Philippine Institute for Development Studies
Surian sa mga Pag-aaral Pangkaunlaran ng Pilipinas

Spatial Stochastic Frontier Models

Erniel B. Barrios and Rouselle F. Lavado

DISCUSSION PAPER SERIES NO. 2010-08

May 2010

For comments, suggestions or further inquiries please contact:

The Research Information Staff, Philippine Institute for Development Studies
5th Floor, NEDA sa Makati Building, 106 Amorsolo Street, Legaspi Village, Makati City, Philippines
Tel Nos: (63-2) 8942584 and 8935705; Fax No: (63-2) 8939589; E-mail: publications@pids.gov.ph
Or visit our website at http://www.pids.gov.ph

SPATIAL STOCHASTIC FRONTIER MODELS

By Erniel B. Barrios (correspondence)
School of Statistics, University of the Philippines, Magsaysay Ave., Diliman, Quezon City, Philippines; erniel.barrios@up.edu.ph, ernielb@yahoo.com
Tel./Fax. No.:(63-2)928-08-81
and
Rouselle F. Lavado
Philippine Institute for Development Studies, NEDA sa Makati Building, 106 Amorsolo St. Legaspi Village, Makati City; rlavado@mail.pids.gov.ph
Tel.: (63-2) 893-9585 to 87 loc 3112; Fax: (63-2) 816-1091

Summary: The stochastic frontier model with heterogeneous technical efficiency explained by exogenous variables is augmented with a sparse spatial autoregressive component for a cross-section data, and a spatial-temporal component for a panel data. An estimation procedure that takes advantage of the additivity of the model is proposed, computational advantages over simultaneous maximum likelihood estimation of all parameters is exhibited. The technical efficiency estimates are comparable to existing models and estimation procedures based on maximum likelihood methods. A spatial or spatial-temporal component can improve estimates of technical efficiency in a production frontier that is usually biased downwards.

Keywords: stochastic frontier models, technical efficiency, spatial externalities, spatial-temporal model, backfitting

1. INTRODUCTION

Traditional econometric modeling aims to explain the output indicator y_t in terms of determinants, say x_t. The error term or the difference between the predicted value $\hat{y}_t$ from the actual value y_t, is attributed to other unaccounted determinants, random errors that cannot be account by x_t through the specified model, or simply due to model misspecification. The equilibrium assumption also implies that the producer always aims to optimize the output, further implying that the error is a random occurrence. In reality, some producers may not be efficient enough in utilizing their factors of production resulting to the discrepancy (inefficiency) between their actual output and the expected optimum output. The discrepancy in output can further be explained by some exogenous factors peculiar in each producer. The result is a stochastic frontier model where the error term in standard econometric specification is decomposed further into those that can be explained by exogenous factors characterizing the producer's inability to maximize the output, and the pure error term. Stochastic Frontier Analysis (SFA) helps explain producer's efficiency and provide an alternative paradigm to econometric analysis when some assumptions fail.

Although the producers by default are the firms, (Amos *et. al.*, 2004), used SFA in studying productivity and technical efficiency of small-scale farmers (producing units). Empirical evidence of the usual assumption that technical efficiency of farmers engaged in mixed crops are generally higher than those propagating only one crop at a time were generated. The notion of a producing unit has been liberally defined in various applications of SFA.

The literature of SFA initially focused on applications in various production setups, but estimates of efficiency were noted to vary with model specification and distributional assumptions on the error terms. The non-robust estimates of efficiency should be addressed though a model specification that describes the production process more realistically.

The literature of statistical modelling on the other hand continues to postulate new models that aim to describe reality in as vivid mathematical abstractions as possible. Spatial and temporal dependence has initially been treated separately. However, as more panel data becomes available and the realization of the potential multiplier effect in information-generation capability of the interaction of space and time, there is a growing interest in spatial-temporal models. A purely spatial model usually has no causative component in it; such models are useful when space-time process has reached temporal equilibrium, or when short-term causal effects are aggregated over a fixed time period (Cressie, 1993). A spatial-temporal model then provides a more flexible alternative to postulate a model.

In the simultaneous treatment of space and time, estimation procedures become more complicated. (Richardson *et. al.*, 1992) estimated a spatial linear model with autocorrelated errors, where the spatial and temporal dependencies of the observations yield a general form of the variance-covariance matrix, hence least squares estimate is weighted by elements of the variance-covariance matrix. Iteratively estimated general least squares (EGLS) method was used to sequentially estimate the parameters and the elements of the variance-covariance matrix. More complicated estimation procedures are proposed continuously, and the trend is to develop simpler, yet competitive estimation procedures. The backfitting algorithm initially proposed to estimate an additive model by (Hastie and Tibshirani, 1990) provides simple alternative to the least square or maximum likelihood-based estimation procedures. The same algorithm has been used to simplify the estimation procedure for a spatial-temporal model, see for example, (Landagan and Barrios, 2006).

We proposed to augment the stochastic frontier model with a sparse autoregressive component for cross-section data, and a spatial-temporal component for a panel data. The backfitting algorithm will be modified in estimating both models.

2. STOCHASTIC FRONTIER ANALYSIS

The extensive literature on SFA has been summarized comprehensively by (Kumbhakar and Lovell, 2000). A cross-sectional production frontier model is given by:

(1) $$y_i = f\left(x_i;\beta\right)\exp\left(v_i\right)TE_i \text{ or } TE_i = \frac{y_i}{f\left(x_i;\beta\right)\exp\left(v_i\right)}$$

where y_i is the single output of producer i, x_i is the vector of inputs used in producing y_i, f is a parametric function, TE$_i$ is the output-oriented technical efficiency of producer i, and v_i is a random error. There is perfect efficiency when TE=1, while inefficiency when

TE<1. The shortfall in production environment characterized by $\exp(v_i)$ varies across producers. Let $TE_i = \exp(-u_i)$, then the production stochastic frontier model becomes $y_i = f(x_i;\beta)\exp(v_i)\exp(-u_i)$, the last two factors are corresponding error components.

For the parametric function f, the literature is dominated among those using the Cobb-Douglas production function family. Recently however, (Henderson and Simar, 2005) considered a nonparametric specification of f, desirable in cases where the modeler is not willing to risk any parametric functional form because of the insufficient knowledge about the phenomenon being modeled. A Bayesian formulation of f was considered by (Koop and Steel, 2004), where contrary to the nonparametric argument, prior knowledge about the efficiency of producers being analyzed are incorporated into the model.

The model is estimated usually via the maximum likelihood (MLE) or its variants. The quantities v_i, u_i, and x_i, are assumed to be independent and v_i is usually assumed to be normally distributed while u_i is the positive half normal distribution to ensure that technical efficiency estimates are between zero and one. Other combination of the distribution of v and u include normal-exponential, normal-truncated normal, and normal-gamma. The nature and relationship between v and u can be enhanced further using mixed model specifications. (Green, 1990) however, observed that estimates of efficiency vary depending on the distributional assumption on v and u. A model specification that best characterize reality can help improve the robustness property of the estimates.

Since the model is postulated in such a way so that efficiency is an upper bound of productive capacity of producers, efficiency estimates are restricted to be biased downwards (inefficient than they really are). The bias is analyzed by (Gijbels et.al., 1999) in the data envelopment analysis (DEA) estimator which is the set under "lowest" concave monotone function covering all the sample points, for a single input and single output case.

For time series data on the other hand, time-invariant or time-varying technical efficiency were considered. The error assumptions also included fixed and random effects. Heteroskedasticity in v and u was also considered, possibly leading to volatility assumption in technical efficiency.

Stochastic frontier models for panel data were postulated with time-invariant technical efficiency assumption, fixed-effects model, random-effects model, or even mixed model. A careful attention on coverage period of the data used in estimation is necessary. (Kumbhakar and Lovell, 2000) warned that the longer the panel, the less likely it becomes that technology remains constant, a serious violation of the assumption. The learning curve of producers is expected to improve over time and inefficiency realized in the distant past will be less likely to repeat in the future. (Battese and Coelli, 1992) postulated the following model

(2) $$y_{it} = f(x_{it};\beta)\exp(v_{it})\exp(-u_{it})$$

(3) $$u_{it} = \exp\{-\gamma(t-T)\}u_i$$

Where $v_{it} \sim NID(0, \sigma_v^2)$ and $u_i \sim NID^+(0, \sigma_u^2)$. Equation (3) characterizes the improving learning curve over time. The likelihood function is easily constructed from the normal and half-normal distributions and maximum likelihood estimators formulated.

One aim in SFA is to explain inefficiency/efficiency in terms of exogenous determinants, (Kumbhakar and Lovell, 2000) summarized some models to explain inefficiency/efficiency of a producer. (Kumbhakar *et. al.*, 1991) assumed a Cobb-Douglas production function $\ln y_i = \ln f(x_i;\beta) + v_i - u_i$ with $u_i = \gamma' z_i + \varepsilon_i$, the exogenous determinant of efficiency is postulated outside the production function, implying additivity of the effect of factors of production and exogenous factors to actual production. (Reifschneider and Stevenson, 1991) generalized the efficiency equation into $u_i = g(z_i;\gamma) + \varepsilon_i$.

The choice of the best way to analyze the effect of exogenous factors depends on adequacy of the underlying assumption associated with the model. Even a nonlinear regression was used in estimation. The resulting estimates of production efficiencies however, are expected to vary according to the postulated model.

3. SPATIAL TEMPORAL MODELS

The classes of spatial temporal models are widely varied, differences usually depends on the nature of the spatial units, measurements of spatial correlations and subsequent specification of the spatial component, and the temporal model. The spatial units can be defined as units arranged in a lattice or some irregularly shaped elements. Specification of the spatial and temporal components determines the nature of the error structure that affects estimation. Parameter estimation varies from simple likelihood based procedures to hierarchical approaches.

A spatial-temporal model, possibly entertaining irregularly shaped spatial units, with temporal observations made at equal intervals of time was postulated by (Landagan and Barrios, 2006) as

(4) $y_{it} = x_{it}\beta + w_{it}\gamma + \varepsilon_{it}$ $\quad i = 1,\ldots,n \quad t = 1,\ldots,T$

where y_{it} is the response variable from location *i* at time *t*, x_{it} is the set of covariates from location *i* at time *t*, w_{it} is set of variables in the neighborhood system of location *i* at time *t*, and ε_{it} are error components. The error component is postulated as $\varepsilon_{it} = \mu_i + \upsilon_{it}$ following a one-way error component with individual effects $\mu_I \sim IID(0,\sigma^2_\mu)$, and the remainder disturbances υ_{it} following a stationary AR(p). The backfitting algorithm introduced by (Hastie and Tibshirani, 1990) for additive models was modified to simultaneously estimate certain group of parameters at some point of the iterative process. The estimated model is superior to some panel data models.

The introduction of the spatial component in the usual time series models will help econometric specification in the context of the new economic geography. The spatial component accounts for some spatial externalities like, industrial clustering, trade agreements, local policies in a decentralized economy, natural resource constraints, and many others.

4. SPARSE SPATIAL AUTOREGRESSION STOCHASTIC FRONTIER

Consider a cross-section data for n producers and the production stochastic frontier model $y_i = f(x_i;\beta)\exp(v_i)\exp(-u_i)$. The inefficiency u_i is not a purely random occurrence but postulated to be influenced by some factors that affect production efficiency. Following (Reifschneider and Stevenson, 1991), with $\ln y_i = \ln f(x_i;\beta) + v_i - u_i$ and $u_i = g(z_i;\gamma) + \varepsilon_i$, we imbed a sparse spatial autoregression (SAR) proposed by (Pace and Barry, 1997) in the production frontier, and the general linear mixed model into the efficiency equation

(5) $$\ln y_i = \ln f(x_i;\beta) + \delta D[\ln y_i - \ln f(x_i;\beta)] + v_i - u_i$$

(6) $$u_i = \frac{1}{1+\exp[-(w_i\varphi + z_i\phi)]} + \varepsilon_i$$

where $\eta_i = (\eta_{i1},...,\eta_{in})'$, $w_i = (w_{i1},...,w_{ia})'$, $z_i = (z_{i1},...,z_{ib})'$ $\varphi = (\varphi_1,...,\varphi_a)'$, $\phi = (\phi_1,...,\phi_b)'$, and $\varepsilon_i = (\varepsilon_{i1},...,\varepsilon_{in})'$. The function f may take a general Cobb-Douglas form or a more general exponential or a non-linear function. To allow dummy variables among the factors of production (zero values for some producers), an exponential function f can be used. δ is the parameter accounting for spatial externalities common among spatial neighbors, this also accounts for clustering or convergence of efficiency among 'neighboring' producers. $D = [(d_{ij})]$ is the spatial weight matrix where

(7) $$d_{ij} = \begin{cases} 1, \text{if unit i and unit j are spatially related} \\ 0, \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}$$

If the observations are arranged so that neighboring units are next to each other, then the matrix D is block-diagonal. w_i is a vector of fixed factors while z_i is a vector of random factors, and $\varepsilon_i = (\varepsilon_{i1},...,\varepsilon_{in})'$ is pure error. The logistic specification of u_i in (6) will ensure positive values for u_i, this will no longer require a positive-valued error distribution like the truncated normal that will only cause computing problems in the iterative process of maximum likelihood estimation. The sparse spatial autoregression in equation (5) accounts for spatial externalities in the production capacity of the producers. Producers in a spatial neighborhood that commonly use inferior technology will all exhibit lower production while those exposed to superior technology will all exhibit higher production. If such space-related production specificity is not accounted into the model, it could manifest into the error term either causing heterogeneity in the variance structure, or into a completely different error distribution.

Spatial externalities help characterize producers' response to various efficiency-inducing interventions. As an example, a targeted development intervention may have been packaged to be site-specific, thereby resulting to producers' adoption exhibiting similarity among neighbors exposed to the same or similar packages. Given the same technology at one point in time, the spatial component will account for that part of efficiency reflecting producers' adoption capacity as facilitated by site-specific support services.

The model described in equations (5) and (6) constitutes an additive model including the components on production function, the effect of spatial externalities, and the logit of efficiency/inefficiency-inducing factors. In an additive model, (Hastie and Tibshirani, 1990), discussed the advantages and optimality of the backfitting algorithm in estimation. Estimation will also use a modified backfitting algorithm similar to (Landagan and Barrios, 2006) and given as follows:

1. Depending on the link function f, ignore u_i in (5) and estimate β using maximum likehood estimation (MLE) or least squares estimation (LSE). Compute the residuals from (5) as $-\hat{u}_i = \ln y_i - \ln f\left(x_i; \hat{\beta}\right)$ or $\hat{u}_i = \ln f\left(x_i; \hat{\beta}\right) - \ln y_i$ containing information on φ, ϕ and δ.
2. Estimate δ from $\hat{u}_i = \delta D\hat{u}_i + \xi_i$ ignoring w_i and z_i. Compute the residuals from $\hat{\hat{u}}_i = \hat{u}_i - \hat{\delta} D\hat{u}_i$, this contain information on φ and ϕ.
3. Estimate φ and ϕ from $\hat{\hat{u}}_i = \dfrac{1}{1+\exp\left[-\left(w_i\varphi + z_i\phi\right)\right]} + \varepsilon_i$, a mixed model logistic regression where elements of u are taken from the residuals in (2).
4. The estimates of technical efficiency is computed from

$$\text{TE}_i = \exp-\left[\frac{1}{1+\exp\left[-\left(w_i\hat{\varphi} + z_i\hat{\phi}\right)\right]}\right] \tag{8}$$

See (Landagan and Barrios, 2006) for details of convergence issues. Hypothesis testing can be carried out through resampling methods. For the parameter estimates, a nonparametric bootstrap can be done in the respective steps (1,2, and 3) above to understand their empirical distribution. For the confidence interval of the estimates of technical efficiency, a jackknife can be applied to equation (8) in step 4.

The spatial autoregression component of the model will reclaim that portion of the residuals from the production function that will otherwise be attributed to production inefficiency. The underestimation of technical efficiency (overestimation of inefficiency) can somehow be eased when a source of discrepancy between output and predicted value of production function is identified, and not all lumped together into inefficiency. The backfitting algorithm and the logistic specification of the inefficiency/efficiency equation will also yield computing advantage over maximum likelihood estimation that will require more complicated distribution (to account for the technical efficiency constraint) and the large set of parameters (production and efficiency determinants simultaneously estimated).

5. SPATIAL TEMPORAL STOCHASTIC FRONTIER

Panel data contains information on both the temporal dependencies and the relationship among units at specific point in time. However, if units were selected at one point in time using a probability sampling procedure, oftentimes, the induced sampling distribution characterizes basic independence of the observations. These are the most common models usually postulated for panel data.

Across units fixing time, dependencies can be defined not only by the sampling distribution by the selection procedure, but also by those exerted by other units within specific neighborhood. Several measures of spatial distance have been proposed in the literature of spatial statistics, from the simplest to the more complicated ones. Depending on the complexity of the model and the problem, simple or complicated measures of spatial distance will be needed.

In stochastic frontier modelling, several models were proposed given a panel data. Assuming constant factor coefficients over time, (Battese and Coelli, 1992) postulated a time-decaying inefficiency (improving learning curve) as

(9) $$y_{it} = f(x_{it};\beta)\exp(v_{it})\exp(-u_{it})$$

(10) $$u_{it} = \exp\{-\gamma(t-T)\}u_i$$

Over time, the producers get to realize their failure to adopt efficient technologies and correct it soon after which, more efficient production process is applied. (Battese and Coelli, 1995) further postulated that inefficiencies are function of some exogenous variables and used the maximum likelihood technique in parameter estimation.

Many stochastic frontier models for panel data failed to account for temporal dependencies (improving learning curve of producers) and spatial externalities (adoption of efficiency-enhancing technologies among the producers in a spatial neighborhood) simultaneously. Ignoring this aspect of the information contained in the panel data will result to inadequate differentiation of the producer's efficiency-inducing potentials, hence, may result to inferior estimates of technical efficiency coefficients.

A spatial-temporal stochastic frontier model is postulated as

(11) $$\ln y_{it} = \ln f(x_{it};\beta) + v_{it} - u_{it}$$

(12) $$v_{it} = \rho v_{it-1} + \psi_{it}$$

(13) $$u_{it} = \frac{1}{1+\exp[-(w_{it}\gamma + z_{it}\phi)]} + \varepsilon_{it}$$

where, the subscript *i* refer to the producer and *t* the time period, hence, y_{it} is the output of producers *i* at time *t*, x_{it} are the factors of production, v_{it} is the autocorrelated pure error, u_{it} are measures of inefficiency, w_{it} are measures of spatial distance, z_{it} are other determinants of inefficiency, ε_{it} and ψ_{it} are a white noise terms, β, γ, ϕ, and ρ are the corresponding parameters. The production structure is assumed to be constant over time, hence reflected in time-independence of β. In a reasonably sized panel, production structure is not expected to change since changes may have been brought by significant technological innovations that can be detected only in a much longer panel. The temporal dependence measured by ρ also assumes homogeneity across producers. The short-term dependency in efficiency indexed by ρ is not expected to exhibit structural changes within a short panel. Unlike (Battese and Coelli, 1995) that specified a non-negative-valued distribution for error terms (hence, complicating the likelihood function), the logit specification in equation (13) will ensure non-negative predicted values of u_{it} that yield estimates of technical efficiency ≤1.

A dynamic production parameter in equation (13) may account for the spatial externalities accounted by the spatial indicator, but will require more complicated estimation procedure. Equation (12) can also be generalized to higher-order AR process, but the time-adjustment process of inefficiency reduction might be contaminated for much longer autoregressions given a short panel.

The additivity of the models presented in equations (11) to (13) will make estimation via the hybrid backfitting algorithm feasible. The estimation algorithm follows:

1. Equations (11) and (12) are combined and ignore $-u_{it}$ to estimate β and ρ simultaneously using generalized least squares. Compute the residuals $\hat{u}_{it} = \ln y_{it} - \ln f\left(x_{it};\hat{\beta}\right) - \hat{\rho}e_{it-1}$, this contains information on γ, ϕ. e_{it-1} is the lagged value of the residuals from the fitted model.
2. Given $\hat{u}_{it}$, fit equation (13) as a general linear model to estimate γ and ϕ.
3. The estimate of technical efficiency is

$$\text{(14)} \qquad \mathrm{TE}_{it} = \exp -\left[\frac{1}{1+\exp\left[-\left(w_{ti}\hat{\gamma} + z_{it}\hat{\phi}\right)\right]}\right]$$

The simultaneous estimation of β and ρ yield optimality over individual estimation in pure backfitting of an additive model. Following, the argument of (Landagan and Barrios, 2006), this will not necessitate further iteration of the algorithm.

The inclusion of autoregression in the error of the production function will account for the producers' learning curve while also accounting for the possible cumulative effect of production errors. The spatial externalities that can vary over time and across spatial neighbors help characterize efficiency/inefficiency differences among the producers.

6. NUMERICAL ILLUSTRATIONS

To illustrate the application of the models and the estimation procedures, two data sets are used. The first is a subset of the 2003 Family Income and Expenditure Survey, considering rural households only. The survey is conducted by the Philippine National Statistics Office, where data on income and expenditures as well as its possible determinants are collected at the household level. The second data set is based on the monitoring of Agrarian Reform Communities by the Philippine Department of Agrarian Reform, collected from the period 2002-2005. The data collected includes an index of sustainable rural development and various factors/indicators needed in the attainment of development.

Total Family Income

The sparse spatial autoregression SFM estimated through the modified backfitting algorithm (Model 1) and the ordinary SFM estimated using maximum likelihood estimation in a truncated normal error distribution (Model 2) are compared. The producing units are the households, the output is total income. There were 15 factors of production, where 9 are continuous variables, while 6 are dummy variables. For the efficiency equation, 14 determinants (7 continuous and 7 dummy indicators) were used to characterize household's income-generating efficiency/inefficiency. The Cobb-Douglas production frontier is specified for both models.

Model 2 requires careful specification of the iterative estimation process since it involves matrices with large dimension in the likelihood function. Model 1 on the other hand, is much easy to handle in the empirical implementation since the factors of production and the factors of efficiency are dealt separately at different steps in the iterative process.

The parameter estimates for both models are similar, indicating that they both estimate similar empirical structures characterizing the income-generation process of rural households (see Table 1 for details). Model 1 yields an average estimate of technical efficiency of 0.9007 (s.d.=0.1203) or about 10% inefficiency in income-generation among the rural households. Model 2 on the other hand, produced an average estimate of technical efficiency of 0.7624 (s.d.=0.1695) or 24% inefficiency. The higher average technical efficiency estimate from Model 1 can be attributed to the significant amount of the residual that is further accounted into the effect of spatial externalities, added to the inefficiency in the case of Model 2. The technical efficiency estimates from Models 1 and 2 yield a correlation of 0.8051, indicating that the models were able to identify the same households as inefficient/efficient. The correlation between technical efficiency estimates with the output (income) is 0.2530 for Model 1, while 0.3708 for Model 2. Both models also yield similar correlation between technical efficiency estimates and the determinants in the efficiency/inefficiency equation.

(INSERT TABLE 1 HERE)

Index of Sustainable Rural Development

The panel data on agrarian reform communities (production unit) is analyzed with an index of sustainable rural development as output and provision of rural infrastructure and support services as input and determinants of the efficiency equation. The data is first analyzed by ignoring the panel, and analyzed separately per year. The Model 1 and Model 2 described in the previous section are also compared. In analyzing the panel, the spatial temporal SFM (Model 3) and the time-varying decay model (Model 4) of (Battese and Coelli, 1992), are also compared.

Models 1 and 2 yield similar parameters estimates, yielding the same signs for the factors of production and the determinants of efficiency/inefficiency for all years. Model 1 generally yield higher estimates of technical efficiency than Model 2, explained by the additional portion of the residual explained by spatial externalities in Model 1 that is contributed to inefficiency in Model 2. The correlation between technical efficiency estimates from Models 1 and 2 has a minimum of 0.6654 to as much as 0.8258, indicating that both models were able to identify the similar group of communities to be efficient/inefficiency in utilizing infrastructure and support services in moving forward to development. Furthermore, Model 1 was able to distinguish the more inefficient communities from the "average" communities, this can facilitate the identification of appropriate interventions (see Figure 1 for details).

(INSERT FIGURE 1 HERE)

The parameter estimates for the factors of production (using the Cobb-Douglas family) are given in Table 2. Similarity in estimates from Models 3 and 4 justifies the assumption of additivity of the model described in equations 11 to 13. Just like in the cross-section data, estimates of technical efficiency from Model 3 are generally higher than those coming from Model 4. Model 3 explained further that portion of the residuals from the production function into spatial externalities and temporal dependencies, lumped together into inefficiencies in Model 4. The correlation between estimates of technical efficiencies from Model 3 and 4 is 0.5512 indicating that both models identified fairly similar communities to be efficient/inefficient. The efficiency estimates from both models also yield similar correlations with the determinants of efficiency.

(INSERT TABLE 2 HERE)

7. CONCLUSIONS

In a stochastic frontier model where specification generally leads to estimates of technical efficiency that is biased downwards, a sparse spatial autoregression component in cross-section data and a spatial-temporal component in panel data can help improve the estimate. A modified backfitting algorithm that take advantage of the additivity of the models can also facilitate computing especially when large set of factors of production and determinants of inefficiency complicates maximum likelihood estimation in truncated distribution. The proposed model and the corresponding estimation procedure yields estimates of technical efficiency that are similar to those obtained from some commonly used methods, being able to identify similarly efficient/inefficient producers.

REFERENCES

Amos T, Chikwendu D, Nmadu J. 2004. Productivity, technical efficiency, and cropping patterns in the savanna zone of Nigeria. *Food, Agriculture and Environment* **2**:173-176.

Battese G, Coelli T. 1992. Frontier production functions, technical efficiency and panel data with application to paddy rice 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.

Cressie N. 1993. *Statistics for Spatial Data*. Wiley: New York

Gijbels I, Mammen E, Park B, Simar L. 1999. On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association* 94:220-228.

Green W. 1990. A gamma-distributed stochastic frontier model. *Journal of Econometrics* 46:141-164.

Hastie T, and Tibshirani R. 1990): *Generalized Additive Models*. Chapman and Hall: New York.

Henderson D, Simar L. 2005. A fully nonparametric stochastic frontier model for panel data. *Working Paper No. 519, Department of Economics, Binghamton University*. http://econ.binghamton.edu/wp05/WP0519.pdf.

Koop G, Steel M. 2004. Bayesian analysis of stochastic frontier models. *ESE Discussion Paper No. 19, Edinburgh School of Economics*, *University of Edinburgh,* http://ideas.repec.org/p/edn/esedps/19.html.

Kumbhakar S, Ghosh S, McGuckin J. 1991. A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms. *Journal of Business and Economic Statistics* 9:279-286.

Kumbhakar S, Lovell K. 2000. *Stochastic Frontier Analysis*. Cambridge University Press: UK.

Landagan O., Barrios E. 2006. An estimation procedure for a spatial-temporal model. In press *Statistics and Probability Letters*, doi:10.1016/j.spl.2006.08.006.

Pace R, Barry R. 1997. Sparse spatial autoregressions. *Statistics and Probability Letters* 33:291-297.

Reifschneider D, Stevenson R. 1991. Systematic departures from the frontier: a framework for the analysis of firm inefficiency. *International EconomicReview* 32:715-723.

Richardson S, Guihenneuc C, Lasserre V. 1992. Spatial linear models with autocorrelated error structure. *The Statistician* 41:539-557.

TABLE 1

PARAMETER ESTIMATES IN A CROSS-SECTION DATA

Variables	Model 1		Model 2	
	Coefficient	p-value	Coefficient	p-value
Factors of Production				
Log(family size)	0.4136	0.000	0.3297	0.000
Whether head is male	-0.0550	0.000	-0.0529	0.000
Log(proportion of employed members)	0.0637	0.000	0.0455	0.000
Whether the spouse is employed	0.1238	0.000	0.1241	0.000
Whether head had no formal education	-0.2025	0.000	-0.1888	0.000
Log(age of head)	0.0769	0.000	0.0781	0.000
Whether roof is made of strong materials	0.0809	0.000	0.0786	0.000
Whether wall is made of strong materials	0.1606	0.000	0.1235	0.000
Whether toilet is hygienic	0.2089	0.000	0.1863	0.000
Log(house and lot value)	0.3149	0.000	0.2661	0.000
Log(income from entrepreneurship)	0.0013	0.089	0.0010	0.156
Log(income from crop production)	-0.0204	0.000	-0.0155	0.000
Log(income from livestock production)	-0.0020	0.015	-0.0007	0.413
Log(income from fishing)	-0.0064	0.000	-0.0048	0.000
Log(income from wholesale and retail trade)	0.0067	0.000	0.0041	0.000
Efficiency Equation				
Agriculture Dependency	-0.1133	0.000	-0.0500	0.000
Savings Rate	-5.3899	0.000	-3.7027	0.000
Expenditure to electricity	0.0003	0.000	0.0001	0.000
Expenditure to Water	0.0004	0.000	0.0002	0.000
Expenditure to fuel	-0.0004	0.000	-0.0003	0.000
Land Fare	-0.0003	0.000	-0.0002	0.000
Expenditure to Telephone	-0.0024	0.000	-0.0008	0.000
Whether engaged in agriculture	-0.0723	0.226	-0.0978	0.008
Whether engaged in crop production	-0.0093	0.003	-0.0014	0.153
Whether engaged in livestock production	-0.0415	0.000	-0.0041	0.153
Whether engaged in fishing	-0.0103	0.068	-0.0026	0.181
Whether engaged in wholesale and retail	0.0674	0.006	0.0188	0.013
Whether engaged in manufacturing	0.0040	0.110	0.0031	0.109

Whether engaged in transportation	-0.3358	0.000	-0.0019	0.407

TABLE 2

PARAMETER ESTIMATES IN A PANEL DATA

Variables	**Model 3**		**Model 4**	
	Coefficient	p-value	Coefficient	p-value
Factors of Production				
Log(Index of basic social services)	0.3283	0.000	0.3051	0.000
Log(Index of organizational maturity)	0.1597	0.000	0.1794	0.000
Log(No.Beneficiaries cultivating the land)	0.0069	0.134	0.0036	0.070
Log(No. of Beneficiaries of agrarian reform)	-0.0006	0.815	0.0013	0.520
Log(Index of land tenure improvement)	0.1829	0.000	0.1531	0.000
Log(Proportion of credit needs served)	0.0146	0.000	0.0179	0.000
Log(Land area covered by agrarian reform)	0.0023	0.697	-0.0113	0.000
Log(Age of the community in the program)	0.0876	0.032	0.0211	0.000

FIGURE 1

BOXPLOTS OF ESTIMATES OF TECHNICAL EFFICIENCY