta_{it} \mid \boldsymbol{\Theta}_{-\eta_{it}}, \mathbf{D}\right) \sim TN\left\{\mathbf{x}_{1,it}\boldsymbol{\beta} - u_i - y_{it} - \sigma^2\delta, \sigma^2\right\} I\left(\eta_{it} > 0\right), \tag{3.4}$$

$$\left(\delta \mid \boldsymbol{\Theta}_{-\delta}, \mathbf{D}\right) \sim G\left(A_\delta + NT, B_\delta + \sum_{i=1}^{N}\sum_{t=1}^{T}\eta_{it}\right), \tag{3.5}$$

$$\left(\boldsymbol{\gamma} \mid \boldsymbol{\Theta}_{-\gamma}, \mathbf{D}\right) \sim TN\left\{\left(\lambda^{-2}\mathbf{X}_2'\mathbf{X}_2 + \mathbf{V}_\gamma^{-1}\right)^{-1}\left(\lambda^{-2}\mathbf{X}_2' \ln\mathbf{u} + \mathbf{V}_\gamma^{-1}\underline{\boldsymbol{\gamma}}\right),\ \left(\lambda^{-2}\mathbf{X}_2'\mathbf{X}_2 + \mathbf{V}_\gamma^{-1}\right)^{-1}\right\}\left[\prod_{s=1}^{S} I\left(L_s \le \gamma_s \le U_s\right)\right] \tag{3.6}$$

$$\left(\lambda^{-2} \mid \boldsymbol{\Theta}_{-\lambda^{-2}}, \mathbf{D}\right) \sim G\left\{\frac{N}{2} + A_\lambda,\ \frac{1}{2}\sum_{i=1}^{N}\left(\ln u_i - \overline{\mathbf{x}}_{2,i}\boldsymbol{\gamma}\right)^2 + B_\lambda\right\}, \tag{3.7}$$

$$p\left(u_i \mid \boldsymbol{\Theta}_{-u_i}, \mathbf{D}\right) \propto \exp\left\{\left(\lambda^{-2}\overline{\mathbf{x}}_{2,i}\boldsymbol{\gamma} - 1\right)\ln u_i - \lambda^{-2}\left(\ln u_i\right)^2/2 - \left(\sigma^{-2}/2\right)\left[Tu_i^2 + 2u_i\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} + \eta_{it}\right)\right]\right\}. \tag{3.8}$$

As before, all these densities are of recognisable forms from which observations can be drawn directly, except for $p\left(u_i \mid \boldsymbol{\Theta}_{-u_i}, \mathbf{D}\right)$, which will require a Metropolis step or the slice sampler. An estimate of the *i*-th firm's permanent inefficiency can be found from the mean of the post burn-in draws from $p\left(u_i \mid \boldsymbol{\Theta}_{-u_i}, \mathbf{D}\right)$. To convert measurement of inefficiency to a measure of efficiency the draws on u_i are transformed to $r_i = \exp\{-u_i\}$, and the mean of the r_i is an estimate of the posterior mean of permanent efficiency. Similarly, draws for η_{it} can be used to find an estimate of transient efficiency for the *i*-th firm in the *t*-th time period, $\exp\{-\eta_{it}\}$. Posterior standard deviations estimated from the draws indicate the reliability of these estimates.

4. A model with full endogeneity and instrumental variables

In the previous two sections endogeneity was modelled as correlation between the inefficiency errors u_i and the inputs. However, in a number of studies (e.g., Kutlu 2010, Karakaplan and Kutlu 2013, Tran and Tsionas 2013) allowance is made for correlations between idiosyncratic error terms and the inputs. In this section we consider a model that, in its most general form, allows for (i) time varying inefficiencies, (ii) correlation between the inputs and both the inefficiency error and the idiosyncratic error, (iii) correlation between the two types of errors, and (iv) the introduction of instrumental variables. We write the general model as

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{1,it}\boldsymbol{\beta} - u_{it} + v_{1,it} \\
H(u_{it}) &= \mathbf{x}_{2,it}\boldsymbol{\gamma} + v_{2,it} \ , \\
\mathbf{x}'_{it} &= \left(\mathbf{I} \otimes \mathbf{z}_{it}\right)\boldsymbol{\pi} + \mathbf{v}_{it}
\end{aligned}
\tag{4.1}
$$

where $\mathbf{z}_{it}$ is a $(1\times m)$ vector of instrumental variables, and $\mathbf{x}_{it}$ is $(1\times k)$ vector of log-inputs, differing from $\mathbf{x}_{1,it}$ and $\mathbf{x}_{2,it}$ in that the latter may contain various transformations of the inputs. The $(mk\times 1)$ vector $\boldsymbol{\pi}$ contains the parameters from the “reduced form” equations for $\mathbf{x}_{it}$ .The error terms $\left(v_{1,it}, v_{2,it}, \mathbf{v}'_{it}\right)$ are assumed to be normally distributed with zero mean, uncorrelated over firms and time, and with endogeneity modelled through the $\left((k+2)\times(k+2)\right)$ covariance matrix

$$
\Sigma = \mathrm{cov}\left(v_{1,it}, v_{2,it}, \mathbf{v}'_{it}\right) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \boldsymbol{\Sigma}_{1v} \\ \sigma_{21} & \sigma_{22} & \boldsymbol{\Sigma}_{2v} \\ \boldsymbol{\Sigma}_{v1} & \boldsymbol{\Sigma}_{v2} & \boldsymbol{\Sigma}_{vv} \end{pmatrix}. \tag{4.2}
$$

Assuming $\boldsymbol{\Sigma}_{1v} \neq \mathbf{0}$ leads to correlation between the inputs and the idiosyncratic error $v_{1,it}$. Correlation between the inefficiency error u_{it} and the inputs arises from both the second equation and from $\boldsymbol{\Sigma}_{2v} \neq \mathbf{0}$. Simplified versions of the model that still allow for endogeneity with respect to u_{it} can be obtained by dropping the inputs from the second equation, making it $H(u_{it}) = \gamma_0 + v_{2,it}$, or by setting $\boldsymbol{\Sigma}_{2v} = \mathbf{0}$. Finally, having $\sigma_{12} \neq 0$ means there can be correlation between the inefficiency and idiosyncratic errors. To exclude this correlation we can set $\sigma_{12} = 0$.

The model in (4.1) extends the models considered in Kutlu (2010) and Tran and Tsionas (2013) in which only correlations between regressors and the conventional error terms are allowed – they have only the first and third equations in (4.1). Kutlu uses a two-step maximum likelihood procedure while Tran and Tsionas propose a GMM estimation method. Karakaplan and Kutlu (2013) consider a model that allows for full endogeneity but it is to some extent different from our model and our estimation methods are very different.

4.1 Estimation: some preliminaries

We consider two methods of estimating the model: maximum simulated likelihood, and Bayesian estimation via Gibbs sampling. As a starting point for deriving the likelihood function for both methods, we write

$$p(y_{it},\mathbf{x}_{it}) = \int_0^{\infty} p(y_{it}\,|\,\mathbf{x}_{it},u_{it})\,p(u_{it}\,|\,\mathbf{x}_{it})\,p(\mathbf{x}_{it})\,du_{it}\,. \qquad (4.3)$$

The likelihood for a single observation is obtained by multiplying together the three densities on the right-hand side of (4.3) and then integrating out u_i. Working in this direction, from the third equation in (4.1), we have the following normal distribution for $p(\mathbf{x}_{it})$

$$p(\mathbf{x}_{it}) = N\left((\mathbf{I}\otimes\mathbf{z}_{it})\boldsymbol{\pi},\ \boldsymbol{\Sigma}_{vv}\right). \qquad (4.4)$$

To find $p(u_{it}\,|\,\mathbf{x}_{it})$, we begin by noting that

$$E\left[H(u_{it})\,|\,\mathbf{x}_{it}\right] = \mathbf{x}_{2,it}\boldsymbol{\gamma} + E\left(v_{2,it}\,|\,\mathbf{v}_{it}\right) \quad \text{and} \quad \mathrm{var}\left[H(u_{it})\,|\,\mathbf{x}_{it}\right] = \mathrm{var}\left(v_{2,it}\,|\,\mathbf{v}_{it}\right),$$

where

$$E\left(v_{2,it}\,|\,\mathbf{v}_{it}\right) = \boldsymbol{\Sigma}_{2v}\boldsymbol{\Sigma}_{vv}^{-1}\left[\mathbf{x}'_{it} - (\mathbf{I}\otimes\mathbf{z}_{it})\boldsymbol{\pi}\right] \quad \text{and} \quad \sigma_{2|v} = \mathrm{var}\left(v_{2,it}\,|\,\mathbf{v}_{it}\right) = \sigma_{22} - \boldsymbol{\Sigma}_{2v}\boldsymbol{\Sigma}_{vv}^{-1}\boldsymbol{\Sigma}_{v2}\,.$$

Thus,

$$p\left(H(u_{it})\,|\,\mathbf{x}_{it}\right) = N\left(\mathbf{x}_{2,it}\boldsymbol{\gamma} + E(v_{2,it}\,|\,\mathbf{v}_{it}),\ \sigma_{2|v}\right),$$

and

$$p\left(u_{it}\,|\,\mathbf{x}_{it}\right) = \frac{dH(u_{it})}{du_{it}}\,N\left(\mathbf{x}_{2,it}\boldsymbol{\gamma} + E(v_{2,it}\,|\,\mathbf{v}_{it}),\ \sigma_{2|v}\right). \qquad (4.5)$$

To obtain $p\left(y_{it}\,|\,\mathbf{x}_{it},u_{it}\right)$, we note that

$$E\left(y_{it}\,|\,\mathbf{x}_{it},u_{it}\right) = \mathbf{x}_{1,it}\boldsymbol{\beta} - u_{it} + E\left(v_{1,it}\,|\,v_{2,it},\mathbf{v}_{it}\right) \quad \text{and} \quad \mathrm{var}\left(y_{it}\,|\,\mathbf{x}_{it},u_{it}\right) = \mathrm{var}\left(v_{1,it}\,|\,v_{2,it},\mathbf{v}_{it}\right),$$

where

$$E\left(v_{1,it}\,|\,v_{2,it},\mathbf{v}_{it}\right) = \begin{bmatrix}\sigma_{12} & \boldsymbol{\Sigma}_{1v}\end{bmatrix}\begin{bmatrix}\sigma_{22} & \boldsymbol{\Sigma}_{2v}\\ \boldsymbol{\Sigma}_{v2} & \boldsymbol{\Sigma}_{vv}\end{bmatrix}^{-1}\begin{bmatrix}H(u_{it}) - \mathbf{x}_{2,it}\boldsymbol{\gamma}\\ \mathbf{x}'_{it} - (\mathbf{I}\otimes\mathbf{z}_{it})\boldsymbol{\pi}\end{bmatrix},$$

and

$$\sigma_{1|2,v} = \mathrm{var}\left(v_{1,it}\,|\,v_{2,it},\mathbf{v}_{it}\right) = \sigma_{11} - \begin{bmatrix}\sigma_{12} & \boldsymbol{\Sigma}_{1v}\end{bmatrix}\begin{bmatrix}\sigma_{22} & \boldsymbol{\Sigma}_{2v}\\ \boldsymbol{\Sigma}_{v2} & \boldsymbol{\Sigma}_{vv}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{21}\\ \boldsymbol{\Sigma}_{v1}\end{bmatrix},$$

leading to

$$p\left(y_{it} \mid \mathbf{x}_{it}, u_{it}\right) = N\left(\mathbf{x}_{1,it}\boldsymbol{\beta} - u_{it} + E\left(v_{1,it} \mid v_{2,it}, \mathbf{v}_{it}\right), \sigma_{1|2,v}\right). \tag{4.6}$$

After combing (4.4), (4.5), and (4.6), the integral with respect to u_{it} is intractable. To overcome this problem we consider in turn maximum simulated likelihood, and then Bayesian estimation.

4.2 Estimation via maximum simulated likelihood

An estimate of the likelihood for a single observation is given by $R^{-1}\sum_{r=1}^{R} p\left(y_{it} \mid \mathbf{x}_{it}, u_{it}^{(r)}\right) p\left(\mathbf{x}_{it}\right)$ where $u_{it}^{(r)}$ is the *r*-th draw $(r = 1, 2, \ldots, R)$ from $p(u_{it} \mid \mathbf{x}_i)$. This draw can be obtained from

$$u_{it}^{(r)} = H^{-1}\left\{\mathbf{x}_{2,it}\boldsymbol{\gamma} + E\left(v_{2,it} \mid \mathbf{v}_{it}\right) + \sqrt{\sigma_{2|1}}\Phi^{-1}\left(\xi_{it}^{(r)}\right)\right\}, \tag{4.7}$$

where $\Phi(\cdot)$ is the standard normal cdf, and the $\xi_{it}^{(r)}$ are independent draws from a uniform $(0,1)$ distribution. The simulated log-likelihood function can be written as

$$\begin{aligned} L = &\sum_{t=1}^{T}\sum_{i=1}^{N}\ln\left\{\frac{1}{R}\sum_{r=1}^{R} N\left(\mathbf{x}_{1,it}\boldsymbol{\beta} - u_{it}^{(r)} + \begin{bmatrix}\sigma_{12} & \boldsymbol{\Sigma}_{1v}\end{bmatrix}\begin{bmatrix}\sigma_{22} & \boldsymbol{\Sigma}_{2v} \\ \boldsymbol{\Sigma}_{v2} & \boldsymbol{\Sigma}_{vv}\end{bmatrix}^{-1}\begin{bmatrix}H\left(u_{it}^{(r)}\right) - \mathbf{x}_{2,it}\boldsymbol{\gamma} \\ \mathbf{x}_{it}' - \left(\mathbf{I} \otimes \mathbf{z}_{it}\right)\boldsymbol{\pi}\end{bmatrix}, \sigma_{1|2,v}\right)\right\} \\ &+ \sum_{t=1}^{T}\sum_{i=1}^{N}\ln\left\{N\left(\left(\mathbf{I} \otimes \mathbf{z}_{it}\right)\boldsymbol{\pi}, \boldsymbol{\Sigma}_{vv}\right)\right\} \end{aligned}. \tag{4.8}$$

Maximizing this likelihood function with respect to the parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\pi}, \boldsymbol{\Sigma})$ gives estimates for these parameters. Using these estimates, and for each $(NT \times 1)$ vector of draws $\mathbf{u}^{(r)}$, we can find an estimate of the conditional density for $\mathbf{y}$ and $\mathbf{X}$ defined as $\hat{p}\left(\mathbf{y}, \mathbf{X} \mid \mathbf{u}^{(r)}\right) = \prod_{t=1}^{T}\prod_{i=1}^{N} \hat{p}\left(y_{it}, \mathbf{x}_{it} \mid u_{it}^{(r)}\right)$. Then, to obtain estimates (predictions) of each of the inefficiency errors u_{it}, we recognize that

$$E\left(u_{it} \mid \mathbf{y}, \mathbf{X}\right) = \int_0^\infty u_{it}\, p\left(\mathbf{u} \mid \mathbf{y}, \mathbf{X}\right) d\mathbf{u} = \frac{\int_0^\infty u_{it}\, p\left(\mathbf{y}, \mathbf{X} \mid \mathbf{u}\right) p(\mathbf{u}) d\mathbf{u}}{p(\mathbf{y}, \mathbf{X})},$$

and estimate this mean using

$$\hat{u}_{it} = \frac{\sum_{r=1}^{R} u_{it}^{(r)} \hat{p}\left(\mathbf{y}, \mathbf{X} \mid \mathbf{u}^{(r)}\right)}{\sum_{r=1}^{R} \hat{p}\left(\mathbf{y}, \mathbf{X} \mid \mathbf{u}^{(r)}\right)}. \tag{4.9}$$

4.3 Bayesian estimation

For Bayesian estimation, we derive the conditional posterior densities for each component of $\boldsymbol{\Theta} = \left(\beta, \gamma, \pi, \mathbf{u}, \boldsymbol{\Sigma}^{-1}\right)$ for use in a Gibbs sampler. Extending the definition of $\mathbf{D}$ to include observations on $\mathbf{z}_{it}$ as well as those for y_{it}, $\mathbf{x}_{1,it}$ and $\mathbf{x}_{2,it}$, we begin with the conditional posterior for $\boldsymbol{\Sigma}^{-1}$. Assuming the Wishart prior $\boldsymbol{\Sigma}^{-1} \sim W(\underline{\mathbf{S}}^{-1}, \underline{d})$, we can show that the conditional posterior for $\boldsymbol{\Sigma}^{-1}$ is also Wishart. That is, $\left(\boldsymbol{\Sigma}^{-1} \mid \boldsymbol{\Theta}_{-\Sigma}, \mathbf{D}\right) \sim W\left(\overline{\mathbf{S}}^{-1}, \overline{d}\right)$, with arguments

$$\overline{\mathbf{S}} = \underline{\mathbf{S}} + \sum_{t=1}^{T}\sum_{i=1}^{N}\begin{bmatrix} y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} + u_{it} \\ H(u_{it}) - \mathbf{x}_{2,it}\boldsymbol{\gamma} \\ \mathbf{x}'_{it} - \left(\mathbf{I} \otimes \mathbf{z}_{it}\right)\boldsymbol{\pi} \end{bmatrix}\begin{bmatrix} y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} + u_{it} \\ H(u_{it}) - \mathbf{x}_{2,it}\boldsymbol{\gamma} \\ \mathbf{x}'_{it} - \left(\mathbf{I} \otimes \mathbf{z}_{it}\right)\boldsymbol{\pi} \end{bmatrix}' \quad \text{and} \quad \overline{d} = \underline{d} + NT.$$

For $\boldsymbol{\beta}$, we assume the normal prior $\boldsymbol{\beta} \sim N\left(\underline{\boldsymbol{\beta}}, \underline{\mathbf{V}}_{\beta}\right)$. Then, using an argument similar to that which resulted in equation (4.6) leads to the normal conditional posterior $\left(\boldsymbol{\beta} \mid \boldsymbol{\Theta}_{-\beta}, \mathbf{D}\right) \sim N\left(\overline{\boldsymbol{\beta}}, \overline{\mathbf{V}}_{\beta}\right)$ where

$$\overline{\boldsymbol{\beta}} = \overline{\mathbf{V}}_{\beta}\left[\underline{\mathbf{V}}_{\beta}^{-1}\underline{\boldsymbol{\beta}} + \sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{x}'_{1,it}\sigma_{1|2,v}^{-1}\left(y_{it} - E\left(v_{1,it} \mid v_{2,it}, \mathbf{v}_{it}\right) + u_{it}\right)\right] \quad \text{and} \quad \overline{\mathbf{V}}_{\beta} = \left(\underline{\mathbf{V}}_{\beta} + \sigma_{1|2,v}^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{x}_{1,it}\mathbf{x}'_{1,it}\right)^{-1}.$$

Assuming a normal prior for $\boldsymbol{\gamma} \sim N\left(\underline{\boldsymbol{\gamma}}, \underline{\mathbf{V}}_{\gamma}\right)$, and a similar trick, leads to the normal conditional posterior $\left(\boldsymbol{\gamma} \mid \boldsymbol{\Theta}_{-\gamma}, \mathbf{D}\right) \sim N\left(\overline{\boldsymbol{\gamma}}, \overline{\mathbf{V}}_{\gamma}\right)$, where $\overline{\mathbf{V}}_{\gamma} = \left(\underline{\mathbf{V}}_{\gamma} + \sigma_{2|1,v}^{-1}\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{x}_{2,it}\mathbf{x}'_{2,it}\right)^{-1}$ and

$$\overline{\boldsymbol{\gamma}} = \overline{\mathbf{V}}_{\gamma}\left[\underline{\mathbf{V}}_{\gamma}^{-1}\overline{\boldsymbol{\gamma}} + \sum_{t=1}^{T}\sum_{i=1}^{N}\mathbf{x}'_{2,it}\sigma_{2|1,v}^{-1}\left(H(u_{it}) - E\left(v_{2,it} \mid v_{1,it}, \mathbf{v}_{it}\right)\right)\right],$$

with

$$E\left(v_{2,it} \mid v_{1,it}, \mathbf{v}_{it}\right) = \begin{bmatrix}\sigma_{21} & \boldsymbol{\Sigma}_{v2}\end{bmatrix}\begin{bmatrix}\sigma_{11} & \boldsymbol{\Sigma}_{1v} \\ \boldsymbol{\Sigma}_{v1} & \boldsymbol{\Sigma}_{vv}\end{bmatrix}^{-1}\begin{bmatrix} y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} \\ \mathbf{x}'_{it} - \mathbf{I} \otimes \mathbf{z}_{it}\end{bmatrix},$$

and

$$\sigma_{2|1,v} = \sigma_{22}^{2} - \begin{bmatrix}\sigma_{21} & \boldsymbol{\Sigma}_{v2}\end{bmatrix}\begin{bmatrix}\sigma_{11} & \boldsymbol{\Sigma}_{1v} \\ \boldsymbol{\Sigma}_{v1} & \boldsymbol{\Sigma}_{vv}\end{bmatrix}^{-1}\begin{bmatrix}\sigma_{12} \\ \boldsymbol{\Sigma}_{v2}\end{bmatrix}.$$

A similar argument can be used for $\boldsymbol{\pi}$. With the normal prior $\boldsymbol{\pi} \sim N\left(\underline{\boldsymbol{\pi}}, \underline{\mathbf{V}}_{\pi}\right)$, we obtain the conditional posterior $\left(\boldsymbol{\pi} \mid \boldsymbol{\Theta}_{-\pi}, \mathbf{D}\right) \sim N\left(\bar{\boldsymbol{\pi}}, \bar{\mathbf{V}}_{\pi}\right)$ where $\bar{\mathbf{V}}_{\pi} = \left(\underline{\mathbf{V}}_{\pi} + \sum_{t=1}^{T}\sum_{i=1}^{N}\left(\mathbf{I} \otimes \mathbf{z}'_{it}\right)\boldsymbol{\Sigma}_{v|1,2}^{-1}\left(\mathbf{I} \otimes \mathbf{z}_{it}\right)\right)^{-1}$, and

$$\bar{\boldsymbol{\pi}} = \bar{\mathbf{V}}_{\pi}\left[\underline{\mathbf{V}}_{\pi}^{-1}\boldsymbol{\pi} + \sum_{t=1}^{T}\sum_{i=1}^{N}\left(\mathbf{I} \otimes \mathbf{z}'_{it}\right)\boldsymbol{\Sigma}_{v|1,2}^{-1}\left[\mathbf{x}'_{it} - E\left(\mathbf{v}_{it} \mid v_{1,it}, v_{2,it}\right)\right]\right],$$

with

$$E\left(\mathbf{v}_{it} \mid v_{1,it}, v_{2,it}\right) = \begin{bmatrix}\boldsymbol{\Sigma}_{v1} & \boldsymbol{\Sigma}_{v2}\end{bmatrix}\begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22}\end{bmatrix}^{-1}\begin{bmatrix} y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} \\ H(u_{it}) - \mathbf{x}_{2,it}\boldsymbol{\gamma}\end{bmatrix},$$

and

$$\boldsymbol{\Sigma}_{v|1,2} = \boldsymbol{\Sigma}_{vv} - \begin{bmatrix}\boldsymbol{\Sigma}_{v1} & \boldsymbol{\Sigma}_{v2}\end{bmatrix}\begin{bmatrix}\sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{\Sigma}_{1v} \\ \boldsymbol{\Sigma}_{2v}\end{bmatrix}.$$

Finally, the conditional posterior distribution for u_{it} can be written as

$$p\left(u_{it} \mid \boldsymbol{\Theta}_{-u_{it}}, \mathbf{D}\right) \propto \frac{dH(u_{it})}{du_{it}} \exp\left\{ -\frac{\left[y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} + u_{it} - E\left(v_{1,it} \mid v_{2,it}, \mathbf{v}_{it}\right)\right]^2}{2\sigma_{1|2,v}} - \frac{\left[H\left(u_{it}\right) - \mathbf{x}_{2,it}\boldsymbol{\gamma} - E\left(v_{2,it} \mid \mathbf{v}_{it}\right)\right]^2}{2\sigma_{2|v}} \right\}$$

With the exception of $p\left(u_{it} \mid \boldsymbol{\Theta}_{-u_{it}}, \mathbf{D}\right)$, all conditional distributions are recognizable distributions from which we can draw observations; for $p\left(u_{it} \mid \boldsymbol{\Theta}_{-u_{it}}, \mathbf{D}\right)$ we can use slice sampling or a Metropolis step. As before, observations from $p\left(u_{it} \mid \boldsymbol{\Theta}_{-u_{it}}, \mathbf{D}\right)$ form the basis of inferences about inefficiencies for the *i*-th firm in the *t*-th time period.

5. An application to Philippines rice data

Since at least the 1970s some studies have reported an inverse relationship between farm size and productivity (efficiency) in developing countries (see e.g., Bardhan 1973 or Sen 1975), or they have argued that such a relationship exists because smaller firms use better-motivated or monitored family labor, whereas bigger farms use less-motivated hired labor. Imperfections in the labor or credit markets have also been put forward as possible reasons for the relationship. However, other studies have either not found such a relationship (e.g., Lamb 2003) or have argued that such observations might be due to omitted variable biases, such as smaller farms having better soil quality. Our purpose here is not to resolve this long-standing issue. Rather, we apply the models discussed in Sections 2 and 3 to a Philippines rice data set and check whether we find any evidence of correlation between the permanent part of inefficiency and land size.

The widely-used Philippines rice data collected by the International Rice Research Institute consist of a panel of 43 Philippine rice farms observed over 8 years from 1990 to 1997 (see Coelli, et al. 2005 for further information). We use the last 4 years of the data because of the time-invariance assumption in one of the models. Following Section 2, the first model we consider is

$$y_{it} = \beta_0 + \beta_1 land_{it} + \beta_2 labor_{it} + \beta_3 fert_{it} + \beta_4 others_{it} - u_i + v_{it}\,, \tag{5.1}$$

where y_{it}, $land_{it}$, $labor_{it}$, $fert_{it}$ and $others_{it}$ are the logarithms of output, land, hired labour, amount of fertilizer and other inputs, respectively. Use of the Cobb-Douglas function for the frontier is in line with several other studies that have used this data set. We assume time-invariant inefficiencies and allow them to be correlated with land size through

$$u_i = LN\left(\gamma_0 + \gamma_1 \overline{land}_i + \gamma_2 \overline{land}_i^2, \lambda^2\right). \tag{5.2}$$

The square of $\overline{land}$ is included to account for a potential nonlinear relationship. The second model we consider is identical except that, following Section 3, an extra inefficiency error, $-\eta_{it}$, is included so that the model has both permanent and transient inefficiencies. It is assumed that η_{it} follows an exponential distribution with parameter δ. So that we can compare estimates obtained with and without the endogeneity assumption, we also estimate these models assuming that $\gamma_1 = \gamma_2 = 0$.

5.1 Prior distributions

The priors that we used are as follows. In all cases, we have $\sigma^{-2} \sim G(0.01, 0.01)$, $p(\boldsymbol{\beta}) \propto 1$, and $\lambda \sim U(0.1, 2)$. For the models with endogeneity, $\boldsymbol{\gamma} \sim TN\left(\underline{\boldsymbol{\gamma}}, \mathbf{V}_\gamma; \mathbf{L}, \mathbf{U}\right)$, with

$$\underline{\boldsymbol{\gamma}} = \begin{pmatrix} -2 \\ 0 \\ 0 \end{pmatrix} \qquad \mathbf{V}_\gamma = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad \mathbf{L} = \begin{pmatrix} -4 \\ -1.5 \\ -1.5 \end{pmatrix} \qquad \mathbf{U} = \begin{pmatrix} 0 \\ 1.5 \\ 1.5 \end{pmatrix}.$$

When we set $\gamma_1 = \gamma_2 = 0$, the first components in these vectors were used for γ_0. For the model with a term for time-varying inefficiency, we used $\delta \sim G(0.1, 0.1)$. With the exception of those for $\boldsymbol{\gamma}$ and λ, these priors can be regarded as noninformative. Before turning to the results, it is worth digressing to discuss the issues involved when choosing prior parameter values for $\boldsymbol{\gamma}$ and λ.

Previous work that assumed u_i is exponential with a constant scale parameter often used a relatively noninformative prior for that parameter such that the median of the resulting prior for efficiency, $r_i = \exp(-u_i)$, is 0.87. See, for example, Koop and Steel (2001). To see how a similar marginal prior for efficiencies can be constructed when u_i follows a lognormal distribution, we first consider setting values $\left(\underline{\gamma}_0, V_{\gamma 0}, L_0, U_0\right)$ for the case where $u_i \sim LN\left(\gamma_0, \lambda^2\right)$. Given a reasonable prior median for efficiency, say r^*, we can set $\underline{\gamma}_0$ to "centre" the distribution for γ_0 around a value that yields an efficiency distribution that has r^* as its median. Now, the median of u_i is $\exp(\gamma_0)$ and the median of r_i is $\exp\{-\exp(\gamma_0)\}$. Thus, a value $\underline{\gamma}_0$ that leads to an efficiency distribution centred around r^* is $\underline{\gamma}_0 = \ln\left(-\ln(r^*)\right)$. If we choose $r^* = 0.87$, then $\underline{\gamma}_0 = -2$ is a suitable value. Values for L_0 and U_0 can be chosen in a similar way. For example, setting $L_0 = -4$ leads to a maximum possible value for median efficiency of 0.982, and setting $U_0 = 0$ leads to a minimum possible value for median efficiency of 0.368. The value for $V_{\gamma 0}$ controls the possible spread of values for γ_0 within the truncation points. In the example that follows, we used $V_{\gamma 0} = 4$ implying that $U_0 = 0$ and $L_0 = -4$ would each be one standard deviation from $\underline{\gamma}_0$ if the distribution was not truncated.

Adding a prior for λ introduces extra prior uncertainty about the distribution of u_i and controls its skewness and variance. In Section 2 we suggested two possible priors: $\lambda^{-2} \sim G\left(A_\lambda, B_\lambda\right)$ and $\lambda \sim U\left(a_\lambda, b_\lambda\right)$. In the first case experimentation suggested that $A_\lambda = B_\lambda = 0.25$ are relatively noninformative, but sufficiently informative to facilitate MCMC convergence. In the second case, and that used in the application, we set $a_\lambda = 0.1$ and $b_\lambda = 2$. To check whether these values and the settings for γ_0 provide for a sufficiently wide range of possible efficiencies, we can consider the efficiencies corresponding to the mean values of u_i at the largest and smallest values for $\left(\gamma_0, \lambda\right)$. At the upper truncation points we find $E\left(u_i \mid \gamma_0 = 0, \lambda^2 = 4\right) = 7.4$, with corresponding efficiency value of $r = \exp(-7.4) = 0.0006$. The lower truncation points lead to $E\left(u_i \mid \gamma_0 = -4, \lambda^2 = 0.01\right) = 0.0185$, which has a corresponding efficiency value of 0.982. Thus, these prior settings accommodate a wide range of efficiency distributions.

In Figure 1 we plot the marginal distributions for u_i that correspond to each of the two prior specifications, with the graphs cut off at a maximum value of 1 (a minimum efficiency value of $\exp(-1) = 0.368$). Recognising that $p(u) = \int p\left(u \mid \gamma_1, \lambda\right) p\left(\gamma_1\right) p\left(\lambda\right) d\gamma_1 d\lambda$, we obtained these plots by averaging $p\left(u \mid \gamma_1, \lambda\right)$ over a large number of draws from the priors $p\left(\gamma_1\right)$ and $p\left(\lambda\right)$. Both have an "exponential like" shape with long tails. The prior parameter settings are such that the uniform prior places a relatively heavier weight on small values of u_i (more efficient firms) and the gamma prior places a relatively heavier weight on larger values of u_i (less efficient firms). Also, one criticism that might be levelled at the assumption of a lognormal distribution for the errors is that, because it does not have a nonzero mode, it does not accommodate a situation where most firms are close to 100% efficient. However, when one allows for uncertainty about the parameters of the lognormal distribution, that situation can be accommodated.

In Table 1 the cdf's and moments of the two marginal distributions for $r_i = \exp\left(-u_i\right)$ are compared with those of the exponential prior distribution used by Koop and Steel (2001). The uniform prior is very similar to the exponential prior, while, as already noted, our choice of hyperparameters for the gamma prior allows for a greater prevalence of relatively inefficient firms.

When we move to the model of interest where $u_i \sim LN\left(\overline{\mathbf{x}}_{2i}\boldsymbol{\gamma}, \lambda^2\right)$, rather than the simple version where $\overline{\mathbf{x}}_{2i}\boldsymbol{\gamma} = \gamma_0$, the priors for the remaining elements in $\boldsymbol{\gamma}$ can be set in a similar way, but the magnitudes of the elements in $\overline{\mathbf{x}}_{2,i}$ will have a bearing on what values of $\boldsymbol{\gamma}$ are likely to produce reasonable distributions for efficiency r_i. Finally, we note that, in the application whose results are reported next, we experimented with less informative priors with no substantial changes in the results.

5.2 Results

Posterior means and standard deviations for the parameters for each of the 4 estimated models are presented in Table 2. They were calculated after performing 600,000 MCMC iterations, discarding the first 100,000 and reserving every 50th draw. The elasticity estimates (β_1 to β_4) all have reasonable magnitudes and are not overly sensitive to the assumed model. The largest differences occur in the coefficients β_1 for land (which is assumed to be the source of any endogeneity), and β_2 for labor (which is highly correlated with land). The estimates for λ are considerably larger in the models without endogeneity, picking up variation in u_i not attributable to $\overline{land}_i$ and $\overline{land}_i^2$. Also noticeable is the large increase in the estimate for σ^{-2} (implying a decrease in the estimate for σ^2) when the extra time-varing inefficiency error is introduced. It suggests that much of the residual variation in the time-invariant inefficiency model is attributable to time-varying inefficiency.

The posterior standard deviations for γ_1 and γ_2 are relatively large, casting doubt on the existence of endogeneity. Nevertheless, to investigate whether there is an inverse relationship between efficiency and land size, we plot the posterior mean for $m_i = \gamma_0 + \gamma_1 \overline{land}_i + \gamma_2 \overline{land}_i^2$ against $\overline{land}_i$ in Figures 2 and 3 for the time-invariant and time-varying models, respectively. Also included are 95% credible bands. In both cases the lines are downward sloping indicating that mean inefficiency decreases with land-size. Thus, we do not find any evidence for the inverse relationship between land-size and efficiency as some have suggested.

Finally, although our results in Table 2 have focused on the parameter estimates, it is worth reporting that there was little difference between the posterior means for the inefficiencies between the models with correlated and uncorrelated effects, but, in line with our observations about σ^{-2}, the total inefficiencies from the time-varying models $(u_i + \eta_{it})$ were substantially bigger than the inefficiencies from the time-invariant models.

6. Conclusions

By transforming the inefficiency error to a normally distributed random term, we have been able to construct a relatively general model for introducing endogeneity into stochastic frontier analysis.

Endogeneity can be introduced through either the mean of the transformed inefficiency error, or the covariance structure of the various errors, or both. The model can accommodate the introduction of instrumental variables, can be used with time-invariant and time-varying inefficiency terms, permits endogeneity with respect to both the inefficiency and idiosyncratic errors, and allows for correlation between these errors. Although our conditional posterior densities were in terms of a general transformation, we focus mainly on a log transformation. Future research can be directed towards other transformations relevant for specific distributions for the inefficiency error. Our application showed some but not strong evidence of endogeneity, highlighted the importance of allowing for time-varying inefficiency, and also suggested that frontier parameters are not overly sensitive to these assumptions.

References

Alvarez, A., Amsler, C., Orea, L., Schmidt, P., 2006. Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. J. Productivity Analysis 25, 201-212.

Bardhan, P., 1973. Size, productivity and returns to scale: an analysis of farm-level data in Indian agriculture. J. Political Economy 81, 1370-1386.

Chamberlain, G., 1984. Panel data, in: Griliches, Z., Intriligator, M., (Eds.), Handbook of econometrics, vol. II. North Holland, Amsterdam, pp. 1247-1318.

Coelli, T., Rao, P., O'Donnell, C., Battese, G., 2005. An introduction to efficiency and productivity analysis, 2nd ed. Kluwer, Boston.

Colombi, R., Kumbhakar, S., Martin, G., Vittadini, G., 2011. A stochastic frontier model with short-run and long-run inefficiency random effects. Discussion paper, Department of Economics and Technology Management, University of Bergamo, Italy.

Filippini, M., Greene, W., 2014. Persistent and transient productive inefficiency: a maximum simulated likelihood approach. Working paper, Swiss Federal Institute of Technology, Zurich.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis 1, 515–533.

Koop, G., Steel, M., 2001. Bayesian analysis of stochastic frontier models, in: Baltagi, B., (Ed.), A companion to theoretical econometrics. Blackwell, Oxford, pp. 520–573.

Karakaplan, M., Kutlu, L., 2013. Handling endogeneity in stochastic frontier analysis: a solution to endogenous education cost frontier models. Discussion paper, Oregon State University.

Kumbhakar, S., Tsionas, E., 2014. Firm heterogeneity, persistent and transient technical inefficiency: a generalized true random-effects model. J. Applied Econometrics 29, 110-132.

Kutlu, L., 2010. Battese-Coelli estimator with endogenous regressors. Economics Letters 109, 79-81.

Lamb, R., 2003. Inverse productivity: land quality, labor markets, and measurement error. J. Development Economics 71, 71-95.

Mundlak, Y., 1978. On the pooling of time series and cross section data. Econometrica 46, 69–85.

Neal, R., 2003. Slice sampling. Annals of Statistics 31, 705–767.

Parmeter, C., Kumbhakar, S., 2014. Efficiency analysis: a primer on recent advances. Foundations and Trends in Econometrics 7, 191-385.

Sen, A., 1975. Employment, technology, and development. Oxford University Press, London.

Sickles, R., 2005. Panel estimators and the identification of firm specific efficiency levels in parametric, semiparametric and nonparametric settings. J. Econometrics 126, 305-334.

Tran, K., Tsionas, E., 2013. GMM estimation of stochastic frontier models with endogenous regressors. Economics Letters 118, 233-236.

Wooldridge, J., 2010. Econometric analysis of cross section and panel data, 2nd ed. MIT Press, Cambridge, MA.

Table 1
CDFs, and moments for marginal efficiency distributions

Prior	Quantile					Mean	Var
	0%	25%	50%	75%	100%		
G prior	0.0000	0.3700	0.8738	0.9816	1.0000	0.6699	0.1481
U prior	0.0000	0.6701	0.8741	0.9554	1.0000	0.7715	0.0636
Koop-Exp	0.0000	0.6671	0.8732	0.9561	1.0000	0.7610	0.0742

Table 2
Posterior means and standard deviations for the parameters of estimated models

Parameter	Time-Invar, with corr		Time-Invar, no corr		Time-Var, with Corr		Time-Var, no corr	
	Mean	St.Devn.	Mean	St.Devn.	Mean	St.Devn.	Mean	St.Devn.
β_0	-0.903	0.403	-0.882	0.412	-0.429	0.407	-0.408	0.419
β_1	0.353	0.107	0.382	0.104	0.388	0.103	0.414	0.099
β_2	0.282	0.102	0.290	0.102	0.229	0.094	0.231	0.093
β_3	0.191	0.063	0.194	0.063	0.189	0.058	0.194	0.057
β_4	0.059	0.035	0.059	0.036	0.063	0.030	0.066	0.030
γ_0	-2.497	0.722	-2.170	0.853	-2.188	0.794	-1.949	0.863
γ_1	-0.585	0.447			-0.436	0.530		
γ_2	-0.217	0.438			-0.396	0.434		
δ					4.172	0.719	4.271	0.758
λ	0.969	1.074	2.576	5.221	1.544	2.002	3.039	3.418
σ^{-2}	9.876	1.242	10.090	1.248	28.430	8.802	29.030	9.842

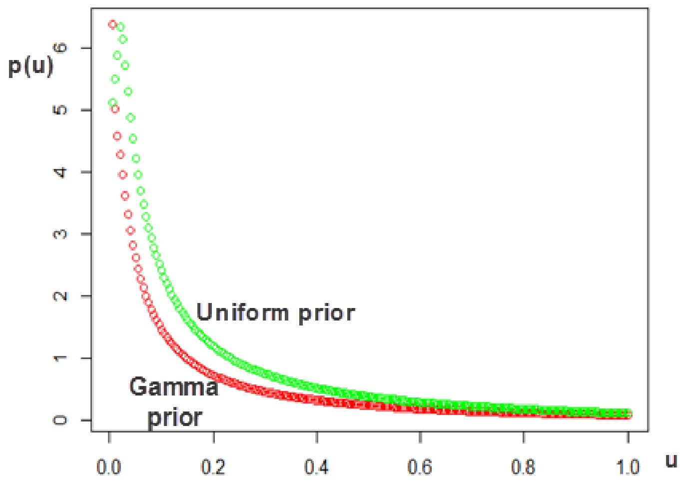

Fig. 1. Marginal distributions for inefficiency errors from two different priors.

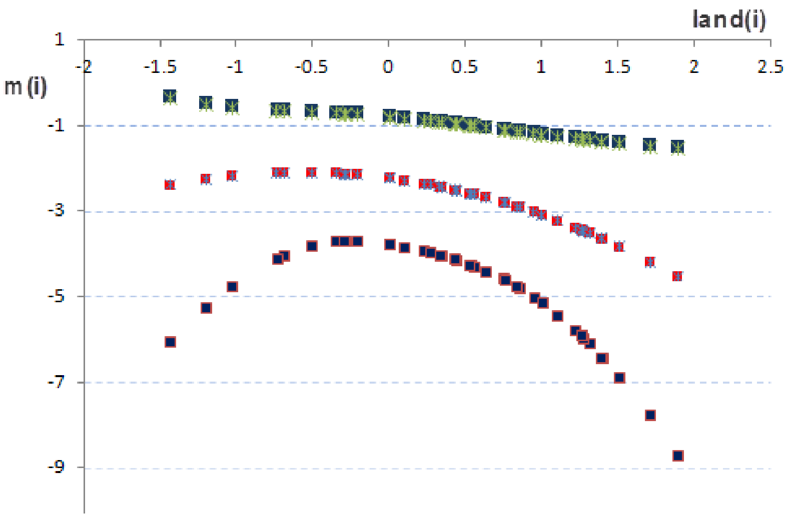

Fig. 2. Posterior mean and credible bands for $m_i = \gamma_0 + \gamma_1 \overline{land}_i + \gamma_2 \overline{land}_i^2$ for time-invariant model.

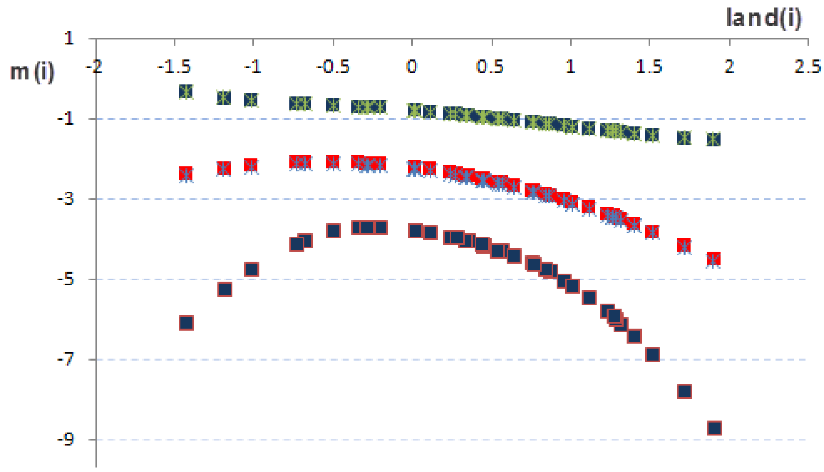

Fig. 3. Posterior mean and credible bands for $m_i = \gamma_0 + \gamma_1 \overline{land}_i + \gamma_2 \overline{land}_i^2$ for time-varying model.

Author/s:
Griffiths, WE; Hajargasht, G

Title:
Some models for stochastic frontiers with endogeneity

Date:
2016-02-01

Citation:
Griffiths, W. E. & Hajargasht, G. (2016). Some models for stochastic frontiers with endogeneity. JOURNAL OF ECONOMETRICS, 190 (2), pp.341-348. https://doi.org/10.1016/j.jeconom.2015.06.012.