Consistent inference in fixed-effects stochastic frontier models*

Federico Belotti
University of Rome Tor Vergata

Giuseppe Ilardi
Bank of Italy

November 7, 2014

Abstract

The classical stochastic frontier panel data models provide no mechanism to disentangle individual time invariant unobserved heterogeneity from inefficiency. Greene (2005a,b) proposed the so-called "true" fixed-effects specification that distinguishes these two latent components while allowing for time varying inefficiency. However, due to the incidental parameters problem, the maximum likelihood estimator proposed by Greene may lead to biased variances estimates. We propose two alternative estimation procedures that, by relying on first difference data transformation, achieve consistency for $n \to \infty$ with fixed T. Furthermore, we generalize Chen et al. (2014) results providing a computationally feasible approach to estimate a "true" fixed-effects model in which the inefficiency can be heterogenous, heteroskedastic and can follow a dynamic process. We investigate the performances of the proposed estimators through a set of Monte Carlo experiments. Our results show good finite sample properties, especially in small samples. An application to hospitals' technical efficiency illustrates the usefulness of the new approach.

Keywords: Stochastic frontiers, Fixed-effects, Panel data, Marginal simulated likelihood, Pairwise differencing.

JEL: C13, C16, C23.

*The views expressed in this paper are those of the authors and do not involve the responsibility of the Bank of Italy. We thank Silvio Daidone and Francesco D'Amico for sharing their data with us. We thank Bill Greene, Vincenzo Atella, Clinton Horrace, Andrea Piano Mortari, Valentina Nigro and two anonymous referees for their helpful comments. We are indebted with Marco Perone Pacifico and Franco Peracchi for valuable advices and discussions. We also thank participants at the XII European Workshop on Efficiency and Productivity Analysis and the 46^{th} Italian Statistical Society Scientific Meeting, and seminar participants at the Bank of Italy and University of Rome Tor Vergata. All remaining errors are ours. An earlier version of this work have circulated with the title "Consistent estimation of the true fixed-effects stochastic frontier model". Corresponding author. Federico Belotti, Faculty of Economics, University of Rome Tor Vergata, via Columbia 2, 00133 Rome, Italy. E-mail: federico.belotti@uniroma2.it, phone/fax: +39 06 7259 5627

1 Introduction

The analysis of efficiency is an important issue in many empirical studies and the Stochastic Frontier (SF) models, introduced by Aigner et al. (1977) and by Meeusen & van den Broeck (1977), represent a popular tool to measure this unobservable economic indicator.[1] Since then, several improvements on SF modeling have led to a large empirical literature, most of which is based on panel data.[2]

Compared to cross-sectional data, longitudinal data has an important advantage as it allows to follow the same unit over time, thereby allowing to control and model unobserved heterogeneity. Depending on how this source of heterogeneity is treated, SF panel data literature can be classified into two major groups of models.[3] The first group treats time invariant heterogeneity as if it was inefficiency, thus not providing any mechanism to disentangle the former from the latter. This group includes, among others, Schmidt & Sickles (1984), Pitt & Lee (1981), Battese & Coelli (1988, 1992, 1995) and Kumbhakar (1990). On the other hand, the second group distinguishes between the aforementioned latent components by separating the inefficiency from the effect of time invariant omitted explanatory variables that are unrelated with the production process but affect the output (Kumbhakar & Hjalmarsson, 1995; Greene, 2005a,b; Wang & Ho, 2010; Emvalomatis, 2012; Chen et al., 2014).

This study reconsiders the estimation of the “true” fixed-effects (TFE) model in Greene (2005a) trying to solve the incidental parameters problem affecting his maximum likelihood dummy variables estimator (MLDVE). As Greene’s simulations suggest, this issue does not affect the frontier coefficients but it leads to inconsistent variances estimates. Since these parameters represent the key ingredients in the post-estimation of inefficiencies, a solution to this issue is crucial in the SF context. To this aim, we propose two alternative estimators that, by relying on a first-difference data transformation, avoid the incidental parameters problem and achieve consistency under both fixed-n and fixed-T asymptotics. The first is based on the marginalization of the inefficiency term via simulation and can be used to estimate normal-half

[1] Production frontiers represent the maximum amount of output that can be obtained from a given level of inputs, while a cost frontier characterizes the minimum expenditure required to produce a bundle of outputs given the prices of the inputs used in its production.

[2] See Kumbhakar & Lovell (2000) or Greene (2008) for recent surveys.

[3] See Greene (2005a,b) for a detailed review on the treatment of unobserved heterogeneity in SF models.

normal and normal-exponential models. The use of the maximum simulated likelihood (MSL) principle imposes a restriction for the model in which the inefficiency (and/or the idiosyncratic error) is allowed to be heteroskedastic: the scale parameter(s) can only be expressed as a function of time invariant exogenous explanatory variables. To overcome this limitation, we propose a second estimation strategy by exploiting the closeness property of the normal-exponential marginal likelihood function when $T = 2$ to define a U-estimator based on all pairwise quasi likelihood contributions.[4]

In a parallel research, Chen et al. (2014) propose a consistent estimator for the TFE normal-half normal homoskedastic model by exploiting the properties of the closed skew normal (CSN) class of distributions (Gonzalez-Farias et al., 2004a).[5] By using our pairwise estimation strategy, we generalize their results providing a computationally feasible approach to estimate a normal-truncated normal model in which the inefficiency can be heterogenous, i.e. its underlying mean can be expressed as a function of exogenous covariates, heteroskedastic and can follow a dynamic process. These extensions may be considered relevant from the methodological point of view since both model parameters and inefficiency estimates may be adversely affected when these features are neglected. Furthermore, they are also important from the empirical perspective because they allow to test specific hypothesis of interest and policy implications, avoiding biased two-step procedures.[6]

The remainder of this paper is organized as follows. Section 2 reviews the TFE model and describes the derivation of the marginal likelihood in the first-difference setup. Section 3 presents our consistent estimators, while Section 4 provides some extensions to the Chen et al. (2014) homoskedastic results. Section 5 investigates the small sample properties of the proposed estimators through a set of Monte Carlo experiments. Section 6, provides an illustration through an application on hospitals technical efficiency. Finally, Section 7 offers some conclusions.

[4]In the rest of this paper, the term *marginal* refers to a model in which the fixed-effects have been wiped-out from the parameter space through a data transformation.

[5] An earlier version of this and Chen et al. (2014) paper has been presented at the *XII* European Workshop on Efficiency and Productivity Analysis.

[6]See Wang & Schmidt (2002) for further details on the problems affecting two-step estimation procedures.

2 The model

Consider the following specification for a fixed-effects stochastic production frontier model

$$
\begin{aligned}
y_{it} &= \alpha_i + \boldsymbol{x}_{it}\boldsymbol{\beta} + \varepsilon_{it}, && (1)\\
\varepsilon_{it} &= v_{it} - u_{it}, && (2)\\
v_{it} &\sim IID\ \mathcal{N}(0, \psi^2), && (3)\\
u_{it} &\sim IID\ \mathcal{F}_u(\sigma), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, && (4)
\end{aligned}
$$

where, for each unit i and period t, y_{it} represents the level of output, $\boldsymbol{x}_{it}$ is a $1 \times k$ vector of exogenous inputs, $\boldsymbol{\beta}$ is a $k \times 1$ vector of technology parameters and α_i is the unit fixed-effect.[7] The composite error term ε_{it} is the difference between the idiosyncratic error v_{it}, and the one-sided disturbance u_{it} which represents inefficiency.[8] As customary in SF literature, we assume that v_{it} and u_{it} are independently distributed. The inefficiency u_{it} is distributed according to $\mathcal{F}_u$, a generic member of the "one-parameter" family of distributions with support defined over $\mathbb{R}^+$ (e.g., half-normal or exponential) and scale parameter σ, while v_{it} is normally distributed.[9] We define σ/ψ as the signal-to-noise ratio (STN), which provides an indication of the relative contributions of u_{it} and v_{it} to the variability of ε_{it}. As already mentioned, there is a philosophical debate in the SF literature about whether or not separate α_i from u_{it}. In what follows, we sidestep this issue and focus instead on how to consistently estimate the parameters of model (1)$-$(4).

Greene (2005a) shows that, by treating the unit-specific intercepts as parameters to be estimated, the maximization of the likelihood function for the model (1)$-$(4) is computationally feasible also in presence of a large number of nuisance parameters ($> 1,000$). However, as Greene's simulations suggest, this approach may lead to inconsistent variance estimates, especially in short panels.[10]

[7] For notational simplicity, we assume balanced panels but all the results can be easily generalized to the unbalanced case.

[8] Notice that when the sign of the last term in (2) is positive, the model describes a stochastic cost frontier.

[9] Contrary to Schmidt & Sickles (1984), Cornwell et al. (1990) and Lee & Schmidt (1993), these distributional assumptions are required here for identification purposes.

[10] The incidental parameters problem is no longer an issue for the MLDVE when $T \to \infty$ with fixed n. The MLDVE shows very good finite sample properties when the longitudinal dimension is large enough ($T > 15$, see

A natural strategy to avoid this issue consists in eliminating the nuisance parameters through a data transformation. This strategy has been followed for instance by Wang & Ho (2010) for a model in which inefficiency is assumed to change deterministically over time, that is $u_{it} = h_{it}u_i$ where $h_{it} = \exp(\boldsymbol{z}_{it}\boldsymbol{\delta})$ and $u_i \sim \mathcal{N}^+(\mu, \sigma_u^2)$. As noted by the authors, this assumption is critical for deriving the marginal likelihood since the distribution of u_i is not affected by the first-difference transformation. In what follows, we do not impose any constraint on the variability of u_{it} at the cost of a more challenging marginal likelihood derivation.

The first-difference transformation implies that model (1)$-$(4) may be rewritten as

$$\begin{aligned}
\Delta\boldsymbol{y}_i &= \Delta X_i\boldsymbol{\beta} + \Delta\boldsymbol{\varepsilon}_i, && (5)\\
\Delta\boldsymbol{\varepsilon}_i &= \Delta\boldsymbol{v}_i - \Delta\boldsymbol{u}_i, && (6)\\
\Delta\boldsymbol{v}_i &\sim IID\ \ \mathcal{N}_{T-1}(\mathbf{0}, \Psi), && (7)\\
\Delta\boldsymbol{u}_i &\sim IID\ \ \mathcal{F}_{\Delta\boldsymbol{u}}(\sigma), \quad i = 1, \ldots, n, && (8)
\end{aligned}$$

where $\Delta\boldsymbol{y}_i = (\Delta y_{i2}, \ldots, \Delta y_{iT})$ with $\Delta y_{it} = y_{it} - y_{it-1}$ and ΔX_i is the $T-1 \times k$ matrix of time-varying covariates with the t-th row denoted by $\Delta\boldsymbol{x}_{it} = (\Delta x_{it1}, \ldots, \Delta x_{itk})$, $\forall\ t = 2, \ldots, T$. The normality assumption for v_{it} implies that $\Delta\boldsymbol{v}_i$ has a $T-1$-variate normal distribution with covariance matrix $\Psi = \psi^2\Lambda_{T-1}$, where Λ_{T-1} is the symmetric tridiagonal $T-1 \times T-1$ matrix

$$\Lambda_{T-1} = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -1 \\ 0 & 0 & \ldots & -1 & 2 \end{pmatrix}. \tag{9}$$

On the other hand, the multivariate distribution of $\Delta\boldsymbol{u}_i$ is generally unknown. Nevertheless, given the independence assumption between $\Delta\boldsymbol{v}_i$ and $\Delta\boldsymbol{u}_i$, the marginal likelihood contribution

Section 5.1).

L_i^* can be defined in general terms as

$$\begin{aligned} L_i^*(\boldsymbol{\theta}) &= \int f(\Delta \boldsymbol{v}_i, \Delta \boldsymbol{u}_i | \boldsymbol{\theta}) \, d\Delta \boldsymbol{u}_i = \int f(\Delta \boldsymbol{v}_i | \boldsymbol{\theta}) f(\Delta \boldsymbol{u}_i | \sigma) \, d\Delta \boldsymbol{u}_i \\ &= \int f(\Delta \boldsymbol{y}_i | \boldsymbol{\beta}, \psi, \Delta X_i, \Delta \boldsymbol{u}_i) f(\Delta \boldsymbol{u}_i | \sigma) \, d\Delta \boldsymbol{u}_i, \end{aligned} \tag{10}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma, \psi)$. The marginalization of $\Delta \boldsymbol{u}_i$ hides two challenges: the multivariate density function $f(\Delta \boldsymbol{u}_i | \sigma)$ is unknown; the integral (10) does not lead to a closed-form expression. In the next Section we present two estimation strategies that overcome these issues.

3 Estimation

3.1 Marginal maximum simulated likelihood estimation

The marginalization in equation (10) can be performed by simulation, in a way similar to the elimination of nuisance parameters in nonlinear random-effects models. We propose to estimate model (5)$-$(8) treating the marginal likelihood function as an expectation with respect to the random vector $\Delta \boldsymbol{u}_i$. Lemma 1 explicitly states the distributional assumptions required to apply the MSL approach.

Lemma 1 *Assume that (i) the distribution of the inefficiency $\mathcal{F}_{\boldsymbol{u}}$ belongs to a one-parameter family of distributions with a support defined over $\mathbb{R}^+$ and scale parameter σ; (ii) $\mathcal{F}_u$ exhibits a scaling property, i.e. $\boldsymbol{u}_i = \sigma \tilde{\boldsymbol{u}}_i$ where the rescaled inefficiency $\tilde{\boldsymbol{u}}_i \sim \mathcal{F}_{\boldsymbol{u}}(1)$. Then the multivariate distribution of the rescaled first-differenced inefficiency $F_{\Delta \tilde{\boldsymbol{u}}}$ is parameter free, such that $\Delta \boldsymbol{u}_i = \sigma \Delta \tilde{\boldsymbol{u}}_i$.*

The first assumption rules out the possibility of using two-parameters distributions (e.g., truncated normal or gamma) for the inefficiency term. While two-parameters flexible distributions have been widely and successfully applied in many studies, the estimation of the location/shape parameter can be hard in small samples. This practical identification issue has been raised for a normal-gamma convolution by Ritter & Simar (1997). Even if less flexible, models based on half-normal and exponential inefficiency are not affected by this practical identification problem and can be very useful in many empirical applications. Furthermore, both half-normal and exponential distributions are characterized by the scaling property and therefore the second assumption does not appear restrictive.

The marginal likelihood contribution L_i^* can be expressed in term of its simulated counterpart as

$$\begin{aligned} L_i^*(\boldsymbol{\theta}) &= \int f(\Delta \boldsymbol{y}_i|\boldsymbol{\theta}, \Delta X_i, \Delta \boldsymbol{u}_i) f(\Delta \boldsymbol{u}|\sigma) d\Delta \boldsymbol{u}_i = && (11)\\ &= \mathbb{E}_{\Delta\tilde{\boldsymbol{u}}}\left[\phi_{T-1}\left(\Delta\boldsymbol{\varepsilon}_i + \sigma\Delta\tilde{\boldsymbol{u}}_i; \mathbf{0}, \Psi\right)\right] = && (12)\\ &\approx \frac{1}{G}\sum_{g=1}^{G}\left[\phi_{T-1}\left(\Delta\boldsymbol{\varepsilon}_i + \sigma\Delta\tilde{\boldsymbol{u}}_{ig}; \mathbf{0}, \Psi\right)\right], && (13) \end{aligned}$$

where $\Delta\boldsymbol{\varepsilon}_i = \Delta\boldsymbol{y}_i - \Delta X_i\boldsymbol{\beta}$, $\phi_{T-1}(.; \boldsymbol{m}, V)$ is the $T-1$-variate Gaussian density with mean vector $\boldsymbol{m}$ and covariance matrix V, and G is the number of draws from the multivariate distribution of the first-differenced rescaled inefficiency $\mathcal{F}_{\Delta\tilde{\boldsymbol{u}}}$. Therefore, all we need is to simulate from the distribution of $\Delta\tilde{\boldsymbol{u}}_i$ and this task is straightforward if one is able to simulate from the univariate distribution of u_{it}. Indeed, the vector $\Delta\tilde{\boldsymbol{u}}_{ig} = (\tilde{u}_{i2g} - \tilde{u}_{i1g}, \ldots, \tilde{u}_{iTg} - \tilde{u}_{i(T-1)g})'$ is, by construction, a valid draw from $\mathcal{F}_{\Delta\tilde{\boldsymbol{u}}}$. The resulting marginal maximum simulated likelihood estimator (MMSLE) is asymptotically equivalent to the marginal maximum likelihood estimator (MMLE) as $G \to \infty$. It then follows that the accuracy of this approach relies on whether G is sufficiently large to guarantee that the average (13) is a good approximation of the expectation (12).

For practical implementation, two issues must be considered: (i) how many draws are needed to obtain a good approximation and (ii) how to simulate them efficiently. On the first point, the literature is heterogenous. As pointed out by Greene (2003), the rule of thumb “the more the better” is not helpful when time (and computational power) becomes a consideration. In fact, the marginal benefit of additional draws eventually becomes nil. The second consideration concerns how to obtain the draws efficiently. Numerous procedures have been recently proposed in the numerical analysis literature to reduce the computational burden related to the use of pseudo-uniform random draws (Morokoff & Caflisch, 1995; Sloan & Woźniakowski, 1998). We propose to use Halton sequences (Halton, 1960), a computationally efficient alternative which has been extensively used for the implementation of the MSL estimation technique (see, among others, Train, 2000; Bhat, 2001; Greene, 2003). This strategy makes the MMSLE manageable even in the case of moderately large sample sizes (e.g., $n = 1000$, $T = 10$). We have conducted a simulation study to assess the minimum number of Halton sequences needed to get a suitable approximation of the expectation (12). The rule of thumb is to use at least 10 sequences for observation (at least 30 in the case of heteroskedastic inefficiency).[11]

It is worth emphasizing that the MMSLE may be directly applied also when the within-group transformation is used to remove the fixed-effects. In this case, the likelihood contribution of each unit is

$$L_i^*(\boldsymbol{\theta}) \approx \frac{1}{G}\sum_{g=1}^{G}\left[\phi_T\left(\nabla \boldsymbol{y}_i - \nabla X_i\boldsymbol{\beta} + \sigma\nabla\tilde{\boldsymbol{u}}_{ig}; \mathbf{0}, \Psi\right)\right], \tag{14}$$

where "∇" denotes within-group transformed variables. Again, the vector $\nabla\tilde{\boldsymbol{u}}_{ig} = (\tilde{u}_{i1g} - \bar{u}_{ig}, \ldots, \tilde{u}_{iTg} - \bar{u}_{ig})'$ is a valid draw from $F_{\nabla\tilde{\boldsymbol{u}}}$, with $\bar{u}_{ig} = \frac{1}{T}\sum_{t=1}^{T}\tilde{u}_{itg}$.

3.1.1 Heteroskedastic case

The estimation procedure described above can be easily generalized to the case of heteroskedastic inefficiency. As already mentioned, this extension is important since both model parameters and inefficiency estimates may be adversely affected by neglected heteroskedasticity.[12] Moreover, the inclusion of explanatory variables correlated with inefficiency but not with unobserved heterogeneity may enhance parameters identification. Unfortunately, introducing heteroskedasticity in $\boldsymbol{u}$ requires the following additional assumption.

Assumption iii) *The scale parameter of the inefficiency distribution $\mathcal{F}_u$ is time invariant for each unit, i.e. $\boldsymbol{\sigma}_i = g(Z_i\boldsymbol{\delta})$ where $g(.)$ is a known positive monotonic function, Z_i is a $T-1 \times s$ matrix of time-invariant covariates and $\boldsymbol{\delta}$ is a $s \times 1$ vector of parameters to be estimated.*

The time invariant nature of Z_i ensures that the simulated draws $\Delta\tilde{u}_{itd}$ are identically distributed for each unit of the panel. The usefulness of this heteroskedastic specification relies on the

[11]Notice that the same approximation can be reached using 300 pseudo-random draws for observation. In our case the computational efficiency compared to pseudo-uniform random draws appears to be at least 10 to 1.

[12] The multi-stage approach by Kumbhakar & Hjalmarsson (1995) represents a simple way to estimate parameters of model (1)−(4). However, in presence of neglected heteroskedasticity driven by variables that are correlated with those included in the frontier's equation, it may lead to biased estimates also in the first stage.

tenability of assumption iii), which in turn is strictly related to the specificity of the empirical application. The marginal likelihood contribution becomes

$$L_i^*(\boldsymbol{\theta}) \quad \approx \quad \frac{1}{G}\sum_{g=1}^{G}\left[\phi_{T-1}\left(\Delta\boldsymbol{\varepsilon}_i + \boldsymbol{\sigma_i}\odot\Delta\tilde{\boldsymbol{u}}_{ig};\mathbf{0},\Psi\right)\right], \tag{15}$$

where the symbol $\odot$ represents the element-wise product.[13] Notice that when the inefficiency is assumed to be heteroskedastic, the STN ratio can be defined as an average, that is $\bar{\lambda} = \frac{1}{n\psi}\sum_{i=1}^{n}\boldsymbol{\sigma}_i$.

3.2 Pairwise difference estimator

In order to relax assumption iii), we propose to exploit the closed-form expression of the integral (10) when the inefficiency is exponentially distributed and $T = 2$.[14] In fact, the difference $\Delta U = U_2 - U_1$ between two independent but not identically distributed exponential random variables $U_t \sim \mathcal{E}(\varsigma_t)$ with $\varsigma_t = \sigma_t^{-1}$, $t = 1, 2$, has the following p.d.f.

$$f(\Delta u|\varsigma_1,\varsigma_2) = \begin{cases} \frac{\varsigma_1\varsigma_2}{\varsigma_1+\varsigma_2}\exp\left(\varsigma_1\Delta u\right), & \text{if} \quad \Delta u < 0, \\ \frac{\varsigma_1\varsigma_2}{\varsigma_1+\varsigma_2}\exp\left(-\varsigma_2\Delta u\right), & \text{otherwise.} \end{cases} \tag{16}$$

Thus, by allowing the scale parameter of the inefficiency distribution to depend on a set of exogenous explanatory variables $\sigma_{it} = \exp(\boldsymbol{z}_{it}\boldsymbol{\gamma})$, the resulting marginal likelihood function for

[13] As discussed in Section 3.1, the within-group transformation can alternatively be used to remove the fixed-effects.

[14] We also derive the p.d.f. of the random vector $\Delta\boldsymbol{u}_i$ for the general case ($T > 2$). However, the subsequent marginalization cannot be performed in closed form. These details are available upon request.

a two-period panel is

$$
\begin{aligned}
L^*(\boldsymbol{\theta}) &= \prod_{i=1}^{n} f(\Delta y_{it}|\boldsymbol{\theta}, \Delta \boldsymbol{x}_{it}) \\
&= \prod_{i=1}^{n} \left\{ \int_{\mathbb{R}} f(\Delta y_{it}|\boldsymbol{\beta}, \psi, \Delta \boldsymbol{x}_{it}, \Delta u_{it}) f(\Delta u|\varsigma_1, \varsigma_2) d\Delta u_{it} \right\} \\
&= \prod_{i=1}^{n} \left\{ \int_{\mathbb{R}} \frac{1}{(4\pi\psi^2)^{T/2}} \exp\left[-\frac{1}{2} \frac{\Delta\varepsilon_{it} - \Delta u_{it}}{2\psi^2} \right] d\Delta u_{it} \right\} \\
&= \prod_{i=1}^{n} \left\{ \frac{1}{(4\pi\psi^2)^{T/2}} \left[\int_{\mathbb{R}_+} \exp\left(-\frac{1}{2} \frac{(\Delta\varepsilon_{it} - \Delta u_{it})^2}{2\psi^2} \right) d\Delta u_{it} + \int_{\mathbb{R}_-} \exp\left(-\frac{1}{2} \frac{(\Delta\varepsilon_{it} - \Delta u_{it})^2}{2\psi^2} \right) d\Delta u_{it} \right] \right\} \\
&= \prod_{i=1}^{n} \frac{\varsigma_{i1}\varsigma_{i2}}{(\varsigma_{i1} + \varsigma_{i2})(4\pi\psi^2)^{T/2}} \left\{ \int_{\mathbb{R}_+} \exp\left[-\frac{1}{2} \left(\frac{(\Delta\varepsilon_{it} - \Delta u_{it})^2}{2\psi^2} + 2\varsigma_{i2}\Delta u_{it} \right) \right] d\Delta u_{it} \right. \\
&\quad \left. + \int_{\mathbb{R}_-} \exp\left[-\frac{1}{2} \left(\frac{(\Delta\varepsilon_{it} - \Delta u_{it})^2}{2\psi^2} - 2\varsigma_{i1}\Delta u_{it} \right) \right] d\Delta u_{it} \right\} \\
&= \prod_{i=1}^{n} \left\{ \frac{\varsigma_{i1}\varsigma_{i2}}{(\varsigma_{i1} + \varsigma_{i2})} \exp\left(\varsigma_{i2}^2\psi^2 - \varsigma_{i2}\Delta\varepsilon_{it} \right) \times \right. \\
&\quad \left. \times \left[\Phi\left(\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi} - \sqrt{2}\varsigma_{i2}\psi \right) + \exp\left[\psi^2(\varsigma_{i1}^2 - \varsigma_{i2}^2) + (\varsigma_{i1} + \varsigma_{i2})\Delta\varepsilon_{it} \right] \Phi\left(-\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi} - \sqrt{2}\varsigma_{i1}\psi \right) \right] \right\}, \qquad (17)
\end{aligned}
$$

where $\Delta\varepsilon_{it} = \Delta y_{it} - \Delta \boldsymbol{x}_{it}\boldsymbol{\beta}$, $\Phi(.)$ is the c.d.f. of a standard Gaussian random variable and $\varsigma_{it} = \exp(-\boldsymbol{z}_{it}\boldsymbol{\gamma})$ with $t = 1, 2$. Notice that the homoskedastic case can be easily obtained by substituting $\varsigma_{i1} = \varsigma_{i2} = \sigma^{-1}$. Similarly, heteroskedasticity in $\boldsymbol{v}$ can be introduced by modeling the variance of the idiosyncratic error.

The marginal likelihood function (17) implies the existence of $H = \binom{T}{2}$ consistent "subsample" estimators. In fact, if we limit the inference to a subsample extracted considering any two waves of the panel, we can still consistently estimate the vector of parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \psi)$. Even if each of these H estimators shares the asymptotic properties of a MMLE over the whole sample, it exploits only a portion of the available information implying high inefficiency in finite samples. Similarly to Abrevaya (1999), we can produce a more efficient estimator by combining the H marginal log-likelihoods in just one objective function. The resulting estimator can be viewed as a quasi MMLE for the whole sample in which the correlation between the subsamples is ignored. As pointed out by Poirier & Ruud (1988), the application of a quasi MLE in a non-linear panel data model in which the dependency structure among observations is mis-specified, often yields a consistent and asymptotically normal estimator.

Before discussing in detail the estimator, let us present this inferential process in a simplified case where only subsamples extracted from consecutive pairs of waves are considered. Similarly to the partial maximum likelihood approach used in Wang et al. (2013), by combining the marginal likelihood function defined in (17) we can define a "split-sample" estimator $\breve{\boldsymbol{\theta}}$ as

$$\breve{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}} \frac{1}{nT^*} \sum_{i=1}^{n} \sum_{t=2}^{T} \log f(\Delta y_{it}|\ \boldsymbol{\theta}, \Delta \boldsymbol{x}_{it}), \tag{18}$$

where $T^* = int(T/2)$. The distinguishing feature of this estimator from the MMLE is that we are *not* assuming a full sample likelihood factorization in term of the product of the subsamples likelihood contributions. Therefore this estimator is clearly not a maximum likelihood estimator (MLE) but, more properly, an M-estimator.

As shown by Honoré & Powell (1994), the split-sample estimator is inefficient compared to the estimator $\tilde{\boldsymbol{\theta}}$ defined as the maximizer of the following objective function

$$U_n(\boldsymbol{\theta}) = n^{-1} \binom{T}{2}^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{s<t} \log f(\Delta_t^s y_i|\ \boldsymbol{\theta}, \Delta_t^s \boldsymbol{x}_i), \tag{19}$$

where $\Delta_t^s y_i = y_{it} - y_{is}$ and $\Delta_t^s \boldsymbol{x}_i = \boldsymbol{x}_{it} - \boldsymbol{x}_{is}$. Following previous literature (Honoré & Powell, 1994; Abrevaya, 1999, among the others), we refer to the maximizer of (19) as pairwise difference estimator (PDE). This estimator maximizes an objective function which involves all pairs of waves, exploiting more information with respect to the split-sample estimator. Moreover, consistently with Abrevaya (1999), we find that the criterion function (19) is much smoother than the MMLE's objective function.

As already discussed, the relevant property of this inferential procedure is that the asymptotics are also determined by $n \to \infty$ with fixed T. In particular, its consistency and asymptotic normality are a direct consequence of consistency and asymptotic normality of the underlying subsample estimators.[15] As the PDE belongs to the class of U-estimators, its asymptotic

[15]A formal proof of the large sample properties may adapt the arguments used by Martins-Filho & Yao (2010) for the normal-half normal cross-sectional case. Details on the conditions for consistency and asymptotic normality of the PDE estimator assuming the validity of the subsamples MMLEs' asymptotic properties are available upon request from the corresponding author.

variance is equal to $A_0^{-1}B_0A_0^{-1}$, where

$$A_0 = -\sum_{t=2}^{T}\sum_{s<t}\mathbb{E}\left(\nabla_{\theta\theta}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)\right), \tag{20}$$

and

$$B_0 = \mathbb{E}\left\{\left[\sum_{t=2}^{T}\sum_{s<t}\nabla_{\boldsymbol{\theta}}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)\right]\left[\sum_{t=2}^{T}\sum_{s<t}\nabla_{\boldsymbol{\theta}}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)\right]'\right\}, \tag{21}$$

with ∇_θ and $\nabla_{\theta\theta}$ denote the vector of first derivatives and the Hessian matrix of the objective function respectively and $\boldsymbol{\theta}_0$ is the true parameter value. Estimation of the asymptotic variance of the PDE is straightforward, because for each pairwise difference we have that

$$-\mathbb{E}[\nabla_{\theta\theta}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)] = \mathbb{E}[\nabla_{\theta}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)\nabla_{\theta}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)'].$$

Therefore, the estimator of the asymptotic variance of $\tilde{\boldsymbol{\theta}}$ is given by

$$\widehat{Avar}(\tilde{\theta}) = n^{-1}\hat{A}_0^{-1}\hat{B}_0\hat{A}_0^{-1}, \tag{22}$$

where

$$\hat{A}_0 = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s<t}\nabla_{\theta}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \tilde{\boldsymbol{\theta}})\nabla_{\theta}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \tilde{\boldsymbol{\theta}}) \tag{23}$$

and

$$\hat{B}_0 = \hat{A}_0 + \frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s<t}\sum_{(k,h)\neq(t,s)}\nabla_{\boldsymbol{\theta}}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \tilde{\boldsymbol{\theta}})\nabla_{\boldsymbol{\theta}}\log f(\Delta_k^h y_i|\Delta_k^h \boldsymbol{x}_i, \tilde{\boldsymbol{\theta}}). \tag{24}$$

Notice that the consistency property of $\tilde{\boldsymbol{\theta}}$ can also be obtained under $T \to \infty$ with fixed n. However, the convergence arguments used above cannot be exploited in a fixed-n asymptotics because of the inconsistency of the subsample MMLEs. Nevertheless, a formal proof may be based on the large sample theory for minimizers of U-processes developed by Honoré & Powell (1994).

3.3 Fixed-effects and inefficiency scores

A fundamental feature of SF models is the estimation of technical (cost) inefficiency. The standard approach is to post-estimate inefficiency exploiting the conditional distribution of u_{it} given ε_{it}. Following Jondrow et al. (1982) (JLMS) a point estimate of u_{it} can be obtained using the mean of the conditional distribution, $\mathbb{E}(u_{it}|\hat{\varepsilon}_{it})$, evaluated at $\hat{\varepsilon}_{it} = y_{it} - \hat{\alpha}_i - \boldsymbol{x}_{it}\hat{\boldsymbol{\beta}}$.

Since we consider a transformed model in which the α_i are ruled out from the parameter space, the estimation of the fixed-effects has to be performed in a second stage. An efficient estimator for α_i can be obtained by maximizing the log-likelihood function of the untransformed model where the other parameters are substituted by a consistent estimates.[16] A simpler alternative when the inefficiency is assumed to be heteroskedastic is given by

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \boldsymbol{x}_{it}\hat{\boldsymbol{\beta}} + \hat{c}_{it}\right) \qquad i = 1, \ldots, n, \tag{25}$$

where $\hat{\boldsymbol{\beta}}$ and $\hat{c}_{it} = \mathbb{E}(u_{it}|\hat{\boldsymbol{\beta}}, \hat{\sigma}_{it})$ are consistent estimates. In particular, $\hat{c}_{it} = \hat{\sigma}_{it}$ when $u_{it} \sim \mathcal{E}(\sigma_{it})$ and $\hat{c}_{it} = \sqrt{2\pi^{-1}}\hat{\sigma}_{it}$ when $u_{it} \sim \mathcal{N}^+(0, \sigma^2_{it})$ ($\hat{\sigma}_{it} = \hat{\sigma}$ in the homoskedastic case).[17] This estimator is equivalent to the mean-adjusted estimator of α_i in the fixed-effects linear model and, therefore, it is consistent as $T \to \infty$.

4 Extensions

In this Section, we show how the pairwise difference estimation strategy can be used to extend the recent work of Chen et al. (2014) to the heteroskedastic normal-truncated normal and dynamic models.[18] In both cases we use the properties of the Closed Skew Normal (CSN) class of distributions (Gonzalez-Farias et al., 2004a). The multivariate CSN distributions have been introduced as a generalization of the Gaussian distribution to model, in a natural way, the skewness feature of the distribution. Thanks to their closeness under marginalization and linear

[16] Although our case is equivalent to the one reported in Section 2.2.2 of Wang & Ho (2010), an explicit formula for α_i cannot be obtained from the first-order condition of the log-likelihood due to the presence of the individual effect also in the arguments of the inverse Mills ratio.

[17] Equation (25) refers to the case of production frontiers. For cost frontiers, the $\hat{c}_{it}$ term enters the expression with a minus sign.

[18] We thank an anonymous referee for suggesting the extensions proposed in this Section.

transformations, this class of distributions naturally applies in the SF context. A comprehensive discussion of the CSN family and its properties can be found in Gonzalez-Farias et al. (2004b).

4.1 Normal-truncated normal model

A p-dimensional random vector $\boldsymbol{Y}$ is distributed according to a CSN distribution with parameters $\boldsymbol{\mu}$, Σ, D, $\boldsymbol{\nu}$ and Δ, denoted by $\boldsymbol{y} \sim CSN_{p,q}(\boldsymbol{\mu}, \Omega, D, \boldsymbol{\nu}, \Delta)$, if it is continuous with the following p.d.f

$$f(\boldsymbol{y}) = C\phi_p\left(\boldsymbol{y}; \boldsymbol{\mu}, \Omega\right) \Phi_q\left(D(\boldsymbol{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta\right), \quad \boldsymbol{y} \in \mathbb{R}^p \tag{26}$$

where $C^{-1} = \Phi_q\left(\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Omega D'\right)^{-1}$ and $\phi_p(.; \boldsymbol{\mu}, \Omega)$ and $\Phi_q(.; \boldsymbol{\nu}, \Delta)$ are the p-dimensional p.d.f. and q-dimensional c.d.f. of the Gaussian distribution, with $p \geq 1$, $q \geq 1$, $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\nu} \in \mathbb{R}^q$, D an arbitrary $q \times p$ matrix, Ω and Δ positive definite matrices of dimensions $p \times p$ and $q \times q$, respectively.

Dominguez-Molina et al. (2004, proposition 13.6.1) prove that the T-dimensional random variable $\boldsymbol{\varepsilon}_i = \boldsymbol{v}_i - \boldsymbol{u}_i$, for $i = 1, \ldots, n$ with $\boldsymbol{v}_i \sim \mathcal{N}_T(0, \Psi)$ and $\boldsymbol{u}_i \sim \mathcal{N}_T^+(\boldsymbol{\nu}, \Sigma)$, is distributed as

$$\boldsymbol{\varepsilon}_i \sim CSN_{T,T}(-\boldsymbol{\nu}, \Omega_*, -\Sigma\Omega_*^{-1}, -\boldsymbol{\nu}, \Delta_*), \quad i = 1, \ldots, n, \tag{27}$$

where $\Omega_* = \Sigma + \Psi$ and $\Delta_* = \Sigma - \Sigma(\Sigma + \Psi)^{-1}\Sigma$.

Similarly to Chen et al. (2014) who have used the within-group transformation in the case of a homoskedastic normal half-normal model, we use the first-difference transformation to eliminate the nuisance parameters in model (1)−(4) where equation (4) is replaced by $u_{it} \sim \mathcal{N}^+(\nu, \sigma^2)$. In what follows we show the derivation of the density of the random vector $\Delta\boldsymbol{\varepsilon}_i = A\boldsymbol{\varepsilon}_i$, where A is the following $T - 1 \times T$ matrix

$$A = \begin{pmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & 0 & \ldots & 0 & -1 & 1 \end{pmatrix}. \tag{28}$$

As noted before, the CSN family is closed to linear transformations. Hence, using the Proposition 2.3.1 of Gonzalez-Farias et al. (2004b), the random vector $\Delta\boldsymbol{\varepsilon}_i$ is distributed as

$$\Delta\boldsymbol{\varepsilon}_i \sim CSN_{T-1,T}(-\Delta\boldsymbol{\nu}, \Omega_A, D_A, -\boldsymbol{\nu}, \Delta_A), \quad i = 1, \ldots, n, \tag{29}$$

where $\Omega_A = A(\Sigma+\Psi)A'$, $D_A = -\Sigma A'(A\Sigma A' + A\Psi A')^{-1}$ and $\Delta_A = \Sigma - \Sigma A'(A(\Sigma+\Psi)A')^{-1}A\Sigma$, with p.d.f. given by[19]

$$f(\Delta\boldsymbol{\varepsilon}_i) \;=\; [\Phi_T(\mathbf{0}; -\boldsymbol{\nu}, \Sigma)]^{-1}\, \phi_{T-1}(\Delta\boldsymbol{\varepsilon}_i; -\Delta\boldsymbol{\nu}, \Omega_A)\, \Phi_T(D_A\Delta\boldsymbol{\varepsilon}_i; -\boldsymbol{\nu}, \Delta_A). \tag{30}$$

By exploiting this result, the marginal maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\nu}}, \widehat{\Sigma}, \widehat{\Psi})$ for the normal-truncated normal "true" fixed-effects model is

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}} \frac{1}{n}\sum_{i=1}^{n} \log f(\Delta\boldsymbol{\varepsilon}_i |\, \boldsymbol{\theta}, \Delta\boldsymbol{x}_{it}). \tag{31}$$

However, the maximization of (31) requires the numerical approximation of a T-dimensional normal integral for each unit in the panel. Regardless of the method used to approximate this integral, the higher its dimension, the higher the computational burden. As pointed out by Kumbhakar & Tsionas (2011), this approximation becomes cumbersome when $T > 5$ but can be handled when the inefficiency is homoskedastic. Indeed, similarly to Chen et al. (2014), the covariance matrix Δ_A in $\Phi_T(.)$ has a special (equicorrelated) structure and the computations may be greatly simplified following the result outlined in Kotz et al. (2000).

This simplification does not apply to the case of heteroskedastic errors, i.e. $\Sigma_i = diag(\sigma^2_{i1}, \ldots, \sigma^2_{iT})$ and $\Psi_i = diag(\psi^2_{i1}, \ldots, \psi^2_{iT})$, since Δ_A has not the aforementioned desirable structure anymore. In order to keep the estimation feasible, we propose to apply the PDE approach using the marginal likelihood function of a two-periods normal-truncated normal model. The marginal likelihood function can be straightforwardly obtained by considering $T = 2$ in equation (30) as

$$L^*(\boldsymbol{\theta}) = \prod_{i=1}^{n} [\Phi_2(\mathbf{0}; \tilde{\boldsymbol{\nu}}_i, \Sigma_i)]^{-1}\, \phi\left(\Delta\varepsilon_{it}; -\Delta\nu_{it}, \xi_i^2\right) \Phi_2\left(\tilde{\boldsymbol{d}}_i; \mathbf{0}, \Xi_i\right), \tag{32}$$

[19]The normal-half normal case can be straightforwardly obtained by replacing $\boldsymbol{\nu} = \mathbf{0}$ in (27).

where the location parameter $\tilde{\boldsymbol{\nu}}_i = \tilde{R}_i \boldsymbol{\tau}$ with $\tilde{R}_i$ a $2 \times r$ matrix of exogenous covariates, $\tilde{\boldsymbol{d}}_i = \{[(\sigma_{i1}^2, -\sigma_{i2}^2)' \xi_i^{-2}] (\Delta\varepsilon_{it} + \Delta\nu_{it})\} + \tilde{\boldsymbol{\nu}}_i$, $\Sigma_i = diag(\sigma_{i1}^2, \sigma_{i2}^2)$, $\xi_i^2 = \sigma_{i1}^2 + \sigma_{i2}^2 + \psi_{i1}^2 + \psi_{i2}^2$, $\sigma_{it} = \exp(\boldsymbol{z}_{it}' \boldsymbol{\gamma})$ and $\psi_{it} = \exp(\boldsymbol{w}_{it}' \boldsymbol{\delta})$, with $\boldsymbol{z}_{it}$ and $\boldsymbol{w}_{it}$ two vectors of explanatory variables, and

$$\Xi_i = \begin{pmatrix} \sigma_{i1}^2 & 0 \\ 0 & \sigma_{i2}^2 \end{pmatrix} - \xi_i^{-2} \begin{pmatrix} \sigma_{i1}^4 & -\sigma_{i1}^2 \sigma_{i2}^2 \\ -\sigma_{i1}^2 \sigma_{i2}^2 & \sigma_{i2}^4 \end{pmatrix}.$$

When $T > 2$, the evaluation of T-dimensional normal integrals, which make problematic the extension of the Chen et al. (2014) approach to the heteroskedastic case, can be replaced by the approximation of $\binom{T}{2}$ 2-dimensional normal integrals whose evaluation has been shown to be accurate and computationally efficient (Genz, 2004). Hence, the PDE objective function is

$$U_n(\boldsymbol{\theta}) = n^{-1} \binom{T}{2}^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{s<t} \log \left\{ [\Phi_2(\mathbf{0}; \tilde{\boldsymbol{\nu}}_i, \Sigma_i)]^{-1} \phi\left(\Delta_t^s \varepsilon_i; -\Delta_t^s \nu_i, \xi_i^2\right) \Phi_2\left(\tilde{\boldsymbol{d}}_i; \mathbf{0}, \Xi_i\right) \right\}, \quad (33)$$

where the pairwise difference operator Δ_t^s is defined in equation (19). Finally, inefficiency estimates can be obtained using the procedure described in Section 3.3 substituting $\hat{c}_{it} = \hat{\nu}_{it} + \sqrt{\frac{2}{\pi}} \hat{\sigma}_{it}$, with $\hat{\nu}_{it} = \tilde{\boldsymbol{r}}_{it} \boldsymbol{\tau}$.

4.2 A dynamic inefficiency model

The sources of inefficiency dynamics may be manifold. First, the dynamics can be linked to parametric functions of time or time varying observable factors that are under the influence of firms. Second, since the production process may be affected by unexpected events, inefficiency can be consider a stochastic variable that randomly varies over time. Third, some inputs are considered "fixed" in the short run because the economic environment places a high costs on adjusting these input levels. In this case, we expect the inefficiency to be persistent, that is inefficiency in one period is influenced by its past levels. The normal truncated-normal model considered in Section 4.1 accommodates the first two sources of dynamics: the inefficiency randomly vary over time and both its location and scale may depend on a set of observable factors. However, an "endogenous" dynamics of inefficiency has not yet been included in the model.

A new generation of dynamic frontier approaches is emerging with the aim of disentangling the long-run from the short-run inefficiency levels (Ahn & Sickles, 2000; Tsionas, 2006; Emvalomatis, 2012). To the best of our knowledge, Emvalomatis (2012) is the only study in which unobserved heterogeneity is separated from a first order autoregressive inefficiency. However, the model is estimated through a Bayesian correlated random effects approach in which a parametric distribution for the unit-specific effects must be specified.

We propose instead to introduce the aforementioned dynamics in a fixed-effects framework, by adequately specifying Σ in (27). In particular, we consider the following heteroskedastic normal-half normal model with $AR(1)$ inefficiencies

$$\begin{aligned} \boldsymbol{v}_i &\sim \mathcal{N}_T(\mathbf{0}, \Psi_i), & (34)\\ \boldsymbol{u}_i &\sim \mathcal{N}_T^+(\mathbf{0}, \Sigma_i), & (35)\\ \Sigma_i &= \frac{1}{1-\rho^2}\Omega_i, & (36)\\ \Psi_i &= diag(\psi_{i1}^2, \ldots, \psi_{iT}^2), & (37) \end{aligned}$$

where $\Omega_i = \{\omega_{its}\}^{t,s=1,\ldots,T}$ with $\omega_{its} = \sigma_{it}\sigma_{is}\rho^{|t-s|}$, and where σ_{it} and ψ_{it} are defined as in Section 4.1. In this representation ρ represents the inefficiency autocorrelation coefficient.

Again, in order to lower the computational burden, we propose to estimate the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \rho)$ by applying the pairwise estimation strategy to the marginal likelihood function of a dynamic two-periods normal-half normal model. The PDE objective function can be easily obtained by substituting in equation (33) the following

$$\begin{aligned} \Sigma_i &= \begin{pmatrix} \sigma_{i1}^2 & \rho\sigma_{i1}\sigma_{i2} \\ \rho\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 \end{pmatrix}, & (38)\\ \xi_i^2 &= \sigma_{i1}^2 + \sigma_{i2}^2 + \psi_{i1}^2 + \psi_{i2}^2 - 2\rho\sigma_{i1}\sigma_{i2}, & (39)\\ \tilde{\boldsymbol{d}}_i &= \{\left[(\sigma_{i1}^2 + \rho\sigma_{i1}\sigma_{i2}, -\sigma_{i2}^2 - \rho\sigma_{i1}\sigma_{i2})'\xi_i^{-2}\right]\Delta\varepsilon_{it}, & (40)\\ \Xi_i &= \begin{pmatrix} \sigma_{i1}^2 & \rho\sigma_{i1}\sigma_{i2} \\ \rho\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 \end{pmatrix} - \xi_i^{-2}\begin{pmatrix} a & b \\ c & d \end{pmatrix}. & (41) \end{aligned}$$

where $a = \sigma_{i1}^4 - 2\rho\sigma_{i1}^3\sigma_{i2} + \rho^2\sigma_{i1}^2\sigma_{i2}^2$, $b = c = \rho\sigma_{i1}^3\sigma_{i2} - (1+\rho^2)\sigma_{i1}^2\sigma_{i2}^2 + \rho\sigma_{i1}\sigma_{i2}^3$, and $d = \sigma_{i2}^4 - 2\rho\sigma_{i1}\sigma_{i2}^3 + \rho^2\sigma_{i1}^2\sigma_{i2}^2$. As discussed in Section 3.3, in order to apply the JLMS estimator we first need an estimate for the α_i's. We propose to use equation (25) with

$$\hat{\boldsymbol{c}}_i = \hat{\Sigma}_{it}\left(\frac{\Phi_T^*(\mathbf{0},\mathbf{0},\hat{\Sigma}_i)}{\Phi_T(\mathbf{0},\mathbf{0},\hat{\Sigma}_i)}\right), \tag{42}$$

where Φ_T^* is the vector of partial derivatives of Φ_T. Then, inefficiency scores can be easily obtained by adapting equation 13.19 of Dominguez-Molina et al. (2004) as

$$\mathbb{E}(\boldsymbol{u}_i|\hat{\boldsymbol{\varepsilon}}_i) = \hat{\Upsilon}_i + \hat{\Delta}_{*i}\frac{\Phi_T^*(\hat{\Upsilon}_i,\mathbf{0},\hat{\Delta}_{*i})}{\Phi_T(\hat{\Upsilon}_i,\mathbf{0},\hat{\Delta}_{*i})}, \tag{43}$$

where $\hat{\Upsilon}_i = -\hat{\Sigma}_i(\hat{\Sigma}_i + \hat{\Psi}_i)^{-1}\hat{\boldsymbol{\varepsilon}}_i$, $\hat{\Delta}_{*i} = \hat{\Sigma}_i - \hat{\Sigma}_i(\hat{\Sigma}_i + \hat{\Psi}_i)^{-1}\hat{\Sigma}_i$.[20]

5 Monte Carlo evidence

In this Section we study the finite sample properties of the MMSLE and PDE via numerical simulations. We start by comparing the PDE and Greene's MLDVE in the case of a heteroskedastic normal-exponential model. This first set of simulations illustrates the consistency property of the PDE and complements previous simulation studies by Greene (2005a,b), Wang & Ho (2010) and Chen et al. (2014) on the adverse effect of the incidental parameters bias. In a second set of simulations, we study the finite sample properties of our MMSLE together with those of the MMLE proposed by Chen et al. (2014) in a homoskedastic normal-half normal setup.[21] This exercise compares two consistent estimators allowing to make some statement on their finite samples efficiency. Finally, we offer some evidence about the performance of the PDE in a normal-half normal production model where inefficiency is assumed to be heteroskedastic and to follow a first-order autoregressive process.

In all these cases, we investigate the effect of different sample sizes ($n = 100, 250$) and panel lengths ($T = 5, 10$). Simulation results are summarized for each set of simulations by

[20] The computation of the inefficiency scores requires in this case the numerical approximation of a T-dimensional integral for each unit in the panel, but this cumbersome approximations are performed in a second stage and not within the optimization algorithm.

[21] Consistently with the terminology used in this paper, we refer to the estimator proposed by Chen et al. (2014) as MMLE (instead of *Within* MLE).

reporting the average bias and Mean Squared Error (MSE) of the estimates, together with the linear and the Spearman rank correlation coefficients between the (true) simulated inefficiencies and the estimated ones. The inefficiency's bias and MSE are computed for each replication over the $N = n \times T$ observations, and then these quantities are averaged over replications, e.g., MSE($\hat{u}_{it}$)= $R^{-1}\sum_{r=1}^{R}(NT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(\mathbb{E}(u_{it}|\hat{\varepsilon}_{it}) - u_{it}^{0})^2$, where $\mathrm{E}(u_{it}|\hat{\varepsilon}_{it})$ is the JLMS estimate and u_{it}^{0} is the simulated inefficiency.

5.1 PDE vs MLDVE

We consider the following heteroskedastic normal-exponential model

$$
\begin{aligned}
y_{it} &= \alpha_i + \beta_1 x_{1it} + \beta_2 x_{2it} + v_{it} - u_{it}, && (44)\\
v_{it} &\sim \mathcal{N}(0, \psi), && (45)\\
u_{it} &\sim \mathcal{E}\left(\sigma_{it}\right), && (46)\\
\sigma_i &= \exp(\gamma_0 + z_i\gamma_1), \quad i = 1, \ldots, n, \;\; t = 1, \ldots, T, && (47)
\end{aligned}
$$

where the fixed-effect parameters $\alpha_1, ..., \alpha_n$ are drawn from a $\mathcal{U}[0,1]$, $x_{1it} \sim \mathcal{N}(\alpha_i, 1)$, $x_{2it} \sim \mathcal{N}(\alpha_i, 4)$ and $z_i \sim \mathcal{N}(0, 0.25)$. We keep fixed in each experiment the values of the frontier parameters $\beta_1 = 0.3$ and $\beta_2 = 0.7$, $\psi = 0.25$ and $\gamma_1 = 1$, while the value of γ_0 varies across the scenarios in order to obtain different STN ratios ($\bar{\lambda} = 1, 1.5, 2, 4, 6$).[22] For instance, we set $\gamma_0 = -1.5$ and $\gamma_0 = -0.8$ to obtain $\bar{\lambda} \approx 1$ and $\bar{\lambda} \approx 2$, respectively. The number of replications for each experiment is $1,000$.

Figure 1 reports a comparison between PDE and MLDVE in terms of the proportion of non-problematic replications, i.e. replications where $\hat{\psi} > 0.001$. The first thing to highlight is the huge gap between the performance of the two estimators, with the MLDVE unable to provide a non-zero estimate of ψ when $\bar{\lambda} \geq 4$ and $T \leq 10$.

We argue that the downward bias of $\hat{\psi}$ has two drivers. First of all, it is due to the incidental parameters problem since the number of replications with non-zero $\hat{\psi}$ increases with larger T's but not with the cross-sectional dimension. Secondly, given that the number of replications

[22] Given the heteroskedastic specification for the inefficiency, see equation (47), the considered STN ratios are actually defined as averages $\bar{\lambda} = \frac{1}{n\psi}\sum_{i=1}^{n}\sigma_i$.

with non-zero $\hat{\psi}$ remains low for large STN ratios (e.g., $\bar{\lambda} > 4$) even in presence of long panels (e.g., $T = 15$), we argue that this source of bias is related to the intrinsic characteristics of the likelihood function in SF models. As shown by Liseo (1990) for the convolution of normal and half-normal distributions, this issue has a simple justification when all realizations are negative (i.e., $\varepsilon_{it} < 0, \forall i, t$). In this case, the likelihood becomes an increasing function of λ implying that the $\hat{\lambda} = \infty$, or equivalently $\hat{\psi} = 0$.[23] Notice that this problematic behavior is not limited to this extreme case. Indeed, it can be often observed in small cross-sectional samples when $\lambda \geq 8$, while our simulations show that the MLDVE suffers from this issue also for smaller values of $\bar{\lambda}$ (Figure 1).[24] This evidence suggests that the incidental parameters problem amplifies this critical behavior of the likelihood function. We find that both MMSLE and PDE are essentially not affected by the aforementioned issue, presumably because $\Delta\boldsymbol{\varepsilon}_i$ is centered at zero.[25]

Following Chen et al. (2014), we discuss the simulation results by distinguishing problematic and non-problematic replications. Since the inference using MLDVE is markedly problematic for large STN ratios, we limit our analysis to $\bar{\lambda} = 1, 2$. Even so, for some configurations we have been forced to increase the number of replications to 10,000 in order to be able to compare the two estimators.

Tables 1 - 4 show that the frontier parameters (β_1 and β_2) are accurately estimated by both estimators in all scenarios. Consistently with Greene (2005b), we find that the incidental parameters bias does not affect the frontier parameter estimates. Thus, the TFE model behaves as a linear panel data model where the bias only affects the variance parameters. In the estimation of the variances, the PDE performs quite well even in the case of small samples ($n = 100, T = 5$), while the MLDVE estimates are accurate only for the non-problematic samples. In these cases, the most favorable for the MLDVE, the PDE properties appear to be in line with those of the MLDVE. On the other hand, in the problematic scenarios, the MLDVE not only systematically underestimates ψ, but also leads to a relevant bias for both γ_0 and γ_1.[26] The PDE's perfor-

[23]As noted by Azzalini & Capitanio (1999), this intrinsic property of the likelihood function cannot be removed by a reparameterization. It is worth noting that some statistical packages reports non-convergence when ψ approaches zero and, as a consequence, this behaviour is often mistakenly considered as a numerical maximization problem.

[24]Simulations for the MLE in the case of normal-half normal and normal-exponential cross-sectional models are available from the authors upon request.

[25]The same is true for the Chen et al. (2014) estimator where the nuisance parameters are eliminated through the within-group transformation.

[26]Interestingly, the PDE exhibits its better finite sample performances in the problematic cases. This behavior

mances improve when n gets larger. The better performances stem from both a smaller bias and smaller variance of the parameter estimates proving evidence of fixed-T consistency. For example, the MSEs of γ_0 and γ_1 decrease from 0.061 and 0.104 to 0.025 and 0.044, respectively, when n increases from 100 to 250 keeping fixed $\bar{\lambda} = 1$ and $T = 5$. The MLDVE behaves quite well in scenarios where T = 10, however the PDE offers performances that are substantially equivalent, in particular when $\bar{\lambda} = 2$.

As for the estimation of the inefficiencies, we do not observe a substantial difference between the two approaches in the non-problematic samples, while the PDE appears to be slightly superior in the problematic ones. Interestingly, an increase in the length of the panel does not produce significant improvements in the correlation between the (true) simulated inefficiencies and the estimated ones. This evidence suggests that, even in short panels, the relative ranking of the sample units is not affected by the inefficiency bias due to the post-estimation of fixed-effects.

5.2 MMSLE vs MMLE

We consider the homoskedastic normal-half normal model investigated by Chen et al. (2014), that is

$$y_{it} = \alpha_i + \beta x_{it} + v_{it} - u_{it}, \tag{48}$$

$$v_{it} \sim \mathcal{N}(0, \psi), \tag{49}$$

$$u_{it} \sim \mathcal{N}^{+}(0, \sigma) \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, \tag{50}$$

where the fixed-effect parameters $\alpha_1, ..., \alpha_n$ are drawn from a standard Gaussian random variable, and $x_{it} = 0.5\alpha_i + \sqrt{0.5^2} w_{it}$ with $w_{it} \sim \mathcal{N}(0, 1)$. For each experiment, we use the same α_i and x_{it} in all replications, thus only u_{it} and v_{it} are redrawn in each replication. We set $\beta = 1$ and consider two different STN ratios ($\lambda = 1, 2$) fixing the variance of the compounded error to unity $\omega_\varepsilon^2 = \left(\frac{\pi-2}{\pi}\right)\sigma^2 + \psi^2 = 1$. This setup implies $\sigma = \psi = 0.85643$ when $\lambda = 1$, and $\sigma = 1.27684$, $\psi = .63842$ in the other case. The analysis is based on $R = 250$ replications for

may be due to the fact that the distinction of problematic and non problematic replications is driven by the performance of the MLDVE in estimating ψ and this is likely to create some advantages for the MLDVE in the non-problematic replications.

each experiment.[27]

The aim of this exercise is to compare in a homoskedastic set up the MMSLE with the estimator recently proposed by Chen et al. (2014). Since Chen et al. (2014) show that their MMLE outperforms the MLDVE, we do not include the latter in the comparison.

Table 5 summarizes the simulation results using the same structure adopted before. Consistently with the evidence reported in Chen et al. (2014) and with the behaviour of the PDE, both MMSLE and MMLE do not show any problem in the estimation of ψ. The main message is that both estimators exhibit consistency with fixed T, showing very similar finite sample properties. Only when $n = 100$ and $T = 5$ the MMSLE seems to be slightly better than the MMLE in estimating σ, but this difference vanishes when the sample size grows. Despite our expectation about the MMLE being the most efficient, it is interesting to note that this is not the case since, at least for the considered Monte Carlo design, the loss of efficiency due to the simulated likelihood approach has approximately the same magnitude of the one resulting from the numerical approximation of T-dimensional normal integrals required by the MMLE.

These results can be taken as evidence that the MMSLE is a viable alternative to the MMLE in a homoskedastic normal-half normal setting. Given the unavailability of a closed form expression for the marginal likelihood function when $T > 2$, to the best of our knowledge the MMSLE remains the most efficient estimator for the homoskedastic normal-exponential case.

5.3 Dynamic PDE

In this last simulation exercise, we illustrate the inferential performance of the PDE in a dynamic setup. In particular, we specify the following heteroskedastic normal-half normal model with $AR(1)$ inefficiencies

$$\boldsymbol{y}_i = \alpha_i \iota_T + \beta \boldsymbol{x}_i + \boldsymbol{v}_i - \boldsymbol{u}_i, \tag{51}$$

$$\boldsymbol{v}_i \sim \mathcal{N}_T(0, \psi^2 I_t), \tag{52}$$

$$\boldsymbol{u}_i \sim \mathcal{N}_T^+\left(\boldsymbol{0}, \frac{1}{1-\rho^2}\Omega_i\right), \quad i = 1, \ldots, n, \tag{53}$$

[27] For the MMSLE we use 30 Halton sequences for observation, while for the MMLE we exploit the equicorrelated structure of the covariance matrix which greatly reduces the computational burden (Section 4.1).

where $\Omega_i = \{\omega_{its}\}^{t,s=1,...,T}$ with $\omega_{its} = \sigma_{it}\sigma_{is}\rho^{|t-s|}$ and $\sigma_{it} = \exp(\gamma_0 + z_{it}\gamma_1)$. The simulation of inefficiency vector $\boldsymbol{u}_i$ is performed using the MCMC approach outlined in Geweke (1991) or in Robert (1995), which uses a Gibbs algorithm for sampling from an arbitrary multivariate truncated normal distribution. The fixed-effect parameters $\alpha_1, ..., \alpha_n$ and z_{it} are drawn from a standard Gaussian random variable while $x_{it} = 0.5\alpha_i + \sqrt{0.5^2}w_{it}$ with $w_{it} \sim \mathcal{N}(0,1)$. We set $\beta = 0.5$, $\psi = 0.5$, $\gamma_0 = -0.5$ and $\gamma_1 = 1$ (this implies $\bar{\lambda} = \frac{1}{nT\psi}\sum_{i=1}^{n}\sum_{t=1}^{T}\sigma_{it} \approx 2$), and consider two different values for the ρ parameter ($\rho = 0.3, 0.7$). The analysis is based on 250 replications for each experiment.

Table 6 clearly shows the consistency property of the PDE. In the "low" autocorrelation case ($\rho = 0.3$), all parameters are accurately estimated in almost all the scenarios. Only when $n = 100$ and $T = 5$, we find a relevant MSE for γ_0 and ρ. An increase in the length of the panel produces significant reductions in both the bias and the MSE. For example, the MSEs of γ_0 and ρ decrease from 0.051 and 0.039 to 0.019 and 0.024 when T increases from 5 to 10. Analogously, a larger cross-sectional dimension yields similar improvements.

In the "high" autocorrelation case ($\rho = 0.7$), we observe less accurate estimates only for γ_0, especially when $n = 100$ and $T = 5$. Differently from the "low" autocorrelation case, we find that the finite sample performances of the estimator are much more affected by an increase in the length of the panel than by an equivalent increase in the cross-sectional dimension.

The inefficiencies are accurately estimated in all the scenarios. We do not find significant improvements when the cross-sectional dimension increases, while the availability of a longer panel provides slightly better results. Similarly, the finite sample properties remain substantially unaffected by changes in ρ. It is worth mentioning the case of $\rho = 0.3$ for which we observe a small increase in the bias of $\mathbb{E}(u|\hat{\varepsilon})$ moving from $T = 5$ to $T = 10$. This seemingly counterintuitive result is linked to the computation of $\hat{\boldsymbol{c}}_i$, which becomes slightly less accurate when T grows.

6 Empirical application

In this Section, we apply the PDE estimator in an empirical study of Italian hospitals activity. Recently, Daidone & D'Amico (2009) found that public and private not-for-profit hospitals are significantly more efficient than private for-profit structures. A previous study of Barbetta et al.

(2007), whose analysis only covers the second part of the nineties, found that not-for-profit hospitals are more efficient than their public counterparts. Our analysis integrates the work of Daidone & D'Amico (2009), to which we refer the reader for further details on data sources and variable definitions. In particular, we investigate how labor structure, ownership and level of specialization affect hospital's technical inefficiency. In the remainder of this Section, we provide a brief discussion of the data, outline the model and, finally, discuss the results.

6.1 Data

The data set consists of a yearly unbalanced panel of Italian hospitals located in the Lazio's region from 2000 to 2005. The panel contains 109 hospitals observed over a 6 years period, for a total of 619 observations.

The output variable is the number of acute patients discharges adjusted for its case-mix complexity through the Diagnosis Related Groups (DRG) weights. In order to represent the multi-output nature of the hospitals industry, we classify adjusted discharges using the information contained in the DRG categories. The classification system of the hospital activity in the study period is made up of 492 DRGs. In order to keep the estimation feasible, discharges have been aggregated into the following five output variables: Complex Surgery (Y_1), Emergency Room Treatments (Y_2), Cancers and HIV (Y_3), General Surgery (Y_4) and General Medicine (Y_5). We consider as inputs the number of beds (X_4), the number of physicians (X_1), nurses (X_2) and other personnel (X_3, comprising teaching and ancillary staff).

One of the policy questions of this empirical application is to assess the role of specialization as a determinant of technical inefficiency. We consider the Gini ratio, which is equal to zero in the case of generalist hospitals with perfect equidistribution of health care services (e.g., polispecialistic medical center) and equal to one in the case of hospitals characterized by a single specialty, as an indicator of the level of specialization. The second research question is about the link between ownership structure and inefficiency. We classify the hospitals in public, private but not-for-profit and private for profit.[28] Finally, in order to investigate the role of the labor force structure, both the nurses/beds and the physicians/beds ratios are included in

[28]Notice that this classification reflects not only the ownership status, but also the founding sources. Public hospitals include hospitals directly managed by Local Health Authorities, independent public hospitals (D.L. 502/92) and assimilated to public structures (L. 833/78).

the inefficiency equation. Descriptive statistics and a brief summary of variable definitions are reported in Table 7.

6.2 Model and estimation

A Stochastic Distance Function is the easiest solution to represent the multi-output production technology in a single SF equation.[29] Following Kumbhakar & Lovell (2000), a stochastic multi-output distance function model for panel data can be defined as

$$(Y_{5it})^{-1} = D\left(\boldsymbol{X}_{it}, \frac{Y^*_{mit}}{Y_{5it}}; \beta\right) \exp\left(u_{it} - v_{it}\right). \tag{54}$$

The dependent variable is the reciprocal of the chosen normalizing output Y_{5it} (General Medicine), while the covariates are the inputs and the remaining normalized outputs ($m = 1, \ldots, 4$). By specifying $D(.)$ as a translog function and adapting the model in order to allow a production frontier interpretation, equation (54) can be written as[30]

$$\begin{aligned} y_{5it} &= \alpha_i + \sum_{m=1}^{4} \delta_m y^*_{mit} + \sum_{k=1}^{4} \beta_k x_{kit} + \frac{1}{2} \sum_{m=1}^{4} \sum_{p=1}^{4} \delta_{mp} y^*_{mit} y^*_{pit} + \frac{1}{2} \sum_{k=1}^{4} \sum_{j=1}^{4} \beta_{kj} x_{kit} x_{jit} + \\ &\quad + \sum_{k=1}^{4} \sum_{m=1}^{4} \beta_k \delta_m x_{kit} y^*_{mit} + \sum_{t=2001}^{2005} d_t + v_{it} - u_{it}, \end{aligned} \tag{55}$$

where y^*_{mit} are the normalized outputs, d_t are year dummies, δ_{mp} and β_{kj} are technological parameters with the index i denoting hospital, t representing time, k and j labeling the input variables and m and p indicating outputs.[31]

The inefficiency u_{it} is assumed to be heteroskedastic with scale parameter $\sigma_{it} = \exp(\boldsymbol{z}_{it}\boldsymbol{\gamma})$, where $\boldsymbol{z}_{it}$ contains both time-varying and time-invariant covariates: the Gini ratio (*Gini*), the nurses and the physicians per bed ratios (*Nurses/Bed* and *Physicians/Bed*), dummies for the quartiles of the number of beds (with the 1^{st} quartile as the base category), ownership status dummies (public hospitals represent the base category), year dummies (with 2000 as the base

[29] For a detailed discussion of distance functions and their properties see Fare et al. (2008).

[30] Equation (55) has been adapted for estimation purposes by transforming the left-hand side of the equation to be y_{5it} rather than $-y_{5it}$. This allows to interpret the estimates as in a standard production frontier model (see, among the others, Morrison Paul et al., 2000).

[31] All output and input variables in equation (55) are expressed in logarithms. Symmetry constraints have been imposed on the interaction terms, i.e $\delta_{mp} = \delta_{pm}$ and $\beta_{kj} = \beta_{jk}$.

year) and geographical dummies for each Local Health Authority in the region.

It is worth to emphasize that the data used in this application mirrors the sample size of one of the Monte Carlo experiments analyzed in Section 5.1 ($n = 100$ and $T = 5$). As reported in figure 1, the MLDVE is frequently affected by the $\hat{\psi} \approx 0$ issue in this case. This issue also arose in estimating model (55). Since also the standard MLE applied on pooled data shows this problematic behavior, for comparison purposes we estimate model (55) using a "pooled" version of the PDE characterized by the following objective function

$$U_n\left(\boldsymbol{\theta}\right) = \binom{N}{2}^{-1} \sum_{i=2}^{N} \sum_{j<i} \log f(\Delta_i^j y|\ \boldsymbol{\theta}, \Delta_i^j \boldsymbol{x}), \tag{56}$$

where $\Delta_i^j y = y_i - y_j$, $\Delta_i^j \boldsymbol{x} = \boldsymbol{x}_i - \boldsymbol{x}_j$ and N is the total number of observations.[32] Moreover, in order to investigate whether distributional assumptions matter, we estimate the model assuming both a normal-exponential (N-E) and a normal-half normal (N-HN) distribution for the composed error.

6.3 Results

Table 8 summarizes technological parameter estimates using elasticities (evaluated at covariates means) of the distance function with respect to outputs and inputs. For the pooled case, the results significantly point to increasing returns to scale. Nurses and beds elasticities are positive and statistically significant, while the one for the ancillary staff is not statistically different from zero. As expected, General Medicine has the highest elasticity value among all the outputs confirming the importance of this department in terms of hospital activity. This evidence, both in terms of elasticities and return to scale, is in line with the estimates reported in Daidone & D'Amico (2009). Also the elasticities from PDE estimates suggest a similar "technological" picture: return to scale are increasing and the number of beds still remains the most important input in the production process. However, the elasticity of the ancillary staff becomes negative and statistically significant when the unobserved fixed-effects are included in the model. Since this auxiliary input is less involved in the production process, this result is not surprising and

[32]Following Honoré & Powell (1994, pp.248-255), we also correspondingly adapt the asymptotic variance-covariance matrix estimator.

suggests that the oversizing of ancillary staff may be considered as a ballast for the production process.

Table 9 reports the inefficiency effects estimates. As a first result, we observe that controlling for time-invariant unobserved heterogeneity greatly affects the magnitude and the significance of the estimates. For instance, looking at the effect of ownership structure, the pooled estimates suggest that private not-for-profit hospitals are less efficient than public ones in the normal-exponential model, while this effect vanishes in the "true" fixed-effects specifications. This evidence may be justified by the fact that private hospitals are subject to caps on production for cost-containment reasons, leading some of them to a suboptimal level of production and, as a consequence, to a higher level of inefficiency. This regulatory constraint has been constant over the study period and, therefore, this feature is likely to be captured by the fixed-effects. Further, the labor force structure have a significant role in explaining inefficiency variability once unobserved heterogeneity is controlled for. In the fixed-effects specifications, these results are robust to the distributions assumed for the composed error.

As expected, the estimated inefficiency scores are on average much larger for the pooled models. This result occurs because the estimate of $\bar{\sigma}$ is much lower in the TFE specifications, where a portion of the variability is captured by the fixed-effects. The last panel of Table 9 reports the Spearman rank correlation coefficients between the estimated inefficiencies. As expected, the pooled model provides a very different inefficiency ranking than the TFE one (the Spearman coefficient is between 0.19 and 0.28), whereas the ranking from the TFE model is basically the same regardless of the distribution assumed for inefficiency (the Spearman coefficient between the normal-exponential and the normal-half normal models is about 0.98).

7 Concluding remarks

This paper reconsiders the estimation of the "true" fixed-effects (TFE) stochastic frontier panel data model of Greene (2005a) aiming to solve the incidental parameters problem affecting his maximum likelihood dummy variables estimator (MLDVE). We propose two alternatives that, by relying on a first-difference data transformation, avoid the incidental parameters problem and achieve consistency under both fixed-n and fixed-T asymptotics. The first is a marginal maximum simulated likelihood estimator (MMSLE) that can be used to estimate homoskedastic normal-half normal and normal-exponential models. In the spirit of Honoré & Powell (1994), the second is a pairwise difference estimator (PDE) defined as the minimizer of U-processes that can be used to estimate heteroskedastic normal-exponential specifications. Furthermore, by exploiting our pairwise differencing strategy, we extend the results of Chen et al. (2014) providing a computationally feasible approach to estimate normal-truncated normal TFE models in which the inefficiency can be heterogenous, heteroskedastic and can follow an AR(1) process.

The finite sample properties of the proposed estimators are investigated through a set of Monte Carlo experiments. Our results suggest that both estimation procedures generally perform well also when both n and T are small, as is commonly the case in economic applications. On the other hand, the Greene (2005a)'s estimator provides unsatisfactory results in many of the considered Monte Carlo scenarios, especially when T is small. Furthermore, we find that our MMSLE is a viable alternative to the Chen et al. (2014) marginal maximum likelihood estimator (MMLE) in a homoskedastic normal-half normal setting. What emerges is that the efficiency loss due to the simulated likelihood approach is in the same order of magnitude of the one resulting from the numerical approximation of T-dimensional normal integrals, as required by the MMLE. Of special note is the good performance of the PDE applied to a heteroskedastic normal-half normal model with AR(1) inefficiencies.

Finally, we apply the PDE to estimate a multi-output stochastic distance function on a panel of italian hospitals. This empirical illustration provides evidence of the PDE usefulness in a setting where standard likelihood-based estimators fail to provide reliable inference. We find that controlling for time-invariant unobserved heterogeneity in the frontier function greatly affects the magnitude and the statistical significance of the so-called inefficiency effects and, as expected, that inefficiency estimates from a pooled model are much higher than those obtained from the TFE one.

References

Abrevaya, J. (1999). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics*, *93*, 203–228.

Ahn, S. C., & Sickles, R. C. (2000). Estimation of long-run inefficiency levels: a dynamic frontier approach. *Econometric Reviews*, *19*, 461–492.

Aigner, D., Lovell, C., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*, 21–37.

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *61*, 579–602.

Barbetta, G. P., Turati, G., & Zago, A. M. (2007). Behavioral differences between public and private not-for-profit hospitals in the italian national health service. *Health Economics*, *16*, 75–96.

Battese, G., & Coelli, T. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, *38*, 387–399.

Battese, G., & Coelli, T. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in india. *Journal of Productivity Analysis*, *3*, 153–169.

Battese, G., & Coelli, T. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. *Empirical Economics*, *20*, 325–332.

Bhat, C. R. (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B: Methodological*, *35*, 677 – 693.

Chen, Y., Wang, H., & Schmidt, P. (2014). Consistent estimation of the fixed effects stochastic frontier model. *Journal of Econometrics*, *181*, 65–76.

Cornwell, C., Schmidt, P., & Sickles, R. (1990). Production frontiers with cross-sectional and time series variation in efficiency levels. *Journal of Econometrics*, *46*, 185–200.

Daidone, S., & D'Amico, F. (2009). Technical efficiency, specialization and ownership form: Evidences from a pooling of italian hospitals. *Journal of Productivity Analysis*, *32*, 203–216.

Dominguez-Molina, J., Gonzalez-Farias, G., & Ramos-Quiroga, R. (2004). Skew-normality in stochastic frontier analysis. In M. Genton (Ed.), *Skew Elliptical Distributions and their Applications: A Journey beyond Normality* chapter 13. (pp. 235–253). Boca Raton, Florida. Chapman and Hall / CRC.

Emvalomatis, G. (2012). Adjustment and unobserved heterogeneity in dynamic stochastic frontier models. *Journal of Productivity Analysis*, *37*, 7–16.

Fare, R., Grosskopf, S., & Lovell, C. A. K. (2008). *Production Frontiers*. Cambridge University Press.

Genz, A. (2004). Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing*, *14*, 251–260.

Geweke, J. (1991). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In E. M. Keramidas (Ed.), *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (pp. 571–578). Interface Foundation of North America, Inc.

Gonzalez-Farias, G., Dominguez-Molina, J., & Gupta, A. (2004a). Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference*, *126*, 521–534.

Gonzalez-Farias, G., Dominguez-Molina, J., & Gupta, A. (2004b). The closed skew normal distribution. In M. Genton (Ed.), *Skew Elliptical Distributions and their Applications: A Journey beyond Normality* chapter 2. (pp. 40–57). Boca Raton, Florida. Chapman and Hall / CRC.

Greene, W. (2003). Simulated likelihood estimation of the normal-gamma stochastic frontier function. *Journal of Productivity Analysis*, *19*, 179–190.

Greene, W. (2005a). reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics*, *126*, 269–303.

Greene, W. (2005b). fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, *23*, 7–32.

Greene, W. (2008). The econometric approach to efficiency analysis. In H. O. Fried, C. A. K. Lovell, & S. S. Schmidt (Eds.), *The Measurement of Productive Efficiency and Productivity Change* chapter 2. (pp. 92–250). Oxford University Press. New York.

Halton, J. (1960). On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, *2*, 84–90.

Honoré, B. E., & Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics*, *64*, 241–278.

Jondrow, J., Lovell, C., Materov, I., & Schmidt, P. (1982). On the estimation of technical efficiency in the stochastic production function model. *Journal of Econometrics*, *19*, 233–238.

Kotz, S., Balakrishnan, N., & Johnson, N. L. (2000). *Continuous Multivariate Distributions, Volume 1, Models and Applications,* 2^{nd} *Edition*. John Wiley & Sons.

Kumbhakar, S. (1990). Production frontiers, panel data and time-varying technical inefficiency. *Journal of Econometrics*, *46*, 201–212.

Kumbhakar, S., & Lovell, C. (2000). *Stochastic frontier analysis*. Cambridge University Press.

Kumbhakar, S. C., & Hjalmarsson, L. (1995). Labour-use efficiency in swedish social insurance offices. *Journal of Applied Econometrics*, *10*, 33–47.

Kumbhakar, S. C., & Tsionas, E. G. (2011). Some recent developments in efficiency measurement in stochastic frontier models. *Journal of Probability and Statistics*, *2011*.

Lee, Y., & Schmidt, P. (1993). A production frontier model with flexible temporal variation in technical inefficiency. In H. Fried, C. Lovell, & S. Schmidt (Eds.), *The measurement of productive efficiency: techniques and applications*. Oxford University Press.

Liseo, B. (1990). The skew-normal class of densities: inferential aspects from a bayesian viewpoint (in italian). *Statistica*, *50*, 59–70.

Martins-Filho, C., & Yao, F. (2010). *A note on some properties of a skew-normal density*. Working Papers 10-10 Department of Economics, West Virginia University.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production function with composed errors. *International Economic Review*, *18*, 435–444.

Morokoff, W. J., & Caflisch, R. E. (1995). Quasi-monte carlo integration. *Journal of Computational Physics*, *122*, 218 – 230.

Morrison Paul, C. J., Johnston, W. E., & Frengley, G. A. G. (2000). Efficiency in new zealand sheep and beef farming: The impacts of regulatory reform. *The Review of Economics and Statistics*, *82*, pp. 325–337.

Pitt, M., & Lee, L. (1981). The measurement and sources of technical inefficiency in the indonesian weaving industry. *Journal of Development Economics*, *9*, 43–64.

Poirier, D. J., & Ruud, P. A. (1988). Probit with dependent observations. *Review of Economic Studies*, *55*, 593–614.

Ritter, C., & Simar, L. (1997). Pitfalls of normal-gamma stochastic frontier models. *Journal of Productivity Analysis*, *8*, 167–182.

Robert, C. (1995). Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. *Statistics and computing*, *5*, 121–125.

Schmidt, P., & Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics*, *2*, 367–74.

Sloan, I. H., & Woźniakowski, H. (1998). When are quasi-monte carlo algorithms efficient for high dimensional integrals? *Journal of Complexity*, *14*, 1 – 33.

Train, K. (2000). *Halton Sequences for Mixed Logit*. Economics Working Papers E00-278 University of California at Berkeley.

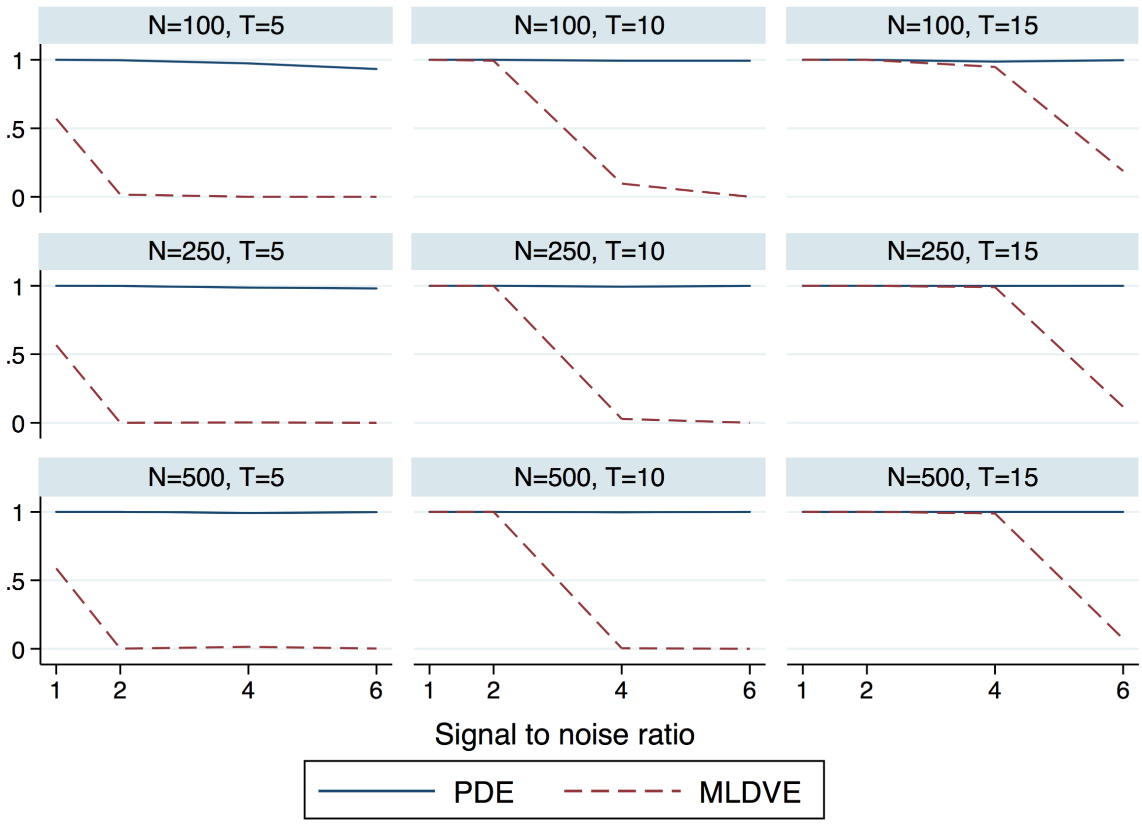
Tsionas, E. G. (2006). Inference in dynamic stochastic frontier models. *Journal of Applied Econometrics*, *21*, 669–676.

Wang, H., & Ho, C. (2010). Estimating fixed-effect panel stochastic frontier models by model transformation. *Journal of Econometrics*, *157*, 286–296.

Wang, H., Iglesias, E. M., & Wooldridge, J. M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics*, *172*, 77 – 89.

Wang, H., & Schmidt, P. (2002). One-step and two-step estimation of the effects of exogenous variables on technical efficiency levels. *Journal of Productivity Analysis*, *18*, 129–144.

Figure 1: PDE *vs* MLDVE: proportion* of replications with non-zero $\hat{\psi}$.



* Computed using the first 1,000 simulated samples.

Table 1: Simulation results for PDE and MLDVE with $n = 100$, $T = 5$.

(a) $\lambda = 1$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	7.9e-04	2.8e-04	0.002	3.7e-04
β_2	1.9e-05	7.0e-05	-1.5e-05	8.8e-05
γ_0	-0.079	0.061	0.217	0.139
γ_1	0.082	0.104	-0.322	0.230
ψ	0.003	7.4e-04	-0.134	0.028
E(u|ε)	-0.015	0.036	0.054	0.060
$\mathrm{r}_{u,\hat{u}}$	0.804		0.764	
	(0.575)		(0.536)	

Non problematic (570)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	3.4e-04	2.6e-04	4.4e-04	2.5e-04
β_2	2.7e-04	7.0e-05	2.8e-04	6.6e-05
γ_0	-0.120	0.067	-0.007	0.042
γ_1	0.115	0.114	-0.081	0.085
ψ	0.007	6.5e-04	-0.047	0.003
E(u|ε)	-0.021	0.036	8.5e-05	0.034
$\mathrm{r}_{u,\hat{u}}$	0.801		0.810	
	(0.573)		(0.574)	

Problematic (430)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	0.001	3.0e-04	0.004	5.3e-04
β_2	-3.2e-04	6.9e-05	-4.0e-04	1.2e-04
γ_0	-0.024	0.054	0.515	0.267
γ_1	0.038	0.091	-0.643	0.422
ψ	-0.001	8.7e-04	-0.250	0.062
E(u|ε)	-0.006	0.036	0.126	0.094
$\mathrm{r}_{u,\hat{u}}$	0.808		0.704	
	(0.578)		(0.485)	

(b) $\lambda = 2$

All replications (10000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	4.2e-05	5.2e-04	0.008	8.8e-04
β_2	-9.0e-05	1.3e-04	6.9e-04	1.8e-04
γ_0	-0.034	0.020	0.140	0.023
γ_1	0.045	0.048	-0.415	0.184
ψ	0.003	0.002	-0.247	0.062
E(u|ε)	-0.013	0.088	0.034	0.100
$\mathrm{r}_{u,\hat{u}}$	0.886		0.876	
	(0.698)		(0.673)	

Non problematic (137)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	3.9e-04	4.7e-04	9.1e-04	4.0e-04
β_2	-8.3e-05	1.2e-04	4.4e-05	1.1e-04
γ_0	-0.193	0.057	-0.133	0.028
γ_1	0.223	0.104	0.041	0.024
ψ	0.035	0.002	-0.035	0.002
E(u|ε)	-0.056	0.091	-0.045	0.064
$\mathrm{r}_{u,\hat{u}}$	0.882		0.914	
	(0.689)		(0.730)	

Problematic (9863)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	3.7e-05	5.2e-04	0.008	8.8e-04
β_2	-9.0e-05	1.3e-04	7.0e-04	1.8e-04
γ_0	-0.032	0.020	0.143	0.023
γ_1	0.042	0.047	-0.421	0.186
ψ	0.002	0.002	-0.250	0.062
E(u|ε)	-0.012	0.088	0.035	0.101
$\mathrm{r}_{u,\hat{u}}$	0.886		0.875	
	(0.698)		(0.672)	

Table 2: Simulation results for PDE and MLDVE with $n = 100$, $T = 10$.

(a) $\lambda = 1$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-5.1e-05	1.2e-04	4.8e-05	1.2e-04
β_2	1.1e-04	2.8e-05	2.0e-04	2.7e-05
γ_0	-0.039	0.029	0.008	0.012
γ_1	0.044	0.054	-0.050	0.031
ψ	6.6e-04	3.6e-04	-0.023	7.4e-04
E(u|ε)	-0.008	0.030	4.9e-04	0.028
$r_{u,\hat{u}}$	0.840		0.848	
	(0.609)		(0.615)	

(b) $\lambda = 2$

All replications (10000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	1.1e-04	2.2e-04	2.3e-04	1.9e-04
β_2	1.0e-04	5.5e-05	1.3e-04	4.6e-05
γ_0	-0.017	0.007	0.026	0.006
γ_1	0.023	0.019	-0.073	0.017
ψ	0.002	8.2e-04	-0.044	0.003
E(u|ε)	-0.007	0.062	0.007	0.050
$r_{u,\hat{u}}$	0.921		0.936	
	(0.748)		(0.768)	

Non problematic (9924)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	1.1e-04	2.2e-04	2.3e-04	1.9e-04
β_2	1.0e-04	5.5e-05	1.1e-04	4.6e-05
γ_0	-0.018	0.007	0.024	0.005
γ_1	0.023	0.019	-0.070	0.016
ψ	0.002	8.1e-04	-0.043	0.002
E(u|ε)	-0.007	0.062	0.006	0.050
$r_{u,\hat{u}}$	0.921		0.936	
	(0.748)		(0.769)	

Problematic (76)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-7.7e-04	2.0e-04	2.7e-04	3.3e-04
β_2	4.3e-04	5.0e-05	0.002	7.7e-05
γ_0	0.026	0.010	0.328	0.109
γ_1	-0.036	0.024	-0.449	0.206
ψ	-0.010	0.001	-0.250	0.062
E(u|ε)	0.004	0.063	0.135	0.105
$r_{u,\hat{u}}$	0.923		0.900	
	(0.751)		(0.721)	

Table 3: Simulation results for PDE and MLDVE with $n = 250$, $T = 5$.

(a) $\lambda = 1$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-5.7e-04	1.1e-04	3.7e-04	1.5e-04
β_2	1.8e-05	2.9e-05	1.7e-04	3.8e-05
γ_0	-0.033	0.025	0.276	0.130
γ_1	0.038	0.044	-0.383	0.217
ψ	0.002	3.2e-04	-0.140	0.029
E(u|ε)	-0.010	0.034	0.064	0.060
$r_{u,\hat{u}}$	0.815		0.769	
	(0.584)		(0.537)	

Non problematic (567)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-6.7e-04	1.1e-04	-7.6e-04	1.1e-04
β_2	-2.6e-04	3.0e-05	-2.6e-04	3.0e-05
γ_0	-0.066	0.026	0.088	0.021
γ_1	0.074	0.047	-0.177	0.055
ψ	0.006	3.0e-04	-0.056	0.003
E(u|ε)	-0.016	0.034	0.015	0.034
$r_{u,\hat{u}}$	0.814		0.818	
	(0.584)		(0.577)	

Problematic (433)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-4.4e-04	1.1e-04	0.002	2.0e-04
β_2	3.9e-04	2.7e-05	7.4e-04	4.9e-05
γ_0	0.010	0.022	0.523	0.274
γ_1	-0.010	0.041	-0.654	0.431
ψ	-0.003	3.4e-04	-0.250	0.062
E(u|ε)	-0.003	0.034	0.129	0.095
$r_{u,\hat{u}}$	0.816		0.704	
	(0.585)		(0.484)	

(b) $\lambda = 2$

All replications (10000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	2.2e-04	2.1e-04	0.006	3.9e-04
β_2	-1.1e-04	5.0e-05	5.4e-04	7.4e-05
γ_0	-0.014	0.006	0.147	0.022
γ_1	0.018	0.016	-0.425	0.185
ψ	0.003	7.6e-04	-0.250	0.062
E(u|ε)	-0.010	0.085	0.036	0.100
$r_{u,\hat{u}}$	0.890		0.878	
	(0.702)		(0.673)	

Non problematic (3)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-0.006	4.3e-04	0.006	4.4e-04
β_2	-0.004	1.1e-04	1.0e-04	4.0e-05
γ_0	-0.104	0.021	-0.014	0.021
γ_1	0.151	0.044	-0.098	0.087
ψ	0.037	0.002	-0.101	0.021
E(u|ε)	-0.046	0.086	-0.021	0.080
$r_{u,\hat{u}}$	0.899		0.908	
	(0.708)		(0.714)	

Problematic (9997)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	2.2e-04	2.1e-04	0.006	3.9e-04
β_2	-1.1e-04	5.0e-05	5.4e-04	7.4e-05
γ_0	-0.014	0.006	0.147	0.022
γ_1	0.018	0.016	-0.425	0.185
ψ	0.002	7.6e-04	-0.250	0.062
E(u|ε)	-0.010	0.085	0.036	0.100
$r_{u,\hat{u}}$	0.890		0.878	
	(0.702)		(0.673)	

Table 4: Simulation results for PDE and MLDVE with $n = 250$, $T = 10$.

(a) $\lambda = 1$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	8.8e-05	4.5e-05	3.4e-05	4.1e-05
β_2	-9.8e-06	1.2e-05	-2.8e-05	1.1e-05
γ_0	-0.011	0.008	0.019	0.005
γ_1	0.015	0.017	-0.057	0.013
ψ	-0.001	1.4e-04	-0.024	6.4e-04
E(u|ε)	-0.004	0.029	0.002	0.027
$\mathrm{r}_{u,\hat{u}}$	0.846		0.853	
	(0.615)		(0.620)	

(b) $\lambda = 2$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-5.8e-05	8.3e-05	-6.7e-05	6.7e-05
β_2	-2.9e-05	2.2e-05	-8.6e-05	1.8e-05
γ_0	-0.007	0.002	0.026	0.002
γ_1	0.011	0.006	-0.069	0.009
ψ	7.2e-06	3.1e-04	-0.043	0.002
E(u|ε)	-0.004	0.060	0.006	0.049
$\mathrm{r}_{u,\hat{u}}$	0.923		0.938	
	(0.751)		(0.771)	

Table 5: Simulation results for MMSLE and MMLE.

(a) $\lambda = 1$

$n = 100$ $T = 5$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	-0.002	0.004	-0.002	0.004
σ	-0.041	0.140	-0.114	0.197
ψ	-0.025	0.011	-0.013	0.014
E(u|ε)	-0.043	0.315	-0.100	0.355
$r_{u,\hat{u}}$	0.472		0.470	
	(0.425)		(0.425)	

$n = 100$ $T = 10$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	-0.002	0.001	-0.002	0.001
σ	-0.054	0.058	-0.048	0.055
ψ	0.008	0.004	0.007	0.004
E(u|ε)	-0.046	0.246	-0.042	0.244
$r_{u,\hat{u}}$	0.506		0.506	
	(0.457)		(0.457)	

$n = 250$ $T = 5$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	0.002	0.001	0.001	0.001
σ	-0.006	0.056	-0.019	0.074
ψ	-0.015	0.006	-0.015	0.006
E(u|ε)	-0.015	0.252	-0.025	0.264
$r_{u,\hat{u}}$	0.475		0.475	
	(0.430)		(0.430)	

$n = 250$ $T = 10$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	7.2e-04	4.4e-04	7.1e-04	4.4e-04
σ	-0.042	0.033	-0.041	0.033
ψ	0.007	0.003	0.006	0.003
E(u|ε)	-0.036	0.225	-0.035	0.225
$r_{u,\hat{u}}$	0.506		0.506	
	(0.457)		(0.457)	

(b) $\lambda = 2$

$n = 100$ $T = 5$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	-0.002	0.004	-0.002	0.004
σ	-0.025	0.028	-0.050	0.061
ψ	-0.006	0.010	4.5e-04	0.012
E(u|ε)	-0.036	0.324	-0.055	0.348
$r_{u,\hat{u}}$	0.707		0.707	
	(0.644)		(0.644)	

$n = 100$ $T = 10$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	-0.001	0.001	-0.001	0.001
σ	-0.051	0.010	-0.007	0.009
ψ	0.036	0.003	0.004	0.003
E(u|ε)	-0.047	0.268	-0.012	0.266
$r_{u,\hat{u}}$	0.752		0.752	
	(0.692)		(0.692)	

$n = 250$ $T = 5$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	0.001	0.001	0.001	0.001
σ	0.002	0.011	0.001	0.012
ψ	-0.011	0.004	-0.011	0.005
E(u|ε)	-0.016	0.304	-0.017	0.305
$r_{u,\hat{u}}$	0.711		0.711	
	(0.651)		(0.651)	

$n = 250$ $T = 10$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	0.001	3.7e-04	8.8e-04	3.6e-04
σ	-0.025	0.004	-0.004	0.004
ψ	0.016	0.001	-2.9e-04	0.001
E(u|ε)	-0.026	0.261	-0.009	0.261
$r_{u,\hat{u}}$	0.752		0.752	
	(0.691)		(0.691)	

Table 6: Simulation results for the dynamic PDE.

(a) $\rho = 0.3$

$n = 100$ $T = 5$

	Bias	MSE
β	-0.002	0.001
γ_0	-0.061	0.051
γ_1	-0.013	0.011
ψ	-0.002	0.001
ρ	0.071	0.039
E(u|ε)	0.028	0.249
$r_{u,\hat{u}}$	0.952	
	(0.781)	

$n = 100$ $T = 10$

	Bias	MSE
β	-7.8e-04	5.5e-04
γ_0	-0.009	0.019
γ_1	-0.004	0.005
ψ	5.4e-04	6.0e-04
ρ	0.034	0.024
E(u|ε)	0.032	0.173
$r_{u,\hat{u}}$	0.970	
	(0.806)	

$n = 250$ $T = 5$

	Bias	MSE
β	-0.001	4.3e-04
γ_0	-0.013	0.016
γ_1	-0.009	0.004
ψ	-0.001	6.2e-04
ρ	0.018	0.023
E(u|ε)	0.018	0.233
$r_{u,\hat{u}}$	0.957	
	(0.784)	

$n = 250$ $T = 10$

	Bias	MSE
β	-6.9e-04	2.1e-04
γ_0	0.009	0.007
γ_1	-0.010	0.002
ψ	0.002	2.5e-04
ρ	0.036	0.011
E(u|ε)	0.037	0.170
$r_{u,\hat{u}}$	0.971	
	(0.810)	

(b) $\rho = 0.7$

$n = 100$ $T = 5$

	Bias	MSE
β	-0.003	0.001
γ_0	-0.220	0.104
γ_1	-0.012	0.007
ψ	-6.3e-04	0.002
ρ	-0.008	0.010
E(u|ε)	-0.123	0.515
$r_{u,\hat{u}}$	0.955	
	(0.800)	

$n = 100$ $T = 10$

	Bias	MSE
β	-9.7e-04	7.2e-04
γ_0	-0.088	0.030
γ_1	-0.010	0.003
ψ	0.002	7.1e-04
ρ	-0.019	0.004
E(u|ε)	-0.050	0.346
$r_{u,\hat{u}}$	0.974	
	(0.838)	

$n = 250$ $T = 5$

	Bias	MSE
β	-0.001	5.2e-04
γ_0	-0.180	0.054
γ_1	-0.012	0.003
ψ	-0.002	7.1e-04
ρ	-0.011	0.004
E(u|ε)	-0.106	0.496
$r_{u,\hat{u}}$	0.959	
	(0.801)	

$n = 250$ $T = 10$

	Bias	MSE
β	-3.9e-04	2.5e-04
γ_0	-0.073	0.012
γ_1	-0.013	0.001
ψ	0.002	3.0e-04
ρ	-0.019	0.002
E(u|ε)	-0.039	0.336
$r_{u,\hat{u}}$	0.975	
	(0.842)	

Table 7: Summary Statistics and brief definition for variables considered in our Hospital Technical Efficiency application. The data set consist of the population of Lazio Hospital (N=109) over the 2000-2005 period (for a total of 619 observations).

Variable	**Definition**	**Mean**	**Std. Dev.**
	Output variables:		
Y_1	Sum of DRG weights related to Complex Surgery	979.15	2846.44
Y_2	Sum of DRG weights related to Emergency Room Treatments	189.87	260.08
Y_3	Sum of DRG weights related to Cancers and HIV	1170.59	2693.84
Y_4	Sum of DRG weights related to General Surgery	3506.41	4382.72
Y_5	Sum of DRG weights related to General Medicine	5358.47	8163.07
	Input variables:		
X_1	No. of Physicians	111.45	189.11
X_2	No. of Nurses	223.45	388.37
X_3	No. of other staff	216.06	424.61
X_4	No. of beds	216.77	315.99
	Inefficiency factors:		
Gini index		0.671	0.13
Nurses/Beds	No. of Nurses / No. of Beds ratio	0.881	0.574
Phys/Beds	No. of Physicians / No. of Beds ratio	0.47	0.333
Public Hospitals	Dummy variable for fully Public Hospitals	0.469	0.50
Not-for-profit Hospitals	Dummy variable for Not-for-profit Hospitals	0.163	0.37
For-profit Hospitals	Dummy variable for For-profit Hospitals	0.369	0.48
Rome	Dummy variable for Hospitals located in the Rome area	0.723	0.45
Viterbo	Dummy variable for Hospitals located in the Viterbo area	0.068	0.25
Rieti	Dummy variable for Hospitals located in the Rieti area	0.029	0.17
Latina	Dummy variable Hospitals located in the Latina area	0.081	0.27
Frosinone	Dummy variable for Hospitals located in the Frosinone area	0.100	0.30
Year 2000	Dummy variable for year 2000	0.169	0.37
Year 2001	Dummy variable for year 2001	0.166	0.37
Year 2002	Dummy variable for year 2002	0.169	0.37
Year 2003	Dummy variable for year 2003	0.172	0.38
Year 2004	Dummy variable for year 2004	0.167	0.37
Year 2005	Dummy variable for year 2005	0.156	0.36

Table 8: Scale and output distance elasticities evaluated at the average values. Comparison between PDE and Pooled PDE. $\epsilon_{\mathbb{Y},\mathbb{X}}$ denotes return to scale estimated by $\sum_{i=1}^{4} \epsilon_{\mathbb{Y},X_i}$.

(a) Exponential

	Pooled PDE			PDE	
	Estimate	Std.error		Estimate	Std.error
$\epsilon_{\mathbb{Y},\mathbb{X}}$	1.041	0.025	$\epsilon_{\mathbb{Y},\mathbb{X}}$	1.103	0.027
$\epsilon_{\mathbb{Y},X_1}$	0.099	0.048	$\epsilon_{\mathbb{Y},X_1}$	0.166	0.017
$\epsilon_{\mathbb{Y},X_2}$	0.195	0.061	$\epsilon_{\mathbb{Y},X_2}$	0.265	0.018
$\epsilon_{\mathbb{Y},X_3}$	0.001	0.037	$\epsilon_{\mathbb{Y},X_3}$	-0.120	0.013
$\epsilon_{\mathbb{Y},X_4}$	0.746	0.066	$\epsilon_{\mathbb{Y},X_4}$	0.791	0.028
$\epsilon_{\mathbb{Y},Y_1}$	-0.009	0.012	$\epsilon_{\mathbb{Y},Y_1}$	0.023	0.007
$\epsilon_{\mathbb{Y},Y_2}$	0.086	0.018	$\epsilon_{\mathbb{Y},Y_2}$	0.116	0.010
$\epsilon_{\mathbb{Y},Y_3}$	0.096	0.015	$\epsilon_{\mathbb{Y},Y_3}$	0.099	0.012
$\epsilon_{\mathbb{Y},Y_4}$	0.216	0.030	$\epsilon_{\mathbb{Y},Y_4}$	0.162	0.012
$\epsilon_{\mathbb{Y},Y_5}$	0.612	0.026	$\epsilon_{\mathbb{Y},Y_5}$	0.601	0.017

(b) Half-normal

	Pooled PDE			PDE	
	Estimate	Std.error		Estimate	Std.error
$\epsilon_{\mathbb{Y},\mathbb{X}}$	1.081	0.060	$\epsilon_{\mathbb{Y},\mathbb{X}}$	1.110	0.045
$\epsilon_{\mathbb{Y},X_1}$	0.367	0.066	$\epsilon_{\mathbb{Y},X_1}$	0.150	0.030
$\epsilon_{\mathbb{Y},X_2}$	0.463	0.093	$\epsilon_{\mathbb{Y},X_2}$	0.292	0.031
$\epsilon_{\mathbb{Y},X_3}$	-0.034	0.052	$\epsilon_{\mathbb{Y},X_3}$	-0.132	0.020
$\epsilon_{\mathbb{Y},X_4}$	0.286	0.101	$\epsilon_{\mathbb{Y},X_4}$	0.800	0.049
$\epsilon_{\mathbb{Y},Y_1}$	-0.005	0.017	$\epsilon_{\mathbb{Y},Y_1}$	0.023	0.011
$\epsilon_{\mathbb{Y},Y_2}$	0.103	0.019	$\epsilon_{\mathbb{Y},Y_2}$	0.115	0.017
$\epsilon_{\mathbb{Y},Y_3}$	0.134	0.020	$\epsilon_{\mathbb{Y},Y_3}$	0.098	0.021
$\epsilon_{\mathbb{Y},Y_4}$	0.203	0.025	$\epsilon_{\mathbb{Y},Y_4}$	0.163	0.021
$\epsilon_{\mathbb{Y},Y_5}$	0.564	0.033	$\epsilon_{\mathbb{Y},Y_5}$	0.601	0.028

Table 9: Estimated inefficiency effects. Comparison between PDE and "Pooled" PDE. $\bar{\sigma} = \frac{1}{N}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\hat{\sigma}_{it}$. Significance levels: "*":$p < 10\%$;"**":$p < 5\%$, "***":$p < 1\%$.

	Pooled PDE				PDE			
	EXP		HN		EXP		HN	
2^{nd} quartile of beds	0.039		0.175		-0.718	***	-0.716	***
	(0.07)		(0.27)		(0.18)		(0.10)	
3^{nd} quartile of beds	-0.096		0.050		-1.250	***	-1.166	***
	(0.09)		(0.36)		(0.23)		(0.15)	
4^{nd} quartile of beds	-0.124		-0.691		-1.728	***	-1.561	***
	(0.11)		(0.43)		(0.28)		(0.19)	
Gini	0.130		-0.629		-0.979		-0.804	
	(0.39)		(0.80)		(0.84)		(0.54)	
Nurses/beds	-0.247		-3.936	***	-0.400	*	-0.438	***
	(0.15)		(0.78)		(0.22)		(0.17)	
Physicians/beds	0.050		-4.014	***	0.444	*	0.509	**
	(0.18)		(1.27)		(0.27)		(0.20)	
Private	-0.084		-0.358		-0.190		-0.209	
	(0.11)		(0.24)		(0.22)		(0.17)	
Not-for-profit	0.279	***	-0.475		-0.030		-0.071	
	(0.08)		(0.42)		(0.24)		(0.18)	
Year 2001	0.001		-0.052		-0.227	*	-0.183	**
	(0.07)		(0.16)		(0.12)		(0.07)	
Year 2002	-0.043		-0.179		-0.355	***	-0.292	***
	(0.07)		(0.16)		(0.14)		(0.09)	
Year 2003	-0.018		0.173		-0.159		-0.110	
	(0.08)		(0.18)		(0.13)		(0.08)	
Year 2004	0.070		0.153		0.058		0.017	
	(0.09)		(0.17)		(0.13)		(0.08)	
Year 2005	0.066		0.394	**	0.303	**	0.239	**
	(0.08)		(0.19)		(0.15)		(0.10)	
Viterbo	0.064		0.447	***	-1.224	***	-1.224	***
	(0.09)		(0.16)		(0.29)		(0.19)	
Latina	0.118		-0.148		-1.084	***	-0.882	***
	(0.11)		(0.17)		(0.28)		(0.23)	
Rieti	-0.016		-0.699		-0.639		-0.691	***
	(0.11)		(1.56)		(0.43)		(0.25)	
Frosinone	-0.029		0.134		-0.741	***	-0.725	***
	(0.11)		(0.14)		(0.22)		(0.16)	
Constant	-0.968	***	2.653	***	-0.142		0.215	
	(0.31)		(0.71)		(0.67)		(0.41)	
$\bar{\sigma}$	0.34		0.38		0.13		0.22	
ψ	0.04		0.23		0.03		0.01	
Estimated technical inefficiencies, $\hat{u}_{it}$								
Mean	0.388		0.327		0.141		0.184	
SD	0.239		1.190		0.156		0.175	
Min	0.002		0.000		0.002		0.001	
Max	1.382		19.640		0.978		1.064	
Spearman correlation								
Pooled PDE (exp)	1.000							
Pooled PDE (hn)	0.194		1.000					
PDE (exp)	0.282		0.167		1.000			
PDE (hn)	0.267		0.198		0.987		1.000	