Maximum likelihood estimation of a stochastic frontier model with residual covariance

Simwaka, Kisu

Reserve Bank of Malawi

27 June 2012

Maximum Likelihood Estimation of a Stochastic Frontier Model with Residual Covariance

By

Kisu Simwaka

Reserve Bank of Malawi

Abstract

In theoretical literature on productivity, the disturbance terms of the stochastic frontier model are assumed to be independent random variables. In this paper, we consider a stochastic production frontier model with residuals that are both spatially and time-wise correlated. We introduce generalizations of the Maximum Likelihood Estimation procedure suggested in Cliff and Ord (1973) and Kapoor (2003). We assume the usual error component specification, but allow for possible correlation between individual specific errors components. The model combines specifications usually considered in the spatial literature with those in the error components literature. Our specifications are such that the model's disturbances are potentially spatially correlated due to geographical or economic activity. For instance, for agricultural farmers, spatial correlations can represent productivity shock spillovers, based on geographical proximity and weather. These spillovers effect estimation of efficiency.

JEL Classification Codes: C21 C23 C24

Keywords: spatial stochastic production frontier models, correlated errors

1.0 Introduction

Estimation of the stochastic frontier function was simultaneously introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977). The production functions specify the maximum potential output levels, given the quantities established for a set of inputs. Aigner et al. (1977) specified a production function with two error terms.

The stochastic frontier methodology has subsequently been extended in many directions using both cross-sectional and panel data. One advantage of using panel data is that it gives opportunity to examine and model behaviour of technical efficiency over time. The earlier models (Pitt and Lee, 1981; Schmidt and Sickles, 1984; Kumbhakar, 1987; among others) treated technical efficiency as time-invariant. Subsequent researchers allowed the technical efficiency to vary over time (Kumbhakar 1990; Cornell, Schmidt, and Sickles, 1993; Lee and Schmidt, 1993; Battesse and Coelli, 1992; and Battese and Coelli 1995). However, none of these allowed error components to be spatially and time-wise correlated.

Meanwhile, interest has been growing for spatial econometrics in recent years. Anselin provides an excellent textbook treatment of the analysis of spatially dependent data. In the agricultural sector, farms in different geographical regions may also differ in their efficiency patterns owing to geographical proximity, differences in education, and access to technology. A common procedure in spatial econometrics is to model interactions between cross-sectional units in terms of some distance measure between them. Thus distance can be modeled by geographic measures as e.g., the physical distance between two regions, or by economic measures as e.g., economic similarities between regions. By far, the most widely used spatial models are variants of the ones considered in Cliff and Ord (1973, 1981). One method of estimation of these models is maximum likelihood, (ML) suggested in Cliff and Ord (1973), and Kapoor (2003). Kelejian and Prucha (1998) suggested an alternative instrumental variable estimation procedure for these models, which is based on a generalized moments (GM) estimator of a parameter in the spatial autoregressive process. Monte Carlo results in Das, Kelejian, and Prucha (2003) suggest that both the GMM and the instrumental variable estimators are “virtually” as efficient as the corresponding ML estimators in small samples.

1.1 Problem Statement

The analysis of spatial processes is of significance in many disciplines, including agriculture and banking. Spatial autocorrelation occurs when population members are related through their geographical positions. In the productivity modelling literature, the disturbance terms of the stochastic frontier model are assumed to be independent random variables. It is well known that inference can be incorrect when the data is characterized by spatial correlation. This is particularly so in the analysis of spatial data when correlation may exist between neighbouring entities. When one begins to look at cross-section of regions, states, etc, these aggregate units may exhibit cross-sectional correlation that has to be dealt with. Clearly, the asymptotic results developed for the spatial models so far are no longer appropriate in the case of heteroscedastic innovations. Ignoring cross-sectional dependence when in fact it exits, results in biased, inconsistent and inefficient estimates of regression coefficients. The nature of the covariance among residuals will usually not be known precisely, but it is often possible to adopt a simple parametric model to describe it. The approach we adopt, through maximum likelihood, is similar to the earlier studies by Cliff and Ord (1973) and Kapoor (2003).

1.2 Objective of the Study

In this paper, we evaluate the maximum likelihood approach for estimating a stochastic production frontier model. We generalize the maximum likelihood estimation procedure suggested in Cliff and Ord (1973) and Kapoor, (2003) to allow for spatial correlation in the context of a stochastic production frontier model. We discuss how the likelihood function may be numerically maximized, giving suitable formula for the derivatives and information matrix.

The stochastic production frontier model presented in this paper differs from the traditional models in two ways. First, we assume the usual error component specification, but allow for possible correlation of the individual specific errors components. Our specifications are such that the model's disturbances are potentially spatially correlated based on geographical or economic activity. For instance, for agricultural farmers, spatial correlations can represent productivity shock spillovers, based on geographical proximity and weather. These spillovers effect estimation of efficiency. In spatial models, interactions between cross sectional units are typically modelled in terms of some measure of distance between them. Second, we assume that error components are time-wise autocorrelated. These specifications merge those typically considered in the spatial literature with those in the error component literature.

In section 2.0, we specify a stochastic production frontier model with spatially correlated residuals and discuss its properties. In Section 3.0, we discuss how the likelihood function may be computed and numerically maximized, giving suitable formulae for the derivatives.

2.0 Model Specification and Assumptions

The stochastic frontier model considered here is specified as follows:

$$Y_{it,N} = x'_{it,N}\beta + v_{it} - u_{it,N} \qquad i=1,\ldots,N;\ \ t=1,\ldots T, \qquad (1.1)$$

where $y_{it,N}$ is the *N* x *1* vector of observations on dependent variable in time period *t,* $x_{it,N}$ represents a vector of exogenous variables in period *t;* β is a vector of parameters of the production function to be estimated, v_{it} is the first component of the error term, representing random effects, and follows a stationary $AR(\lambda)$ process; $u_{it,N}$ is the second component of the error term, representing technical efficiency in production $u_{it,N}$; assumed to be *i.i.d* and follow a truncated normal distribution (with truncations at zero) and mean $z_{it}\delta$ and variance, $\sigma^2 . u_{it}s$ are assumed to be a function of a set of independent variables, the $z_{it}s$ and unknown set of coefficients, δ.

We follow Battese and Coelli (1995) who specify the technical inefficiency effects as follows:

$$u_{it} = z_{it}\delta + w_{it} \qquad (1.2)$$

where z_{it} represent independent variables that determine technical inefficiency, δ is an *(M * 1)* representing coefficients, and w_{it} s represent technical efficiency.

A popular approach to model spatial dependence is that of Cliff and Ord (1973, 1981). We follow this approach and specify v_{it} in each time period $t = 1,\ldots.,T$ as a first order spatial autoregressive process.

$$v_{it,N} = \lambda W_N v_{it,N} + \varepsilon_{it,N} \qquad (1.3)$$

where λ is the scalar spatial autoregressive coefficient which is assumed to lie in the parameter space $|\lambda|<1$; the matrix W is an *N* x *N* spatial weight matrix of constants, which represents the degree of potential interactions between neighbouring locations, whose diagonal elements are zero and off-diagonal elements are non-zero, w_{ij} are chosen to reflect the degree of dependence between the error of unit *i* and the error of unit *j*; and $\varepsilon_{it,N}=[\varepsilon_{1t,N},...,\varepsilon_{NT,N}]'$ is an N x 1 vector of residuals in period *t*. In the following analysis, we maintain that the weighting matrix W_N does not change over time.

Combining observations in (1.1) and (1.3) we have

$$y_N = X_N\beta + v_N - u_N \tag{1.4}$$

and

$$v_N = \lambda(I_T \otimes W_N)v_N + \varepsilon_N \tag{1.5}$$

where $y_{it,N} = [y_{1t,N}',..., y_{NT,N}']'$, $X_{it,N} = [X_{1t,N})',..., X_{Nt,N})']'$, $v_N = [v_{1t,N}(1)',..., v_{N,TN}']'$, and $\varepsilon_N = [\varepsilon_N(1)',...,\varepsilon(T)']'$.

To allow for errors to be correlated over time, we assume the following error component structure for the vector of errors ε_N

$$\varepsilon_N = (e_T \otimes I_N)\upsilon_N + \omega_N \tag{1.6}$$

where ε_T represents the vector of unit specific error components, I_N is an identity matrix of order *N*, and $\mu_N = [\mu_{1,N},...,\mu_{N,N}]'$ represents the vector of unit specific error components, and $\omega_N = [\omega_{1,N},...,\omega_{T,N}]'$ where $\omega_{it,N} = [\omega_{1t,N},...,\omega_{NT,N}]'$ contains the error components that vary over both the cross-sectional units and time periods. In scalar notations, we have

$$\varepsilon_{it,N} = \upsilon_{i,N} + \omega_{it,N} \qquad\qquad i = 1,...,N\,;\; t = 1,...,T$$

It must be noted that the specification of ε_{it} corresponds to that of the classical-one way error component literature. In contrast, however, we group the data by time periods rather than units because this grouping is more convenient for modeling spatial correlation through 1.2.

We retain the following assumptions from the classical error component literature

Assumption 1: Let T be a fixed positive integer, and for the error components assume :(a) For all $1 \le t \le T$ and $1 \le N$; $N \ge 1$ the error components $\omega_{it,N}$ are independently and identically distributed with zero mean and variance σ_{ω}^2, $0 < \sigma_{\omega}^2 < b_{\omega} < \infty$ and finite fourth moments. In addition, for each $N \ge 1$ and $1 \le t \le T$, $1 \le i \le N$ the error components. (b) For all $1 \le i \le N$; $N \ge 1$ the unit specific error components $\upsilon_{i,N}$ are identically distributed with zero means and variance σ_{υ}^2, $0 < \sigma_{\upsilon}^2 < b_{\upsilon} < \infty$ and finite fourth moments. In addition, for each $N \ge 1$ and $1 \le i \le N$ the unit specific error $\upsilon_{i,N}$ are independently distributed; (c) The process $\mu_{it,}$ and $\omega_{it,}$ are independent.

Assumption 2 (a) All diagonal elements of W_N are zero; (b)$|\lambda|$<1; (c) The matrix $I_N - \lambda W_N$ is non-singular for all$|\lambda|$<1.

In scalar notation, the specification in (1.3) is

$$v_{it,N} = \lambda \Sigma_{j=1}^{N} w_{ij,N} v_{jt,N} + \varepsilon_{it,N}, \qquad \text{i=1,…,N; t=1,…,}T$$

where $w_{ij,N}$ is the *(i, j)-th* element of the weighting matrix W_N. The non-zero weights $w_{ij,N}$ are usually specified to be those that correspond to units that are significantly related. Such units are said to be neighbours. For instance, if the cross-sectional units are geographical regions, one can make $w_{ij,N} \ne 0$ if the i*-th* and j*-th* regions are neighbouring, and $w_{ij,N} = 0$ otherwise. For reasonable time periods, one can assume that this relationship remains unchanged i.e.., $w_{ij,N}$ is constant through time.

2.1 Assumption Implications

Given the above assumptions, and in line with Kapoor (2003), it follows from (1.6) that $E(\varepsilon_N) = 0$ and the innovations of $\varepsilon_{it,N}$ are autocorrelated over time, but are not spatially correlated across units, and the covariance vector matrix of the vector of ε_N is

$$\begin{aligned} E(\varepsilon_N \varepsilon'_N) = \Omega_{\varepsilon,N} &= \sigma_{\upsilon}^2 (J_T \otimes I_N) + \sigma_{\omega}^2 I_{NT} \\ &= \sigma_v^2 Q_{0,N} + \sigma_1^2 Q_{1,N} \qquad (1.7) \end{aligned}$$

where $\sigma_1^2 = T\sigma_{\upsilon}^2 + \sigma_{\omega}^2$ and

$$Q_{1,N} = \frac{J_T}{T} \otimes I_N \text{ ; and } Q_{0,N} = I_{NT} - Q_{1,N} = (I_T - \frac{J_T}{T}) \otimes I_N \qquad (1.8)$$

where $J_T = e_T e'_T$ is a T x T matrix of unit elements, and I_K is an identity matrix of order K. The matrices $Q_{0,N}$ and $Q_{1,N}$ are standard transformation matrices utilized in error component literature (see Baltagi 2005).The matrices $Q_{0,N}$ and $Q_{1,N}$ are symmetric idempotent, orthogonal and orthogonal to each other. Furthermore $Q_{0,N} + Q_{1,N}$.

$$Q_{1,N}Q_{1,N} = (\frac{J_T}{T} \otimes I_N)(\frac{J_T}{T} \otimes I_N) = (\frac{J_T J_T}{T^2} \otimes I_N) = \frac{e_T e_{TT}^{'} e_T^{'}}{T^2} \otimes I_N$$

$$= (\frac{e_T e_T^{'}}{T} \otimes I_N) = (\frac{J_T}{T} \otimes I_N) = Q_{1,N} \qquad (1.9)$$

$$Q_{0,N}Q_{0,N} = (I_{NT} - Q_{1,N})(I_{NT} - Q_{1,N})$$

$$= I_{NT} - Q_{1,N} - Q_{1,N} + Q_{1,N}Q_{1,N}$$

$$= I_{NT} - Q_{1,N} - Q_{1,N} + Q_{1,N} = I_{NT} - Q_{1,N} = Q_{0,N},$$

$$Q_{0,N}Q_{1,N} = (I_{NT} - Q_{1,N})Q_{1,N} = Q_{1,N} - Q_{1,N} = 0,$$

$$Q_{0,N} + Q_{1,N} = I_{NT}$$

Observe that the elements $Q_{0,N}$ and $Q_{1,N}$ are uniformly bounded by 1. It will be necessary to prove that for any N x N matrix A_N, we have

$$(I_T \otimes A_N)Q_{0,N} = Q_{0,N}(I_T \otimes A_N),$$

$$(I_T \otimes A_N)Q_{1,N} = Q_{1,N}(I_T \otimes A_N) \qquad (1.10)$$

The proof is contained in Kapoor (2003):

Given that $Q_{0,N} = (I_T - \frac{J_T}{T}) \otimes I_N$ and $Q_{1,N} = (I_T - \frac{J_T}{T}) \otimes I_N$

We have $(I_T \otimes A_N)Q_{0,N} = (I_T - \frac{J_T}{T}) \otimes A_N = ((I_T - \frac{J_T}{T}) \otimes I_N)(I_T \otimes A_N) = Q_{0,N}(I_T \otimes A_N)$.

In like manner, $(I_T \otimes A_N)Q_{1,N} = \frac{J_T}{T} \otimes A_N = (\frac{J_T}{T} \otimes I_N)(I_T \otimes A_N) = Q_{1,N}(I_T \otimes A_N)$

From 1.5, it follows that

$$v_N = [I_T \otimes (I_N - \lambda W_N)^{-1}]\varepsilon_N \qquad (1.11)$$

Thus $Ev_N = 0$ and considering (1.7) and (1.10)

$$Ev_N v_N^{'} = \Omega_{v,N}(\lambda) = [I_T \otimes (I_N - \lambda W_N)^{-1}]\Omega_{\varepsilon,N}[I_N - \lambda W_N^{'})^{-1}]$$

$$=\Omega_{\varepsilon,N}[I_T \otimes (I_N-\lambda W_N)^{-1}(I_N-\lambda W_N^{'})^{-1}] \tag{1.12}$$

Please note that in general, the elements of $(I_N-\lambda W_N)^{-1}$ will depend on the sample size of the cross-sectional units *N*. As a result, the elements of v_N will depend on *N* and therefore form a triangular array. In general, the elements of $\Omega_{v,N}(\lambda)$ will depend on *N*. Additionally, the elements of v_N are heteroskedastic, and spatially correlated, as well as correlated over time. In the following sections we explore the estimation strategies for the parameters of the model considered in (1.4), (1.5) and (1.6).

3.0 Quasi-Maximum Likelihood Estimation

Maximum likelihood (ML) estimation is a well-known parametric method of inference in statistics. It has been frequently been suggested as a way of estimating covariance parameters in spatial Gaussian processes. We follow quasi-maximum likelihood (ML) estimation approach used in Cliff and Ord (1973) and Kapoor (2003).

Remember our model in stacked form, from (1.4)-(1.6)

$$y_N = X_N\beta + v_N - u_N$$

$$v_N = \lambda(I_T \otimes W_N)v_N + \varepsilon_N$$

$$\varepsilon_N = (e_T \otimes I_N)\upsilon_N + \omega_N$$

Assuming $\varepsilon_N \sim N(0,\Omega_{u,N}(\lambda))$ we have $v_N \sim N(0,\Omega_{v,N}(\lambda))$

Thus $y_N \sim N(X\beta,\Omega_{u,N}(\lambda))$ (1.13)

By substituting (1.7) and (1.8) into (1.12) we get

$$\Omega_{v,N}(\lambda)=\sigma_\omega^2[I_T-\frac{J_T}{T})\otimes(I_N-\lambda W_N)^{-1}(I_N-\lambda W_N^{'})^{-1}]+\sigma_1^2[\frac{J_T}{T}\otimes(I_N-\lambda W_N)^{-1}(I_N-\lambda W_N^{'})^{-1}]$$

$$=[(\sigma_\omega^2(I_T-\frac{J_T}{T})+\sigma_1^2\frac{J_T}{T})\otimes(I_N-\lambda W_N)^{-1}(I_N-\lambda W_N^{'})^{-1}] \tag{1.14}$$

and hence

$$\det(\Omega_{v,N}(\lambda))=\det[(\sigma_\omega^2[I_T-\frac{J_T}{T})+\sigma_1^2\frac{J_T}{T})\otimes(I_N-\lambda W_N)^{-1}(I_N-\lambda W_N^{'})^{-1}]$$

$$= \det[(\sigma_\omega^2[I_T - \frac{J_T}{T}) + \sigma_1^2 \frac{J_T}{T})]^N [\det(I_N - \lambda W_N)^{-2T} \quad (1.15)$$

Given (1.7) and (1.8), then it follows that

$$\Omega_{\varepsilon,N}^{-1} = \sigma_\omega^{-2} Q_{0,N} + \sigma_1^{-2} Q_{1,N} \quad (1.16)$$

thus from (1.12),

$$\Omega_{\omega,N}^{-1}(\lambda) = [I_T \otimes (I_N - \lambda W_N^{'})(I_N - \lambda W_N)]\Omega_{\varepsilon,N}^{-1}$$

$$= [\sigma_\omega^{-2}(I_T - \frac{J_T}{T}) + \sigma_1^{-2} \frac{J_T}{T}) \otimes (I_N - \lambda W_N^{'})(I_N - \lambda W_N)] \quad (1.17)$$

Assuming (1.15) the likelihood function for the model in (1.4)-(1.6) is given by

$$L = (2\pi)^{-NT/2} \,|\det(\Omega_{\omega,N}^{-1}(\lambda))|^{1/2} \exp(-\frac{1}{2}[y_N - X\beta]^{'} \Omega_{\omega,N}^{-1}(\lambda)[y_N - X\beta])$$

$$= (2\pi)^{-NT/2} \,|\det(\sigma_\omega^2(I_T - \frac{J_T}{T}) + \sigma_1^2 \frac{J_T}{T})]|^{-N/2} * \quad (1.18)$$
$$|\det(I_N - \lambda W_N)^T \exp(-\frac{1}{2}[y_N - X_N\beta]^{'} \Omega_{\omega,N}^{-1}(\lambda)[y_N - X_N\beta])$$

Substituting (1.15) and (1.17) into (1.19) and then taking the logs gives us the log likelihood function

$$\ln(L) = -\frac{NT}{2}\ln(2\pi) - \frac{N}{2}\ln|\det(\sigma_\omega^2(I_T - \frac{J_T}{T}) + \sigma_1^2 \frac{J_T}{T})| +$$

$$T\ln|\det(I_N - \lambda W_N| - \frac{1}{2}[y_N - X_N\beta]^{'}(\sigma_\omega^{-2}(I_T - \frac{J_T}{T}) + \sigma_1^2 \frac{J_T}{T}) \otimes \quad (1.19)$$
$$(I_N - \lambda W_N)(I_N - \lambda W_N)][y_N - X_N\beta]$$

Equation (1.19) represents quasi-Maximum Likelihood estimators. The computation of quasi-Maximum Likelihood estimators involves repeated evaluation of the determinant of the $N \times N$ matrix $I_N - \lambda W_N$. In order to reduce the computational burden, Ord (1975) suggested that $\ln|\det(I_N - \lambda W_N|$ in (1.19) be determined as $\ln|\det(I_N - \lambda W_N| = \Sigma_{i=1}^N \ln(|1 - \lambda\psi_i|$ where ψ_i denotes the i-th eigenvalue of W_N. Since W_N is a known matrix its eigenvalues have to be computed only once at the outset of the numerical optimization procedure employed in finding the quasi Maximum Likelihood estimates and not repeatedly at each of the necessary numerical iterations.

References

Aigner, D., C.A.K. Lovell, and P. Schmidt. "Formulation and Estimation of Stochastic Frontier Production Function Models." *Journal of Econometrics* 6 (July 1977): 21-37.

Baltagi, B.H., 2005. Econometric Analysis of Panel Data. Wiley, New York

Battese, G.E., and T.J. Coelli. (992). "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India." *Journal of Productivity Analysis* 3 (1992): 149-165.

Battese, G E & Coelli, T J, (1995). "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data," Empirical Economics, Springer, vol. 20(2), pages 325-32.

Cliff, A. and J.K Ord. 1973. Spatial autocorrelation. London: Pion.

Cliff, A. and J.K. Ord. 1981. Spatial processes: Models and applications. London: Pion.

Cornwell, C., P. Schmidt, and R.C. Sickles.(1990) "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels." *Journal of Econometrics* 46: 185-200.

Das, D., Kelejian, H.H., Prucha, I.R., 2003. Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances. Papers in Regional Science 82, 1–26.

Kapoor, M. (2003). Panel Data Models with Spatial Correlation: Estimation Theory and an Empirical Investigation of the US Wholesale Gasoline Industry, Department of Economics, University of Maryland, USA.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99–121.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533

Kelejian, H.H. and Prucha, I.R., (2007). HAC Estimation in a Spatial Framework. Journal of Econometrics 140, 131-154

Kumbhakar, S.C. (1990). "Production Frontiers, Panel Data, and Time-Varying Technical Inefficiency." *Journal of Econometrics* 46: 201-211.

Lee, Y.H., and P. Schmidt. "A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency." The Measurement of Productive Efficiency: Techniques and Applications. H.O. Fried, C.A.K. Lovell, and S.S. Schmidt, eds., pp. 237-255. New York, Oxford: Oxford University Press, 1993.

Meeusen, W., and J. van den Broeck. (1977) "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error." *International Economic Review* 18: 435-444

Pitt, M.M. and L. Lee. (1981), "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry", *Journal of Development Economics*, 9, 43-64.

Schmidt, P., and R.C. Sickles. (1984)"Production Frontiers and Panel Data." Journal of Business and Economic Statistics 2: 367-374.