European Journal of Operational Research 255 (2016) 272–287

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Interfaces with Other Disciplines

When, where and how to estimate persistent and transient efficiency in stochastic frontier panel data models

Oleg Badunenko[a,*], Subal C. Kumbhakar[b,c,1]

[a] Faculty of Management, Economics and Social Sciences, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany
[b] Department of Economics, State University of New York, Binghamton, NY 13902-6000, USA
[c] University of Stavanger Business School, Stavanger, Norway

ARTICLE INFO

Article history:
Received 22 January 2016
Accepted 27 April 2016
Available online 3 May 2016

Keywords:
(D) Production/Cost Function
Heterogeneity
Inefficiency
Closed-skew normal distribution
Simulated maximum likelihood

ABSTRACT

In this paper we examine robustness of a recently developed panel data stochastic frontier model that allows for both persistent and transient (also known as long-run and short-run or time-invariant and time-varying) inefficiency along with random firm-effects (heterogeneity) and noise. We address some concerns that the practitioners might have about this model. First, given that there are two random time-invariant components (persistent inefficiency and firm-effects) the concern is whether the model can accurately identify them, and if so how precisely can the model estimate them? Second, there are two time-varying random components (transient inefficiency and noise), and the concern is whether the model can separate noise from transient inefficiency, and if so how precisely can the model estimate transient inefficiency? Third, how well are persistent and transient inefficiency estimated under different scenarios, viz., under different configurations of the variance parameters of the four random components? Given that the model is quite complex, relatively new and becoming quite popular in the panel efficiency literature, we feel that there is need for a detailed simulation study to examine when, where and how one can use this model with confidence to estimate persistent and transient inefficiency.

1. Introduction

Now-a-days panel data models are extensively used in almost every area of microeconometric applications. Efficiency modeling is not an exception to this trend. In efficiency estimation panel models are increasingly used by the academicians (since the early 1980s) as well as in regulatory agencies in more recent years. For example, the regulators in the UK not only use panel data to increase number of observations and precision of the parameters in cases where cross-sectional units are small, many of them also use state-of-the-art efficiency models. Office of the Rail and Road in the UK use panel data models to examine both persistent and transient inefficiency. Other UK regulatory agencies (Ofwat, Ofgem, Royal Mail, etc.) are also interested in separating persistent inefficiency from firm-heterogeneity (often related to special factors), in addition to examining time-varying inefficiency in both price setting and merger cases.

In examining efficiency regulators often give special allowance to some companies because of their special production conditions, locations, etc. The allowance is somewhat ad hoc because it is not estimated from any formal economic/econometric model and therefore there is no way of knowing whether it captures firm heterogeneity or persistent inefficiency or both. It is perhaps better to decide on the special factor allowance from a formal model so that it becomes transparent to all the firms that are being regulated. Since the regulators all over the world use carrots and stick principle, it is also desirable for both the regulators and the firms being regulated to know whether carrots and sticks are equally applicable to persistent and transient inefficiency. For this, one needs to know where and when persistent inefficiency can be accurately estimated. The other important issue in regulatory cases is whether the regulated firms are improving their efficiency over time to attain the benchmark (catch-up effect). In doing so one has to estimate time-varying (transient) inefficiency and again it is important to know that the transient component is estimated accurately, so that no undue burden is placed on the firms being regulated in achieving a target that is not estimated accurately.

To address these issues the stochastic frontier model that was introduced recently (Colombi, Kumbhakar, Martini, & Vittadini, 2014; Kumbhakar, Lien, & Hardaker, 2014; Tsionas & Kumbhakar, 2014) has four components, viz., persistent and transient inefficiency, random firm-effects (firm heterogeneity) and noise. Because of the complexity of the model different estimation methods are

* Corresponding author. Tel.: +49 221 470 1285; fax: +49 221 470 2313.
E-mail addresses: obadunen@uni-koeln.de (O. Badunenko), kkar@binghamton.edu (S.C. Kumbhakar).
[1] Tel.: +1 607 777 4762; fax: +1 607 777 2681.

http://dx.doi.org/10.1016/j.ejor.2016.04.049

proposed. For example, Colombi et al. (2014) used a full maximum likelihood method, Kumbhakar et al. (2014) used a multi-step approach, Tsionas and Kumbhakar (2014) used a Bayesian approach and finally Filippini and Greene (2016) used the simulated maximum likelihood (SML) approach. In our simulations we use the SML method to estimate the models designed to address the concern 'when, where and how one can use this model with confidence to estimate persistent and transient inefficiency'.

Our results show that the reliability of estimation of persistent and transient technical efficiency critically depends on three estimated parameters, viz., (i) the ratio of the variance parameter in persistent technical inefficiency to the variance parameter in random effects, (ii) the ratio of the variance parameter in transient technical inefficiency to the variance parameter in noise, and (iii) the ratio of the variance parameter in persistent technical inefficiency to the variance parameter in transient technical inefficiency. Specifically, the estimator does a good job estimating persistent technical efficiency (transient technical efficiency) for relatively large values of the first (second) ratio. The third ratio plays a corrective role in the accuracy of the estimates.

It is important to note that in nearly all the cases the estimator can estimate either persistent or transient technical efficiency reliably. Only in the first and second cases when the variance parameters are relatively high (high variation of persistent technical inefficiency relative to random effects and high variation of transient technical inefficiency relative to noise) the estimator provides accurate technical efficiency estimates of both persistent and transient technical efficiency. If both ratios are relatively low, the estimator cannot be trusted for estimating either type of technical efficiency.

It is worth emphasizing that in practice the ratios are not known. Using four empirical examples, we provide a simple guide on how to judge the reliability of the obtained estimates. We show that the estimator is not consistent with some of the data sets which are used in efficiency analysis using other restrictive models.

The rest of the paper is organized as follows: Section 2 provides a description of the estimator. Section 3 provides the details of the Monte Carlo study as well as the results of the simulations. In Section 4 we apply the estimator on eight data sets to see how it performs in practice and relate the empirical results to those from simulations. The last section concludes.

2. Stochastic frontier model for panel data

The stochastic frontier (SF) model originally proposed by Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977) has traveled a long way since its inception. The panel version of the standard 1977 SF model (without any amendments) can be written as

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it} - \mathsf{p} \cdot u_{it} \tag{1a}$$

$$= \boldsymbol{x}_{it}\boldsymbol{\beta} + \epsilon_{it}, \tag{1b}$$

where $i = 1, \ldots, n$ denotes observation and $t = 1, \ldots, T_i$ denotes time period. In a SF frontier model, the outcome variable y_{it} is the logarithm of output, $\boldsymbol{x}_{it}$ is the row vector of a constant, logarithms of the input variables and possibly other observed covariates that include environmental variables that are not primary inputs but nonetheless affect output. p is a known parameter to distinguish between production and cost function models, viz.,

$$\mathsf{p} = \begin{cases} 1 & \text{for a stochastic production frontier model} \\ -1 & \text{for a stochastic cost frontier model.} \end{cases} \tag{2}$$

The random noise term v_{it} is assumed to be *i.i.d.* normal with zero mean and variance σ_v^2. Similarly, $u_{it} \geq 0$ is the time-varying technical inefficiency term which is assumed to be *i.i.d.* as half normal, that is, $u_{it} = |U_{it}|$, where U_{it} is *i.i.d.* normal with zero mean and variance σ_u^2. Note that this model is simply a pooled cross-sectional model with the additional subscript t which is redundant because of the *i.i.d.* nature of both noise and inefficiency.

In several papers, Kumbhakar (1991), Kumbhakar and Heshmati (1995), Kumbhakar and Hjalmarsson (1993, 1995) interpreted $u_{it} \geq 0$ as time-varying technical inefficiency and added an extra component $u_{0i} \geq 0$ to represent persistent inefficiency (in addition to the noise term v_{it}). In other words, in the models used by Kumbhakar and coauthors in the 1990s inefficiency is decomposed into two parts: persistent and time-varying, u_{0i} and u_{it}. The persistent component is consistent with the models used in the 1980s (Battese & Coelli, 1988; Kumbhakar, 1987; Pitt & Lee, 1981; Schmidt & Sickles, 1984), whereas the time-varying component is consistent with the models developed in the 1990s (Battese & Coelli, 1992; Cornwell, Schmidt, & Sickles, 1990; Kumbhakar, 1990) in which u_{it} is allowed to vary over time either by assuming it to be *i.i.d.* over i and t or making its mean/variance parameter a function of other exogenous variables varying over i and t. Quantifying the magnitude of persistent inefficiency is important, especially in short panels, because it reflects the effects of inputs like management (Mundlak, 1961) as well as other unobserved factors that vary across firms but not over time. Thus, unless there is a change in something that affects the management practices at the firm level (such as changes in ownership or new government regulations), it is unlikely that persistent inefficiency will change. Alternatively, time varying efficiency can change over time without operational changes in the firm.

There is, however, a philosophical question about interpreting u_{0i} as persistent inefficiency. Should one view it as the persistent inefficiency as in Kumbhakar (1991), Kumbhakar and Heshmati (1995), Kumbhakar and Hjalmarsson (1993, 1995) or as firm-heterogeneity that captures the effects of (unobserved) time-invariant covariates that has nothing to do with inefficiency? Mester (1997) for example argues that the estimates of efficiencies in stochastic frontier model are biased if heterogeneity is ignored. If one treats u_{0i}, $i = 1, \ldots, N$ as a random variable representing firm heterogeneity and is uncorrelated with $\boldsymbol{x}_{it}$ then the above three-component model becomes the 'true random-effects' (TRE) model (Greene, 2005).[2] Bos, Koetter, Kolari, and Kool (2009) account for sample heterogeneity by shifting the underlying technology. Lee (2010) also estimate different frontiers, but this is not the same as accounting for unobserved heterogeneity. Thus, the difference between the TRE and the models proposed by Kumbhakar and co-authors mentioned above is in the interpretation of the 'time-invariant' term, u_{0i}, i.e., whether it is persistent inefficiency or firm-effects.

2.1. Model that accounts for heterogeneity and persistent inefficiency

Recently Colombi et al. (2014), Kumbhakar et al. (2014), Tsionas and Kumbhakar (2014) introduced a model that split the error term into four components. The first component captures firms' latent heterogeneity (Greene, 2005) and the second component captures long-run (persistent) inefficiency as in Kumbhakar and Heshmati (1995), Kumbhakar and Hjalmarsson (1993, 1995), both of which are time-invariant. The third component captures short-run/transient/time-varying inefficiency while the last component captures random shocks. Both the third and fourth components

[2] Kumbhakar and Wang (2005) developed a similar model in which the firm-effects are treated as fixed but they modeled time-varying inefficiency in more general terms by allowing factors that can affect it.