



Efficient estimation of true fixed-effects stochastic frontier models

Ruggero Bellio ¹ · Luca Grassetto ¹

Accepted: 9 May 2024
© The Author(s) 2024

Abstract

Fixed-effects modeling has become the method of choice in several panel data settings, including models for stochastic frontier analysis. A notable instance of stochastic frontier panel data models is the true fixed-effects model, which allows disentangling unit heterogeneity from efficiency evaluations. While such a model is theoretically appealing, its estimation is hampered by incidental parameters. This note proposes a simple and rather general estimation approach where the unit-specific intercepts are integrated out of the likelihood function. We apply the theory of composite group families to the model of interest and demonstrate that the resulting integrated likelihood is a marginal likelihood with desirable inferential properties. The derivation of the result is provided in full, along with some connections with the existing literature and computational details. The method is illustrated for three notable models, given by the normal-half normal model, the heteroscedastic exponential model, and the normal-gamma model. The results of simulation experiments highlight the properties of the methodology.

Keywords Fixed effects · Marginal likelihood · Panel data · Stochastic frontier

JEL classification C5 · C13 · C46

1 Introduction

Fixed-effects models are prominent in modern panel data econometrics since they adjust for unit heterogeneity, potentially correlated with the covariates. The robustness of fixed-effects models is rather important in stochastic frontier analysis, which explains the popularity of the *true fixed-effects stochastic frontier model* (Greene 2005a, b). For a production frontier model, the specification of the true fixed-effects model is given by

$$y_{it} = \alpha_i + \mathbf{x}_{it}^\top \boldsymbol{\beta} - u_{it} + v_{it}, \quad (1)$$

where i is the index for the unit of the panel ($i = 1, \dots, n$) and t is the index for time ($t = 1, \dots, T$), with T assumed to be the same across units without loss of generality. Here y_{it} is the output and \mathbf{x}_{it} is a vector of exogenous inputs, both possibly in log form, $\boldsymbol{\beta}$ is a vector of coefficients and α_i the

unit fixed effect. In the following, we will also employ \mathbf{y}_i and \mathbf{X}_i to denote the vector of outputs and the matrix of inputs for the i -th panel member, respectively. Furthermore, u_{it} is the one-sided inefficiency term, and v_{it} is the symmetric idiosyncratic error, independent of u_{it} . A simple change in the sign of u_{it} is required for a cost frontier model specification. We assume conditional independence (given α_i and $\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}$) of observations within the same panel member and independence across units. Finally, we denote by $\boldsymbol{\theta}$ the set of structural model parameters, given by $\boldsymbol{\beta}$ plus all the parameters entering the distribution of the error term $\varepsilon_{it} = v_{it} - u_{it}$.

The introduction of the unit-specific intercepts α_i is the key to disentangle heterogeneity from efficiency, but at the same time, when the intercepts are estimated jointly with the other model parameters, the incidental parameters issue arises; see, for example, Bartolucci et al. (2016). In short, α_i is consistently estimated only for $T \rightarrow \infty$. For finite T , the resulting bias accumulates across panel members, and it propagates to $\boldsymbol{\theta}$, thus affecting the efficiency evaluation for settings with short panels. Therefore, some caution is required when estimating the structural parameters.

✉ Luca Grassetto
luca.grassetto@uniud.it

¹ Department of Economics and Statistics, University of Udine, Via Tomadini, 30/a, Udine 33100 UD, Italy

There are some proposals in the econometric literature that overcome the incidental parameters problem for some notable true fixed-effects model specifications, but they all rely on specific distributional properties. This note aims to present a simple and \sqrt{n} -consistent estimation method that could be effectively applied to a broad array of distributions for ε_{it} . Moreover, it provides an efficient estimator which is at least as efficient as existing consistent estimators.

The outline of this note is as follows. Section 2 reviews the econometric literature on consistent estimation of true fixed-effects models. The proposal is presented in Section 3, where mathematical support is provided. Some numerical studies are reported in Section 4, including an application to the challenging normal-gamma model. Section 5 provides a brief discussion.

2 Background

A marginal maximum likelihood estimation for model (1) is developed in Chen et al. (2014) paralleling the usual derivation of the within estimator for linear models. In particular, the authors consider the deviation of each term of the model from the unit mean and then maximize the within likelihood defined on the $(T-1)$ deviations of the error term for the i -th unit. The method is specific to the normal-half normal model where the components of ε_{it} are half normally distributed $u_{it} \sim N^+(0, \sigma_u^2)$, and normally distributed $v_{it} \sim N(0, \sigma_v^2)$, respectively. Using the results from the multivariate closed skew normal distribution (see González-Farías et al. 2004, and the references therein), they show that neglecting a constant additive term, the within log-likelihood function is

$$\begin{aligned} \ell_W(\boldsymbol{\theta}) = & \sum_{i=1}^n \log \phi_{T-1}(\tilde{\mathbf{y}}_i^* - \tilde{\mathbf{X}}_i^* \boldsymbol{\beta}; \mathbf{0}_{T-1}, \sigma^2 (\mathbf{I}_{T-1} - \frac{\mathbf{E}_{T-1}}{T})) \\ & + \sum_{i=1}^n \log \Phi_T \left(-\frac{\lambda}{\sigma} (\tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \boldsymbol{\beta}); \mathbf{0}_T, \mathbf{I}_T + \frac{\lambda^2}{T} \mathbf{E}_T \right), \end{aligned} \quad (2)$$

where $\phi_m(\cdot)$ and $\Phi_m(\cdot)$ denote the density function and the CDF of the m -dimensional normal distribution, $\sigma^2 = \sigma_u^2 + \sigma_v^2$, and $\lambda = \sigma_u/\sigma_v$. Further, the notation $\tilde{\mathbf{X}}$ denotes the within transformation of the m -dimensional vector \mathbf{x} , namely the deviation of the m values of \mathbf{x} from its the sample mean, whereas the notation $\tilde{\mathbf{X}}^*$ denotes that only the first $m-1$ such values are retained. Finally, $\mathbf{0}_m$ is a vector of m zeros, \mathbf{I}_m is the identity matrix of size m , and \mathbf{E}_m is a $m \times m$ matrix of ones. The expression (2) is free of α_i , and its maximizer defines the maximum marginal likelihood estimator (MMLE), which is consistent for increasing n regardless of the value of the panel length T . Wang and Ho (2010) define a similar estimator, but for a different model

where $u_{it} = u_i h_{it}$, with h_{it} being a non-random function of some covariates, and assuming u_i half normally distributed. See also Kutlu et al. (2020).

Another contribution is Wikström (2015), who defines a method-of-moments estimator and provides a simulation study for the same model considered by Chen et al. (2014). The proposal has some interesting properties, such as its logical simplicity. On the other hand, it is likely to be less efficient than the MMLE since the latter is based on a marginal likelihood (see Pace and Salvani 1997, Chap. 4), and the method of moments is at times prone to numerical instabilities, as testified by the “problematic replications” reported in Wikström (2015).

Belotti and Ilardi (2018) provide some notable estimation strategies. Among other things, they note that the proposal in Chen et al. (2014), exploiting the closure property of the skew normal distribution, could be applied to some extensions of the normal-half normal model, including the truncated normal distribution for the inefficiency term, heteroscedasticity, and autocorrelation between different inefficiencies. They also note that while the T -dimensional normal CDF involved in the log-likelihood (2) can be reduced to a one-dimensional integral, the within log-likelihood for such extensions would involve an irreducible T -dimensional normal CDF, whose computation for large panel lengths may become cumbersome. Hence, they propose another estimator, the pairwise difference estimator (PDE).

The PDE is the maximizer of an objective function involving the distribution of all the differences between pairs of observations belonging to the same panel member, much in the spirit of composite likelihood estimation (Varin et al. 2011). The results reported by Belotti and Ilardi (2018) illustrate the \sqrt{n} -consistency of the proposal, which is, however, feasible only for truncated normal or exponential inefficiency terms.

Finally, we note that Belotti and Ilardi (2018) propose another estimator, the marginal maximum simulated likelihood estimator (MMSLE), which approximates the MMLE by simulation, also covering some heteroscedastic extensions not considered by Chen et al. (2014). The simulations in Belotti and Ilardi (2018) show that the MMSLE is a good approximation to the MMLE for the normal-half normal model. At any rate, while theoretically interesting, its Monte Carlo nature makes it less appealing than the original MMLE.

3 Estimation via marginal likelihood

A general solution for consistent estimation of the structural parameters $\boldsymbol{\theta}$ proceeds by noting that given the sample for the i -th unit, α_i acts as a location parameter.

We start by considering the order statistics for the i -th panel member $(y_{i(1)}, \dots, y_{i(T)})$, from which we define the following data vector

$$s_i = s_i(y_{i1}, \dots, y_{iT}) = (y_{i(2)} - y_{i(1)}, \dots, y_{i(T)} - y_{i(1)}). \quad (3)$$

It is straightforward to verify that the distribution of s_i is free of the unit-specific intercept α_i , which is eliminated by pivoting on the smallest observation. This suggests the following marginal likelihood for θ

$$L_M(\theta) = \prod_{i=1}^n f_S(s_i; \theta),$$

where $f_S(s_i; \theta)$ is the density of s_i . The remaining task is to obtain an analytic expression for $f_S(s_i; \theta)$. To this end, we rely on the theory of composite group families. In particular, we follow the same steps as Pace and Salvani (1997, Theorem 7.5)¹ to express the marginal likelihood (aside from a constant factor) as

$$L_M(\theta) = \prod_{i=1}^n f_S(s_i; \theta) = \prod_{i=1}^n \int_{-\infty}^{\infty} \prod_{t=1}^T f_Y(y_{it}; \alpha_i, \theta) d\alpha_i, \quad (4)$$

where $f_Y(y_{it}; \alpha_i, \theta)$ is the density function of y_{it} , from which we readily define the marginal log-likelihood $\ell_M(\theta) = \log L_M(\theta)$. The expression (4) greatly facilitates the task of obtaining the marginal likelihood based on (3), which would be otherwise rather involved to obtain directly. The only issue is the approximation of the one-dimensional integrals in (4), for which we provide some guidelines in what follows. Moreover, note that the expression (4) requires the density $f_Y(y_{it}; \alpha_i, \theta)$ of the response variable y_{it} , that in most cases is not difficult to obtain; at any rate, the example in section 4.3 illustrates how to proceed when this is not the case.

Since $\ell_M(\theta)$ is a marginal log-likelihood based on n independent data vectors, it can be treated as a bona fide log-likelihood function for θ . Hence its maximizer is \sqrt{n} -consistent and the inverse of its negative Hessian evaluated at the maximizer is a valid estimate of the sampling variance. Note that for an incidental parameter of a different nature than a location parameter, the elimination through integration as in (4) may not provide any improvement on the incidental parameter bias; see Schumann et al. (2021) and the references therein. We note in passing that the marginal likelihood (4) is an instance of a class of integrated likelihood functions studied in Arellano and Bonhomme (2009). A connection between the results presented here and the theory of the latter paper is illustrated in Appendix B.

The marginal likelihood in (4) provides a rather general and relatively simple estimation approach without needing

specific distributional properties. As a general fact in the frontier specification, the integrals in (4) do not have a closed-form solution, and they must be approximated by numerical methods. For panel data settings, however, the n integrals must be approximated with high accuracy; otherwise, the error in the approximation accumulates across panel groups. For this reason, we endorse the use of adaptive Gaussian quadrature (Liu and Pierce 1994) with Q quadrature points, which gives the following approximation to $\ell_M(\theta)$ (neglecting a constant additive term)

$$\hat{\ell}_M(\theta) = \sum_{i=1}^n \log \left[\text{se}_i(\theta) + \sum_{q=1}^Q w_q \exp \left\{ z_q^2 + \sum_{t=1}^T \log f_Y(y_{it}; z_q^*, \theta) \right\} \right], \quad (5)$$

where z_q and w_q are the Gauss-Hermite quadrature nodes and weights. Furthermore, z_q^* are the adjusted nodes $z_q^* = \sqrt{2} \text{se}_i(\theta) z_q + \hat{\alpha}_i(\theta)$, where $\hat{\alpha}_i(\theta)$ is the mode of $\sum_{t=1}^T \log f_Y(y_{it}; \alpha_i, \theta)$ in α_i , and $\text{se}_i(\theta)$ is the reciprocal of the squared root of the second derivative of $-\sum_{t=1}^T \log f_Y(y_{it}; \alpha_i, \theta)$ with respect to α_i , evaluated at $\hat{\alpha}_i(\theta)$.

The important advantage of the approximation (5) is that, by tuning Q , we can keep the approximation error under strict control. Indeed, having a \sqrt{n} -consistent estimator as a target, it seems reasonable to aim for an approximation error smaller than the inferential error $n^{-1/2}$. Following Stringer and Bilodeau (2022), we should choose Q at least equal to $\lceil (3/2) \log_T(n) - 2 \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. For instance, for $T=5$, which is the smallest panel length considered in the aforementioned stochastic frontier literature, the latter formula implies that around 10 quadrature points are sufficient for n as large as 100,000. All the results of the next section were obtained with $Q=25$, well above the required number, and indeed, we cross-checked the high accuracy of the approximation by comparison with the more computationally expensive Gauss-Kronrod quadrature.

4 Some notable models

In this section, we apply the Integrated Likelihood Estimator (ILE) defined by the maximizer of (4) to some models of interest.

4.1 The normal-half normal model

A fundamental model is the normal-half normal model, studied by both Chen et al. (2014) and Belotti and Ilardi (2018). For the case where $u_{it} \sim N^+(0, \sigma_u^2)$ and $v_{it} \sim N(0, \sigma_v^2)$, we followed these authors and considered the same single-input simulation scenario. Since we found that the ILE matches the MMLE for all the model parameters with high numerical accuracy, we do not report any

¹ Reported in Appendix A

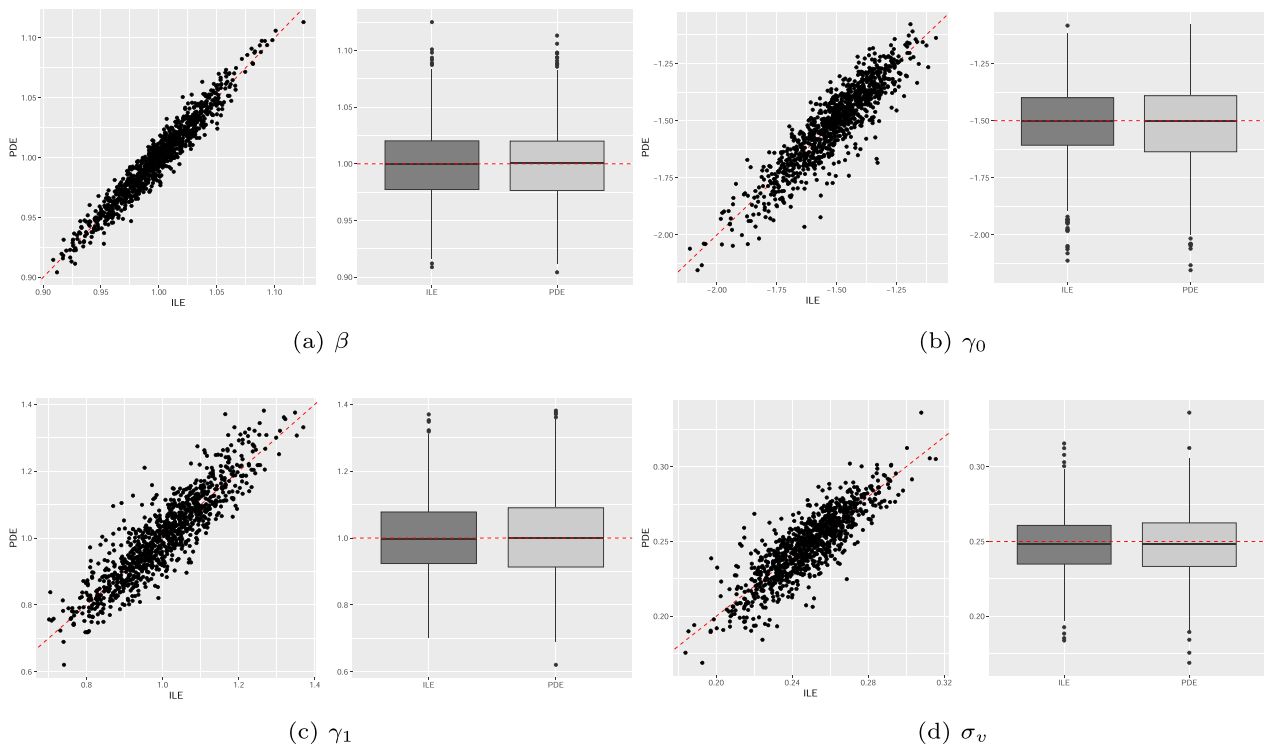


Fig. 1 Heteroscedastic exponential model. Scatterplots of PDE vs ILE results and boxplots of the two sets of estimates for the setting reported in Table 1 in Belotti and Ilardi (2018), with bisector line added (scatterplots) and horizontal line corresponding to the true parameter value (boxplots)

results here. The equivalence between ILE and MMLE is encouraging since the latter is a benchmark estimator for the normal-half normal TFE model. Notice that the same equivalence occurs in the case of panel data models with fixed effects and normal errors (Schumann et al. 2021).

We end this example by noting that for the extension to truncated normal inefficiency coupled with heteroscedasticity, the computation of MMLE would require the evaluation of the T -dimensional normal CDF, with a computational burden increasing with the group size. On the contrary, the marginal likelihood (4) would retain the same order of complexity.

4.2 The heteroscedastic exponential model

This is another example considered in Belotti and Ilardi (2018), for which the PDE provides good estimation performance. Here we assume $u_{it} \sim \mathcal{E}(\sigma_{iu})$ and $v_{it} \sim N(0, \sigma_v^2)$, where $\sigma_{iu} = \exp(\gamma_0 + z_i \gamma_1)$. With such a specification, and after setting $r_{it} = y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}$, the density of each observation of the i -th panel member becomes

$$f_Y(y_{it}; \alpha_i, \boldsymbol{\theta}) = \frac{1}{\sigma_{iu}} \exp \left\{ \frac{r_{it}}{\sigma_{iu}} + \log \Phi \left(-\frac{r_{it}}{\sigma_v} - \frac{\sigma_v}{\sigma_{iu}} \right) + \frac{\sigma_v^2}{2\sigma_{iu}^2} \right\}.$$

Table 1 Estimated power of the test of exponentiality

Setting	Power
$n = 100, T = 5$	0.22
$n = 250, T = 5$	0.57
$n = 500, T = 5$	0.83
$n = 250, T = 10$	0.99
$n = 500, T = 10$	1.00

We focus here on some of the simulation experiments reported in Belotti and Ilardi (2018) and compared our implementation of the ILE with the PDE; the latter was computed using the Stata code made available by Belotti et al. (2013).

Plots in Fig. 1 display the results based on 1000 simulated datasets for the same simulation scenario of section 6.1 in Belotti and Ilardi (2018), with the setting corresponding to their Table 1, namely $n = 100, T = 5$, and “signal to noise ratio” set to 1; this setting was chosen since it is the most challenging, but the results for the other settings are rather similar and are reported in the Supplementary Material. The two estimators may differ on a sample-by-sample basis, yet they exhibit very little bias (if any). The ILE slightly outperforms the PDE in terms of MSE

since it is based on a marginal likelihood rather than a pairwise likelihood, yet the difference in efficiency is small.

4.3 The normal-gamma model

As mentioned at the outset, the important property of the proposal (4) is its broad generality, and to illustrate this point, we consider here the normal-gamma model. First introduced by Greene (1990), this model has been somewhat underused in applications since it involves computations that are notoriously challenging; see Ritter and Simar (1997), Tsionas (2000), Greene (2003) and the references therein. To the best of our knowledge, the normal-gamma model has never been adopted for the true fixed-effects stochastic frontier model. While the results of Ritter and Simar (1997) illustrate the difficulties that may arise in estimating the efficiency parameters, they also suggest that estimation issues are much less severe with large sample sizes. Since, in the setting of interest, we can glean information about the efficiency parameter from n panel units, the total number of observations may be large enough for meaningful estimation. A similar positive note comes from the Bayesian inferential approach in Tsionas (2000), where it is reported that for a normal-gamma model with a total sample size of 1000, “results are satisfactory”.

The model assumes $u_{it} \sim \mathcal{G}(\gamma, \lambda)$ and $v_{it} \sim N(0, \sigma_v^2)$. The model density requires the numerical approximation of an integral; letting again $r_{it} = y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}$, the density of each observation of the i -th panel member is

$$f_Y(y_{it}; \alpha_i, \boldsymbol{\theta}) = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left\{-\frac{(r_{it} + z)}{2\sigma_v^2} - \frac{z}{\lambda}\right\} \frac{z^{\gamma-1}}{\lambda^\gamma \Gamma(\gamma)} dz. \tag{6}$$

For the ILE, this implies that a double integral is required, and a meticulous numerical implementation is called for. Our strategy entails coding the normal-gamma convolution integral (6) in C++ using adaptive Gauss-Kronrod quadrature with a large number of nodes, plus other steps required for successful implementation of the approximation (5); further details are provided in the Supplementary Material. A fact worth mentioning is that the numerical optimization of the log-likelihood (5) is best achieved with the re-parameterization in which γ and σ_v are replaced by the mean and the standard deviation of the gamma term, namely

$$\mu_G = \gamma \lambda, \quad \sigma_G = \sqrt{\sigma_v^2 + \gamma \lambda^2}.$$

The fact that very high accuracy is required for both the inner and outer integrals involved in the log-likelihood function makes its optimization rather expensive; while other models are estimated in a matter of seconds, the

estimation of the normal-gamma model may take several minutes.

Figure 2 summarizes the estimation results for 1000 data sets at each of four different settings, with the true efficiency parameter values set at $\gamma = 2, \lambda = 1$ and $\sigma_v = 0.25$. This choice of parameters avoids the “relatively difficult” case where $\gamma < 1$ (Tsionas 2000), for which, in our settings, the optimization becomes far too slow for a Monte Carlo study. Note that the minimum total sample size considered is 1250: compared to the previous example, the results for the setting $n = 100$ and $T = 5$ are not reported, since the estimation turned out to be unreliable.

Despite the challenging estimation, the results are reasonably satisfactory since the distribution of each of the estimators has a median matching the true parameter value in all cases, and it becomes more concentrated around it when the total sample size increases. Though the two settings $n = 500, T = 5$ and $n = 250, T = 10$ have the same total sample size, the latter shows slightly better results.

Another useful point concerns the possibility of distinguishing the gamma model from simpler ones. A natural competitor is the normal-exponential, which is obtained as a special case of the model of section 4.2. Table 1 reports the estimated power for a fixed level $\alpha = 0.05$ of the log-likelihood ratio test for the hypothesis $\gamma = 1$, i.e., the test of exponentiality

$$W = 2 \left\{ \widehat{\ell}_M(\widehat{\boldsymbol{\theta}}) - \widehat{\ell}_M(\widehat{\boldsymbol{\theta}}_0) \right\},$$

where $\widehat{\boldsymbol{\theta}}$ is the estimator from the normal-gamma model and $\widehat{\boldsymbol{\theta}}_0$ the estimator from the normal-exponential model. This is a power computation since the data were generated under the normal-gamma model with $\gamma = 2$.

The results suggest that while the setting $n = 100$ and $T = 5$ is very far from being suitable to distinguish the gamma model from the exponential one, things get progressively better for the four settings reported in Fig. 2.

5 Discussion

The estimator based on the marginal likelihood obtained by integrating out the panel-specific intercepts in the true fixed-effects stochastic frontier model represents a simple and efficient methodology. The proposal recovers the existing consistent estimators whenever available and can provide a useful solution in other cases. It seems worth noting that the same strategy where group-specific intercepts are integrated out may be relevant to other fixed-effects panel data models with continuous response as an alternative to differencing, providing a \sqrt{n} -consistent estimator and avoiding any issue related to the size of T .

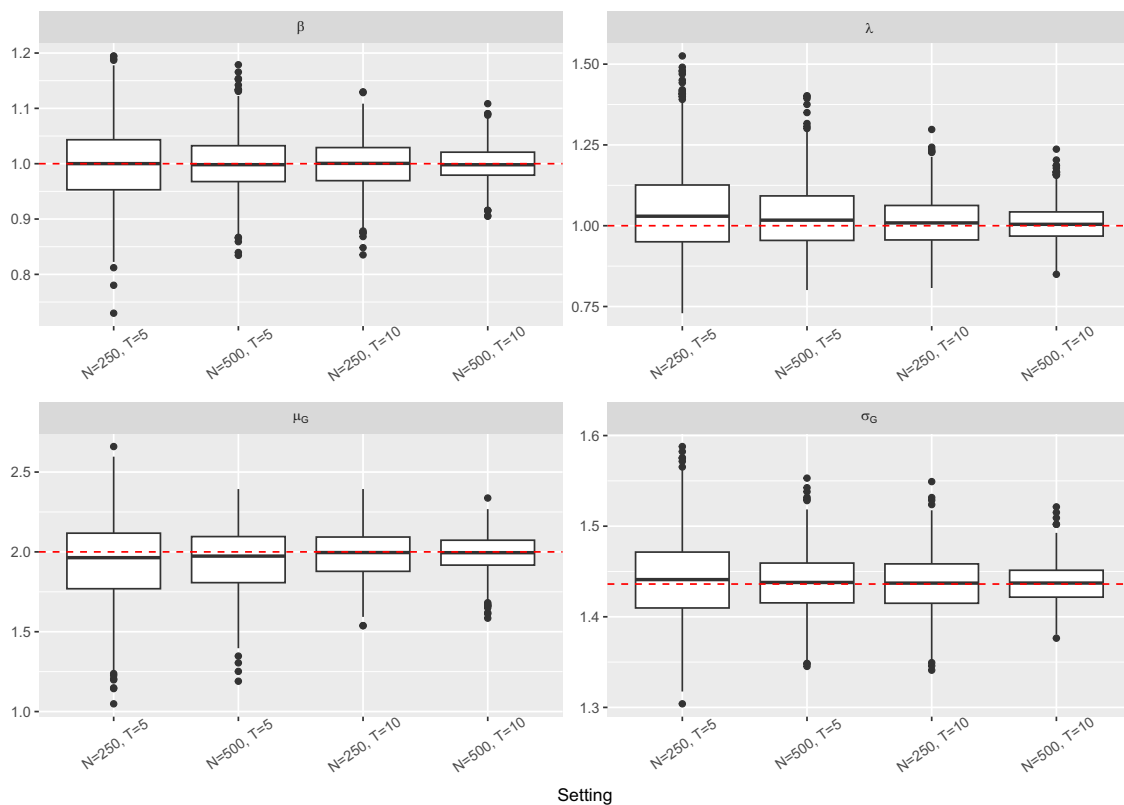


Fig. 2 Normal-gamma model. Boxplots of ILE results in four different settings, with horizontal lines corresponding to the true parameter value

The ILE may also be useful in connection with four-component panel stochastic frontier models; see Kumbhakar and Lai (2022) for a recent review. Such models include two random group-specific effects in place of α_i in (1), namely $\alpha_i = \tau_i - \eta_i$, where τ_i is a group effect and η_i is the persistent inefficiency components. Such an approach allows for more informative efficiency evaluations than the true fixed-effects setting, at the expense of stronger distributional specifications. For those researchers willing to adopt the four-component specification, a recommended approach may be to supplement the analysis with the estimates of several alternative models along the lines of what was done by Kumbhakar et al. (2014). We endorse the inclusion of the true fixed-effects model among the array of considered specifications since its robust nature will provide some coverage against input endogeneity issues. A simple comparison of the estimates of structural parameters and estimated inefficiencies across several models could then provide some valuable information. To this end, the availability of resources for true-fixed effects analysis, including software, may represent a useful addition.

In closing, among possible directions for extension of the current paper, we mention the adoption of other distributions for the inefficiency term or the inclusion of semiparametric effects for the production or cost function, along the lines of the proposal in Bellio and Grassetti (2011). The implementation of the latter, however, would require special care.

6 Software

A public repository with R code for fitting the models presented in this note is available at github.com/lucagragetti/ilsfa.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1007/s1123-024-00725-3>.

Acknowledgements A very preliminary version of this work was presented at the 35th International Workshop on Statistical Modelling (IWSM), held online in July 2020. The authors are very grateful to Luigi Pace for insightful comments. We are also grateful to the Editor, the Associate Editor, and two Reviewers for valuable comments and suggestions. This work was supported by the Departmental Strategic Plan (PSD) of the University of Udine, Department of Economics and Statistics (2022-2025).

Author contributions Ruggero Bellio mainly contributed to the theoretical investigation and model specification (sections 1, 3, and the appendices) and contributed to the code writing. Literature review, further code writing (software repository), and analysis were performed by Luca Grassetti (sections 2 and 4). Ruggero Bellio wrote the first draft of the manuscript, and Luca Grassetti performed text review and editing. All authors read and approved the final manuscript and contributed to the conclusions.

Funding Open access funding provided by Università degli Studi di Udine within the CRUI-CARE Agreement.

Conflict of interest The authors declare no competing interests.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

7 Appendix A: Inference based on the integrated likelihood

We need to show that marginal likelihood based on s_i in expression (3) of the main paper equals the integrated likelihood

$$f_S(s_i; \theta) = \int_{-\infty}^{\infty} \prod_{t=1}^T f_Y(y_{it}; \alpha_i, \theta) d\alpha_i, \tag{A.1}$$

To this end, we follow the same steps of the proof of Theorem 7.5 of Pace and Salvani (1997), which covers an even more complex setting for location and scale transformations.

Before proceeding, note that since α_i is a location parameter and $\mathbf{x}_{it}^\top \boldsymbol{\beta}$ acts on the location as well, we can define $f_0(\cdot)$ such that

$$f_Y(y_{it}; \alpha_i, \theta) = f_0(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi}), \tag{A.2}$$

where $\boldsymbol{\psi}$ denotes the set of structural parameters without $\boldsymbol{\beta}$, i.e., $\theta = (\boldsymbol{\beta}, \boldsymbol{\psi})$.

First, we perform a sufficiency reduction to the order statistic for the i -th unit in the panel $(y_{i(1)}, \dots, y_{i(T)})$, having density

$$f_{Y_{\text{ord}}}(y_{i(1)}, \dots, y_{i(T)}; \alpha_i, \theta) = T! \prod_{t=1}^T f_0(y_{i(t)} - \alpha_i - \mathbf{x}_{i(t)}^\top \boldsymbol{\beta}; \boldsymbol{\psi}). \tag{A.3}$$

Since the distribution of s_i does not depend on α_i , without losing generality we may assume $\alpha_i = 0$, and the distribution (A.3) becomes

$$f_{Y_{\text{ord}}}(y_{i(1)}, \dots, y_{i(T)}; 0, \theta) = T! \prod_{t=1}^T f_0(y_{i(t)} - \mathbf{x}_{i(t)}^\top \boldsymbol{\beta}; \boldsymbol{\psi}).$$

The next step is to consider the one-to-one transformation

$$\begin{cases} z_i = y_{i(1)} \\ s_{i2} = y_{i(2)} - y_{i(1)} \\ \vdots \\ s_{iT} = y_{i(T)} - y_{i(1)} \end{cases}$$

having inverse

$$\begin{cases} y_{i(1)} = z_i \\ y_{i(2)} = s_{i2} + z_i \\ \vdots \\ y_{i(T)} = s_{iT} + z_i \end{cases},$$

and unitary Jacobian. It follows that

$$f_{Z,S}(z_i, s_{i1}, \dots, s_{iT}; \theta) = T! f_0(z_i - \mathbf{x}_{i(1)}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) \prod_{t=2}^T f_0(s_{it} + z_i - \mathbf{x}_{i(t)}^\top \boldsymbol{\beta}; \boldsymbol{\psi}),$$

from which we can integrate out z_i to obtain the marginal distribution of s_i

$$f_S(s_i; \theta) = T! \int_{-\infty}^{\infty} f_0(z_i - \mathbf{x}_{i(1)}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) \prod_{t=2}^T f_0(s_{it} + z_i - \mathbf{x}_{i(t)}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) dz_i.$$

Furthermore, note that the constant $T!$ can be omitted in the likelihood definition, and then we proceed with another change of variable $u_i = z_i - y_{i(1)}$, for $i = 1, \dots, n$, having again unitary Jacobian. With some simple algebra, we obtain

$$f_S(s_i; \theta) = \int_{-\infty}^{\infty} \prod_{t=1}^T f_0(y_{i(t)} + u_i - \mathbf{x}_{i(t)}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) du_i.$$

A further change of variable $\alpha_i = -u_i$ finally gives the result (A.1), using again the identity (A.2), so that

$$f_S(s_i; \theta) = \int_{-\infty}^{\infty} \prod_{t=1}^T f_0(y_{i(t)} - \alpha_i - \mathbf{x}_{i(t)}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) d\alpha_i = \int_{-\infty}^{\infty} \prod_{t=1}^T f_Y(y_{it}; \alpha_i, \theta) d\alpha_i.$$

8 Appendix B: Relation to the results in Arellano and Bonhomme (2009)

The integrated likelihood (4) corresponds to the choice of the uniform prior according to Arellano and Bonhomme (2009). The fact that it corresponds to the marginal likelihood obtained from the distribution of (3) implies that the uniform prior resolves the incidental parameter issue, and this, in turn, must imply that condition (11) in the cited

article holds. In what follows, we illustrate that this is the case.

As reminded by the authors, to prove this fact, it suffices that $\rho_i(\boldsymbol{\theta}, \alpha_i)$ does not depend on α_i , where $\rho_i(\boldsymbol{\theta}, \alpha_i)$ is the following vector

$$\rho_i(\boldsymbol{\theta}, \alpha_i) = -\frac{i_{\boldsymbol{\theta}\alpha_i}(\boldsymbol{\theta}, \alpha_i)}{i_{\alpha_i\alpha_i}(\boldsymbol{\theta}, \alpha_i)},$$

with $i_{\boldsymbol{\theta}\alpha_i}(\boldsymbol{\theta}, \alpha_i)$ and $i_{\alpha_i\alpha_i}(\boldsymbol{\theta}, \alpha_i)$ denoting the $\boldsymbol{\theta}\alpha_i$ -block and the $\alpha_i\alpha_i$ -block of the expected Fisher information matrix, respectively.

The fact that the expected Fisher information does not depend on α_i is straightforward to verify. To this aim, let us consider the log-likelihood function for the i -th unit corresponding to the TFE model (1), using again the identity (A.2)

$$\ell_i(\alpha_i, \boldsymbol{\theta}) = \sum_{t=1}^T \log f_Y(y_{it}; \alpha_i, \boldsymbol{\theta}) = \sum_{t=1}^T \log f_0(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi}).$$

From this, we obtain the score function for α_i

$$\frac{\partial \ell_i(\alpha_i, \boldsymbol{\theta})}{\partial \alpha_i} = \sum_{t=1}^T \frac{-f'_0(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi})}{f_0(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi})},$$

where $f'_0(z; \boldsymbol{\psi}) = \partial f_0(z; \boldsymbol{\psi}) / \partial z$, and then the $\alpha_i\alpha_i$ -block of the observed information matrix

$$-\frac{\partial^2 \ell_i(\alpha_i, \boldsymbol{\theta})}{\partial \alpha_i^2} = \sum_{t=1}^T \frac{-f''_0(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) f_0(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) - (f'_0(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi}))^2}{f_0^2(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi})},$$

where again $f''_0(z; \boldsymbol{\psi}) = \partial^2 f_0(z; \boldsymbol{\psi}) / \partial z^2$. Note that we can write

$$-\frac{\partial^2 \ell_i(\alpha_i, \boldsymbol{\theta})}{\partial \alpha_i^2} = \sum_{t=1}^T q(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi}),$$

for a suitable $q(\cdot; \boldsymbol{\psi})$ function. Then

$$i_{\alpha_i\alpha_i}(\boldsymbol{\theta}, \alpha_i) = \sum_{t=1}^T \int_{-\infty}^{\infty} q(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) f_0(y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}; \boldsymbol{\psi}) dy_{it},$$

from which we make a change of variable $r_{it} = y_{it} - \alpha_i - \mathbf{x}_{it}^\top \boldsymbol{\beta}$, obtaining

$$i_{\alpha_i\alpha_i}(\boldsymbol{\theta}, \alpha_i) = \sum_{t=1}^T \int_{-\infty}^{\infty} q(r_{it}; \boldsymbol{\psi}) f_0(r_{it}; \boldsymbol{\psi}) dr_{it},$$

which shows that $i_{\alpha_i\alpha_i}(\boldsymbol{\theta}, \alpha_i)$ is a function of $\boldsymbol{\psi}$ only. Similarly, we obtain that the $i_{\boldsymbol{\theta}\alpha_i}(\boldsymbol{\theta}, \alpha_i)$ depends on $\boldsymbol{\psi}$ only, and so does $\rho_i(\boldsymbol{\theta}, \alpha_i)$.

References

- Arellano M, Bonhomme S (2009) Robust priors in nonlinear panel data models. *Econometrica* 77:489–536
- Bartolucci F, Bellio R, Salvan A, Sartori N (2016) Modified profile likelihood for fixed-effects panel data models. *Econ Rev* 35:1271–1289
- Bellio R, Grassetto L (2011) Semiparametric stochastic frontier models for clustered data. *Comput Stat Data Anal* 55:71–83
- Belotti F, Daidone S, Ilardi G, Atella V (2013) Stochastic frontier analysis using Stata. *Stata J* 13:719–758
- Belotti F, Ilardi G (2018) Consistent inference in fixed-effects stochastic frontier models. *J Econ* 202:161–177
- Chen Y-Y, Schmidt P, Wang H-J (2014) Consistent estimation of the fixed effects stochastic frontier model. *J Econ* 181:65–76
- González-Farías G, Domínguez-Molina A, Gupta AK (2004) Additive properties of skew normal random vectors. *J Stat Plan Inference* 126:521–534
- Greene WH (1990) A gamma-distributed stochastic frontier model. *J Econom* 46:141–163
- Greene WH (2003) Simulated likelihood estimation of the normal-gamma stochastic frontier function. *J Prod Anal* 19:179–190
- Greene WH (2005a) Fixed and random effects in stochastic frontier models. *J Prod Anal* 23:7–32
- Greene WH (2005b) Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *J Econ* 126:269–303
- Kumbhakar SC, Lai H-P (2022) Recent advances in the panel stochastic frontier models: Heterogeneity, endogeneity and dependence. *Int J Empir Econ* 01:2250002
- Kumbhakar SC, Lien G, Hardaker JB (2014) Technical efficiency in competing panel data models: a study of Norwegian grain farming. *J Prod Anal* 41:321–337
- Kutlu L, Tran KC, Tsionas MG (2020) Unknown latent structure and inefficiency in panel stochastic frontier models. *J Prod Anal* 54:75–86
- Liu Q, Pierce DA (1994) A note on Gauss-Hermite quadrature. *Biometrika* 81:624–629
- Pace L, Salvan A (1997) *Principles of statistical inference: from a neo-fisherian perspective* (World Scientific, Singapore).
- Ritter C, Simar L (1997) Pitfalls of normal-gamma stochastic frontier models. *J Prod Anal* 8:167–182
- Schumann M, Severini TA, Tripathi G (2021) Integrated likelihood based inference for nonlinear panel data models with unobserved effects. *J Econ* 223:73–95
- Stringer A, Bilodeau B (2022) Fitting generalized linear mixed models using adaptive quadrature. *arXiv*. 2202.07864.
- Tsionas EG (2000) Full likelihood inference in normal-gamma stochastic frontier models. *J Prod Anal* 13:183–205
- Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Stat Sin* 21:5–42
- Wang H-J, Ho C-W (2010) Estimating fixed-effect panel stochastic frontier models by model transformation. *J Econ* 157:286–296
- Wikström D (2015) Consistent method of moments estimation of the true fixed effects model. *Econ Lett* 137:62–69