

Interpreting Stochastic Frontier Model as Random Coefficient Model and Vice Versa

Spyros Missiakoulis¹

Abstract

This note explores the relationship between the stochastic frontier model and the random coefficient regression model. It shows how to interpret the former as a special case of the latter and vice versa.

JEL classification numbers: C13, C51, D24

Keywords: Stochastic production frontier, Random coefficient regression, Composite error, Technical inefficiency

1. The Stochastic Frontier Production Model

The stochastic frontier production model introduced by Aigner, Lovell and Schmidt (1977) and independently by Meeusen and Broeck (1977) relates the production function

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad (1)$$

with the composite error specification

$$\varepsilon_i = v_i - u_i, \quad (2)$$

where y_i measures the output, x_{ij} measures the j^{th} input, β_0 is the unknown intercept, β_j is the j^{th} unknown slope parameter, v_i represents the usual two-sided distribution term with $v_i \sim NID(0, \sigma_v^2)$, and u_i , which is distributed independently of v_i and satisfies $u_i \geq 0$, represents the technical errors of the firm. The most commonly assumed distribution of u_i is the (positive) half-normal. This is not a restrictive assumption as any other suitable (positive) distribution of u_i may be considered. The technical inefficiency relative to the stochastic frontier is then given by u_i . Model (1) with (2) will be referred to as the ALS model. Excellent reviews of the ALS model and the frontier literature, in general, may be found in Forsund *et al.* (1980) and Kumbhakar and Lovell (2000).

All unknown parameters of the ALS model are estimated either by Corrected Ordinary Least Squares, or Maximum Likelihood (Aigner *et al.* 1977; Olson *et al.* 1980; Lee 1983; Greene 1990; Greene 2010), with each method having its pros and cons.

The main weakness of the ALS model was the impossibility to determine the individual technical inefficiencies. Although the mean technical inefficiency over the

¹ Philips University, Nicosia, Cyprus. email: s.missiakoulis@gmail.com

sample is estimable, it was not possible to decompose each residual into its two components. Jondrow *et al.* (1982) suggested a solution to this problem by considering the expected value of u_i , conditional upon ε_i . Under the half-normal assumption for u_i , individual coefficients are estimated either by

$$u_i = \frac{\sigma_u \sigma_v}{\sqrt{\sigma_u^2 + \sigma_v^2}} \left(\frac{f(z_i)}{1 - F(z_i)} \right), \quad (3)$$

where

$$z_i = \frac{\varepsilon_i \sigma_u}{\sigma_v \sqrt{\sigma_u^2 + \sigma_v^2}},$$

and $f(\cdot)$ and $F(\cdot)$ are the standard normal density and distribution functions respectively; or by

$$\begin{aligned} u_i &= \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \varepsilon_i & \text{if } \varepsilon_i \leq 0 \\ &= 0 & \text{if } \varepsilon_i > 0. \end{aligned} \quad (4)$$

The former is based on the **mean** of u_i conditional on ε_i , and the latter is based on the **mode** of u_i conditional on ε_i .

2. The Random Coefficient Model

The conventional random coefficient regression model is given by

$$y_i = \beta_{i0} + \sum_{j=1}^k \beta_{ij} x_{ij}, \quad (5)$$

with

$$\beta_{ij} = \beta_j + w_{ij} \quad j = 0, 1, \dots, k, \quad (6)$$

where all the w 's are assumed as $w_{ij} \sim NID(0, \sigma_j^2)$, $j=0, \dots, k$. Note, if it is desired, an error term may be also introduced in (5). We may now rewrite (5) as

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \quad (7)$$

where now

$$\varepsilon_i = w_{i0} + \sum_{j=1}^k x_{ij} w_{ij}. \quad (8)$$

Maximum Likelihood and Generalized Least Squares methods have been suggested for the estimation of β_j , $j=0, \dots, k$ and their corresponding variances (Swamy, 1971; Raj and Ullah, 1981; Kreft and de Leeuw 1998).

3. The equivalence of the two models

Consider now the random coefficient model in (5) or (7) with only one difference, namely, w_{i0} in (6) or (8) is not normally distributed but follows the one-sided distribution of $-u_i$ in (2). In other words, by setting $w_{i0} = -u_i$ and $\sum x_{ij} w_{ij} = v_i$ the model in (7) and (8) is nothing else but the ALS model. Of course, we may argue the other

way. That is, if we assume that $u_i = -w_{i0}$ and $v_i = \sum x_{ij}w_{ij}$, the ALS model is a special case of the random coefficient regression model. Since each w_{ij} , $j=1, \dots, k$, is normally distributed, then v_i is also normally distributed.

Griffiths (1972) showed that under the assumption, each residual is allocated between its components in the same proportion as the variance of ε_i is allocated between the variance of each error component, GLS estimates of each individual coefficient are minimum variance, linear, unbiased estimates.

Adopting Griffiths' assumption we may construct technical inefficiencies as

$$u_i = -\frac{\text{Var}(u_i)}{\text{Var}(u_i)+\text{Var}(v_i)} \varepsilon_i = -\frac{(\pi-2)\sigma_u^2}{(\pi-2)\sigma_u^2+\pi\sigma_v^2} \varepsilon_i \quad \text{if } \varepsilon_i \leq 0$$

$$= 0 \quad \text{if } \varepsilon_i > 0, \quad (9)$$

If, however, we also adopt the ALS interpretation that the relative variability of the two sources of random error is a function of σ_u^2 and σ_v^2 and not of $\text{Var}(u_i)$ and $\text{Var}(v_i)$ then (9) is written as

$$u_i = \frac{\sigma_u^2}{\sigma_u^2+\sigma_v^2} \varepsilon_i \quad \text{if } \varepsilon_i \leq 0$$

$$= 0 \quad \text{if } \varepsilon_i > 0, \quad (10)$$

which is exactly the same as (4).

Although this note focuses on the frontier production model, its generalization for all frontier models, i.e. cost or profit frontiers, is trivial.

References

- [1] Aigner, D.J., Lovell, C.A.K. and Schmidt, P. (1977). Formulation and Estimation of Stochastic Frontier Production Models. *Journal of Econometrics*, 6, pp. 21-37.
- [2] Forsund, F.R., Lovell, C.A.K. and Schmidt, P. (1980). A survey of frontier production functions and their relationship to efficiency measurements. *Journal of Econometrics*, 13, pp. 5-26.
- [3] Greene, W.H. (1990). A gamma distributed stochastic frontier model. *Journal of Econometrics*, 46, pp. 141-164.
- [4] Greene, W.H. (2010). A stochastic frontier model with correction for sample selection. *Journal of productivity analysis*, 34, pp. 15-24.
- [5] Griffiths, W.E. (1972). Estimation of actual response coefficients in the Hildreth-Houck random coefficient model. *Journal of the American Statistical Association*, 67, pp. 633-635.
- [6] Jondrow, J., Lovell, C.A.K., Materov, I.S. and Schmidt, P. (1982). On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics*, 19, pp. 233-238.
- [7] Kreft, I.G. & de Leeuw, J. (1998). Varying and random coefficient models. In *Introducing multilevel modeling*, SAGE Publications, pp. 35-56.
- [8] Kumbhakar, S. and Lovell, C.A.K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.
- [9] Lee, L.F. (1983). On maximum likelihood estimation of stochastic frontier production models. *Journal of Econometrics*, 23, pp. 269-274.
- [10] Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed errors. *International Economic Review*, 18, pp. 435-444.

- [11] Olson, J.A., Schmidt, P. and Waldman M. (1980). A Monte Carlo study of estimators of stochastic frontier production functions. *Journal of Econometrics*, 13, pp. 67-82.
- [12] Raj, B. and Ullah, A. (1981). *Econometrics: A Varying Coefficient Approach*, Croom Helm, London.
- [13] Swamy, P.A.V.B. (1971). *Statistical Inference in Random Coefficient Regression Models*, Springer-Verlag, Berlin.