Some Models for Stochastic Frontiers with Endogeneity

William E. Griffiths and Gholamreza Hajargasht

Department of Economics
University of Melbourne, Australia

Jan 2, 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 several directions: First an extra one-sided error term is added to allow for time-varying efficiencies. Next, 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. Finally, we show how the analysis can be extended to a nonparametric technology using Bayesian penalised splines. An application of the first and second models to Philippines rice data is provided. A limited Monte Carlo experiment is used to investigate the consequences of ignoring correlation between the effects and the regressors, and choosing the wrong functional form for the technology.

Keywords: Technical Efficiency, Penalised Splines, Gibbs Sampling,

*JEL Classification*s: C11, D24, C23, C12

Corresponding Author

William Griffiths
Economics Department, University of Melbourne
Vic 3010, Australia
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 several 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. Our final extension is to show how the frontier function can be modelled nonparametrically using splines estimated within a Bayesian framework. Although estimation of stochastic frontier models with nonparametric elements is popular – see Parmeter and Kumbhakar (2014) for a review – there are few (if any) studies that have considered both nonparametrics and endogeneity.

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. In Section 5 the analysis is extended to a nonparametric frontier modelled using splines. An application using Philippine rice data and the models from Sections 2 and 3 is provided in Section 6. A Monte Carlo experiment, designed to examine the effects of ignoring endogeneity and/or misspecifying the frontier function, is given in an Appendix.

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 "$+u_i$".

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 $H(u_i)$, 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 $u_i = \exp\left\{\bar{\mathbf{x}}_{2,i}\boldsymbol{\gamma}\right\}u_i^*$ 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

[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 $F(u_i)$.

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 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 6.

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 $\overline{\mathbf{x}}_{2,i}$. The joint posterior kernel for $\boldsymbol{\Theta} = \left(\boldsymbol{\beta}, \sigma^{-2}, \boldsymbol{\gamma}, \lambda^{-2}, \mathbf{u}\right)$ is

$$
\begin{aligned}
p\left(\boldsymbol{\Theta} \mid \mathbf{y}, \mathbf{X}, \mathbf{X}_2\right) &\propto p\left(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^{-2}, \mathbf{u}\right) p\left(\mathbf{u} \mid \mathbf{X}_2, \boldsymbol{\gamma}, \lambda^{-2}\right) p(\boldsymbol{\beta})\, p\left(\sigma^{-2}\right) p(\boldsymbol{\gamma})\, p\left(\lambda^{-2}\right) \\
&\propto \left(\sigma^{-2}\right)^{NT/2 + A_\sigma - 1} \exp\left\{-\frac{\sigma^{-2}}{2}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{1,it}\boldsymbol{\beta} + u_i\right)^2 + 2B_\sigma\right]\right\}\left(\lambda^{-2}\right)^{N/2 + A_\lambda - 1}\left[\prod_{i=1}^{N} u_i^{-1}\right] \qquad (2.3) \\
&\exp\left\{-\left[\frac{\lambda^{-2}}{2}\sum_{i=1}^{N}\left(\ln u_i - \overline{\mathbf{x}}_{2,i}\boldsymbol{\gamma}\right)^2 + 2B_\lambda\right]\right\}\exp\left\{-\frac{1}{2}\left(\boldsymbol{\gamma} - \underline{\boldsymbol{\gamma}}\right)' \mathbf{V}_{\boldsymbol{\gamma}}^{-1}\left(\boldsymbol{\gamma} - \underline{\boldsymbol{\gamma}}\right)\right\}\left[\prod_{s=1}^{S} I\left(L_s \le \gamma_s \le U_s\right)\right]
\end{aligned}
$$

where $I(L_s \le \gamma_s \le U_s)$ 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 $\lambda \sim U(a_\lambda, b_\lambda)$, then the joint posterior density can be obtained from (2.3) by setting $A_\lambda = 1, B_\lambda = 0$, and including the indicator function $I(a_\lambda \le \lambda \le b_\lambda)$. Using (2.3), the following conditional posterior densities can be derived:

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

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

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

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

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

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

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

All the densities in (2.4)-(2.8) are recognised densities which are straightforward to draw from, with the exception of $p\left(u_i \mid \boldsymbol{\Theta}_{-u_i}\right)$. 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

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

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 $H(u_{it}) = x_{2,it}\boldsymbol{\gamma} + e_{it}$ 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,

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

Here, u_i represents permanent and η_{it} transient inefficiencies. 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 $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \sigma^{-2}, \boldsymbol{\gamma}, \lambda^{-2}, \delta, \mathbf{u}, \boldsymbol{\eta})$.

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

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

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

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

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

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

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

As before, all these densities are of recognisable forms from which observations can be drawn directly, except for $p\left(u_i \mid \boldsymbol{\Theta}_{-u_i}\right)$,which will require a Metropolis step or the slice sampler.

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

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

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 both methods, we derive the likelihood function by writing

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

From the third equation in (4.1), we have

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

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

$$E\left[H\left(u_{it}\right) \mid \mathbf{x}_{it}\right] = \mathbf{x}_{2,it}\boldsymbol{\gamma} + E\left(v_{2,it} \mid \mathbf{v}_{it}\right) \qquad \operatorname{var}\left[H\left(u_{it}\right) \mid \mathbf{x}_{it}\right] = \operatorname{var}\left(v_{2,it} \mid \mathbf{v}_{it}\right)$$

where

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

Thus,

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

and

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

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

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

where

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

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

leading to

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

The likelihood for a single observation is obtained by multiplying together (4.4), (4.5), and (4.6), and integrating out u_{it}. Since such an integral is intractable, we consider maximum simulated likelihood and 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

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

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 $\hat{p}\left(\mathbf{y},\mathbf{X}\,|\,\mathbf{u}^{(r)}\right) = \prod_{t=1}^{T}\prod_{i=1}^{N}\hat{p}\left(y_{it},\mathbf{x}_{it}\,|\,u_{it}^{(r)}\right)$. Then, to obtain estimates (predictions) of each of the inefficiency errors u_{it}, we recognize that

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

and estimate this mean using

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

4.3 Bayesian estimation

For Bayesian estimation, we derive the conditional posterior densities for each component of $\boldsymbol{\Theta} = \left(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\pi},\mathbf{u},\boldsymbol{\Sigma}^{-1}\right)$ for use in a Gibbs sampler. 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, $\left(\boldsymbol{\Sigma}^{-1}\,|\,\boldsymbol{\Theta}_{-\boldsymbol{\Sigma}}\right) \sim W\left(\overline{\mathbf{S}}^{-1},\overline{d}\right)$, with arguments

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

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

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

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 $\left(\boldsymbol{\gamma}\,|\,\boldsymbol{\Theta}_{-\boldsymbol{\gamma}}\right) \sim N\left(\bar{\boldsymbol{\gamma}}, \bar{\mathbf{V}}_{\boldsymbol{\gamma}}\right)$, where

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

with

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

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

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

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

with

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

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

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

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

With the exception of $p\left(u_{it}\,|\,\boldsymbol{\Theta}_{-u_{it}}\right)$, all conditional distributions are recognizable distributions from which we can draw observations; for $p\left(u_{it}\,|\,\boldsymbol{\Theta}_{-u_{it}}\right)$ we can use slice sampling or a Metropolis step.

5. Extension to a nonparametric technology

A possible limitation of the models described so far is the parametric nature of the production technology. Although many of the so-called flexible functional forms such as translog and generalised Leontief fit nicely within the framework of these models, it is useful to examine how one might incorporate a nonparametric function for the production frontier. Studies that have done so, but without allowing for endogeneity, include Fan et al. (1996), Kumbhakar et al. (2007), and Martins-Filho and Yao (2015). Parmeter and Kumbhakar (2014) review nonparametric contributions and other recent advances. In this Section we introduce spline-based nonparametric modelling of the frontier in the context of the model studied in Section 3, where both transient and permanent inefficiency errors are included, with the permanent inefficiency errors correlated with the inputs. In principle, a similar extension can be made to the model in Section 4, but the level of computational complexity increases considerably.

Consider the model

$$\begin{aligned} y_{it} &= f(\mathbf{x}_{it}) - u_i - \eta_{it} + v_{it} \\ u_i &\sim LN\left(\overline{\mathbf{x}}_{2i}\boldsymbol{\gamma}, \lambda^2\right) \end{aligned} \tag{5.1}$$

where f is modelled nonparametrically[2]. We use penalised low-ranked splines to model f because they have a nice Bayesian counterpart, they can be used effectively to estimate our model, and they have good properties (see e.g., Ruppert et al. 2003, Hajarghast 2006, Claeskens et al. 2008, Chib et al. 2009). One can replace the nonparametric function with terms that are linear in the parameters and place Gaussian priors on these parameters. For expositional convenience we begin by explaining the procedure for a univariate regressor x_{it}. Extension to various multivariate settings such as partially linear, additive, additive with interactions, and fully nonparametric models is straightforward and presented later. The polynomial spline representation of the function f is[3]

[2] We can consider a nonparametric specification in the second equation similar to that in the first equation; the nature of the conditional posterior densities required for Gibbs sampling does not change, but, from a practical standpoint, estimation becomes much more challenging.

[3] Following Ruppert et al. (2003), we write this and subsequent equations as an equality, although it needs to be recognised that the spline formulation is an approximation not an exact representation of any function *f*.

$$y_{it} = \beta_0 + \beta_1 x_{it} + \cdots + \beta_p x_{it}{}^p + \sum_{k=1}^{K} w_k (x_{it} - \kappa_k)_+^p - u_i - \eta_{it} + v_{it} \tag{5.2}$$

where $\{\kappa_1, \kappa_2, \cdots, \kappa_K\}$ are some chosen points in the range of x known as knots, and $(x_{it} - \kappa_k)_+^p$ is a polynomial term equal to zero when $x_{it} < \kappa_k$; the β_j and w_k are coefficients to be estimated[4]. A polynomial degree of $p = 2$ is typically adequate – one less than the polynomial degree is the degree of differentiability at all points. To avoid overfitting, a restriction is imposed on the magnitude of the elements of $\mathbf{w}' = (w_1, w_2, \cdots, w_K)$, either of the form $\mathbf{w}'\mathbf{w} \leq C$, leading to penalised least squares or penalised likelihood estimators, or by assuming $\mathbf{w} \sim N(\mathbf{0}, \tau^2 \mathbf{I})$, leading to Bayesian or non-Bayesian mixed model estimators. To put (5.2) in matrix notation convenient for describing Bayesian estimation, we write

$$\mathbf{y} = \mathbf{X}_0 \boldsymbol{\beta}_0 + \mathbf{X}_1 \mathbf{w} - \mathbf{u} \otimes \mathbf{i}_T - \boldsymbol{\eta} + \mathbf{v} = \mathbf{X}\boldsymbol{\beta} - \mathbf{u} \otimes \mathbf{i}_T - \boldsymbol{\eta} + \mathbf{v} \tag{5.3}$$

where $\mathbf{y}$ and $\mathbf{v}$ are NT-dimensional vectors containing the y_{it} and v_{it}, respectively, $\mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1]$ is an $[NT \times (p+1+K)]$ matrix with $\mathbf{x}_{it} = \left(1, x_{it}, \cdots, x_{it}^p, (x_{it} - \kappa_1)_+^p, \cdots, (x_{it} - \kappa_K)_+^p\right)$ in its it-th row, $\boldsymbol{\beta}' = (\boldsymbol{\beta}_0', \mathbf{w}') = (\beta_0, \cdots, \beta_p, w_1, \cdots, w_K)$, $\mathbf{u}' = (u_1, u_2, \cdots, u_n)$ and $\mathbf{i}_T$ is a T-dimensional vector of ones. Here and in what follows we use the notation $\mathbf{x}_{it}$ to denote the regressors required for the spline formulation rather than a vector of logs of inputs as was the case in equation (5.1). When more than one regressor is introduced, the nonparametric function f is again replaced by $\mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\mathbf{w}$ with a similar penalty on $\mathbf{w}$, but, as we see below, the definitions of $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{x}_{it}$ change. Thus, Bayesian estimation using (5.3) is equally applicable to the case of multivariate regressors, providing suitable changes are made for $\mathbf{X}_0$, $\mathbf{X}_1$ and $\mathbf{x}_{it}$.

Assuming $\mathbf{w} \sim N(\mathbf{0}, \tau^2 \mathbf{I})$, and a gamma prior for $\tau^{-2} \sim G(A_1, B_1)$ and forming the posterior kernel leads to a Gibbs sampler that is very similar to that in equations (3.2)-(3.8). The differences are (1) $\mathbf{x}_{1,it}$ is replaced with $\mathbf{x}_{it}$, (2) the conditional posterior for $\boldsymbol{\beta}$ changes to

[4] Full details about modelling with polynomial splines including guidelines for the choice of knots can be found in Ruppert et al. (2003).

$$\left(\boldsymbol{\beta} \mid \boldsymbol{\Theta}_{-\boldsymbol{\beta}}\right) \sim N\left\{\left(\mathbf{X}'\mathbf{X}+\left(\tau^{-2}/\sigma^{-2}\right)\mathbf{K}\right)^{-1}\mathbf{X}'(\mathbf{y}+\mathbf{u}\otimes\mathbf{i}_T+\boldsymbol{\eta}),\ \sigma^2\left(\mathbf{X}'\mathbf{X}+\left(\tau^{-2}/\sigma^{-2}\right)\mathbf{K}\right)^{-1}\right\} \tag{5.4}$$

and (3) for the extra parameter τ^{-2}, we have the conditional posterior density

$$\left(\tau^{-2} \mid \boldsymbol{\Theta}_{-\tau^{-2}}\right) \sim G\left(A_1+K/2,\ \ B_1+\boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta}/2\right) \tag{5.5}$$

where $\mathbf{K}=\begin{bmatrix}\mathbf{0}_{p+1} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_K\end{bmatrix}$.

5.1 Multivariate cases

It is well-known that multivariate nonparametric estimation suffers from a curse of dimensionality: as the dimension of the regression function increases, estimation becomes less precise and sometimes impractical. Thus, we first describe some restricted multivariate models, and then move on to the general case. We start with a *partially linear* model and continue with an *additive model*. Additive models may lack flexibility in that they ignore interaction between variables. A model which includes interaction terms, known as the *additive with interactions model*, is also briefly discussed. Finally, we discuss modelling of general nonparametric multivariate functions using multivariate spline bases.

The regression function for a partially linear model can be written as

$$f(\mathbf{x}_{it}) = \mathbf{q}_{it}\boldsymbol{\alpha} + f(x_{it}) \tag{5.6}$$

where $\mathbf{q}_{it}$ is a row vector of explanatory variables exclusive of the variable x_{it} that we are treating nonparametrically, and $\boldsymbol{\alpha}$ is a conformable vector of unknown parameters. The regression spline representation of the model is

$$f(\mathbf{x}_{it}) = \mathbf{q}_{it}\boldsymbol{\alpha} + \beta_0 + \beta_1 x_{it} + \cdots + \beta_p {x_{it}}^p + \sum_{k=1}^{K} w_k \left(x_{it}-\kappa_k\right)_+^p \tag{5.7}$$

or, in matrix notation, $f(\mathbf{x}) = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\mathbf{w}$, where $\mathbf{X}_0$ and $\mathbf{X}_1$ are matrices with rows $\mathbf{x}_{0,it} = \left(\mathbf{q}_{it}, 1, x_{it}, \cdots, {x_{it}}^p\right)$ and $\mathbf{x}_{1,it} = \left[\left(x_{it}-\kappa_1\right)_+^p, \cdots, \left(x_{it}-\kappa_K\right)_+^p\right]$, $\boldsymbol{\beta}_0' = \left(\boldsymbol{\alpha}, \beta_0, \beta_1, \cdots, \beta_p\right)$ and $\mathbf{w}' = \left(w_1, \cdots, w_K\right)$. Again, we use $\mathbf{w} \sim N\left(0, \tau^2\mathbf{I}\right)$ to impose a penalty on the magnitude of w_k.

The additive model in its simplest form with two components can be represented as

$$f(\mathbf{x}_{it}) = f_1(x_{1i}) + f_2(x_{2i}) \tag{5.8}$$

where f_1 and f_2 are univariate functions of unknown forms. In a stochastic production frontier, with the variables in logs, equation (5.8) can be considered a generalisation of a Cobb-Douglas function. The penalised spline formulation of the additive model with linear basis functions (i.e., $p=1$) is

$$f(\mathbf{x}_{it}) = \beta_0 + \beta_1 x_{1it} + \sum_{k=1}^{K_1} w_{1k}(x_{1,it} - \kappa_{1k})_+ + \beta_2 x_{2,it} + \sum_{k=1}^{K_2} w_{2k}(x_{2,it} - \kappa_{2k})_+ \tag{5.9}$$

In this case, we define $\boldsymbol{\beta}_0' = (\beta_0, \beta_1, \beta_2)$, $\mathbf{x}_{0,it} = (1, x_{1it}, x_{2it})$, $\mathbf{w}' = \left(w_{11}, \cdots, w_{1K_1}, w_{21}, \cdots, w_{2K_2}\right)$, and $\mathbf{x}_{1,it} = \left[(x_{1it} - \kappa_{11})_+, ..., \left(x_{1it} - \kappa_{1K_1}\right)_+, (x_{2it} - \kappa_{21})_+, ..., \left(x_{2it} - \kappa_{2K_2}\right)_+\right]$, leading again to the standard model $\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\mathbf{w} - \mathbf{u} \otimes \mathbf{i} + \mathbf{v}$ which can be estimated using a Bayesian approach. A more general prior is needed for $\mathbf{w}$, namely

$$\mathbf{w} \sim N\left(\mathbf{0}, \begin{bmatrix} \tau_1^2 \mathbf{I}_{K_1} & \mathbf{0} \\ \mathbf{0} & \tau_2^2 \mathbf{I}_{K_2} \end{bmatrix}\right) \tag{5.10}$$

A weakness of the purely additive model is that interactions between independent variables are completely ignored. It is possible to allow for second-order interactions. In the simple case with three regressors this results in the following model

$$f(x_1, x_2, x_3) = f_0 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_{12}(x_1, x_2) + f_{13}(x_1, x_3) + f_{23}(x_2, x_3) \tag{5.11}$$

The observation subscript *it* has been omitted for notational convenience. This model is referred to as an *additive model with interactions* (see e.g., Sperlich et al. 2002). Parametric models with interactions are common in economics, but may lead to incorrect conclusions if the assumed parametric form is incorrect. In production economics parametric forms which are special cases of an additive model with interactions are the translog, generalised Leontief, and generalised quadratic. Bivariate functions such as $f_{12}(x_1, x_2)$ can be modelled using the multivariate approach to which we now turn.

Estimation of general multivariate nonparametric functions requires multivariate spline bases. At least two approaches to creating multivariate splines have been proposed: one is by forming tensor products of univariate spline bases, and the other is by using radial basis splines (both are discussed in

Ruppert et al 2003). Here, we briefly discuss radial basis splines which is the method we suggest and employ in our Monte Carlo experiments reported in the Appendix.

If $\mathbf{x} \in \Re^d$ and $\boldsymbol{\kappa}_1, \cdots, \boldsymbol{\kappa}_K$ are some chosen knots in $\Re^d$, the multivariate function $f(\mathbf{x}_{it})$ can be approximated by

$$f(\mathbf{x}_{it}) = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + \sum_{k=1}^{K} w_k C\left(\| \mathbf{x}_{it} - \boldsymbol{\kappa}_k \|\right) \tag{5.12}$$

with

$$C\left(\| \mathbf{r} \|\right) = \begin{cases} \| \mathbf{r} \|^{2m-d} & \text{for } d \text{ odd} \\ \| \mathbf{r} \|^{2m-d} \ln\left(\| \mathbf{r} \|\right) & \text{for } d \text{ even} \end{cases} \tag{5.13}$$

where $\| \mathbf{r} \|$ is the Euclidean norm of the vector $\mathbf{r}$; m is chosen such that $2m - d > 0$ where $2m - d$ is the degree of smoothness. To write (5.12) in a form analogous to what we specified earlier, we define $\mathbf{X}_0$ as a matrix with $\left(1, \mathbf{x}_{it}\right)$ as its it-th row, $\mathbf{X}_1$ as a matrix with $C\left(\| \mathbf{x}_{it} - \boldsymbol{\kappa}_k \|\right)$ as its (it,k)-th element, $\boldsymbol{\beta}_0' = \left(\beta_0, \boldsymbol{\beta}'\right)$, and $\mathbf{w} = (w_1, ..., w_K)'$. With this notation we can write the model as $\mathbf{y} = \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1\mathbf{w} - \mathbf{u} \otimes \mathbf{i}_T + \mathbf{v}$. However, an adjustment is necessary to accommodate a penalty matrix. A suitable penalty matrix that has been suggested in the literature (Ruppert et al 2003) is

$$\boldsymbol{\Omega} = C\left(\begin{bmatrix} 0 & \| \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_2 \| & \cdots & \| \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_{K-1} \| & \| \boldsymbol{\kappa}_1 - \boldsymbol{\kappa}_K \| \\ \| \boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_1 \| & 0 & \cdots & \| \boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_{K-1} \| & \| \boldsymbol{\kappa}_2 - \boldsymbol{\kappa}_K \| \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \| \boldsymbol{\kappa}_{K-1} - \boldsymbol{\kappa}_1 \| & \| \boldsymbol{\kappa}_{K-1} - \boldsymbol{\kappa}_2 \| & \cdots & 0 & \| \boldsymbol{\kappa}_{K-1} - \boldsymbol{\kappa}_K \| \\ \| \boldsymbol{\kappa}_K - \boldsymbol{\kappa}_1 \| & \| \boldsymbol{\kappa}_K - \boldsymbol{\kappa}_2 \| & \cdots & \| \boldsymbol{\kappa}_K - \boldsymbol{\kappa}_{K-1} \| & 0 \end{bmatrix} \right) \tag{5.14}$$

The matrix $\boldsymbol{\Omega}$ will not be positive definite and cannot be used as a proper covariance matrix required for Bayesian estimation. However, a positive definite approximation given by

$$\boldsymbol{\Omega}^* = \left(\boldsymbol{\Omega}^{1/2}\right)' \left(\boldsymbol{\Omega}^{1/2}\right) \tag{5.15}$$

has been suggested.[5] The penalty constraint $\mathbf{w} \sim N\left(0, \tau^2 \boldsymbol{\Omega}^{*-1}\right)$ can be introduced using the transformations $\mathbf{X}_1^* = \mathbf{X}_1 \boldsymbol{\Omega}^{*-1/2}$, and $\mathbf{w}^* = \boldsymbol{\Omega}^{*1/2}\mathbf{w}$, so that a suitable form of the model which introduces the penalty in the same way as specified in the previous section is

$$\begin{aligned} \mathbf{y} &= \mathbf{X}_0\boldsymbol{\beta}_0 + \mathbf{X}_1^*\mathbf{w}^* - \mathbf{u} \otimes \mathbf{i}_T + \mathbf{v} \\ \mathbf{w} &\sim N(\mathbf{0}, \tau^2\mathbf{I}) \end{aligned} \tag{5.16}$$

Further information on multivariate splines and radial bases can be found in Ruppert et al. (2003) or Nychka (2000).

6. 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

$$y_{it} = \beta_0 + \beta_1 land_{it} + \beta_2 labor_{it} + \beta_3 fert_{it} + \beta_4 others_{it} - u_i + v_{it} \tag{6.1}$$

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

[5] $\boldsymbol{\Omega}^{1/2} = \mathbf{U}\mathbf{D}^{1/2}\mathbf{V}'$ where $(\mathbf{U}, \mathbf{D}, \mathbf{V})$ is obtained from singular value decomposition of $\boldsymbol{\Omega}$.

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

$$u_i = LN\left(\gamma_0 + \gamma_1 \overline{land}_i + \gamma_2 \overline{land}_i^2, \lambda^2\right) \tag{6.2}$$

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 $\gamma_1 = \gamma_2 = 0$.

6.1 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 $r_i = \exp(-u_i)$ 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.

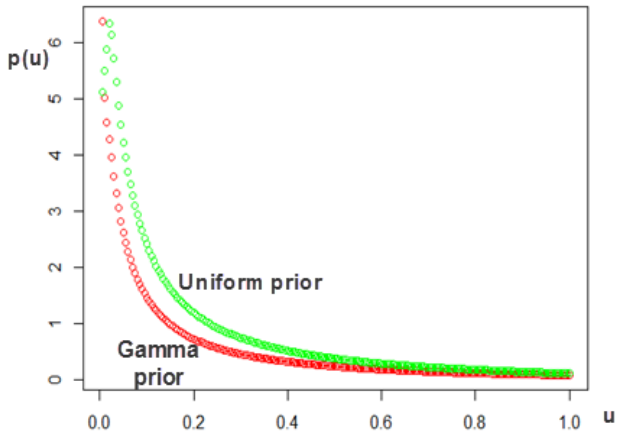


Figure 1. Marginal distributions for inefficiency errors from two different priors.

Table 1. CDFs, and moments for marginal efficiency distributions

	Quantile						
	0%	**25%**	**50%**	**75%**	**100%**	**Mean**	**Var**
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

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.

6.2 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.

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

	Time-Invar-Corr		**Time-Invar**		**Time-Var-Corr**		**Time-Var**	
	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

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.

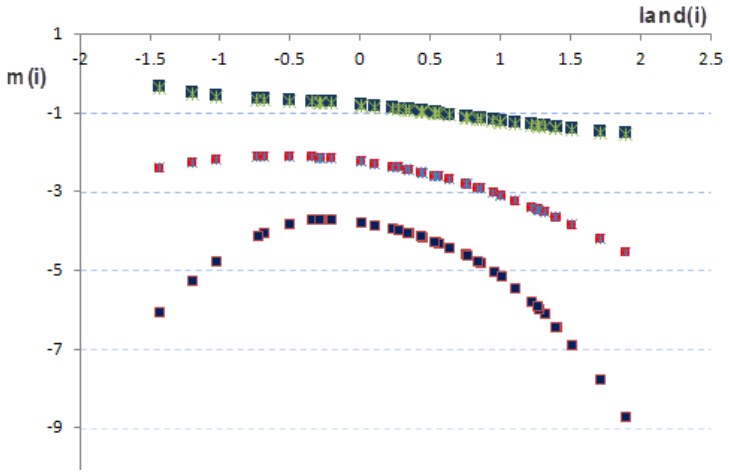


Figure 2. Posterior mean and credible bands for $m_i = \gamma_0 + \gamma_1 \overline{land}_i + \gamma_2 \overline{land}_i^2$ for time-invariant model

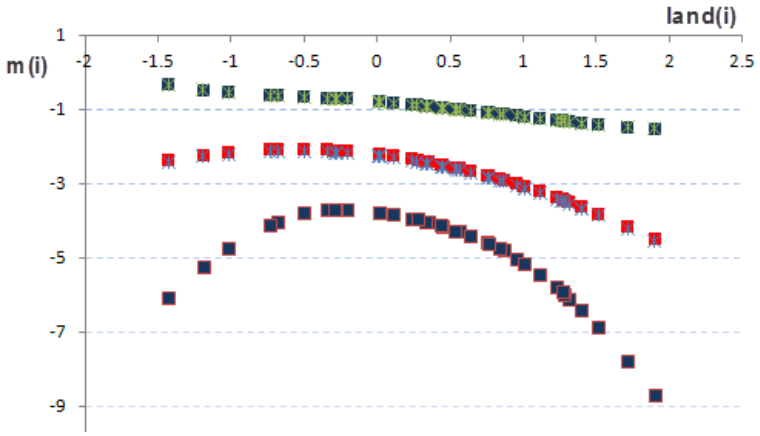


Figure 3. Posterior mean and credible bands for $m_i = \gamma_0 + \gamma_1 \overline{land}_i + \gamma_2 \overline{land}_i^2$ for time-varying model

7. 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, allows for correlation between these errors, and is readily extended to a nonparametric frontier using splines. 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.

Results from a Monte Carlo experiment investigating the impact of ignoring endogeneity and choosing the incorrect functional form suggest that the nonparametric model and corresponding estimation procedure are robust. They successfully eliminate biases that are present if the wrong functional form is chosen or if endogeneity is ignored, and any improvement in results from modelling the correct functional form, and/or correctly ignoring endogeneity when it is not present, is small.

References

Alvarez A., C. Amsler, L. Orea and P. Schmidt, (2006). Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics, *Journal of Productivity Analysis*, Springer, 25, 201-212.

Bardhan, P. (1973). Size, Productivity and Returns to Scale: an Analysis of Farm-Level Data in Indian Agriculture, *Journal of Political Economy* , 81, 1370-86.

Chamberlain, G. (1984). "Panel Data" in: Grilliches and Intriligator ed. *Handbook of Econometrics Vol. II*, Chapter 22, 1247-1318.

Chib, S., E. Greenberg and I. Jeliazkov (2009). Estimation of Semiparametric Models in the presence of Sample Selection and Endogeniety, *Journal of Computational and Graphical Statistics*, 18, 321-348

Claeskens, G., T. Krivobokova T. and J. Opsomer (2008). Asymptotic Properties of Penalized Spline Estimators, *Biometrika*, 96, 529-544.

Coelli, T., P. Rao, C. O'Donnell, and G. Battese (2005). *An Introduction to Efficiency and Productivity Analysis*. 2nd edition. Boston: Kluwer Academic Publishers.

Colombi R, S. Kumbhakar, G. Martin and G.Vittadini (2011),. A Stochastic Frontier Model with Short-run and Long-run Inefficiency Random Effects, Department of Economics and Technology Management,University of Bergamo, Italy.

Fan, Y., Q. Li, and A. Weersink (1996). Semiparametric Estimation of Stochastic Production Frontier Models, *Journal of Business and Economic Statistics*, 14, 460-68.

Filippini M. and W. Greene (2014). Persistent and Transient Productive Inefficiency: A Maximum Simulated Likelihood Approach, *Unpublished Manuscript.*

Gelman A. (2006). Prior Distributions for Variance Parameters in Hierarchical Models. *Bayesian Analysis*, 1, 515–533.

Griffin, J., and M. Steel (2007). Bayesian stochastic frontier analysis using WinBUGS, *Journal of Productivity Analysis*, 27, 163-176.

Hajargasht, G. (2006). *Non- and Semiparametric Stochastic Frontiers: A Penalized Spline Approach*, Unpublished PhD Thesis, University of Queensland.

Koop, G., and M. Steel (2001). Bayesian Analysis of Stochastic Frontier Models. b. Baltagi, ed. *A Companion to Theoretical Econometrics*, Blackwell, Oxford: Blackwell Publishers, pp. 520–573.

Karakaplan, M.U. and L. Kutlu (2013). Handling Endogeneity in Stochastic Frontier Analysis, *Unpublished Manuscript*.

Kumbhakar, S., B. Park, L. Simar and E. Tsionas, (2007). Nonparametric Stochastic Frontiers: A Local Maximum Likelihood Approach, *Journal of Econometrics*, 137, 1-27

Kumbhakar, S. and E. Tsionas, (2014). Firm Heterogeneity, Persistent and Transient Technical Inefficiency: A Generalized True Random-Effects Model, *Journal of Applied Econometrics*, 29, 110-132.

Kutlu, L. (2010), "Battese-Coelli Estimator with Endogenous Regressors," *Economics Letters*, 109, 79-81.

Lamb, R. L. 2003. "Inverse Productivity: Land Quality, Labor Markets, and Measurement Error." *Journal of Development Economics* , 71, 71-95.

Martins-Filho, C. and F. Yao, (2015). Semiparametric Stochastic Frontier Estimation via Profile Likelihood, *Econometric Reviews, 34,413-451*

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

Nychka, D. (2000). Spatial Process Estimates as Smoothers, in *Smoothing and Regression* edited by M. Schimek, Wiley, New York, 393-424.

Nychka, D., Haaland, P., O'Connell, M. and Ellner, S. (1998). FUNFITS, data analysis and statistical tools for estimating functions, in *Case Studies in Environmental Statistics (Lecture Notes in Statistics, vol 132)*, pp. 159-179. New York, Springer-Verlag.

Parmeter, C. and S. Kumbhakar (2014). Efficiency Analysis: A Primer on Recent Advances, *Foundations and Trends in Econometrics*, 7, Issue 3-4.

Ruppert, D., M. Wand and R. Carroll. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

Sen, A. (1975). *Employment, Technology, and Development*. London: Oxford University Press.

Sickles, R., (2005). Panel Estimators and the Identification of Firm Specific Efficiency Levels in Parametric, Semiparametric and Nonparametric Settings, *Journal of Econometrics*, 126, 305-334

Sperlich, S., D. Tjøstheim and L. Yang (2002). Nonparametric Estimation and Testing of Interaction in Additive Models. *Econometric Theory*, 18, 197-251.

Tran, K. and E. Tsionas (2013), "GMM Estimation of Stochastic Frontier Models with Endogenous Regressors," *Economics Letters*, 118, 233-236.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, second edition. Cambridge, MA: MIT Press.

Appendix A Monte Carlo experiment

Based on the model in Section 2, we use a Monte Carlo experiment to examine the performance of Bayesian estimators that allow and do not allow for a nonparametric frontier and do and do not allow for endogeneity. Two main issues are investigated: (1) the effect of correlation between the

regressors and the inefficiency errors, and (2) the effect of choosing an incorrect functional form for the technology.

A.1 Design of the experiment

We consider a bivariate stochastic frontier model of the following form

$$y_{it} = f(x_{1,it}, x_{2,it}) - u_i + v_{it}$$

where the v_{it} are i.i.d. $N\left(0, 0.1^2\right)$. Values for $x_{1,it}$ and $x_{2,it}$ are the logs of area and labour in the Philippine rice data set. This data set has observations on 43 firms for 8 years; we used the last 4 years, giving us $T = 4$ and $n = 43$. For f we consider the following two cases depicted in Figures A1(a) and A1(b).

Translog model: $f(x_1, x_2) = x_1 - 0.1x_1^2 + 1.5x_2 - 0.25x_2^2$

Arctan model: $f(x_1, x_2) = x_1 - 0.1x_1^2 + \arctan(2x_2 - 0.5) + 0.25x_2$

The first case is a translog function commonly used in production economics; the second function has been chosen so that it is not of translog form with respect to x_2, but is still a reasonable candidate for a production function. The constant term in both these functions is zero; it was nevertheless treated as an unknown parameter to be estimated.

Three alternative assumptions were made about the inefficiency error u_i, each one representing a different degree of correlation between u_i and $x_{1,it}$.

Case 1: no correlation $u_i \sim LN\left(-2, 0.75^2\right)$

Case 2: moderate correlation $u_i \sim LN\left(-1 - 0.5\bar{x}_{1,i}, 0.75^2\right)$

Case 3: high correlation $u_i \sim LN\left(-0.75 - 0.8\bar{x}_{1,i}, 0.75^2\right)$

In the original data set, x_1 is the log of the area sown to rice and so cases 2 and 3 suggest that larger firms tend to be more efficient. The correlation between u_i and $x_{1,it}$ is zero for case 1, around –0.3 for case 2 and around –0.6 for case 3. A seemingly good strategy for setting larger correlations between u_i and $x_{1,it}$ is to set smaller values for λ. However, we discovered that if the variation in u_i becomes too small, it can be difficult to obtain MCMC convergence because of high correlation

between u_i and the constant term in the technology function. Increasing the absolute magnitude of γ_2, the coefficient of $\bar{x}_{1,i}$, will also increase the correlation, but in this case one must be wary of producing an unreasonable distribution of efficiencies. A further way of increasing the correlation between u_i and $x_{1,it}$ is to reduce the time variation in $x_{1,it}$. This strategy, alongside an increase in γ_2, was used for case 3. New values $x^*_{1,it}$ were generated using $x^*_{1,it} = \bar{x}_{1,i} + e_{it}$ where the e_{it}'s were independent random draws from $U(-0.1, 0.1)$. At the overall average of $x_{1,it}$, all cases lead to a median of technical efficiency between 0.87 and 0.90 and mean technical efficiency between 0.84 to 0.87. For the smallest firm, median efficiency is 0.69 and for the largest firm it is 0.97.

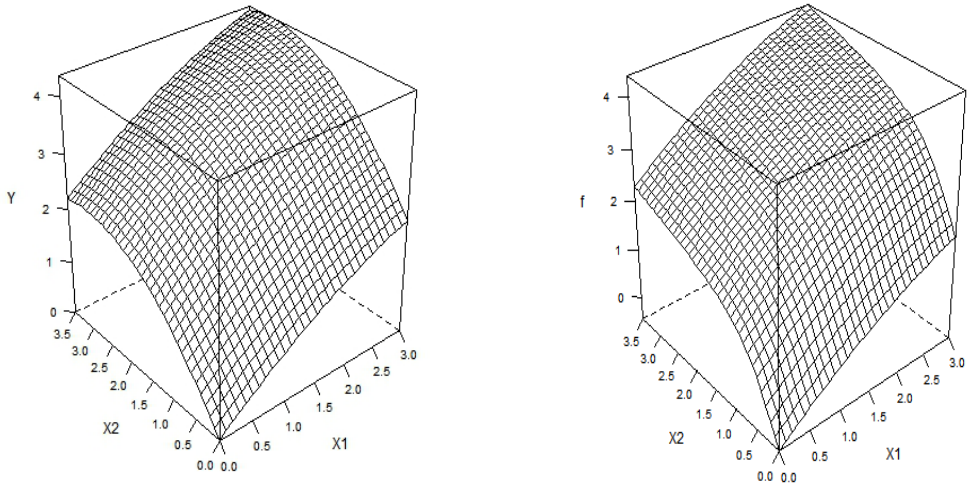


Figure A1 (a) Translog function (b) Arctan function

A.2 Estimation

Two alternative technology functions, coupled with 3 levels of endogeneity for the inefficiency error, give us a total of 6 data generating processes (DGPs). Four different estimators were applied to each of these processes, each one reflecting a different set of assumptions about the functional form for *f* or the data generating process for u_i:

1. A translog function with no correlation between u_i and $\left(x_{1,it}, x_{2,it}\right)$.
2. A translog function with the assumption u_i is related to $\left(x_{1,it}, x_{2,it}\right)$.

3. A nonparametric function with no correlation between u_i and $\left(x_{1,it}, x_{2,it}\right)$.

4. A nonparametric function with the assumption u_i is related to $\left(x_{1,it}, x_{2,it}\right)$.

For the nonparametric estimators we chose 20 knots located using a space filling algorithm for which publicly available software exists (Nychka et al., 1998).

The DGP/estimator combinations allow us to investigate the effect of choosing an incorrect functional form, the effect of ignoring endogeneity, the consequences of allowing for endogeneity when it is not present, and the effect of using a nonparametric approach when a parametric translog function is adequate. The estimation procedures we described can be readily adapted to handle a parametric technology function and/or a model with no endogeneity.

The settings for the prior distributions for the fourth estimator (nonparametric with endogeneity) were $\sigma^{-2} \sim G(0.01, 0.01)$, $\tau^{-2} \sim G(0.01, 0.01)$, $\lambda \sim U(0.1, 2)$, $p(\beta) \propto 1$, and $\boldsymbol{\gamma} \sim TN\left(\underline{\boldsymbol{\gamma}}, \mathbf{V}_\gamma; \mathbf{L}, \mathbf{U}\right)$ with

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

The relevant components of these priors are also used for the other three estimators. Only the first elements in $\underline{\boldsymbol{\gamma}}, \mathbf{V}_\gamma, \mathbf{L}$ and $\mathbf{U}$ are relevant if no endogeneity is assumed, and the prior for τ^{-2} is no longer necessary for estimating the translog functions.

Given the relatively demanding nature of the nonparametric estimators, and that we are carrying out MCMC estimation within each replication of the Monte Carlo experiment, we chose a relatively small number of replications, 200 for the translog specifications and 100 for the nonparametric specifications. All computations were performed using R and WinBUGS[6]. For all Bayesian estimations we used 20000 MCMC iterations and discarded the first 10000 as the burn in period.

A.3 Results

Estimates likely to be of interest are those for the inefficiency errors and the elasticities of output with respect to each of the inputs which, given that output and inputs are measured in terms of logs, will be given by the derivatives $\partial f / \partial x_1$ and $\partial f / \partial x_2$. Calculation of the derivatives is

[6] See Griffin and Steel (2007) for the use of WINBUGS to estimate stochastic frontier models.

straightforward for the translog estimations. To calculate the required derivatives with respect to x_1 and x_2 at any point $\mathbf{x}^0$ for the nonparametric specification we first let the *k*-th knot for both variables be given by $\boldsymbol{\kappa}_k = \left(\kappa_k^{(1)}, \kappa_k^{(2)}\right)$. Then, using the fact that $\mathbf{w} = \boldsymbol{\Omega}^{*-1/2}\mathbf{w}^*$, we have

$$\left.\frac{\partial f(\mathbf{x})}{\partial x_1}\right|_{\mathbf{x}^0} = \beta_1 + \sum_{k=1}^{K}\left(x_1 - \kappa_k^{(1)}\right)\left(2\ln \| \mathbf{x} - \boldsymbol{\kappa}_k \| + 2\right) w_k \Bigg|_{\mathbf{x}^0}$$

The derivative with respect to x_2 can be obtained in a similar manner. Note that these derivatives are linear with respect to the parameters and thus their posterior means can be obtained by calculating them at posterior means of the parameters.

The bias and variability of the estimators was assessed by plotting the posterior means for u_i, $\partial f / \partial x_{1it}$ and $\partial f / \partial x_{2it}$ against their true values, for all data points, and for all replications of the Monte Carlo experiment. These plots appear in Figures A2 to A7. Observations scattered evenly around the 45 degree lines suggest no bias. A concentration of observations above or below a 45 degree line depicts bias.

We first consider the outcomes when the 4 estimators are applied to observations from the translog model without correlation. From the plots in Figure A2 we see that the mean-squared errors (MSEs) for $\hat{u}_i$ are not very sensitive to the estimation technique. The nonparametric estimator with the same provision for correlation does worse than the correct translog specification, but not a great deal worse. Perversely, allowing for correlation when none exists does better than assuming no correlation, although the differences are not substantial. The same conclusion is not true for the derivatives. In this case both bias and MSE from the nonparametric specification are worse than for the correctly specified translog. There is little difference between the plots for $\hat{u}_i$ from the 4 estimators. As expected, the spreads of the estimated derivatives are greater for the nonparametric estimates, but there are no substantial biases except at the endpoints.

When we move to a translog DGP with moderate correlation between u_i and x_{1it} (Figure A3), the effect of not allowing for correlation becomes evident. The MSE's of the estimators for both u_i and the derivatives are larger when correlation is ignored, although the effect is mainly on u_i and the

derivative with respect to x_{1i}. Biases appear in the estimated derivatives for both x_{1i} and x_{2i} for estimators that ignore the correlation. We lose in terms of both MSE and bias by using the nonparametric estimator relative to the "correct" parametric translog estimator; a large part of this loss is attributable to bias in estimation at the end points.

The biases and MSE differences become much more pronounced when the level of correlation is increased. This is particularly evident in Figure A4 where the estimates that ignore the correlation are concentrated above the 45 degree line.

Figures A5, A6 and A7 provide the results for an arctan DGP with varying degrees of correlation between u_i and x_{1it}. In this case we are able to compare the performance of estimates from the incorrect translog functional form with those from the nonparametric estimator. Looking first at the case with no correlation (Figure A5), we find that, if a translog model is estimated, there will be significant biases in all quantities of interest, including efficiency effects and derivatives with respect to x_1 and x_2, no matter whether correlations are assumed or not. These biases are not present for the nonparametric estimator, except at the end points for the derivative with respect to x_1. This indicates that choosing a wrong functional form could have serious consequences for estimation. It is interesting, however, that the MSE for the translog estimator of the derivative with respect to x_1 is smaller than its nonparametric counterpart. In terms of what we lose by assuming correlation when none exists, the results are mixed, but it does appear that the losses, if there are any, are not great.

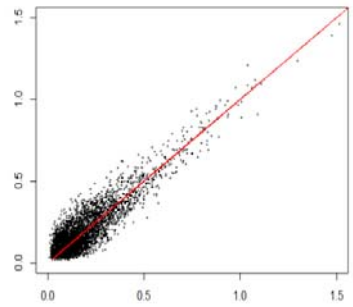
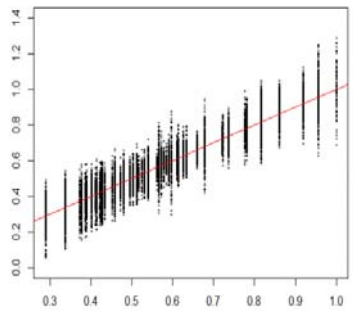
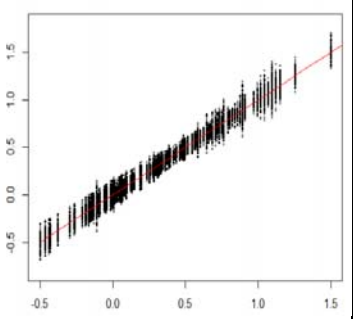
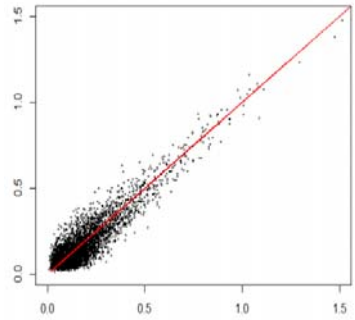
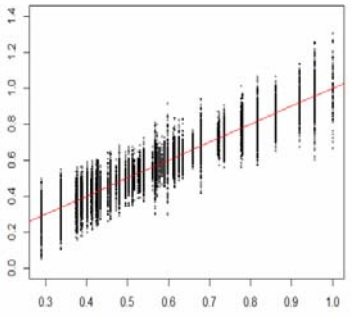
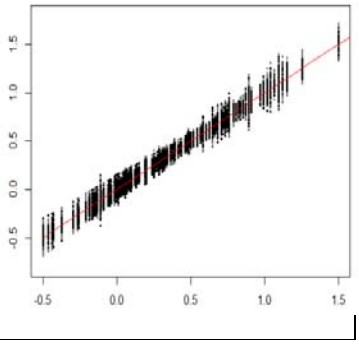
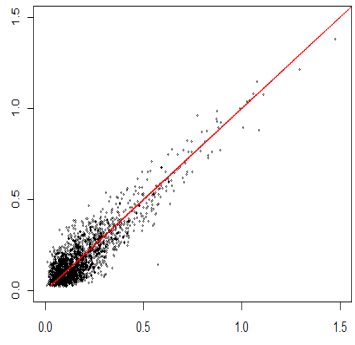
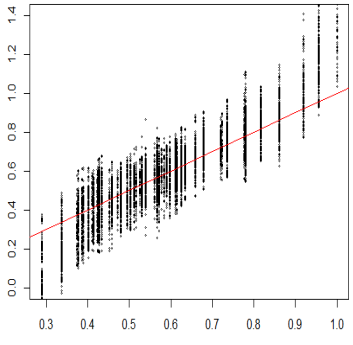
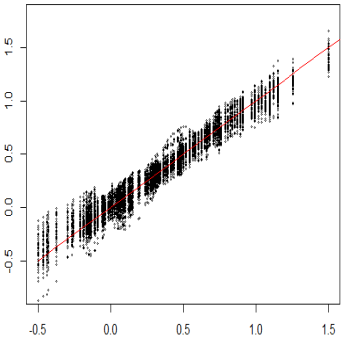
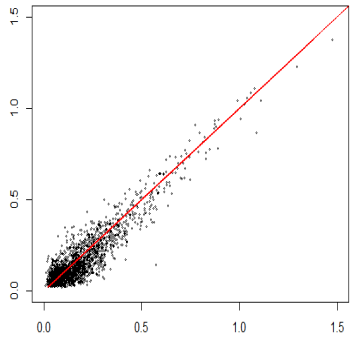
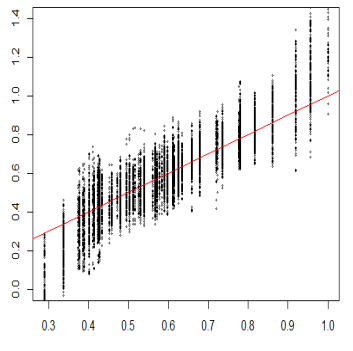
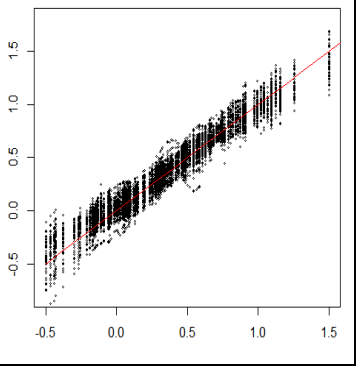
Moving to the arctan DGP with correlation (Figures A6 and A7), we find that the nonparametric estimator that allows for correlation is uniformly better than that which does not, but the same relativity does not hold for the translog function estimators. Allowing for correlation when you have the wrong functional form can be worse than ignoring that correlation even when it is present.

Collecting all this information, the results from the Monte Carlo experiment can be summarised as follows:

1. If there are strong correlations between the efficiency effects and the regressors, ignoring that correlation leads to significant biases in the estimates for elasticities and the efficiencies.

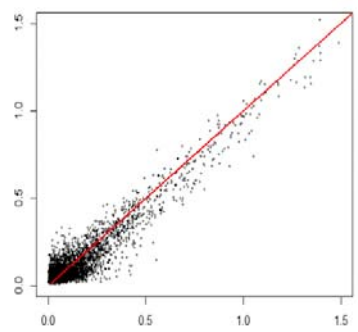
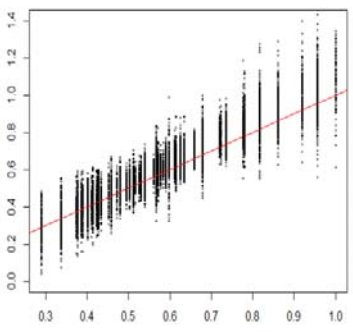
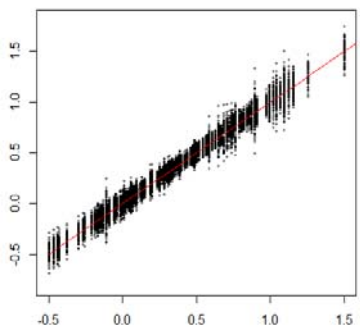
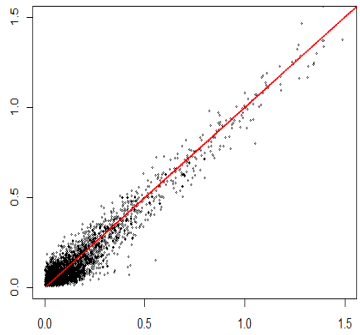
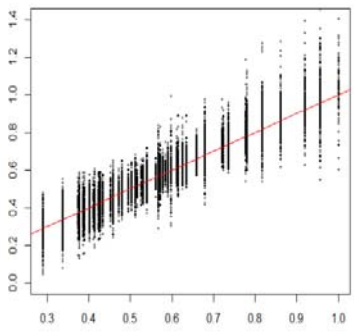
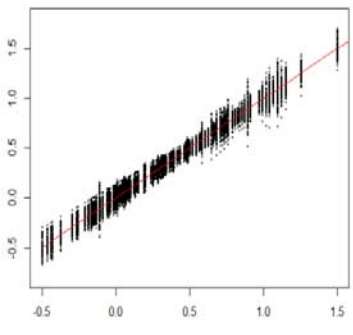
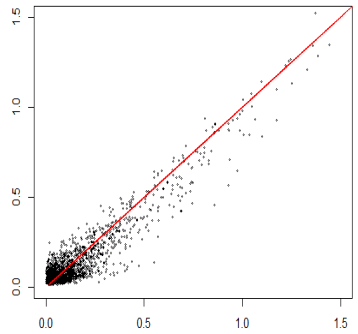
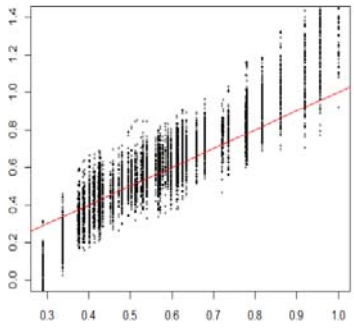
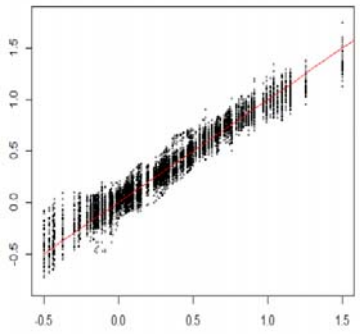
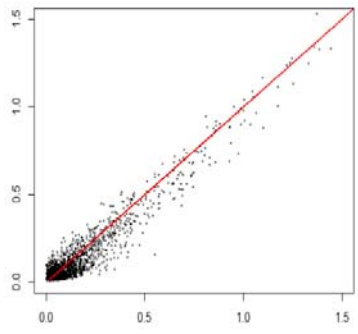
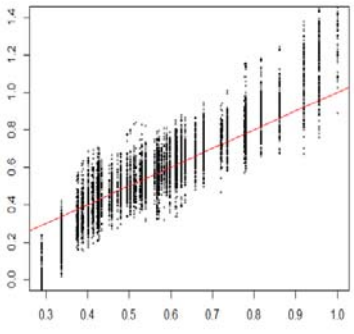
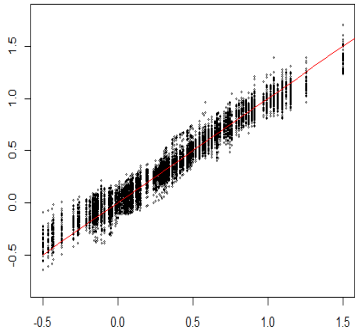
2. If the correlations are not strong, the bias will be small. This is somewhat expected because the efficiency effects are usually small and we have assumed time-invariant assumptions over 4 years.
3. The proposed models allowing for correlations work equally well whether there is correlation in the true data generating process or not.
4. Choosing the wrong functional form can lead to significant biases in all quantities of interests and this effect is more dramatic than ignoring correlations.
5. The nonparametric estimators have a consistent performance no matter what the true functional form for the technology is.

Figure A2 Estimates from DGP: Translog model without Correlation

Estimated Model	u	Elasticity-x1	Elasticity-x2
Translog without Correlation			
Translog with Correlation			
Nonpara without Correlation			
Nonpara with Correlation			

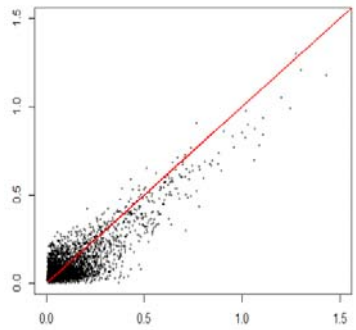
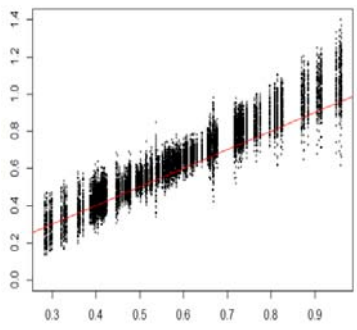
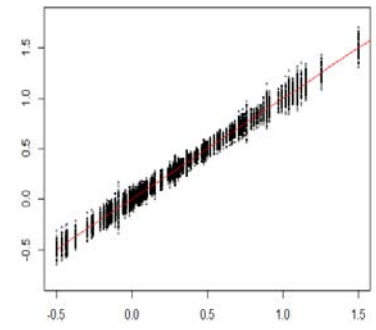
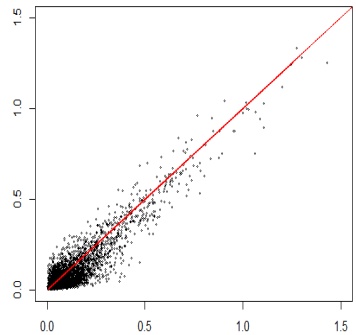
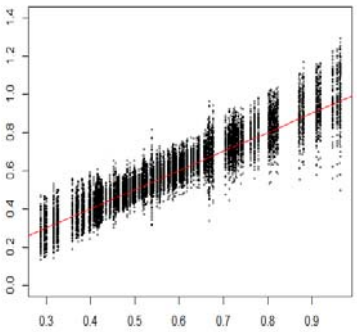
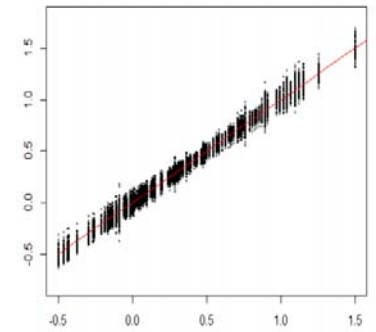
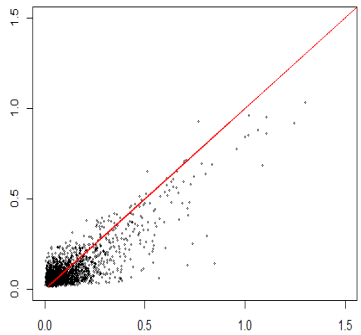
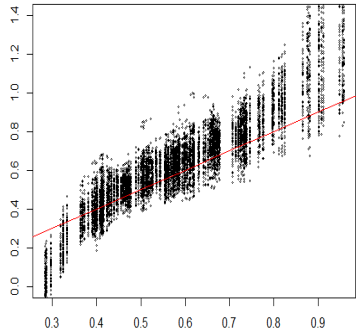
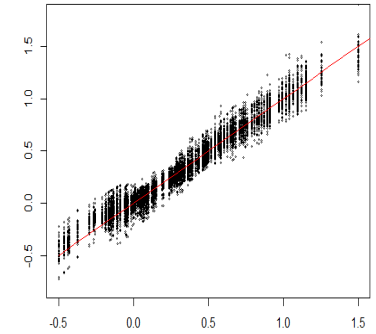
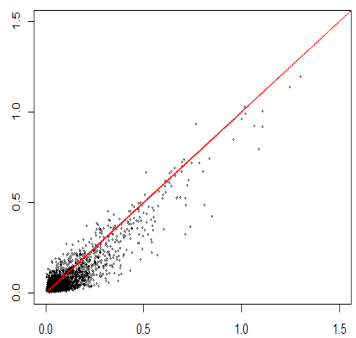
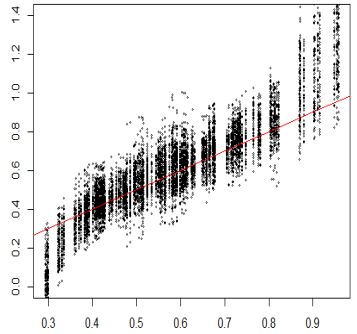
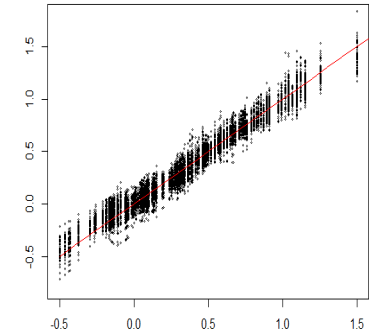
True values are given on the horizontal axes and estimates on the vertical axes.

Figure A3 Estimates from DGP: Translog Model with Moderate Correlation

Estimated Model	u	Elasticity-x1	Elasticity-x2
Translog without Correlation			
Translog with Correlation			
Nonpara without Correlation			
Nonpara with Correlation			

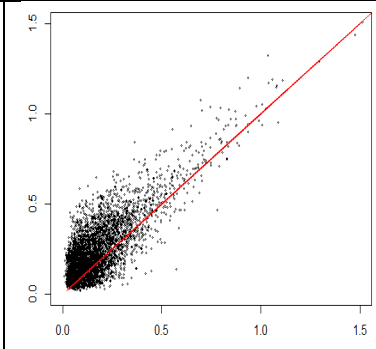
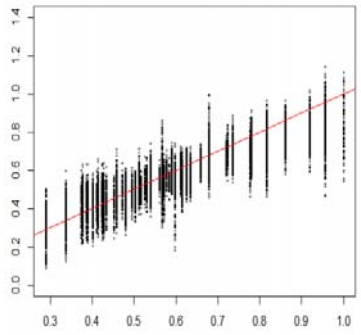
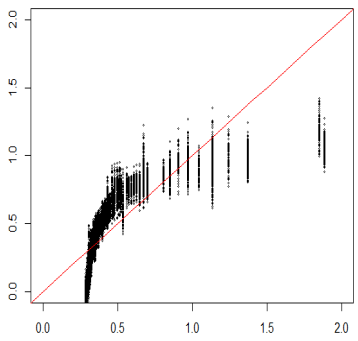
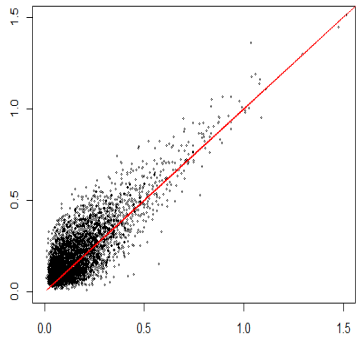
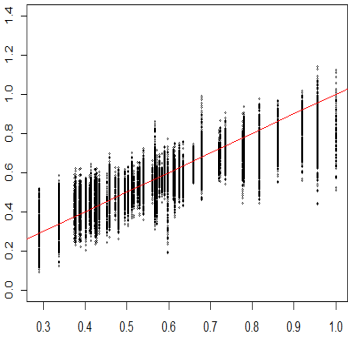
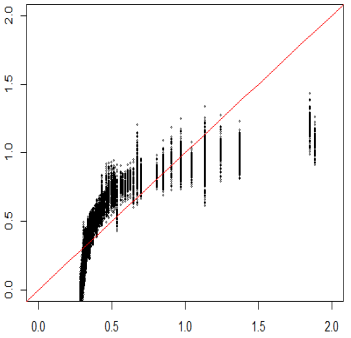
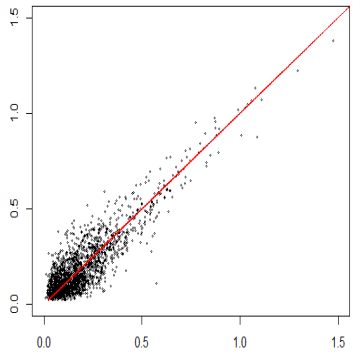
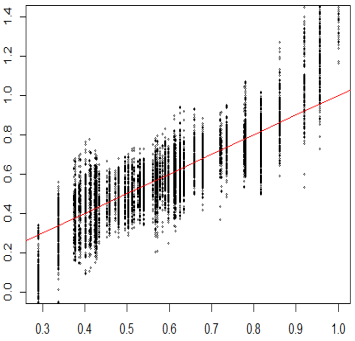
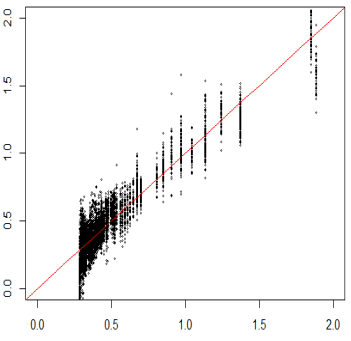
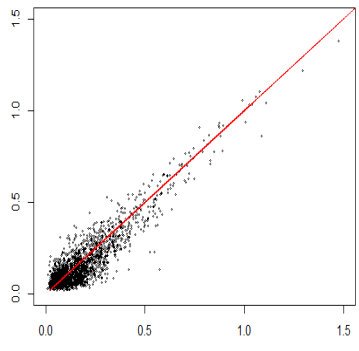
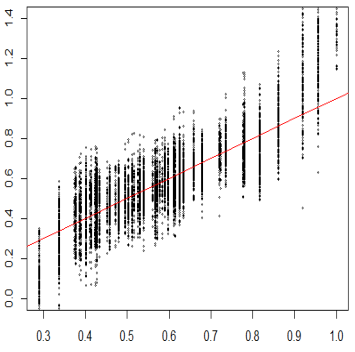
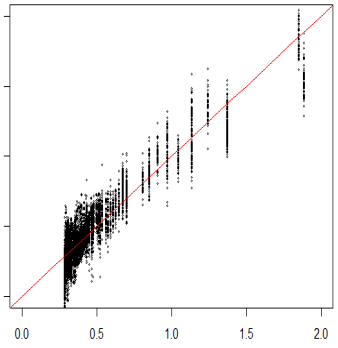
True values are given on the horizontal axes and estimates on the vertical axes.

Figure A4. Estimates from DGP: Translog Model with High Correlation

Estimated Model	u	Elasticity-x1	Elasticity-x2
Translog without Correlation			
Translog with Correlation			
Nonpara without Correlation			
Nonpara with Correlation			

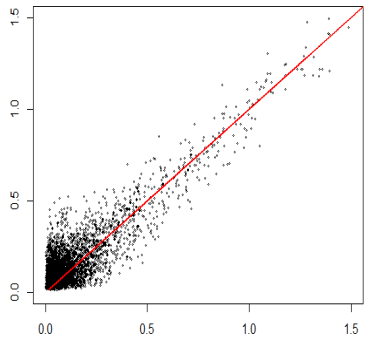
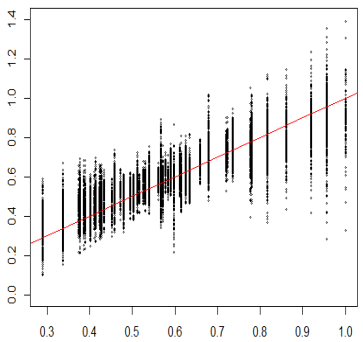
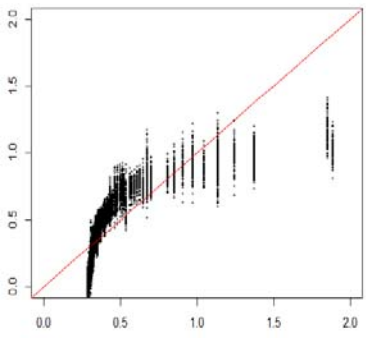
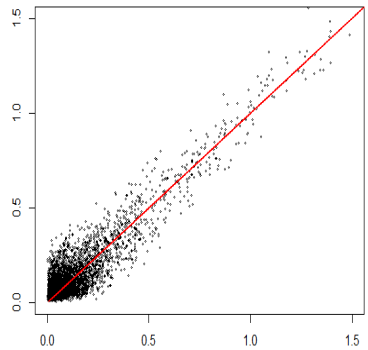
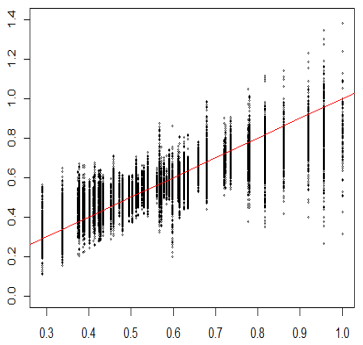
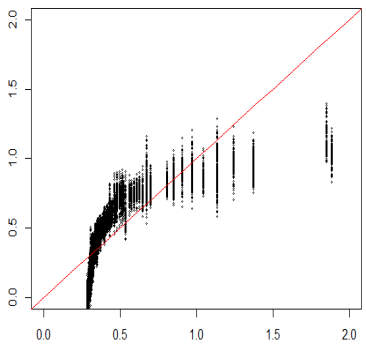
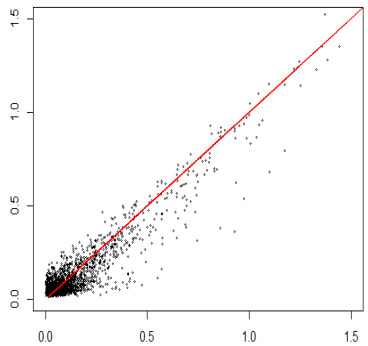
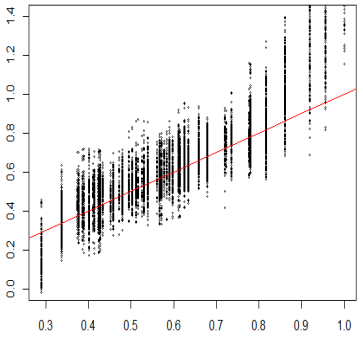
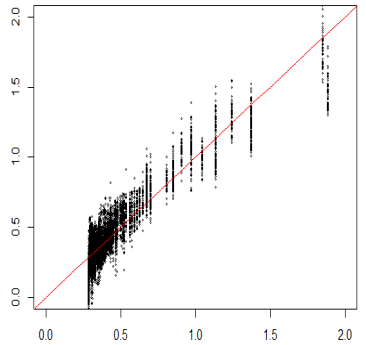
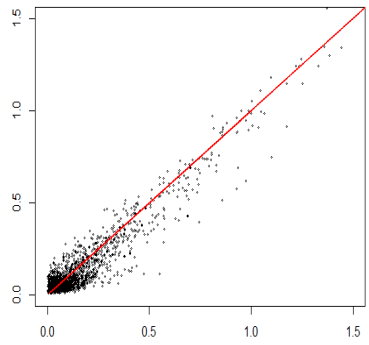
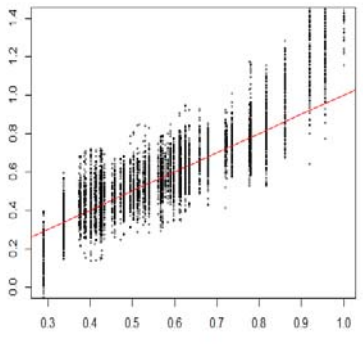
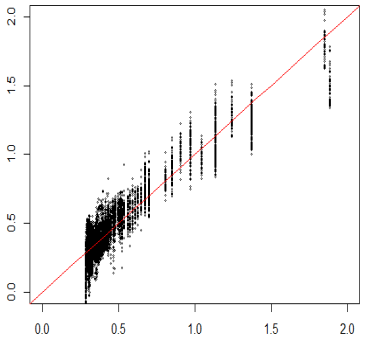
True values are given on the horizontal axes and estimates on the vertical axes.

Figure A5. Estimates from DGP: Arctan Model without Correlation

Estimated Model	u	Elasticity-x1	Elasticity-x2
Translog No Correlation			
Translog with Correlation			
Nonpara without Correlation			
Nonpara with Correlation			

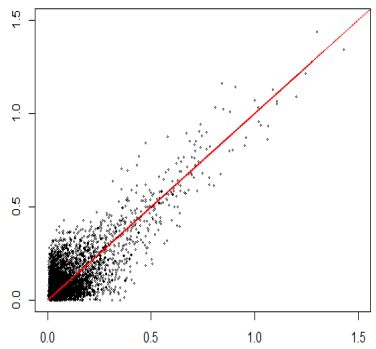
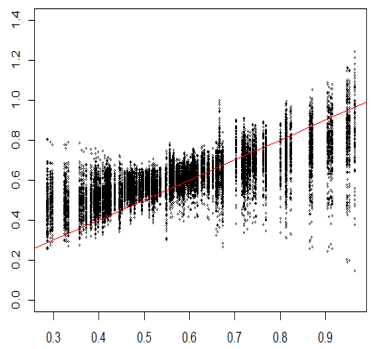
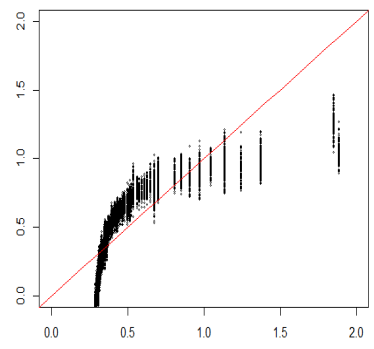
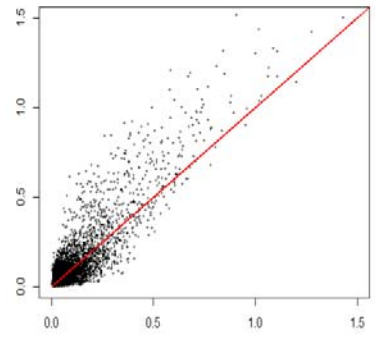
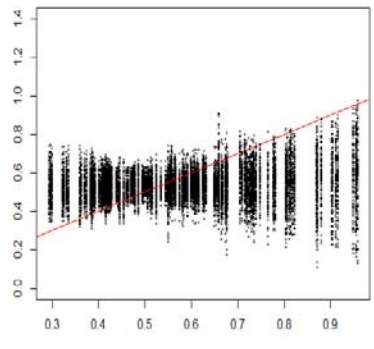
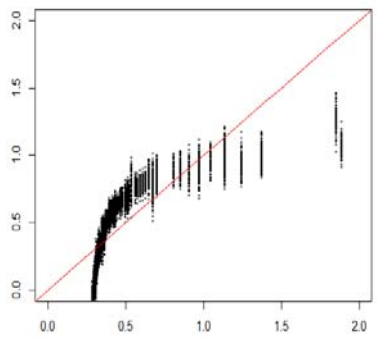
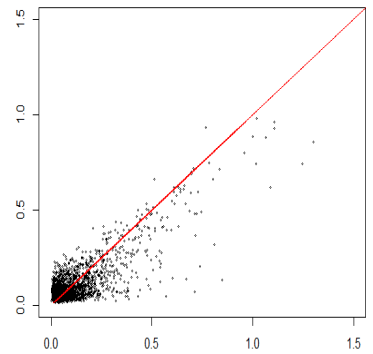
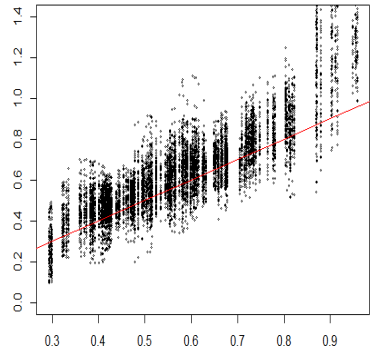
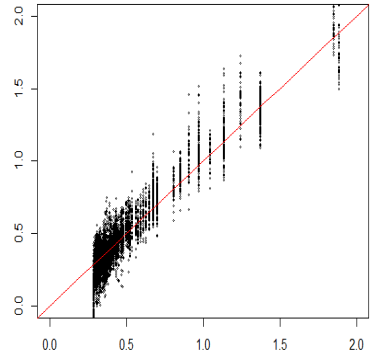
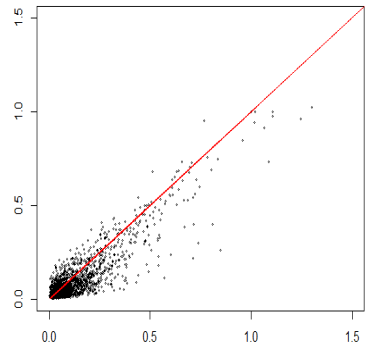
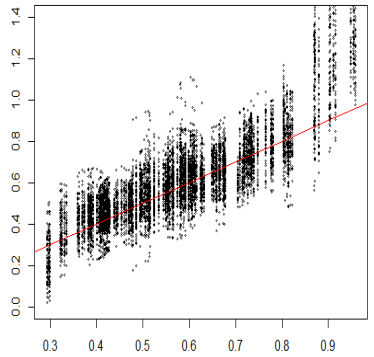
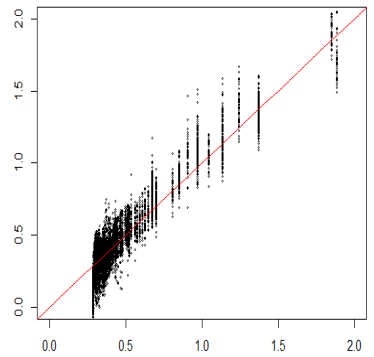
True values are given on the horizontal axes and estimates on the vertical axes.

Figure A6. Estimates from DGP: Arctan Model with Moderate Correlation

Estimated Model	u	Elasticity-x1	Elasticity-x2
Translog No Correlation			
Translog with Correlation			
Nonpara without Correlation			
Nonpara with Correlation			

True values are given on the horizontal axes and estimates on the vertical axes.

Figure A7. Estimates from DGP: Arctan Model with High Correlation

Estimated Model	u	Elasticity-x1	Elasticity-x2
Translog No Correlation			
Translog with Correlation			
Nonpara without Correlation			
Nonpara with Correlation			

True values are given on the horizontal axes and estimates on the vertical axes.