BANCA D'ITALIA
EUROSISTEMA

Temi di Discussione

(Working Papers)

Consistent inference in fixed-effects stochastic frontier models

by Federico Belotti and Giuseppe Ilardi

October 2017

Number 1147

BANCA D'ITALIA
EUROSISTEMA

Temi di discussione

(Working papers)

Consistent inference in fixed-effects stochastic frontier models

by Federico Belotti and Giuseppe Ilardi

Number 1147 - October 2017

ISSN 1594-7939 (print)
ISSN 2281-3950 (online)

Printed by the Printing and Publishing Division of the Bank of Italy

CONSISTENT INFERENCE IN FIXED-EFFECTS STOCHASTIC FRONTIER MODELS

by Federico Belotti* and Giuseppe Ilardi**

Abstract

The classical stochastic frontier panel data models provide no mechanism for disentangling individual time-invariant unobserved heterogeneity from inefficiency. Greene (2005a, b) proposed the 'true' fixed-effects specification, which distinguishes these two latent components while allowing for time-variant inefficiency. However, due to the incidental parameters problem, the maximum likelihood estimator proposed by Greene may lead to biased variance estimates. We propose two alternative estimation procedures that, by relying on a first-difference data transformation, achieve consistency when n goes to infinity with fixed T. Furthermore, we extend the approach of Chen et al. (2014) by providing a computationally feasible solution for estimating models in which inefficiency can be heteroskedastic and may follow a first-order autoregressive process. We investigate the finite sample behavior of the proposed estimators through a set of Monte Carlo experiments. Our results show good finite sample properties, especially in small samples. We illustrate the usefulness of the new approach by applying it to the technical efficiency of hospitals.

JEL Classification: C13, C16, C23.
Keywords: stochastic frontiers, fixed-effects, panel data, pairwise differencing.

Contents

* University of Rome Tor Vergata, Department of Economics and Finance.

** Bank of Italy, Directorate General for Economics, Statistics and Research, Statistical Analysis Directorate.

1 Introduction

The[1] analysis of efficiency is an important issue in many empirical studies and the Stochastic Frontier (SF) model, introduced by Aigner et al. (1977) and by Meeusen & van den Broeck (1977), represents a popular tool to measure this unobservable economic indicator.[2] Since then, several improvements on SF modeling have led to a large empirical literature, most of which is based on panel data.[3]

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.[4] The first group treats time invariant heterogeneity as if it were 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; Colombi et al., 2014; Chen et al., 2014).

This study reconsiders the estimation of the "true" fixed-effects (TFE) model in Greene (2005a) addressing the incidental parameters problem that affect his maximum likelihood dummy variables estimator (MLDVE). As Greene's simulations suggest, this issue does not affect the frontier coefficients but leads to inconsistent variance estimates. Since these parameters represent the key ingredients in the post-estimation of inefficiencies, a solution to this issue is crucial in the

[1] An earlier version of this paper circulated under the title "Consistent Estimation of the true fixed fixed-effects stochastic frontier models". We are indebted with Marco Perone Pacifico, Franco Peracchi and Andrea Piano Mortari for valuable advices and discussions. We thank Silvio Daidone and Francesco D'Amico for sharing their data with us. We also thank Bill Greene, Vincenzo Atella, Clinton Horrace, Santiago Pereda Fernandez, Valentina Nigro and the participants at the XII European Workshop on Efficiency and Productivity Analysis and at the 46^{th} Italian Statistical Society Scientific Meeting for their helpful comments. Any errors and omissions are the sole responsibility of the authors. The opinions expressed are those of the authors and do not necessarily reflect the position of Bank of Italy.

[2] 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.

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

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

SF context. In a parallel research,[5] Chen et al. (2014) propose a consistent marginal maximum likelihood estimator (MMLE) for the TFE model exploiting a within-group data transformation and the properties of the closed skew normal (CSN) class of distributions (Gonzalez-Farias et al., 2004a).[6] The authors study the normal-half normal model and acknowledge the possibility of allowing for a truncated normal inefficiency. However, they restrict both error components to be homoskedastic, a strong assumption that could heavily affect the inference in a SF framework (Kumbhakar & Lovell, 2000).

In this paper we propose two alternative consistent estimators which extend the Chen et al. (2014) results in different directions. The first estimator is based on the marginalization of the inefficiency term via simulation and can be used to estimate both homoskedastic and heteroskedastic normal-half normal and normal-exponential models. The second is a U-estimator based on all pairwise quasi-likelihood contributions constructed exploiting the analytical expression available for the marginal likelihood function when $T = 2$. This strategy allow us to provide a computationally feasible approach to estimate normal-half normal, normal-exponential and normal-truncated normal models in which both error components can be heteroskedastic. Furthermore, for the normal-half normal and normal-truncated normal models, the inefficiency is also allowed to follow a first-order autoregressive 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 hypotheses of interest and policy implications, avoiding biased two-step procedures (Wang & Schmidt, 2002).

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) results. In Section 5 we discuss whether the fixed-effects may also capture part of the inefficiency, providing new insight from a statistical perspective to this relevant issue. Section 6 investigates the small sample properties of the proposed estimators through a set of Monte Carlo

[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] In the spirit of Cornwell & Schmidt (1992), the term *marginal* refers to a model in which the nuisance parameters have been eliminated through a data transformation. Consistently, we then refer to the Chen et al. (2014) estimator as MMLE instead of *Within* MLE.

experiments. Section 7, provides an illustration through an application on hospitals technical efficiency. Finally, Section 8 offers some conclusions.

2 The model

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

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

where, for each unit i and period t, y_{it} represents the log of output, $\boldsymbol{x}_{it}$ is a $1 \times k$ vector of functions 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 symmetric idiosyncratic error v_{it} and the one-sided disturbance u_{it}, which represents inefficiency.[8] The v_{it} disturbance is assumed to be normally distributed with zero mean and variance ψ^2 while u_{it} is assumed to be distributed independently of v_{it} according to the $\mathcal{F}_u$ distribution. In what follows, we are particularly concerned with the cases in which $\mathcal{F}_u$ is exponential with scale parameter σ, $u_{it} \sim \mathcal{E}(\sigma)$, half normal with zero mean and variance σ^2, $u_{it} \sim \mathcal{N}^+(0, \sigma^2)$, and truncated normal with mean ν and variance σ^2, $u_{it} \sim \mathcal{N}^+(\nu, \sigma^2)$.[9] We define $\lambda = \sigma/\psi$ 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}. There is a philosophical debate in the SF literature about whether or not to separate α_i from u_{it}. We sidestep this issue and focus instead on how to consistently estimate the parameters of model (1)-(4). Nevertheless, Section 5 provides a discussion of some relevant cases in which the fixed-effects may capture part of the inefficiency.

[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.

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]

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 $u_i \sim \mathcal{N}^+(\nu, \sigma^2)$. 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 $\boldsymbol{u}$ over time at the cost of a more challenging marginal likelihood derivation.

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

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

where $\Delta\boldsymbol{y}_i = (\Delta y_{i2}, \dots, \Delta y_{iT})'$ 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 $\Delta\boldsymbol{x}_{it} = (\Delta x_{it1}, \dots, \Delta x_{itk})$, $\forall\ t = 2, \dots, T$. The normality assumption for v_{it} implies that $\Delta\boldsymbol{v}_i$ has a $T-1$-variate normal distribution with

[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 Section 6.1).

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 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 $\Delta \boldsymbol{u}_i$.

In order to apply the MSL approach, we make the following assumptions: *i*) $\mathcal{F}_u$ 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 $\tilde{\boldsymbol{u}}_i$ does not depend on the unknown parameter σ.

The first assumption is needed to derive a valid draw to be used for the estimation and rules out the possibility of using two-parameters distributions (e.g., truncated normal or gamma). 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 both the normal-gamma and normal-truncated normal convolutions by Ritter & Simar (1997) and Almanidis & Sickles (2012), respectively. 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 terms of its simulated counterpart as

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

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 $\mathcal{F}_{\Delta\tilde{\boldsymbol{u}}}$.[11] Therefore, all we need is to simulate from the distribution of $\Delta\tilde{\boldsymbol{u}}_i$ and this task is straightforward if we are able to simulate from the univariate distribution of u_{it}. Indeed, if the two previous assumptions hold, the vector $\Delta\tilde{\boldsymbol{u}}_{ig} = (\tilde{u}_{i2g} - \tilde{u}_{i1g}, \dots, \tilde{u}_{iTg} - \tilde{u}_{i(T-1)g})'$ is, by construction, a valid draw from $\mathcal{F}_{\Delta\tilde{\boldsymbol{u}}}$. Following Gouriéroux & Monfort (1991), under the regularity conditions ensuring the large sample properties of the MMLE, the resulting marginal maximum simulated likelihood estimator (MMSLE) is consistent and asymptotically equivalent to the MMLE as $n \to \infty$ and $G \to \infty$ with $\sqrt{n}/G \to 0$.[12] 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).

[11] Clearly, since $\tilde{\boldsymbol{u}}$ does not depend on σ, the same is true also for the distribution of $\Delta\tilde{\boldsymbol{u}}$.

[12] A discussion of the conditions required for the consistency of the MMLE for the normal-exponential model are given in Appendix A.2.

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, 2005a,b). This strategy makes the MMSLE manageable even in the case of moderately large sample sizes (e.g., $n = 1000$, $T = 10$). Our simulations suggest that the minimum number of Halton sequences needed to get a suitable approximation of the expectation (12) is 10 sequences per observation.[13]

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

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

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. This extension relaxes the i.i.d. assumption allowing for an additional source of (deterministic) variability in the inefficiency term. This model feature is relevant since both model parameters and inefficiency estimates may be adversely affected by neglected heteroskedastic-

[13]Notice that the same approximation can be reached using 100 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.

ity.[14] Moreover, the inclusion of explanatory variables correlated with inefficiency but not with unobserved heterogeneity may enhance parameters identification (Amsler & Schmidt, 2015).

A possible specification for the unit- and time-specific scale parameter may be $\sigma_{it} = g(\boldsymbol{z}_{it}\boldsymbol{\delta})$ where $g(.)$ is a known positive monotonic function, $\boldsymbol{z}_{it}$ is a $1 \times s$ vector of exogenous covariates and $\boldsymbol{\delta}$ is a $s \times 1$ vector of parameters to be estimated. The simulated likelihood contribution becomes

$$L_i^*(\boldsymbol{\theta}) \quad \approx \quad \frac{1}{G}\sum_{g=1}^{G} \phi_{T-1}\left(\Delta\boldsymbol{\varepsilon}_i + \Delta\boldsymbol{\eta}_i; \mathbf{0}, \Psi\right), \tag{15}$$

where $\Delta\boldsymbol{\eta}_i = \Delta\left(\boldsymbol{\sigma}_i \odot \tilde{\boldsymbol{u}}_{ig}\right)$ with $\boldsymbol{\sigma}_i = (\sigma_{i1}, \ldots, \sigma_{iT})'$, $\tilde{\boldsymbol{u}}_{ig} = (\tilde{u}_{i1g}, \ldots, \tilde{u}_{iTg})'$ is a draw from $\mathcal{F}_{\tilde{\boldsymbol{u}}}$ and the symbol $\odot$ represents the element-wise product.[15] In this case, we can define the average STN ratio $\bar{\lambda} = \frac{1}{nT\psi}\sum_{i=1}^{n}\sum_{t=1}^{T}\sigma_{it}$. Notice that allowing for heteroskedastic inefficiency makes the MMSLE far more computationally intensive (the rule of thumb is to use at least 50 sequences per observation). As shown in the next Section, this shortcoming can be overcome at the cost of a (negligible) loss in terms of efficiency.

3.2 Pairwise difference estimator

As a computationally feasible alternative, we propose to exploit the closed-form expression of the integral (10) when the inefficiency is exponentially distributed and $T = 2$.[16] Indeed, while we were able to derive the p.d.f. of the random vector $\Delta\boldsymbol{u}_i$ also when $T > 2$, the subsequent marginalization needed to derive the p.d.f. of the random vector $\Delta\boldsymbol{\varepsilon}_i$ cannot be performed in closed form.[17] Nevertheless, when $T = 2$, 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 $\varsigma_t = \sigma_t^{-1}$, $t = 1, 2$,

[14] 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.

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

[16] The half normal and truncated normal cases will be discussed in Section 4.

[17] See Appendix A.1.

is asymmetric Laplace distributed with 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 a two-period panel is

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

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

$$\varsigma_{it}=\exp(-\boldsymbol{z}_{it}\boldsymbol{\gamma}), \tag{18}$$

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 easily introduced by modeling the variance of the idiosyncratic error.

The marginal likelihood function (17) implies the existence of $H=\binom{T}{2}$ consistent MMLEs, one for each "subsample" extracted considering two waves of the panel. In fact, if we restrict the inference to one of these subsamples, 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 applied on 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.

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

$$\breve{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \frac{1}{nT^*} \sum_{i=1}^{n} \sum_{t=2}^{T} \log f(\Delta y_{it} |\ \boldsymbol{\theta}, \Delta \boldsymbol{x}_{it}), \tag{19}$$

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 terms of the product of the subsamples likelihood contributions.

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

$$U_n(\boldsymbol{\theta}) = n^{-1} \binom{T}{2}^{-1} \sum_{i=1}^{n} \sum_{t=2}^{T} \sum_{s<t} \log f(\Delta_t^s y_i |\ \boldsymbol{\theta}, \Delta_t^s \boldsymbol{x}_i), \tag{20}$$

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 (20) 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.[18]

As already mentioned, 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 is a direct

[18]Consistently with Abrevaya (1999), we find that the criterion function (20) is much smoother than the MMLE's objective function.

consequence of the consistency of the underlying subsamples MMLEs.[19] Under the assumptions ensuring the asymptotic normality of PDE and recognizing that this estimator belongs to the class of U-estimators, its asymptotic variance is equal to $A_0^{-1}B_0A_0^{-1}$, where

$$A_0 = -\sum_{t=2}^{T}\sum_{s<t}\mathbb{E}\left[\nabla_{\boldsymbol{\theta\theta}}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)\right], \tag{21}$$

and

$$B_0 = \mathbb{E}\left\{\left[\sum_{t=2}^{T}\sum_{s<t}\nabla_{\boldsymbol{\theta}}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)\right]\left[\sum_{t=2}^{T}\sum_{s<t}\nabla_{\boldsymbol{\theta}}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \boldsymbol{\theta}_0)\right]'\right\}, \tag{22}$$

where $\nabla_{\boldsymbol{\theta}}$ and $\nabla_{\boldsymbol{\theta\theta}}$ denote the vector of first and second derivatives of the objective function respectively, and $\boldsymbol{\theta}_0$ is the true parameters vector.[20] Estimation of the asymptotic variance of the PDE is straightforward since, for each pairwise difference, we have that

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

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

$$\widehat{Avar}(\tilde{\theta}) = n^{-1}\hat{A}_0^{-1}\hat{B}_0\hat{A}_0^{-1}, \tag{23}$$

where

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

and

$$\hat{B}_0 = \hat{A}_0 + \frac{1}{n}\sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s<t}\sum\sum_{(k,h)\neq(t,s)}\nabla_{\boldsymbol{\theta}}\log f(\Delta_t^s y_i|\Delta_t^s \boldsymbol{x}_i, \tilde{\boldsymbol{\theta}})\nabla_{\boldsymbol{\theta}}\log f(\Delta_k^h y_i|\Delta_k^h \boldsymbol{x}_i, \tilde{\boldsymbol{\theta}}). \tag{25}$$

[19] A discussion of the conditions required for the consistency of the subsamples MMLEs and of the PDE for the normal-exponential model are given in Appendix A.2.

[20] Even if we do not establish the asymptotic normality of the PDE, it is worth noting that our Monte Carlo simulations show that the ratio between the average standard errors obtained from the proposed covariance matrix estimator and the standard deviation over replications of the estimated coefficients is close to one.

Notice that the consistency property of $\tilde{\boldsymbol{\theta}}$ can also be obtained when $T \to \infty$ with fixed n. However, due to the inconsistency of the subsamples MMLEs, a formal proof needs to be adapted from 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.[21] A simpler alternative when the inefficiency is assumed to be heteroskedastic is given by

$$\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - \boldsymbol{x}_{it}\hat{\boldsymbol{\beta}} + \hat{c}_{it}\right) \qquad i = 1, \ldots, n, \tag{26}$$

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 $u_{it} \sim \mathcal{N}^+(0, \sigma_{it}^2)$ ($\hat{\sigma}_{it} = \hat{\sigma}$ in the homoskedastic case).[22] This estimator is equivalent to the mean-adjusted estimator of α_i in the fixed-effects linear model and, therefore, 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. Furthermore, we propose a variant of the TFE model in which the inefficiency is assumed to be heteroskedastic

[21] 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.

[22] Equation (26) refers to the case of production frontiers. For cost frontiers, the $\hat{c}_{it}$ term enters the expression with a minus sign.

and follows a first-order autoregressive process. In both cases we exploit 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 its closeness under marginalization and linear 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). For all the specifications covered in this Section, we do not provide any formal proof of, nor claim, consistency and/or asymptotic normality. Nonetheless, the Monte Carlo experiments reported in Section 6, which cover also the specifications considered in this Section, show that both the bias and the variance of the PDE tend to zero as the sample size increases.

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

$$f(\boldsymbol{y}) = C\phi_p\left(\boldsymbol{y}; \boldsymbol{\mu}, \Omega\right)\Phi_q\left(D(\boldsymbol{y}-\boldsymbol{\mu}); \boldsymbol{\nu}, \Delta\right), \quad \boldsymbol{y} \in \mathbb{R}^p, \tag{27}$$

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

$$\boldsymbol{\varepsilon}_i \sim CSN_{T,T}(-\boldsymbol{\nu}, \Omega_*, -\Sigma\Omega_*^{-1}, -\boldsymbol{\nu}, \Delta_*), \quad i = 1, \ldots, n, \tag{28}$$

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 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

$$A = \begin{pmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & 0 & \dots & 0 & -1 & 1 \end{pmatrix}. \tag{29}$$

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

$$\Delta\boldsymbol{\varepsilon}_i \sim CSN_{T-1,T}(-\Delta\boldsymbol{\nu}, \Omega_A, D_A, -\boldsymbol{\nu}, \Delta_A), \quad i = 1, \dots, n, \tag{30}$$

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[23]

$$f(\Delta\boldsymbol{\varepsilon}_i) \;=\; \left[\Phi_T\left(\mathbf{0}; -\boldsymbol{\nu}, \Sigma\right)\right]^{-1} \phi_{T-1}\left(\Delta\boldsymbol{\varepsilon}_i; -\Delta\boldsymbol{\nu}, \Omega_A\right) \Phi_T\left(D_A\Delta\boldsymbol{\varepsilon}_i; -\boldsymbol{\nu}, \Delta_A\right). \tag{31}$$

By exploiting this result, the MMLE for the normal-truncated normal "true" fixed-effects model is

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^{n} \log f(\Delta\boldsymbol{\varepsilon}_i |\, \boldsymbol{\theta}, \Delta\boldsymbol{x}_{it}). \tag{32}$$

In general, the maximization of (32) 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 (Chen et al., 2014). Indeed, the covariance matrix Δ_A in $\Phi_T(.)$ has a special (equicorrelated) structure and the computations may be greatly

[23] The normal-half normal case can be straightforwardly obtained by replacing $\boldsymbol{\nu} = \mathbf{0}$ in (28).

simplified following the result outlined in Kotz et al. (2000).

This simplification does not apply to the case of heteroskedastic errors, i.e. $\Sigma_i = diag(\sigma_{i1}^2, \dots, \sigma_{iT}^2)$ and $\Psi_i = diag(\psi_{i1}^2, \dots, \psi_{iT}^2)$, since Δ_A has not the aforementioned desirable structure anymore. In order to keep the estimation feasible, we propose to apply the pairwise difference approach using the marginal likelihood function of a two-period normal-truncated normal model. The marginal likelihood function can be straightforwardly obtained by considering $T = 2$ in equation (31) as

$$L^*(\boldsymbol{\theta}) = \prod_{i=1}^{n} [\Phi_2(\mathbf{0}; -\tilde{\boldsymbol{\nu}}_i, \Sigma_i)]^{-1} \phi\left(\Delta\varepsilon_{it}; -\Delta\nu_{it}, \xi_i^2\right) \Phi_2\left(\tilde{\boldsymbol{d}}_i; \mathbf{0}, \Xi_i\right), \tag{33}$$

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, $\tilde{\boldsymbol{d}}_i = \{\left[(\sigma_{i1}^2, -\sigma_{i2}^2)'\xi_i^{-2}\right](\Delta\varepsilon_{it} + \Delta\nu_{it})\} + \tilde{\boldsymbol{\nu}}_i$, $\Sigma_i = diag(\sigma_{i1}^2, \sigma_{i2}^2)$, $\xi_i^2 = \sigma_{i1}^2 + \sigma_{i2}^2 + \psi_{i1}^2 + \psi_{i2}^2$, $\sigma_{it} = \exp(\boldsymbol{z}_{it}\boldsymbol{\gamma})$ 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 makes 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

$$U_n(\boldsymbol{\theta}) = n^{-1}\binom{T}{2}^{-1} \sum_{i=1}^{n}\sum_{t=2}^{T}\sum_{s<t} \log\left\{[\Phi_2(\mathbf{0}; -\tilde{\boldsymbol{\nu}}_i, \Sigma_i)]^{-1} \phi\left(\Delta_t^s\varepsilon_i; -\Delta_t^s\nu_i, \xi_i^2\right) \Phi_2\left(\tilde{\boldsymbol{d}}_i; \mathbf{0}, \Xi_i\right)\right\}, \tag{34}$$

where the pairwise difference operator Δ_t^s is defined in equation (20). 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 $\hat{\nu}_{it} = \tilde{\boldsymbol{r}}_{it}\hat{\boldsymbol{\tau}}$.

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 firms' control.

Second, since the production process may be affected by unexpected events, inefficiency can be considered a stochastic variable that randomly varies over time. Third, some inputs are considered "fixed" in the short-run because the economic environment places high adjustment costs. In this case, we expect the inefficiency to be persistent, that is the 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 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 (28). In particular, we consider the following heteroskedastic normal-half normal model with $AR(1)$ inefficiencies

$$\begin{aligned} \boldsymbol{v}_i &\sim \mathcal{N}_T(\mathbf{0}, \Psi_i), && (35)\\ \boldsymbol{u}_i &\sim \mathcal{N}_T^+(\mathbf{0}, \Sigma_i), && (36)\\ \Sigma_i &= \frac{1}{1-\rho^2}\Omega_i, && (37)\\ \Psi_i &= diag(\psi_{i1}^2, \ldots, \psi_{iT}^2), && (38) \end{aligned}$$

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 (34) the following

$$\Sigma_i = \begin{pmatrix} \sigma_{i1}^2 & \rho\sigma_{i1}\sigma_{i2} \\ \rho\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 \end{pmatrix}, \tag{39}$$

$$\xi_i^2 = \sigma_{i1}^2 + \sigma_{i2}^2 + \psi_{i1}^2 + \psi_{i2}^2 - 2\rho\sigma_{i1}\sigma_{i2}, \tag{40}$$

$$\tilde{\boldsymbol{d}}_i = \{\left[(\sigma_{i1}^2 + \rho\sigma_{i1}\sigma_{i2}, -\sigma_{i2}^2 - \rho\sigma_{i1}\sigma_{i2})'\xi_i^{-2}\right] \Delta\varepsilon_{it}, \tag{41}$$

$$\Xi_i = \begin{pmatrix} \sigma_{i1}^2 & \rho\sigma_{i1}\sigma_{i2} \\ \rho\sigma_{i1}\sigma_{i2} & \sigma_{i2}^2 \end{pmatrix} - \xi_i^{-2} \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \tag{42}$$

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 (26) with

$$\hat{\boldsymbol{c}}_i = \hat{\Sigma}_{it} \left(\frac{\Phi_T^*(\mathbf{0}, \mathbf{0}, \hat{\Sigma}_i)}{\Phi_T(\mathbf{0}, \mathbf{0}, \hat{\Sigma}_i)} \right), \tag{43}$$

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

$$\mathbb{E}(\boldsymbol{u}_i|\hat{\boldsymbol{\varepsilon}}_i) = \hat{\Upsilon}_i + \hat{\Delta}_{*i}\frac{\Phi_T^*(\hat{\Upsilon}_i, \mathbf{0}, \hat{\Delta}_{*i})}{\Phi_T(\hat{\Upsilon}_i, \mathbf{0}, \hat{\Delta}_{*i})}, \tag{44}$$

where $\hat{\Upsilon}_i = -\hat{\Sigma}_i(\hat{\Sigma}_i + \hat{\Psi}_i)^{-1}\hat{\boldsymbol{\varepsilon}}_i$, $\hat{\Delta}_{*i} = \hat{\Sigma}_i - \hat{\Sigma}_i(\hat{\Sigma}_i + \hat{\Psi}_i)^{-1}\hat{\Sigma}_i$.[24]

5 Unobserved heterogeneity and inefficiency

A TFE model can be very useful in practice to separate the inefficiency term from the effect of relevant time invariant unobserved features of the production environment that are beyond the control of the firm.[25] Following the example of Amsler & Schmidt (2015), suppose that the firms are farms, then, these factors could be represented by soil quality or microclimate. In these circumstances, a relevant issue is whether the fixed-effects may capture also part of the

[24] 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.

[25] Here, as in the rest of the paper, we are still assuming that the inefficiency is time varying.

inefficiency. This section aims to provide new insight to this issue from a statistical perspective. As formally proved in the Appendix A.2, even thought the pairwise estimation strategy is based on a first difference transformation of the data, the information contained in Δu_{it} allow the estimation of all the relevant features characterizing the distribution of the inefficiency. Despite this fixed-effects killing transformation, the PDE allows to consistently estimate also the parameters associated with time invariant inefficiency factors (see the simulations results reported in Section 6). However, there are cases in which the fixed-effects may capture part of the inefficiency. To see this point, let us consider the following competing model specification proposed by Colombi et al. (2014)

$$y_{it} \quad = \quad \alpha_i + \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it} - \tilde{u}_i - \tilde{u}_{it}, \tag{45}$$

in which the inefficiency is represented by two (additive) components: one is time invariant, $\tilde{u}_i$, the other is time varying, $\tilde{u}_{it}$. In this case, applying any fixed-effects killing transformation will also eliminate $\tilde{u}_i$. Although model (45) looks like a generalization of model (1)-(2), the two models are based on different distributional assumption on the inefficiency term, u_{it} for model (1)-(2) and $u_{it} = \tilde{u}_i + \tilde{u}_{it}$ in model (45). Indeed, Colombi et al. (2014) assume that $\tilde{u}_i \sim \mathcal{N}^+(0, \sigma_1^2)$ and $\tilde{u}_{it} \sim \mathcal{N}^+(0, \sigma_2^2)$ and the sum of independent half normals is not half normally distributed.[26] Moreover, Colombi et al. (2014) also assume that $\tilde{u}_i$ is independent from $\tilde{u}_{it}$, which is a rather strong assumption. As stated in Kumbhakar & Hjalmarsson (1995) and Colombi et al. (2014), the main advantage is to allow for a unit specific long-run component in a simple homoskedastic set up. As also discussed before (see Section 4.2), inefficiencies can be persistent due to an autoregressive structure or can fluctuate around a production unit's specific mean due to observable time invariant factors affecting its moments (e.g., ownership, organizational structure, location).[27] In this paper we show how both cases can be easely modeled allowing the inefficiency to be simultaneously heteroskedastic and dynamic. We believe that, in general, it is not possible to formally test which of the two competing specifications is supported by the

[26]This statistical fact is true also for the exponential and truncated normal distributions.

[27]Notice that the effect of these time-invariant factors is identified even if they are correlated with the unobserved heterogeneity. In a model with only time invariant inefficiency, Amsler & Schmidt (2015) proved that exogenous explanatory variables correlated with the time-invariant inefficiency u_i but not with unobserved heterogeneity α_i can be used to separately identify these two latent components without any distributional assumptions.

data. However, the Colombi et al. (2014) specification shares a set of parameters with the Chen et al. (2014) model (i.e., the frontier parameters and those corresponding to the terms $\tilde{u}_{it}$ and v_{it}). Then, under the assumption that the Colombi et al. (2014) model is correctly specified, the MMLE of Chen et al. (2014) is consistent in estimating these common parameters. This suggests that the validity of Colombi et al. (2014) specification may be easily verified by using a Hausman (1978) test comparing the common parameters estimates. Clearly, a rejection of the null hypothesis does not imply necessarily that the correct model is the one proposed by Chen et al. (2014), but only that the assumptions of the Colombi et al. (2014) model are not supported by the data. Summarizing, unless the inefficiency can be decomposed additively in a time invariant and a time varying components, the estimators proposed in this paper can be used to successfully separate unobserved heterogeneity from inefficiency.

6 Monte Carlo evidence

In this Section we study the finite sample properties of the MMSLE and PDE via numerical simulations.[28] 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. This exercise compares two consistent estimators allowing to make some statement on their finite sample 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 ($T = 5, 10$). All simulation designs have a common base: *i*) the fixed-effect parameters $\alpha_1, ..., \alpha_n$ are drawn from a standard Gaussian random variable; *ii*) one explanatory variable is

[28]A Stata command implementing the methods described in this paper is available. It can be installed from within Stata by typing `net install sftfe, from(http://www.econometrics.it/stata)`. Please, notice that the command cannot be downloaded from the website but just directly installed through the Stata software.

used $x_{it} = 0.5\alpha_i + \sqrt{0.5^2}w_{it}$ with $w_{it} \sim \mathcal{N}(0,1)$; *iii*) 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; *vi*) the number of replications, $R = 1000$.

Simulation results are summarized for each set of simulations by reporting the average bias and Mean Squared Error (MSE) of the estimates, together with the linear ($r_{u,\hat{u}}$) 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 (true) inefficiency.

6.1 PDE vs MLDVE

We consider the following heteroskedastic normal-exponential model

$$
\begin{aligned}
y_{it} &= \alpha_i + \beta x_{it} + v_{it} - u_{it}, && (46)\\
v_{it} &\sim \mathcal{N}(0, \psi^2), && (47)\\
u_{it} &\sim \mathcal{E}(\sigma_i), && (48)\\
\sigma_i &= \exp(\gamma_0 + z_i\gamma_1), \quad i = 1, \ldots, n, \quad t = 1, \ldots, T, && (49)
\end{aligned}
$$

where $z_i \sim \mathcal{N}(0, 0.0625)$. It is worth noting that we have deliberately specified a time invariant inefficiency factor. This allows to show that, despite the first-difference transformation, the PDE is able to consistently estimate the parameters associated with this kind of factors. We keep fixed in each experiment the values of the frontier parameter $\beta = 1$, $\psi = 0.25$ and $\gamma_1 = 1$, while the value of γ_0 varies across scenarios in order to obtain different STN ratios ($\bar{\lambda} = 1, 1.5, 2, 4, 6$).[29] 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.

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 result to highlight is the huge gap between the performance of the two estimators, with the MLDVE unable to

[29]Given the heteroskedastic specification for the inefficiency, see equation (49), the considered STN ratios are actually defined as averages $\bar{\lambda} = \frac{1}{n\psi}\sum_{i=1}^{n}\sigma_i$.

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 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 $\hat{\lambda} = \infty$, or equivalently $\hat{\psi} = 0$.[30] 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).[31] 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.[32]

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 parameter (β) is 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 ($n = 100, T = 5$),

[30] 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 report non-convergence when ψ approaches zero and, as a consequence, this behaviour is often mistakenly considered as a numerical maximization problem.

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

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

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.[33] The PDE's performances 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 0.066 and 0.111 to 0.022 and 0.042, 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.

6.2 MMSLE vs MMLE

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

$$\begin{aligned} y_{it} &= \alpha_i + \beta x_{it} + v_{it} - u_{it}, && (50) \\ v_{it} &\sim \mathcal{N}(0, \psi^2), && (51) \\ u_{it} &\sim \mathcal{N}^+\left(0, \sigma^2\right) \quad i = 1, \ldots, n, \quad t = 1, \ldots, T. && (52) \end{aligned}$$

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

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

$\lambda = 1$, and $\sigma = 1.27684$, $\psi = .63842$ in the other case.[34]

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.

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 only estimator for the homoskedastic normal-exponential TFE model.

6.3 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

$$\begin{aligned}
\boldsymbol{y}_i &= \alpha_i \iota_T + \beta \boldsymbol{x}_i + \boldsymbol{v}_i - \boldsymbol{u}_i, && (53)\\
\boldsymbol{v}_i &\sim \mathcal{N}_T(0, \psi^2 I_t), && (54)\\
\boldsymbol{u}_i &\sim \mathcal{N}_T^+\left(\mathbf{0}, \frac{1}{1-\rho^2}\Omega_i\right), \quad i = 1, \dots, n, && (55)
\end{aligned}$$

where $\Omega_i = \{\omega_{its}\}^{t,s=1,\dots,T}$ with $\omega_{its} = \sigma_{it}\sigma_{is}\rho^{|t-s|}$ and $\sigma_{it} = \exp(\gamma_0 + z_{it}\gamma_1)$ with $z_{it} \sim \mathcal{N}(0,1)$. 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. We set $\beta = 0.5$, $\psi = 0.5$, $\gamma_0 = -0.5$ and $\gamma_1 = 1$

[34] For the MMSLE we use 30 Halton sequences for observation while, following Chen et al. (2014), we exploit the results of Kotz et al. (2000) for the MMLE implementation.

(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 ($\rho = 0.3, 0.7$).

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 0.054 and 0.074 to 0.018 and 0.032 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.

7 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.

7.1 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.[35] Finally, in order to investigate the role of the labor force structure, both the nurses/beds and the physicians/beds ratios are included in the inefficiency equation. Descriptive statistics and a brief summary of variable definitions are reported in Table 7.

7.2 Model and estimation

A Stochastic Distance Function is the easiest solution to represent the multi-output production technology in a single SF equation.[36] Following Kumbhakar & Lovell (2000), a stochastic multi-

[35]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).

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

output distance function model for panel data can be defined as

$$(Y_{5it})^{-1} = D\left(\boldsymbol{X}_{it}, \frac{Y^*_{mit}}{Y_{5it}}; \beta\right) \exp\left(u_{it} - v_{it}\right). \tag{56}$$

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 (56) can be written as[37]

$$\begin{aligned} y_{5it} &= \alpha_i + \sum_{m=1}^{4} \delta_m y^*_{mit} + \sum_{k=1}^{4} \beta_k x_{kit} + \frac{1}{2}\sum_{m=1}^{4}\sum_{p=1}^{4} \delta_{mp} y^*_{mit} y^*_{pit} + \frac{1}{2}\sum_{k=1}^{4}\sum_{j=1}^{4} \beta_{kj} x_{kit} x_{jit} + \\ &\quad + \sum_{k=1}^{4}\sum_{m=1}^{4} \beta_k \delta_m x_{kit} y^*_{mit} + \sum_{t=2001}^{2005} d_t + v_{it} - u_{it}, \end{aligned} \tag{57}$$

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.[38]

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 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 6.1 ($n = 100$ 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 (57). Since also the standard MLE applied on pooled data shows this problematic behavior, for comparison purposes we estimate model (57) using a "pooled" version

[37] Equation (57) 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).

[38] All output and input variables in equation (57) are expressed in logarithms. Symmetry constraints have been imposed on the interaction terms, i.e $\delta_{mp} = \delta_{pm}$ and $\beta_{kj} = \beta_{jk}$.

of the PDE characterized by the following objective function

$$U_n(\boldsymbol{\theta}) = \binom{N}{2}^{-1} \sum_{i=2}^{N} \sum_{j<i} \log f(\Delta_i^j y | \, \boldsymbol{\theta}, \Delta_i^j \boldsymbol{x}), \tag{58}$$

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.[39] 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.

7.3 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 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-

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

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).

8 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 the normal-exponential model and fixed-T asymptotics. The first is a marginal maximum simulated likelihood estimator (MMSLE) that can be used to estimate both homoskedastic and heteroskedastic normal-half normal and normal-exponential models. In the spirit of Honoré & Powell (1994), the second is a U-estimator based on all pairwise quasi-likelihood contributions constructed exploiting the analytical expression available for the marginal likelihood function when $T = 2$. This strategy allow us to provide a computationally feasible approach to estimate normal-half normal, normal-exponential and normal-truncated normal models in which both error components can be heteroskedastic. Furthermore, for the normal-half normal and normal-truncated normal models, the inefficiency is also allowed to follow a first-order autoregressive 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. 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.

Almanidis, P., Qian, J., & Sickles, R. (2014). Stochastic frontier models with bounded inefficiency. In R. C. Sickles, & W. C. Horrace (Eds.), *Festschrift in Honor of Peter Schmidt* (pp. 47–81). Springer New York.

Almanidis, P., & Sickles, R. (2012). The skewness issue in stochastic frontiers models: Fact or fiction? In I. Van Keilegom, & P. W. Wilson (Eds.), *Exploring Research Frontiers in Contemporary Statistics and Econometrics* (pp. 201–227). Physica-Verlag HD.

Amsler, C., & Schmidt, P. (2015). Distinguishing heterogeneity and inefficiency in a panel data stochastic frontier model. *Mimeo*, .

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.

Colombi, R., Kumbhakar, S., Martini, G., & Vittadini, G. (2014). A stochastic frontier model with short-run and long-run inefficiency random effects. *Journal of Productivity Analysis*, (pp. 123–136).

Cornwell, C., & Schmidt, P. (1992). Models for which the mle and the conditional mle coincide. *Empirical Economics*, *17*, 67–75.

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.

Gouriéroux, C., & Monfort, A. (1991). Simulation based econometrics in models with heterogeneity. *Annales d'Economie et de Statistique*, *20/21*, 69–107.

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.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*, 1251–1271.

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.

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.

Newey, W., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle, & D. McFadden (Eds.), *Handbook of Econometrics* chapter 36. (pp. 2111–2245). volume 4.

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

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.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, *39*, 577–591.

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).

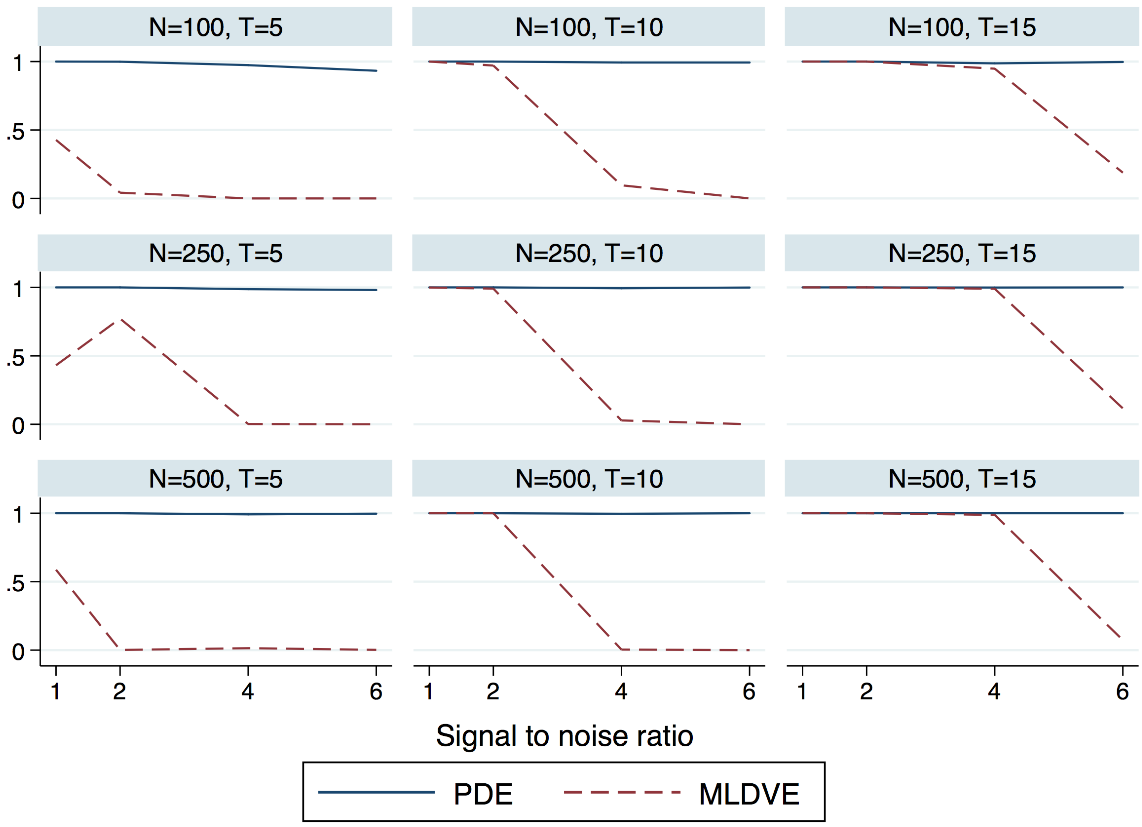

Table 1: Simulation results for PDE and MLDVE with $n = 100$, $T = 5$. The Spearman rank correlation coefficient is in parentheses.

(a) $\lambda = 1$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-0.001	3.5e-04	0.002	6.1e-04
γ_0	-0.075	0.066	0.296	0.174
γ_1	0.081	0.111	-0.408	0.287
ψ	0.002	7.5e-04	-0.164	0.037
E(u|ε)	-0.014	0.036	0.074	0.069
$r_{u,\hat{u}}$	0.805		0.745	
	(0.577)		(0.523)	

Non problematic (427)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-0.003	3.7e-04	-0.002	3.8e-04
γ_0	-0.124	0.073	-9.0e-05	0.045
γ_1	0.131	0.124	-0.083	0.091
ψ	0.007	6.9e-04	-0.049	0.003
E(u|ε)	-0.021	0.036	0.002	0.035
$r_{u,\hat{u}}$	0.804		0.811	
	(0.576)		(0.576)	

Problematic (573)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-3.4e-04	3.4e-04	0.004	7.9e-04
γ_0	-0.038	0.060	0.517	0.269
γ_1	0.043	0.102	-0.651	0.433
ψ	-0.002	7.9e-04	-0.250	0.062
E(u|ε)	-0.008	0.036	0.127	0.095
$r_{u,\hat{u}}$	0.806		0.696	
	(0.578)		(0.484)	

(b) $\lambda = 2$

All replications (10000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-5.0e-04	6.0e-04	0.042	0.004
γ_0	-0.036	0.019	0.152	0.026
γ_1	0.044	0.047	-0.430	0.195
ψ	0.004	0.002	-0.249	0.062
E(u|ε)	-0.014	0.088	0.039	0.103
$r_{u,\hat{u}}$	0.886		0.872	
	(0.699)		(0.668)	

Non problematic (293)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	0.002	5.9e-04	0.075	0.009
γ_0	-0.039	0.021	0.150	0.034
γ_1	0.052	0.050	-0.400	0.197
ψ	0.002	0.002	-0.229	0.056
E(u|ε)	-0.014	0.088	0.043	0.105
$r_{u,\hat{u}}$	0.885		0.870	
	(0.698)		(0.666)	

Problematic (9707)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-5.8e-04	6.0e-04	0.041	0.004
γ_0	-0.035	0.019	0.152	0.025
γ_1	0.044	0.047	-0.430	0.194
ψ	0.004	0.002	-0.250	0.062
E(u|ε)	-0.014	0.088	0.039	0.103
$r_{u,\hat{u}}$	0.886		0.872	
	(0.699)		(0.668)	

Table 2: Simulation results for PDE and MLDVE with $n = 100$, $T = 10$. The Spearman rank correlation coefficient is in parentheses.

(a) $\lambda = 1$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-5.5e-04	1.6e-04	-5.3e-04	1.5e-04
γ_0	-0.052	0.030	-1.3e-04	0.014
γ_1	0.058	0.053	-0.041	0.032
ψ	0.003	3.6e-04	-0.022	6.8e-04
E(u|ε)	-0.010	0.030	-4.3e-04	0.028
$r_{u,\hat{u}}$	0.838		0.846	
	(0.606)		(0.613)	

(b) $\lambda = 2$

All replications (10000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-9.8e-04	2.8e-04	-3.2e-04	2.5e-04
γ_0	-0.023	0.008	0.031	0.008
γ_1	0.027	0.019	-0.081	0.022
ψ	0.005	8.6e-04	-0.047	0.004
E(u|ε)	-0.009	0.062	0.009	0.051
$r_{u,\hat{u}}$	0.920		0.934	
	(0.747)		(0.766)	

Non problematic (9710)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-7.3e-04	2.8e-04	-4.9e-04	2.4e-04
γ_0	-0.023	0.008	0.022	0.005
γ_1	0.027	0.019	-0.070	0.016
ψ	0.005	8.6e-04	-0.041	0.002
E(u|ε)	-0.009	0.062	0.005	0.050
$r_{u,\hat{u}}$	0.920		0.935	
	(0.747)		(0.767)	

Problematic (290)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-0.009	3.6e-04	0.006	8.6e-04
γ_0	-0.022	0.008	0.353	0.126
γ_1	0.024	0.016	-0.462	0.218
ψ	0.006	9.0e-04	-0.250	0.062
E(u|ε)	-0.007	0.067	0.153	0.112
$r_{u,\hat{u}}$	0.919		0.899	
	(0.749)		(0.721)	

Table 3: Simulation results for PDE and MLDVE with $n = 250$, $T = 5$. The Spearman rank correlation coefficient is in parentheses.

(a) $\lambda = 1$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	3.8e-05	1.4e-04	0.008	6.5e-04
γ_0	-0.029	0.022	0.412	0.218
γ_1	0.029	0.042	-0.529	0.340
ψ	-0.001	3.2e-04	-0.198	0.047
E(u|ε)	-0.010	0.034	0.101	0.080
$r_{u,\hat{u}}$	0.815		0.731	
	(0.585)		(0.507)	

Non problematic (431)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	2.8e-04	1.3e-04	0.008	7.8e-04
γ_0	-0.047	0.024	0.265	0.143
γ_1	0.049	0.044	-0.359	0.210
ψ	6.9e-04	3.0e-04	-0.131	0.026
E(u|ε)	-0.013	0.034	0.063	0.060
$r_{u,\hat{u}}$	0.813		0.771	
	(0.585)		(0.541)	

Problematic (569)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-1.5e-04	1.4e-04	0.008	5.5e-04
γ_0	-0.015	0.021	0.522	0.274
γ_1	0.014	0.041	-0.659	0.438
ψ	-0.003	3.4e-04	-0.250	0.062
E(u|ε)	-0.007	0.034	0.129	0.095
$r_{u,\hat{u}}$	0.816		0.701	
	(0.586)		(0.482)	

(b) $\lambda = 2$

All replications (10000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	1.4e-04	2.8e-04	0.119	0.016
γ_0	-0.013	0.006	0.200	0.042
γ_1	0.017	0.016	-0.459	0.215
ψ	-3.9e-04	7.6e-04	-0.246	0.060
E(u|ε)	-0.009	0.085	0.063	0.116
$r_{u,\hat{u}}$	0.891		0.865	
	(0.704)		(0.655)	

Non problematic (7700)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	3.9e-04	2.7e-04	0.128	0.018
γ_0	-0.011	0.006	0.209	0.046
γ_1	0.015	0.016	-0.463	0.219
ψ	-5.9e-04	7.3e-04	-0.245	0.060
E(u|ε)	-0.009	0.085	0.067	0.118
$r_{u,\hat{u}}$	0.891		0.863	
	(0.704)		(0.652)	

Problematic (2300)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-7.0e-04	3.1e-04	0.089	0.010
γ_0	-0.018	0.007	0.169	0.030
γ_1	0.021	0.017	-0.445	0.202
ψ	2.4e-04	8.5e-04	-0.249	0.062
E(u|ε)	-0.010	0.084	0.049	0.108
$r_{u,\hat{u}}$	0.891		0.870	
	(0.703)		(0.663)	

Table 4: Simulation results for PDE and MLDVE with $n = 250$, $T = 10$. The Spearman rank correlation coefficient is in parentheses.

(a) $\lambda = 1$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-1.1e-04	5.8e-05	-8.5e-05	5.6e-05
γ_0	-0.016	0.010	0.016	0.006
γ_1	0.018	0.016	-0.056	0.014
ψ	2.2e-04	1.4e-04	-0.023	7.2e-04
E(u|ε)	-0.005	0.029	0.002	0.028
$\mathrm{r}_{u,\hat{u}}$	0.845		0.852	
	(0.614)		(0.617)	

(b) $\lambda = 2$

All replications (1000)

	PDE		MLDVE	
	Bias	MSE	Bias	MSE
β_1	-2.6e-04	1.1e-04	5.8e-04	1.6e-04
γ_0	-0.007	0.003	0.034	0.006
γ_1	0.009	0.007	-0.080	0.015
ψ	6.8e-04	3.1e-04	-0.047	0.003
E(u|ε)	-0.005	0.061	0.009	0.051
$\mathrm{r}_{u,\hat{u}}$	0.922		0.937	
	(0.750)		(0.769)	

Table 5: Simulation results for MMSLE and MMLE. The Spearman rank correlation coefficient is in parentheses.

(a) $\lambda = 1$

$n = 100$ $T = 5$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	-0.005	0.003	-0.002	0.002
σ	0.054	0.077	0.085	0.060
ψ	-0.042	0.011	-0.050	0.012
E(u|ε)	0.033	0.270	0.057	0.259
$r_{u,\hat{u}}$	0.481		0.479	
	(0.432)		(0.431)	

$n = 100$ $T = 10$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	-0.002	0.001	-6.5e-04	0.001
σ	0.010	0.032	0.014	0.030
ψ	-0.011	0.004	-0.014	0.004
E(u|ε)	0.004	0.229	0.007	0.227
$r_{u,\hat{u}}$	0.504		0.504	
	(0.455)		(0.454)	

$n = 250$ $T = 5$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	0.002	0.001	-1.7e-04	0.001
σ	0.009	0.037	0.035	0.036
ψ	-0.013	0.005	-0.021	0.005
E(u|ε)	-0.004	0.237	0.017	0.237
$r_{u,\hat{u}}$	0.477		0.477	
	(0.431)		(0.431)	

$n = 250$ $T = 10$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	5.0e-04	5.2e-04	0.001	5.1e-04
σ	-0.007	0.018	-0.004	0.019
ψ	-0.001	0.002	-0.003	0.002
E(u|ε)	-0.010	0.214	-0.007	0.214
$r_{u,\hat{u}}$	0.507		0.507	
	(0.457)		(0.457)	

(b) $\lambda = 2$

$n = 100$ $T = 5$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	-0.001	0.003	0.004	0.003
σ	-0.018	0.029	-0.008	0.028
ψ	-0.003	0.010	-0.012	0.012
E(u|ε)	-0.030	0.322	-0.022	0.323
$r_{u,\hat{u}}$	0.712		0.712	
	(0.649)		(0.649)	

$n = 100$ $T = 10$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	1.7e-04	0.001	7.0e-04	0.001
σ	-0.037	0.009	-0.006	0.008
ψ	0.023	0.003	-6.9e-04	0.003
E(u|ε)	-0.037	0.269	-0.012	0.269
$r_{u,\hat{u}}$	0.750		0.749	
	(0.690)		(0.689)	

$n = 250$ $T = 5$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	0.002	0.001	0.004	0.001
σ	-0.008	0.011	-0.007	0.011
ψ	-6.4e-04	0.004	-0.002	0.004
E(u|ε)	-0.024	0.305	-0.023	0.305
$r_{u,\hat{u}}$	0.711		0.711	
	(0.650)		(0.650)	

$n = 250$ $T = 10$

	MMSLE		MMLE	
	Bias	MSE	Bias	MSE
β	9.0e-04	5.0e-04	8.5e-04	5.0e-04
σ	-0.023	0.003	-0.001	0.003
ψ	0.015	0.001	-9.1e-04	0.001
E(u|ε)	-0.026	0.262	-0.008	0.261
$r_{u,\hat{u}}$	0.751		0.751	
	(0.691)		(0.691)	

Table 6: Simulation results for the dynamic PDE. The Spearman rank correlation coefficient is in parentheses.

(a) $\rho = 0.3$

$n = 100$ $T = 5$

	Bias	MSE
β	-0.001	0.002
γ_0	-0.065	0.054
γ_1	-0.020	0.012
ψ	-0.006	0.002
ρ	0.033	0.074
E(u|ε)	0.010	0.246
$r_{u,\hat{u}}$	0.953	
	(0.780)	

$n = 100$ $T = 10$

	Bias	MSE
β	-6.1e-04	7.4e-04
γ_0	0.001	0.018
γ_1	-0.015	0.005
ψ	-1.0e-03	6.4e-04
ρ	0.026	0.032
E(u|ε)	0.033	0.174
$r_{u,\hat{u}}$	0.969	
	(0.809)	

$n = 250$ $T = 5$

	Bias	MSE
β	-7.1e-04	6.8e-04
γ_0	-0.014	0.016
γ_1	-0.007	0.004
ψ	-0.002	6.4e-04
ρ	0.007	0.031
E(u|ε)	0.012	0.236
$r_{u,\hat{u}}$	0.957	
	(0.783)	

$n = 250$ $T = 10$

	Bias	MSE
β	-3.7e-04	2.5e-04
γ_0	0.007	0.007
γ_1	-0.009	0.002
ψ	0.001	2.4e-04
ρ	0.037	0.013
E(u|ε)	0.036	0.170
$r_{u,\hat{u}}$	0.970	
	(0.810)	

(b) $\rho = 0.7$

$n = 100$ $T = 5$

	Bias	MSE
β	-0.003	0.002
γ_0	-0.212	0.100
γ_1	-0.014	0.007
ψ	-0.006	0.002
ρ	-0.006	0.010
E(u|ε)	-0.118	0.515
$r_{u,\hat{u}}$	0.957	
	(0.800)	

$n = 100$ $T = 10$

	Bias	MSE
β	-0.003	9.0e-04
γ_0	-0.074	0.025
γ_1	-0.015	0.003
ψ	6.8e-04	7.9e-04
ρ	-0.022	0.004
E(u|ε)	-0.043	0.342
$r_{u,\hat{u}}$	0.973	
	(0.841)	

$n = 250$ $T = 5$

	Bias	MSE
β	-0.002	8.0e-04
γ_0	-0.185	0.053
γ_1	-0.010	0.003
ψ	-6.5e-04	7.5e-04
ρ	-0.009	0.004
E(u|ε)	-0.109	0.498
$r_{u,\hat{u}}$	0.959	
	(0.801)	

$n = 250$ $T = 10$

	Bias	MSE
β	-0.001	3.0e-04
γ_0	-0.077	0.013
γ_1	-0.013	0.001
ψ	0.002	3.0e-04
ρ	-0.017	0.002
E(u|ε)	-0.041	0.339
$r_{u,\hat{u}}$	0.974	
	(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	

A Appendix

A.1 The derivation of the marginal likelihood function when inefficiency is exponentially distributed

In order to derive the marginal likelihood in the general case ($T \geq 2$), we have to preliminary derive the distribution of $\Delta \boldsymbol{u}_i$. This multivariate distribution can be obtained from a marginalization of the linear combination $\boldsymbol{w}_i = A\boldsymbol{u}_i$, where the $T \times T$ matrix A is given by

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

The matrix A corresponds to a convenient full rank transformation, since its inverse $L = A^{-1}$ is the lower triangular matrix with all elements equal to one. By using the Jacobian transformation and by exploiting that $u_{it} \sim \mathcal{E}(\sigma)$, we can derive the p.d.f. of the vector $\boldsymbol{w}_i$ as

$$\begin{aligned} f_W(\boldsymbol{w_i}) &= f_U(\boldsymbol{u_i}) \times I(\boldsymbol{u_i} > \mathbf{0}) = \\ &= \frac{1}{\sigma^T} \exp\left(-\frac{1}{\sigma}\iota_T' A^{-1}\boldsymbol{w_i}\right) \times I(A^{-1}\boldsymbol{w_i} > \mathbf{0}) = \\ &= \frac{1}{\sigma^T} \exp\left(-\frac{1}{\sigma}\iota_T' L\boldsymbol{w_i}\right) \times I(L\boldsymbol{w_i} > \mathbf{0}) = \\ &= \frac{1}{\sigma^T} \exp\left(-\frac{1}{\sigma}\sum_{t=1}^{T}(T+1-t)w_{it}\right) \times I\left(\bigcap_{t=1}^{T}\left\{\sum_{j=1}^{t} w_{ij} > 0\right\}\right), \end{aligned} \tag{A.2}$$

where $I(.)$ is the indicator function and ι_T is a T-vector with elements all equal to one. Hence, the joint density of $\Delta \boldsymbol{u}_i = (\Delta u_{i2}, \dots, \Delta u_{iT})' = (w_{i2}, \dots, w_{iT})'$ is given by

$f_{\Delta U}(w_2,...,w_T) = f_{\Delta U}(\boldsymbol{w}_{-1})=$

$$
\begin{aligned}
&= \int_{\bigcap_{t=1}^{T}\left\{\sum_{j=1}^{t} w_{ij}>0\right\}} \frac{1}{\sigma^T}\exp\left(-\frac{1}{\sigma}\sum_{t=1}^{T}(T+1-t)w_{it}\right)dw_1 = \\
&= \int_{\bigcap_{t=1}^{T}\left\{\sum_{j=1}^{t} w_{ij}>0\right\}} \frac{1}{\sigma^T}\exp\left[-\frac{1}{\sigma}\left(Tw_{1t}+\sum_{t=2}^{T}(T+1-t)w_{it}\right)\right]dw_1 = \\
&= \int_{\bigcap_{t=1}^{T}\left\{\sum_{j=1}^{t} w_{ij}>0\right\}} \frac{1}{\sigma^T}\exp\left[-\frac{1}{\sigma}\left(Tw_{1t}+\iota_{T-1}'L_{T-1}\boldsymbol{w}_{-1}\right)\right]dw_1 = \\
&= \frac{1}{\sigma^{T-1}}\exp\left(-\frac{1}{\sigma}\iota_{T-1}'L_{T-1}\boldsymbol{w}_{-1}\right)\int_{\bigcap_{t=1}^{T}\left\{\sum_{j=1}^{t} w_{ij}>0\right\}} \frac{1}{\sigma}\exp\left(-\frac{1}{\sigma}Tw_{1t}\right)dw_1 = \\
&= \frac{1}{T\sigma^{T-1}}\exp\left(-\frac{1}{\sigma}\iota_{T-1}'L_{T-1}\boldsymbol{w}_{-1}\right)\exp\left(-\frac{T}{\sigma}\min\left\{0,w_2,w_2+w_3,\ldots,\sum_{t=2}^{T}w_t\right\}\right) = \\
&= \frac{1}{T\sigma^{T-1}}\exp\left[-\frac{1}{\sigma}\left(\iota_{T-1}'L_{T-1}\boldsymbol{w}_{-1}+T\min\left\{0,w_2,w_2+w_3,\ldots,\sum_{t=2}^{T}w_t\right\}\right)\right]. \qquad \text{(A.3)}
\end{aligned}
$$

Finally, the integration of $\Delta\boldsymbol{u}_i$ from the joint distribution $f(\Delta\boldsymbol{y}_i,\Delta\boldsymbol{u}_i|\boldsymbol{\theta},\Delta X_i,\Delta\boldsymbol{u}_i)$ produces the following marginal likelihood

$L(\Delta\boldsymbol{y}_i|\ \boldsymbol{\theta},\Delta\boldsymbol{x}_i) =$

$$
\begin{aligned}
&= \int f(\Delta\boldsymbol{y}_i,\Delta\boldsymbol{u}_i|\boldsymbol{\theta},\Delta X_i,\Delta\boldsymbol{u}_i)dF_{\Delta\boldsymbol{u}_i} = \\
&= \int \frac{1}{(2\pi)^{T/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\Delta\boldsymbol{y}_i-\Delta X_i\boldsymbol{\beta}\mp\Delta\boldsymbol{u}_i)'\Sigma^{-1}(\Delta\boldsymbol{y}_i-\Delta X_i\boldsymbol{\beta}\mp\Delta\boldsymbol{u}_i)\right\}dF_{\Delta\boldsymbol{u}_i} = \\
&= \int \frac{1}{(2\pi)^{T/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\Delta\boldsymbol{y}_i-\Delta X_i\boldsymbol{\beta}\mp\Delta\boldsymbol{u}_i)'\Sigma^{-1}(\Delta\boldsymbol{y}_i-\Delta X_i\boldsymbol{\beta}\mp\Delta\boldsymbol{u}_i)\right\}\times \\
&\qquad \times\frac{1}{T\sigma^{T-1}}\exp\left[-\frac{1}{\sigma}\left(\iota_{T-1}'L_{T-1}\Delta\boldsymbol{u}_i-T\min\left\{0,\Delta u_{i2},\Delta u_{i2}+\Delta u_{i3},\ldots,\sum_{t=2}^{T}\Delta u_{it}\right\}\right)\right]d\Delta\boldsymbol{u}_i = \\
&= \int \frac{1}{(2\pi)^{T/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(\Delta\boldsymbol{u}_i-\Delta\boldsymbol{\mu}_i)'\Sigma^{-1}(\Delta\boldsymbol{u}_i-\Delta\boldsymbol{\mu}_i)\right\}\frac{1}{T\sigma^{T-1}}\exp\left(-B_{\Delta\boldsymbol{u}_i}\Delta\boldsymbol{u}_i\right)d\Delta\boldsymbol{u}_i = \\
&= \frac{\exp\left\{-\frac{1}{2}(\Delta\boldsymbol{\mu}_i)'\Sigma^{-1}\Delta\boldsymbol{\mu}_i\right\}}{T(2\pi)^{(T-1)/2}\sigma^{T-1}|\Sigma|^{1/2}}\times \\
&\qquad \times\int\exp\left\{-\frac{1}{2}\left[(\Delta\boldsymbol{u}_i)'\Sigma^{-1}\Delta\boldsymbol{u}_i-2\left((\Delta\boldsymbol{\mu}_i)'\Sigma^{-1}-B_{\Delta\boldsymbol{u}_i}\right)\Delta\boldsymbol{u}_i\right]\right\}d\Delta\boldsymbol{u}_i. \qquad \text{(A.4)}
\end{aligned}
$$

where $\Delta\boldsymbol{\mu}_i = -(\Delta\boldsymbol{y}_i-\Delta X_i\beta)$, $B_{\Delta\boldsymbol{u}_i} = \iota_{T-1}'L_{T-1}-T\boldsymbol{J}_{\Delta\boldsymbol{u}_i}$ and $\boldsymbol{J}_{\Delta\boldsymbol{u}_i}$ a $T-1$-dimensional vector in which the first $(g-1)$ elements are equal to 1 and the $T-g$ remaining elements equal to zero, with g the position of the minimum value in the vector $(0,\Delta u_{i2},\Delta u_{i2}+\Delta u_{i3},\ldots,\sum_{t=2}^{T}\Delta u_{it})$.

Then, by using the following identity

$$-\frac{1}{2}\left(\boldsymbol{x}'\Omega^{-1}\boldsymbol{x} - 2\boldsymbol{\delta}'\boldsymbol{x}\right) = \frac{1}{2}\boldsymbol{\delta}'\Omega\boldsymbol{\delta} - \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\xi})'\Omega^{-1}(\boldsymbol{x}-\boldsymbol{\xi}),$$

where $\boldsymbol{\xi} = \Omega\boldsymbol{\delta}$, we can rewrite equation (A.4) as

$$\begin{aligned} L(\Delta\boldsymbol{y}_i|\ \boldsymbol{\theta}, \Delta X_i) = & \\ = & \frac{1}{T\sigma^{T-1}}\int_{\mathbb{R}^{T-1}} \exp\left\{-2\left(\Delta\boldsymbol{\mu}_i' B_{\Delta\boldsymbol{u}_i}\right)\right\} \times \\ & \times \frac{1}{(2\pi)^{(T-1)/2}|\Sigma|^{1/2}} \exp\left\{\frac{1}{2}\left[\left(\Delta\boldsymbol{u}_i - \boldsymbol{\zeta}_{\Delta\boldsymbol{u}_i}\right)'\Sigma^{-1}\left(\Delta\boldsymbol{u}_i - \boldsymbol{\zeta}_{\Delta\boldsymbol{u}_i}\right)\right]\right\} d\Delta\boldsymbol{u}_i, \qquad \text{(A.5)} \end{aligned}$$

where $\boldsymbol{\zeta}_{\Delta\boldsymbol{u}_i} = \Delta\boldsymbol{\mu}_i - \Sigma_i B_{\Delta\boldsymbol{u}_i}$. The second term in the integral (A.5) is not the p.d.f. of a multivariate normal distribution because the vector $B_{\Delta\boldsymbol{u}_i}$, as well as $\boldsymbol{\zeta}_{\Delta\boldsymbol{u}_i}$, assumes T different values $\tilde{B}_{S_1}, ..., \tilde{B}_{S_T}$ in $\mathbb{R}^{T-1}$. If we partition $\mathbb{R}^{T-1}$ into these T implicit subregions $S_1, \ldots, S_T$, we can rewrite (A.5) as

$$L^*(\boldsymbol{\theta}) = \prod_{i=1}^{n}\left\{\frac{1}{T\sigma^{T-1}}\sum_{k=1}^{T}\left[\exp\left(-2\Delta\boldsymbol{\mu}_i'\tilde{B}_{S_k}\right)\int_{S_k}\phi_{T-1}\left(\Delta\boldsymbol{u}_i; \tilde{\boldsymbol{\zeta}}_{S_k}, \Sigma_i\right)d\Delta\boldsymbol{u}_i\right]\right\}. \qquad \text{(A.6)}$$

The integral in (A.6) cannot be analytically evaluated if $T > 2$ due to the irregularity of the integration domains. As shown in Section 3.2, when $T = 2$ the two supports S_1 and S_2 are respectively the positive and the negative real line, and a closed-form expression for the marginal likelihood exists.

A.2 Consistency

In this appendix we establish the consistency of the PDE in the case of a correctly specified heteroskedastic normal-exponential model.[40]

Appendix A.2.1 gives and checks the conditions required for consistency of the subsamples MMLEs, while appendix A.2.2 shows how, for the case of the PDE, this property is naturally inherited from the subsamples MMLEs.

To make the asymptotic arguments formal, we distinguish between the true value $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$ a generic point in the parameter space Θ. $\hat{\boldsymbol{\theta}}$ indicates the MMLE for the subsample composed by a pair of generic waves of the panel, while $\tilde{\boldsymbol{\theta}}$ denotes the PDE.

[40]The homoskedastic case is available from the authors upon request.

A.2.1 Consistency of the MMLE

In what follows, we consider the following heteroskedastic normal-exponential model in first-differences for a two-period panel

$$\begin{aligned} \Delta y_i &= \Delta \boldsymbol{x}_i \boldsymbol{\beta} + \Delta \varepsilon_i, & \text{(A.7)} \\ \Delta \varepsilon_i &= \Delta v_i - \Delta u_i, & \text{(A.8)} \end{aligned}$$

where $v_{it} \sim \mathcal{N}(0, \psi^2)$ and $u_{it} \sim \mathcal{E}(\sigma_t)$ with $t = 1, 2$. Furthermore we denote with $Q_n(\boldsymbol{\theta})$ the corresponding marginal log-likelihood

$$Q_n(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta})$$

where

$$\begin{aligned} \ell_i(\boldsymbol{\theta}) = \log \Bigg\{ & \frac{1}{(\sigma_1 + \sigma_2)} \exp\left(\frac{\psi^2}{\sigma_2^2} - \frac{\Delta \varepsilon_{it}}{\sigma_2}\right) \times \\ & \left[\Phi\left(\frac{\Delta \varepsilon_{it}}{\sqrt{2}\psi} - \frac{\sqrt{2}\psi}{\sigma_2}\right) + \exp\left[\psi^2 \left(\frac{\sigma_2^2 - \sigma_1^2}{\sigma_1^2 \sigma_2^2}\right) + \left(\frac{\sigma_1 + \sigma_2}{\sigma_1 \sigma_2}\right) \Delta \varepsilon_{it}\right] \Phi\left(-\frac{\Delta \varepsilon_{it}}{\sqrt{2}\psi} - \frac{\sqrt{2}\psi}{\sigma_1}\right)\right] \Bigg\}. \end{aligned}$$

Given this setup, the following Theorem states the assumptions required for the MMLE to be consistent.

Theorem 1 *In a correctly specified heteroskedastic normal-exponential model, if (i) $\boldsymbol{\theta}_0$ is an interior point of a compact set Θ; (ii) $\mathbb{E}\Delta X' \Delta X = P$, where ΔX, the $n \times k$ matrix of time-varying covariates, is a finite p.d. matrix; (iii) $\mathbb{E}\left[\ell_i(\boldsymbol{\theta})^2\right] < \infty$; then $\hat{\boldsymbol{\theta}}$ exists and is unique with probability approaching one as $n \to \infty$ and $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$.*

An important remark is that the assumptions in Theorem 1 allow for general types of heteroskedasticity as the one given, for example, in equation (18). In order to have a more compact notation but without loss of generality, we are assuming here a special form of heteroskedasticity in which the variance of the inefficiency is fixed across units but varies over time.[41] However, the results of Theorem 1 can be generalized at the expense of more complex notation to go beyond this special form, provided that we extend assumptions such as (iii) to allow for a more general pattern of heteroskedasticity.

Proof: By Newey & McFadden (1994), for consistency it is sufficient to verify the following conditions:

1. $\boldsymbol{\theta}_0$ is an interior point of a compact set Θ;

[41] Using the same notation of equation (18), the parameters σ_t are in one-to-one correspondence with γ_t with $t = 1, 2$ since $\sigma_t = \exp(\boldsymbol{z}_{it}\boldsymbol{\gamma})$ where $\boldsymbol{z}_{it} = (d_{i1}, d_{i2})$, with $d_{i1} = 1(t = 1)$ and $d_{i2} = 1(t = 2)$. Notice that the homoskedastic case can be easily obtained by substituting $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

2. $Q(\boldsymbol{\theta})$ attains a unique maximum over the compact set Θ at $\boldsymbol{\theta}_0$;

3. the sequence of random functions $\{Q_n(\boldsymbol{\theta})\}$ converges in probability uniformly on Θ to a continuous function $Q(\boldsymbol{\theta})$.

We have already assumed condition (i). It is a standard assumption for optimization estimators and can be considered quite innocuous here since the vector of incidental parameters $\boldsymbol{\alpha}$ is excluded from $\boldsymbol{\theta}$.

The second condition is the identification condition for maximum likelihood estimation. In our case, identification can be proved following two steps. The first considers the frontier parameters $\boldsymbol{\beta}$ while the second is related to the error components variances σ_1, σ_2 and ψ^2. As in Almanidis et al. (2014), the identification can be proved following Theorem 4 of Rothenberg (1971).

As for the first step, model (A.7-A.8) can be considered as a standard linear panel data model, thus assumption (ii) of Theorem 1 ensures that the OLS applied on first-differenced data is a consistent estimator of $\boldsymbol{\beta}$.

The identification of the variance parameters comes from the fact that they can be expressed as a function of population moments. Under the distributional assumptions made for the error terms, the moment generating function (*mgf*) of the convolution $\Delta\boldsymbol{\varepsilon}$ is the product of the *mgf* of its Gaussian and asymmetric Laplace components. Then, we have that

$$M_{\Delta\boldsymbol{\varepsilon}}(s) = \frac{\exp\left(\frac{\psi^2 s^2}{2}\right)}{\sigma_1\sigma_2(\sigma_1^{-1} - s)(\sigma_2^{-1} + s)}. \tag{A.9}$$

By exploiting (A.9), it is easy shown that the first three population central moments of $\Delta\boldsymbol{\varepsilon}$ are given by

$$\begin{aligned} m_1 &= \sigma_1 - \sigma_2 && \text{(A.10)} \\ m_2 &= 2\varphi^2 + 2\sigma_1^2 - 2\sigma_1\sigma_2 + 2\sigma_2^2 && \text{(A.11)} \\ m_3 &= 6\sigma_1\varphi^2 - 6\sigma_2\varphi^2 + 6\sigma_1^3 - 6\sigma_1^2\sigma_2 + 6\sigma_1\sigma_2^2 - 6\sigma_2^3 && \text{(A.12)} \end{aligned}$$

Then, we have that

$$\begin{aligned} 3m_1m_2 &= 3(\sigma_1 - \sigma_2)(2\varphi^2 + 2\sigma_1^2 - 2\sigma_1\sigma_2 + 2\sigma_2^2) \\ &= 6\varphi^2\sigma_1 - 6\varphi^2\sigma_2 + 6\sigma_1^3 - 6\sigma_1^2\sigma_2 - 6\sigma_1^2\sigma_2 + 6\sigma_1\sigma_2^2 + 6\sigma_1\sigma_2^2 - 6\sigma_2^3 \\ &= 6\varphi^2\sigma_1 - 6\varphi^2\sigma_2 + 6\sigma_1^3 - 12\sigma_1^2\sigma_2 + 12\sigma_1\sigma_2^2 - 6\sigma_2^3 && \text{(A.13)} \end{aligned}$$

and that

$$m_3 - 3m_1m_2 = 6\sigma_1^2\sigma_2 - 6\sigma_1\sigma_2^2. \tag{A.14}$$

From equation (A.14), we can write

$$\frac{m_3 - 3m_1m_2}{6} = \sigma_1^2\sigma_2 - \sigma_1\sigma_2^2 = (\sigma_1 - \sigma_2)\sigma_1\sigma_2. \tag{A.15}$$

Then, since $m_1 = \sigma_1 - \sigma_2$ we have that

$$\begin{aligned} \frac{m_3 - 3m_1m_2}{6m_1} &= \sigma_1\sigma_2 && \text{(A.16)} \\ \sigma_2^2 + m_1\sigma_2 - \frac{m_3 - 3m_1m_2}{6m_1} &= 0. && \text{(A.17)} \end{aligned}$$

By denoting with $c = \frac{m_3 - 3m_1m_2}{6m_1}$, we can write

$$\sigma_2^2 + m_1\sigma_2 - c = 0 \tag{A.18}$$

The admitted root of (A.18) is then

$$\sigma_2 = -\frac{m_1}{2} + \sqrt{m_1^2/4 + c}. \tag{A.19}$$

By substituting (A.19) in (A.10) we get

$$\sigma_1 = \frac{m_1}{2} + \sqrt{m_1^2/4 + c}. \tag{A.20}$$

Finally, to find and expression for φ^2, we can substitute (A.20) and (A.19) in (A.11)

$$\begin{aligned} \varphi^2 &= \frac{m_2}{2} - m_1^2 - c = m_2 - m_1^2 - \frac{m_3}{6m_1} \\ &= Var - \frac{m_3}{6m_1}. && \text{(A.21)} \end{aligned}$$

Equations (A.19), (A.20) and (A.21) define a new method of moments estimator for the variance parameters of model (A.7)-(A.8). It then follows that the parameters of model (A.7)-(A.8) are *globally* identified.

The only other primitive conditions needed for consistency are those for continuity and uniform convergence in probability of the limiting objective function. As for the former, it can be easily checked by inspection since $\ell_i(\boldsymbol{\theta})$ is composed by continuous functions (logarithm, exponential and standard normal c.d.f.) on the same open set. As far as the uniform convergence in probability is concerned, since we are under the i.n.i.d. case, the Markov ULLN can be exploited. Under the assumptions in Theorem 1, especially assumption (iii), and continuity of the limiting objective function, we only need to check that $|\ell_i(\boldsymbol{w}_i, \boldsymbol{\theta})| < c(\boldsymbol{w}_i)$, $\forall \boldsymbol{\theta} \in \Theta$ with

$\mathbb{E}[c(\boldsymbol{w}_i)] < \infty$ and $\boldsymbol{w}_i = (\Delta y_{it}, \Delta \boldsymbol{x}_{it})$. To this aim, rewrite the marginal log-likelihood function as

$$
\begin{aligned}
& \left|\log\left\{\frac{1}{(\sigma_1+\sigma_2)}\exp\left(\frac{\psi^2}{\sigma_2^2}-\frac{\Delta\varepsilon_{it}}{\sigma_2}\right)\times\right.\right. \\
& \left.\left.\left[\Phi\left(\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}-\frac{\sqrt{2}\psi}{\sigma_2}\right)+\exp\left[\psi^2\left(\frac{\sigma_2^2-\sigma_1^2}{\sigma_1^2\sigma_2^2}\right)+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)\Delta\varepsilon_{it}\right]\Phi\left(-\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}-\frac{\sqrt{2}\psi}{\sigma_1}\right)\right]\right\}\right| \qquad (A.22) \\
< \; & |\log(\sigma_1+\sigma_2)| + \left|\frac{\psi^2}{\sigma_2^2}-\frac{\Delta\varepsilon_{it}}{\sigma_2}\right| + \\
& \left|\log\left[\Phi\left(\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}-\frac{\sqrt{2}\psi}{\sigma_2}\right)+\exp\left[\psi^2\left(\frac{\sigma_2^2-\sigma_1^2}{\sigma_1^2\sigma_2^2}\right)+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)\Delta\varepsilon_{it}\right]\Phi\left(-\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}-\frac{\sqrt{2}\psi}{\sigma_1}\right)\right]\right|. \qquad (A.23)
\end{aligned}
$$

We then need to show that each term of expression (A.23) has a data-dependent upper bound with finite expectation. In what follow, we use (A.11) and the following result

$$
\begin{aligned}
& E\,|\Delta\varepsilon_{it}| = E\,|v_2 - v_1 + u_1 - u_2| \\
< \; & E|v_2| + E|v_1| + E|u_1| + E|u_2| = \psi\sqrt{\frac{8}{\pi}} + \sigma_1 + \sigma_2 \\
< \; & 2\psi + \sigma_1 + \sigma_2. \qquad (A.24)
\end{aligned}
$$

Since Θ is a compact set, $\sigma_1, \sigma_2 > 0$, then there exists two finite constants c_1 and c_2 such that $sup_{\boldsymbol{\theta}\in\Theta}|\log(\sigma_1+\sigma_2)| < c_1$ and

$$
\begin{aligned}
& E\left(sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{\psi^2}{\sigma_2^2}-\frac{\Delta\varepsilon_{it}}{\sigma_2}\right|\right) \\
< \; & sup_{\boldsymbol{\theta}\in\Theta}\left[\frac{\psi^2}{\sigma_2^2}+E\left|\frac{\Delta\varepsilon_{it}}{\sigma_2}\right|\right] \\
< \; & sup_{\boldsymbol{\theta}\in\Theta}\left[\frac{\psi^2}{\sigma_2^2}+\frac{2\psi+\sigma_1+\sigma_2}{\sigma_2}\right] < c_2. \qquad (A.25)
\end{aligned}
$$

As for the third term of expression (A.23), we need to show that

$$
\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\log[g(\boldsymbol{\theta})]| < \infty, \qquad (A.26)
$$

where

$$
g(\boldsymbol{\theta}) = \Phi\left(\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}-\frac{\sqrt{2}\psi}{\sigma_2}\right)+\exp\left[\psi^2\left(\frac{\sigma_2^2-\sigma_1^2}{\sigma_1^2\sigma_2^2}\right)+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)\Delta\varepsilon_{it}\right]\Phi\left(-\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}-\frac{\sqrt{2}\psi}{\sigma_1}\right).
$$

The modulus of a logarithmic function diverges to $+\infty$ as the argument goes to zero or to $+\infty$. Moreover the modulus of a logarithmic function is monotonically decreasing in the interval

$(0, 1]$ and monotonically increasing in $[1, +\infty)$. Therefore in order to prove (A.26) $\forall \boldsymbol{\theta} \in \Theta$ is sufficient to prove that

$$\mathbb{E} \left|\log \left[g_1(\boldsymbol{\theta})\right]\right| < +\infty \quad \text{and } \mathbb{E} \left|\log \left[g_2(\boldsymbol{\theta})\right]\right| < +\infty \quad \forall \boldsymbol{\theta} \in \Theta, \tag{A.27}$$

with

$$g_1(\boldsymbol{\theta}) < g(\boldsymbol{\theta}) < g_2(\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta} \in \Theta. \tag{A.28}$$

Since $\Phi(.) \in (0, 1)$, we can define $g_1(\boldsymbol{\theta})$ and $g_2(\boldsymbol{\theta})$ as

$$g_1(\boldsymbol{\theta}) = \Phi\left(\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi} - \frac{\sqrt{2}\psi}{\sigma_2}\right),$$

and

$$g_2(\boldsymbol{\theta}) = 1 + \exp\left[\psi^2\left(\frac{\sigma_2^2 - \sigma_1^2}{\sigma_1^2\sigma_2^2}\right) + \left(\frac{\sigma_1 + \sigma_2}{\sigma_1\sigma_2}\right)\Delta\varepsilon_{it}\right].$$

As for $g_1(\boldsymbol{\theta})$, it is well known that the derivative $\partial \log \Phi(z)/\partial z = \phi(z)/\Phi(z)$. By denoting $z = \left(\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi} - \frac{\sqrt{2}\psi}{\sigma_2}\right)$, a mean-value expansion around $\psi = \sigma_2 = 0$ gives

$$\begin{aligned}
|\log \Phi(z)| &= |\log \Phi(0)| + \frac{\phi(z')}{\Phi(z')}|z|, \qquad \forall \boldsymbol{\theta} \in \Theta, \quad z' = \delta z \text{with} \delta \in (0,1) \\
&\leq |\log \Phi(0)| + c_3\left(1 + \delta\left|\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi} - \frac{\sqrt{2}\psi}{\sigma_2}\right|\right)\left|\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi} - \frac{\sqrt{2}\psi}{\sigma_2}\right| \\
&\leq |\log \Phi(0)| + c_3\left(1 + \delta\left|\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}\right|\right)\left|\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}\right| \\
&\leq |\log \Phi(0)| + c_3\left[\left|\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}\right| + \left(\frac{\Delta\varepsilon_{it}}{\sqrt{2}\psi}\right)^2\right],
\end{aligned} \tag{A.29}$$

with c_3 a finite constant. Thus, since Θ is a compact set, we have that

$$\begin{aligned}
\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\log \Phi(z)| &\leq |\log \Phi(0)| + c_3\left[\frac{1}{\sqrt{2}\psi}\mathbb{E}\left|\Delta\varepsilon_{it}\right| + \frac{1}{2\psi^2}\mathbb{E}\Delta\varepsilon_{it}^2\right] \\
&\leq |\log \Phi(0)| + c_3\left[\frac{2\psi + \sigma_1 + \sigma_2}{\sqrt{2}\psi} + \frac{2\psi^2 + 2\sigma_1^2 + 2\sigma_2^2 - 2\sigma_1\sigma_2}{2\psi^2}\right] \\
&\leq |\log \Phi(0)| + c_3\left[c_4 + c_5\right] < +\infty,
\end{aligned} \tag{A.30}$$

where c_4 and c_5 are two finite constants such that $c_4 > \frac{2\psi+\sigma_1+\sigma_2}{\sqrt{2}\psi}$ and $c_5 > \frac{2\psi^2+2\sigma_1^2+2\sigma_2^2-2\sigma_1\sigma_2}{2\psi^2}$.

As for $g_2(\boldsymbol{\theta})$, given that $\log(1+\exp(z)) < 1+|z| \ \forall z \in \mathbb{R}$, we have that

$$\begin{aligned}
|\log[g_2(\boldsymbol{\theta})]| &= \log\left[1+\exp\left[\psi^2\left(\frac{\sigma_2^2-\sigma_1^2}{\sigma_1^2\sigma_2^2}\right)+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)\Delta\varepsilon_{it}\right]\right] \\
&< 1+\left|\psi^2\left(\frac{\sigma_2^2-\sigma_1^2}{\sigma_1^2\sigma_2^2}\right)+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)\Delta\varepsilon_{it}\right| \\
&< 1+\left|\psi^2\left(\frac{\sigma_2^2-\sigma_1^2}{\sigma_1^2\sigma_2^2}\right)\right|+\left|\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)\Delta\varepsilon_{it}\right| \\
&< 1+\frac{\psi^2}{\sigma_1^2}+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)|\Delta\varepsilon_{it}|, \ \forall\boldsymbol{\theta}\in\Theta. \qquad \text{(A.31)}
\end{aligned}$$

Therefore we have that

$$\begin{aligned}
\mathbb{E}\,|\log[g_2(\boldsymbol{\theta})]| &< \mathbb{E}\left[1+\frac{\psi^2}{\sigma_1^2}+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)|\Delta\varepsilon_{it}|\right] \\
&= 1+\frac{\psi^2}{\sigma_1^2}+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)\mathbb{E}\,|\Delta\varepsilon_{it}| \\
&< 1+\frac{\psi^2}{\sigma_1^2}+\left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)(2\psi+\sigma_1+\sigma_2), \ \forall\boldsymbol{\theta}\in\Theta. \qquad \text{(A.32)}
\end{aligned}$$

Again since Θ is a compact set, $\sigma_1, \sigma_2, \psi \in [0,+\infty)$, then there exist two finite constants c_6 and c_7 such that

$$c_6 > \frac{\psi^2}{\sigma_1^2} \text{ and } c_7 > \left(\frac{\sigma_1+\sigma_2}{\sigma_1\sigma_2}\right)(2\psi+\sigma_1+\sigma_2),$$

thus

$$\mathbb{E}\sup_{\boldsymbol{\theta}\in\Theta}|\log[g_2(\boldsymbol{\theta})]| < 1+c_6+c_7 < +\infty.$$

A.2.2 Consistency of the PDE

Theorem 2 *If (i) $\boldsymbol{\theta}_0$ is an interior point of a compact set Θ and (ii) $\hat{\boldsymbol{\theta}}_n^{(t,s)} \xrightarrow{p} \boldsymbol{\theta}_0$, $\forall(t,s)$ with $t=1,...,T-1$, $s=2,...,T$ and $t<s$ then $\tilde{\boldsymbol{\theta}}$ exists and is unique with probability approaching one as $n\to\infty$ and $\tilde{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$.*

This theorem states that a sufficient condition for the consistency of the PDE is the consistency of the subsamples MMLEs.

Proof:
Since we have already assumed condition (i), the proof of Theorem 2 requires to check the validity of the following two conditions:

1. the sequence of random functions $\{U_n(\boldsymbol{\theta})\}$ converges in probability uniformly on Θ to a continuous function $U(\boldsymbol{\theta})$;

2. U attains a unique maximum on Θ at the true value $\boldsymbol{\theta}_0$.

The marginal log-likelihood for the subsample obtained considering two generic waves t and s with $t < s$ is given by

$$Q_n^{(t,s)}(\boldsymbol{\theta}) = n^{-1} \sum_{i=1}^{n} \ell_i^{(t,s)}(\boldsymbol{\theta}),$$

then the objective function of the PDE can be expressed as

$$U_n(\boldsymbol{\theta}) = \binom{T}{2}^{-1} \sum_{t=1}^{T} \sum_{s<t} Q_n^{(t,s)}(\boldsymbol{\theta}).$$

As already shown in Section A.2.1, a necessary condition for the consistency of $\hat{\boldsymbol{\theta}}_n^{(t,s)}$ is that $Q_n^{(t,s)}(\boldsymbol{\theta})$ converges in probability uniformly on Θ to the following continuous function

$$Q^{(t,s)}(\boldsymbol{\theta}) = \mathbb{E}[\ell_i^{(t,s)}(\boldsymbol{\theta})].$$

Given that $\hat{\boldsymbol{\theta}}_n^{(t,s)} \xrightarrow{p} \boldsymbol{\theta}_0$, this convergence condition clearly holds. By definition, $U(\boldsymbol{\theta})$ can be rewritten as

$$\begin{aligned} U(\boldsymbol{\theta}) &= \mathbb{E}\left[\binom{T}{2}^{-1} \sum_{t=1}^{T} \sum_{s<t} \ell_i^{(t,s)}(\boldsymbol{\theta})\right] \\ &= \binom{T}{2}^{-1} \sum_{t=1}^{T} \sum_{s<t} \mathbb{E}[\ell_i^{(t,s)}(\boldsymbol{\theta})] \\ &= \binom{T}{2}^{-1} \sum_{t=1}^{T} \sum_{s<t} Q^{(t,s)}(\boldsymbol{\theta}). \end{aligned} \tag{A.33}$$

Since the expectation is a linear operator and each term $Q_n^{(t,s)}(\boldsymbol{\theta})$ in $U_n(\boldsymbol{\theta})$ converges in probability (uniformly on Θ) to its population counterpart $Q^{(t,s)}(\boldsymbol{\theta})$, $U_n(\boldsymbol{\theta})$ converges in probability (uniformly on Θ) to $U(\boldsymbol{\theta})$.

The second condition requires that $\boldsymbol{\theta}_0$ is the unique maximizer of $U(\boldsymbol{\theta})$. Since $U(\boldsymbol{\theta})$ is an average of all $Q^{(t,s)}(\boldsymbol{\theta})$, then it is sufficient to show that $\boldsymbol{\theta}_0$ is the unique global maximizer of each $Q^{(t,s)}(\boldsymbol{\theta})$. The consistency of the MMLEs implies that $\boldsymbol{\theta}_0$ uniquely maximizes each $Q^{(t,s)}(\boldsymbol{\theta})$, which completes the proof. □

RECENTLY PUBLISHED "TEMI" (*)

N. 1121 – *The financial stability dark side of monetary policy*, by Piergiorgio Alessandri, Antonio Maria Conti and Fabrizio Venditti (June 2017).

N. 1122 – *Large time-varying parameter VARs: a non-parametric approach*, by George Kapetanios, Massimiliano Marcellino and Fabrizio Venditti (June 2017).

N. 1123 – *Multiple lending, credit lines, and financial contagion*, by Giuseppe Cappelletti and Paolo Emilio Mistrulli (June 2017).

N. 1124 – *Law enforcement and political participation: Italy, 1861-65*, by Antonio Accetturo, Matteo Bugamelli and Andrea Lamorgese (July 2017).

N. 1125 – *The consequences of public employment: evidence from Italian municipalities*, by Marta Auricchio, Emanuele Ciani, Alberto Dalmazzo and Guido de Blasio (July 2017).

N. 1126 – *The cyclicality of the income elasticity of trade*, by Alessandro Borin, Virginia Di Nino, Michele Mancini and Massimo Sbracia (July 2017).

N. 1127 – *Human capital and urban growth in Italy, 1981-2001*, by Francesco Giffoni, Matteo Gomellini and Dario Pellegrino (July 2017).

N. 1128 – *The double bind of asymmetric information in over-the-counter markets*, by Taneli Mäkinen and Francesco Palazzo (July 2017).

N. 1129 – *The effects of central bank's verbal guidance: evidence from the ECB*, by Maddalena Galardo and Cinzia Guerrieri (July 2017).

N. 1130 – *The Bank of Italy econometric model: an update of the main equations and model elasticities*, by Guido Bulligan, Fabio Busetti, Michele Caivano, Pietro Cova, Davide Fantino, Alberto Locarno, Lisa Rodano (July 2017).

N. 1131 – *Venture capitalists at work: what are the effects on the firms they finance?*, by Raffaello Bronzini, Giampaolo Caramellino and Silvia Magri (September 2017).

N. 1132 – *Low frequency drivers of the real interest rate: a band spectrum regression approach*, by Fabio Busetti and Michele Caivano (September 2017).

N. 1133 – *The real effects of relationship lending*, by Ryan Banerjee, Leonardo Gambacorta and Enrico Sette (September 2017).

N. 1134 – *Credit demand and supply: a two-way feedback relation*, by Ugo Albertazzi and Lucia Esposito (September 2017).

N. 1135 – *Legislators' behaviour and electoral rules: evidence from an Italian reform*, by Giuseppe Albanese, Marika Cioffi and Pietro Tommasino (September 2017).

N. 1136 – *Macroeconomic effects of non-standard monetary policy measures in the euro area: the role of corporate bond purchases*, by Anna Bartocci, Lorenzo Burlon, Alessandro Notarpietro and Massimiliano Pisani (September 2017).

N. 1137 – *On secular stagnation and low interest rates: demography matters*, by Giuseppe Ferrero, Marco Gross and Stefano Neri (September 2017).

N. 1138 – *Tony Atkinson and his legacy*, by Rolf Aaberge, François Bourguignon, Andrea Brandolini, Francisco H. G. Ferreira, Janet C. Gornick, John Hills, Markus Jäntti, Stephen P. Jenkins, Eric Marlier, John Micklewright, Brian Nolan, Thomas Piketty, Walter J. Radermacher, Timothy M. Smeeding, Nicholas H. Stern, Joseph Stiglitz and Holly Sutherland (September 2017).

N. 1139 – *Credit misallocation during the European financial crisis*, by Fabiano Schivardi, Enrico Sette and Guido Tabellini (September 2017).

N. 1140 – *Natural rates across the Atlantic*, by Andrea Gerali and Stefano Neri (September 2017).

(*) Requests for copies should be sent to:
Banca d'Italia – Servizio Studi di struttura economica e finanziaria – Divisione Biblioteca e Archivio storico – Via Nazionale, 91 – 00184 Rome – (fax 0039 06 47922059). They are available on the Internet www.bancaditalia.it.

2015

AABERGE R. and A. BRANDOLINI, *Multidimensional poverty and inequality,* in A. B. Atkinson and F. Bourguignon (eds.), Handbook of Income Distribution, Volume 2A, Amsterdam, Elsevier, **TD No. 976 (October 2014).**

ALBERTAZZI U., G. ERAMO, L. GAMBACORTA and C. SALLEO, *Asymmetric information in securitization: an empirical assessment*, Journal of Monetary Economics, v. 71, pp. 33-49, **TD No. 796 (February 2011).**

ALESSANDRI P. and B. NELSON, *Simple banking: profitability and the yield curve,* Journal of Money, Credit and Banking, v. 47, 1, pp. 143-175, **TD No. 945 (January 2014).**

ANTONIETTI R., R. BRONZINI and G. CAINELLI, *Inward greenfield FDI and innovation,* Economia e Politica Industriale, v. 42, 1, pp. 93-116, **TD No. 1006 (March 2015).**

BARONE G. and G. NARCISO, *Organized crime and business subsidies: Where does the money go?,* Journal of Urban Economics, v. 86, pp. 98-110, **TD No. 916 (June 2013).**

BRONZINI R., *The effects of extensive and intensive margins of FDI on domestic employment: microeconomic evidence from Italy*, B.E. Journal of Economic Analysis & Policy, v. 15, 4, pp. 2079-2109, **TD No. 769 (July 2010).**

BUGAMELLI M., S. FABIANI and E. SETTE, *The age of the dragon: the effect of imports from China on firm-level prices*, Journal of Money, Credit and Banking, v. 47, 6, pp. 1091-1118, **TD No. 737 (January 2010).**

BULLIGAN G., M. MARCELLINO and F. VENDITTI, *Forecasting economic activity with targeted predictors,* International Journal of Forecasting, v. 31, 1, pp. 188-206, **TD No. 847 (February 2012).**

BUSETTI F., *On detecting end-of-sample instabilities,* in S.J. Koopman, N. Shepard (eds.), Unobserved Components and Time Series Econometrics, Oxford, Oxford University Press, **TD No. 881 (September 2012).**

CESARONI T., *Procyclicality of credit rating systems: how to manage it,* Journal of Economics and Business, v. 82. pp. 62-83, **TD No. 1034 (October 2015).**

CIARLONE A., *House price cycles in emerging economies,* Studies in Economics and Finance, v. 32, 1, **TD No. 863 (May 2012).**

CUCINIELLO V. and F. M. SIGNORETTI, *Large bank, loan rate markup and monetary policy,* International Journal of Central Banking, v. 11, 3, pp. 141-177, **TD No. 987 (November 2014).**

DE BLASIO G., D. FANTINO and G. PELLEGRINI, *Evaluating the impact of innovation incentives: evidence from an unexpected shortage of funds,* Industrial and Corporate Change, v. 24, 6, pp. 1285-1314, **TD No. 792 (February 2011).**

DEPALO D., R. GIORDANO and E. PAPAPETROU, *Public-private wage differentials in euro area countries: evidence from quantile decomposition analysis,* Empirical Economics, v. 49, 3, pp. 985-1115, **TD No. 907 (April 2013).**

DI CESARE A., A. P. STORK and C. DE VRIES, *Risk measures for autocorrelated hedge fund returns,* Journal of Financial Econometrics, v. 13, 4, pp. 868-895, **TD No. 831 (October 2011).**

FANTINO D., A. MORI and D. SCALISE, *Collaboration between firms and universities in Italy: the role of a firm's proximity to top-rated departments,* Rivista Italiana degli economisti, v. 1, 2, pp. 219-251, **TD No. 884 (October 2012).**

FRATZSCHER M., D. RIMEC, L. SARNOB and G. ZINNA, *The scapegoat theory of exchange rates: the first tests,* Journal of Monetary Economics, v. 70, 1, pp. 1-21, **TD No. 991 (November 2014).**

NOTARPIETRO A. and S. SIVIERO, *Optimal monetary policy rules and house prices: the role of financial frictions,* Journal of Money, Credit and Banking, v. 47, S1, pp. 383-410, **TD No. 993 (November 2014).**

RIGGI M. and F. VENDITTI, *The time varying effect of oil price shocks on euro-area exports,* Journal of Economic Dynamics and Control, v. 59, pp. 75-94, **TD No. 1035 (October 2015).**

TANELI M. and B. OHL, *Information acquisition and learning from prices over the business cycle,* Journal of Economic Theory, 158 B, pp. 585–633, **TD No. 946 (January 2014).**

2016

ALBANESE G., G. DE BLASIO and P. SESTITO, *My parents taught me. evidence on the family transmission of values,* Journal of Population Economics, v. 29, 2, pp. 571-592, **TD No. 955 (March 2014).**

ANDINI M. and G. DE BLASIO, *Local development that money cannot buy: Italy's Contratti di Programma,* Journal of Economic Geography, v. 16, 2, pp. 365-393, **TD No. 915 (June 2013).**

BARONE G. and S. MOCETTI, *Inequality and trust: new evidence from panel data,* Economic Inquiry, v. 54, pp. 794-809, **TD No. 973 (October 2014).**

BELTRATTI A., B. BORTOLOTTI and M. CACCAVAIO, *Stock market efficiency in China: evidence from the split-share reform,* Quarterly Review of Economics and Finance, v. 60, pp. 125-137, **TD No. 969 (October 2014).**

BOLATTO S. and M. SBRACIA, *Deconstructing the gains from trade: selection of industries vs reallocation of workers,* Review of International Economics, v. 24, 2, pp. 344-363, **TD No. 1037 (November 2015).**

BOLTON P., X. FREIXAS, L. GAMBACORTA and P. E. MISTRULLI, *Relationship and transaction lending in a crisis*, Review of Financial Studies, v. 29, 10, pp. 2643-2676, **TD No. 917 (July 2013).**

BONACCORSI DI PATTI E. and E. SETTE, *Did the securitization market freeze affect bank lending during the financial crisis? Evidence from a credit register,* Journal of Financial Intermediation , v. 25, 1, pp. 54-76, **TD No. 848 (February 2012).**

BORIN A. and M. MANCINI, *Foreign direct investment and firm performance: an empirical analysis of Italian firms,* Review of World Economics, v. 152, 4, pp. 705-732, **TD No. 1011 (June 2015).**

BRAGOLI D., M. RIGON and F. ZANETTI, *Optimal inflation weights in the euro area,* International Journal of Central Banking, v. 12, 2, pp. 357-383, **TD No. 1045 (January 2016).**

BRANDOLINI A. and E. VIVIANO, *Behind and beyond the (headcount) employment rate,* Journal of the Royal Statistical Society: Series A, v. 179, 3, pp. 657-681, **TD No. 965 (July 2015).**

BRIPI F., *The role of regulation on entry: evidence from the Italian provinces,* World Bank Economic Review, v. 30, 2, pp. 383-411, **TD No. 932 (September 2013).**

BRONZINI R. and P. PISELLI, *The impact of R&D subsidies on firm innovation,* Research Policy, v. 45, 2, pp. 442-457, **TD No. 960 (April 2014).**

BURLON L. and M. VILALTA-BUFI, *A new look at technical progress and early retirement,* IZA Journal of Labor Policy, v. 5, **TD No. 963 (June 2014).**

BUSETTI F. and M. CAIVANO, *The trend–cycle decomposition of output and the Phillips Curve: bayesian estimates for Italy and the Euro Area,* Empirical Economics, V. 50, 4, pp. 1565-1587, **TD No. 941 (November 2013).**

CAIVANO M. and A. HARVEY, *Time-series models with an EGB2 conditional distribution,* Journal of Time Series Analysis, v. 35, 6, pp. 558-571, **TD No. 947 (January 2014).**

CALZA A. and A. ZAGHINI, *Shoe-leather costs in the euro area and the foreign demand for euro banknotes,* International Journal of Central Banking, v. 12, 1, pp. 231-246, **TD No. 1039 (December 2015).**

CESARONI T. and R. DE SANTIS, *Current account "core-periphery dualism" in the EMU,* The World Economy, v. 39, 10, pp. 1514-1538, **TD No. 996 (December 2014).**

CIANI E., *Retirement, Pension eligibility and home production,* Labour Economics, v. 38, pp. 106-120, **TD No. 1056 (March 2016).**

CIARLONE A. and V. MICELI, *Escaping financial crises? Macro evidence from sovereign wealth funds' investment behaviour,* Emerging Markets Review, v. 27, 2, pp. 169-196, **TD No. 972 (October 2014).**

CORNELI F. and E. TARANTINO, *Sovereign debt and reserves with liquidity and productivity crises,* Journal of International Money and Finance, v. 65, pp. 166-194, **TD No. 1012 (June 2015).**

D'AURIZIO L. and D. DEPALO, *An evaluation of the policies on repayment of government's trade debt in Italy,* Italian Economic Journal, v. 2, 2, pp. 167-196, **TD No. 1061 (April 2016).**

DE BLASIO G., G. MAGIO and C. MENON, *Down and out in Italian towns: measuring the impact of economic downturns on crime,* Economics Letters, 146, pp. 99-102, **TD No. 925 (July 2013).**

DOTTORI D. and M. MANNA, *Strategy and tactics in public debt management,* Journal of Policy Modeling, v. 38, 1, pp. 1-25, **TD No. 1005 (March 2015).**

ESPOSITO L., A. NOBILI and T. ROPELE, *The management of interest rate risk during the crisis: evidence from Italian banks,* Journal of Banking & Finance, v. 59, pp. 486-504, **TD No. 933 (September 2013).**

LIBERATI D., M. MARINUCCI and G. M. TANZI, *Science and technology parks in Italy: main features and analysis of their effects on hosted firms*, Journal of Technology Transfer, v. 41, 4, pp. 694-729, **TD No. 983 (November 2014).**

MARCELLINO M., M. PORQUEDDU and F. VENDITTI, *Short-Term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility,* Journal of Business & Economic Statistics , v. 34, 1, pp. 118-127, **TD No. 896 (January 2013).**

RODANO G., N. SERRANO-VELARDE and E. TARANTINO, *Bankruptcy law and bank financing,* Journal of Financial Economics, v. 120, 2, pp. 363-382, **TD No. 1013 (June 2015).**

ZINNA G., *Price pressures on UK real rates: an empirical investigation,* Review of Finance,v. 20, 4, pp. 1587-1630, **TD No. 968 (July 2014).**

2017

ADAMOPOULOU A. and G.M. TANZI, *Academic dropout and the great recession,* Journal of Human Capital, V. 11, 1, pp. 35–71, **TD No. 970 (October 2014).**

ALBERTAZZI U., M. BOTTERO and G. SENE, *Information externalities in the credit market and the spell of credit rationing,* Journal of Financial Intermediation, v. 30, pp. 61–70, **TD No. 980 (November 2014).**

ALESSANDRI P. and H. MUMTAZ, *Financial indicators and density forecasts for US output and inflation,* Review of Economic Dynamics, v. 24, pp. 66-78, **TD No. 977 (November 2014).**

BARBIERI G., C. ROSSETTI and P. SESTITO, *Teacher motivation and student learning,* Politica economica/Journal of Economic Policy, v. 33, 1, pp.59-72, **TD No. 761 (June 2010).**

BENTIVOGLI C. and M. LITTERIO, *Foreign ownership and performance: evidence from a panel of Italian firms,* International Journal of the Economics of Business, v. 24, 3, pp. 251-273, **TD No. 1085 (October 2016).**

BRONZINI R. and A. D'IGNAZIO, *Bank internationalisation and firm exports: evidence from matched firm-bank data,* Review of International Economics, v. 25, 3, pp. 476-499 **TD No. 1055 (March 2016).**

BRUCHE M. and A. SEGURA, *Debt maturity and the liquidity of secondary debt markets,* Journal of Financial Economics, v. 124, 3, pp. 599-613, **TD No. 1049 (January 2016).**

BURLON L., *Public expenditure distribution, voting, and growth,* Journal of Public Economic Theory,, v. 19, 4, pp. 789–810, **TD No. 961 (April 2014).**

BURLON L., A. GERALI, A. NOTARPIETRO and M. PISANI, *Macroeconomic effectiveness of non-standard monetary policy and early exit. a model-based evaluation,* International Finance, v. 20, 2, pp.155-173, **TD No. 1074 (July 2016).**

BUSETTI F., *Quantile aggregation of density forecasts,* Oxford Bulletin of Economics and Statistics, v. 79, 4, pp. 495-512, **TD No. 979 (November 2014).**

CESARONI T. and S. IEZZI, *The predictive content of business survey indicators: evidence from SIGE,* Journal of Business Cycle Research, v.13, 1, pp 75–104, **TD No. 1031 (October 2015).**

CONTI P., D. MARELLA and A. NERI, *Statistical matching and uncertainty analysis in combining household income and expenditure data,* Statistical Methods & Applications, v. 26, 3, pp 485–505, **TD No. 1018 (July 2015).**

D'AMURI F. and J. MARCUCCI, *The predictive power of google searches in forecasting unemployment,* International Journal of Forecasting, v. 33, 4, pp. 801-816, **TD No. 891 (November 2012).**

DE BLASIO G. and S. POY, *The impact of local minimum wages on employment: evidence from Italy in the 1950s,* Journal of Regional Science, v. 57, 1, pp. 48-74, **TD No. 953 (March 2014).**

DEL GIOVANE P., A. NOBILI and F. M. SIGNORETTI, *Assessing the sources of credit supply tightening: was the sovereign debt crisis different from Lehman?,* International Journal of Central Banking, v. 13, 2, pp. 197-234, **TD No. 942 (November 2013).**

DELLE MONACHE D. and I. PETRELLA, *Adaptive models and heavy tails with an application to inflation forecasting,* International Journal of Forecasting, v. 33, 2, pp. 482-501, **TD No. 1052 (March 2016).**

DEL PRETE S., M. PAGNINI, P. ROSSI and V. VACCA, *Lending organization and credit supply during the 2008–2009 crisis,* Economic Notes, v. 46, 2, pp. 207–236, **TD No. 1108 (April 2017).**

LOBERTO M. and C. PERRICONE, *Does trend inflation make a difference?,* Economic Modelling, v. 61, pp. 351–375, **TD No. 1033 (October 2015).**

MANCINI A.L., C. MONFARDINI and S. PASQUA, *Is a good example the best sermon? Children's imitation of parental reading,* Review of Economics of the Household, v. 15, 3, pp 965–993, **D No. 958 (April 2014).**

MEEKS R., B. NELSON and P. ALESSANDRI, *Shadow banks and macroeconomic instability,* Journal of Money, Credit and Banking, v. 49, 7, pp. 1483–1516, **TD No. 939 (November 2013).**

MICUCCI G. and P. ROSSI, *Debt restructuring and the role of banks' organizational structure and lending technologies,* Journal of Financial Services Research, v. 51, 3, pp 339–361, **TD No. 763 (June 2010).**

MOCETTI S., M. PAGNINI and E. SETTE, *Information technology and banking organization,* Journal of Journal of Financial Services Research, v. 51, pp. 313-338, **TD No. 752 (March 2010).**

MOCETTI S. and E. VIVIANO, *Looking behind mortgage delinquencies,* Journal of Banking & Finance, v. 75, pp. 53-63, **TD No. 999 (January 2015).**

NOBILI A. and F. ZOLLINO, *A structural model for the housing and credit market in Italy,* Journal of Housing Economics, v. 36, pp. 73-87, **TD No. 887 (October 2012).**

PALAZZO F., *Search costs and the severity of adverse selection,* Research in Economics, v. 71, 1, pp. 171-197, **TD No. 1073 (July 2016).**

PATACCHINI E. and E. RAINONE, *Social ties and the demand for financial services,* Journal of Financial Services Research, v. 52, 1–2, pp 35–88, **TD No. 1115 (June 2017).**

PATACCHINI E., E. RAINONE and Y. ZENOU, *Heterogeneous peer effects in education,* Journal of Economic Behavior & Organization, v. 134, pp. 190–227, **TD No. 1048 (January 2016).**

SBRANA G., A. SILVESTRINI and F. VENDITTI, *Short-term inflation forecasting: the M.E.T.A. approach,* International Journal of Forecasting, v. 33, 4, pp. 1065-1081, **TD No. 1016 (June 2015).**

SEGURA A. and J. SUAREZ, *How excessive is banks' maturity transformation?,* Review of Financial Studies, v. 30, 10, pp. 3538–3580, **TD No. 1065 (April 2016).**

VACCA V., *An unexpected crisis? Looking at pricing effectiveness of heterogeneous banks,* Economic Notes, v. 46, 2, pp. 171–206, **TD No. 814 (July 2011).**

VERGARA CAFFARELI F., *One-way flow networks with decreasing returns to linking,* Dynamic Games and Applications, v. 7, 2, pp. 323-345, **TD No. 734 (November 2009).**

ZAGHINI A., *A Tale of fragmentation: corporate funding in the euro-area bond market,* International Review of Financial Analysis, v. 49, pp. 59-68, **TD No. 1104 (February 2017).**

FORTHCOMING

ADAMOPOULOU A. and E. KAYA, *Young Adults living with their parents and the influence of peers,* Oxford Bulletin of Economics and Statistics, **TD No. 1038 (November 2015).**

ALBANESE G., G. DE BLASIO and P. SESTITO, *Trust, risk and time preferences: evidence from survey data,* International Review of Economics, **TD No. 911 (April 2013).**

BOFONDI M., L. CARPINELLI and E. SETTE, *Credit supply during a sovereign debt crisis,* Journal of the European Economic Association, **TD No. 909 (April 2013).**

CASIRAGHI M., E. GAIOTTI, L. RODANO and A. SECCHI, *A "Reverse Robin Hood"? The distributional implications of non-standard monetary policy for Italian households,* Journal of International Money and Finance, **TD No. 1077 (July 2016).**

D'AMURI F., *Monitoring and disincentives in containing paid sick leave,* Labour Economics, **TD No. 787 (January 2011).**

FEDERICO S. and E. TOSTI, *Exporters and importers of services: firm-level evidence on Italy,* The World Economy, **TD No. 877 (September 2012).**

GIACOMELLI S. and C. MENON, *Does weak contract enforcement affect firm size? Evidence from the neighbour's court,* Journal of Economic Geography, **TD No. 898 (January 2013).**

NATOLI F. and L. SIGALOTTI, *Tail co-movement in inflation expectations as an indicator of anchoring,* International Journal of Central Banking, **TD No. 1025 (July 2015).**

RIGGI M., *Capital destruction, jobless recoveries, and the discipline device role of unemployment,* Macroeconomic Dynamics, **TD No. 871 (July 2012).**

SEGURA A., *Why did sponsor banks rescue their SIVs?,* Review of Finance, **TD No. 1100 (February 2017).**