Abstract:
The paper considers estimation of a panel data model with disturbances that are autocorrelated across cross sectional units. It is assumed that the disturbances are spatially correlated based on some geographic or economic proximity measure. If the time dimension of the data is large, feasible and efficient estimation proceeds by random effects. For the case where the time dimension is small (the usual panel data case), we develop a generalized moments estimation approach that is a generalization of a cross sectional model due to Kelejian and Prucha (1999). We apply this approach in a stochastic frontier framework to a panel of Indonesian rice farms where spatial correlations are based on geographic proximity, altitude and weather. The correlations represent productivity shock spillovers across the rice farms in different villages on the island of Java. Using a Moran I test statistic, we demonstrate empirically that productivity shock spillovers may exist in this (and perhaps other) data sets, and that these spillovers have profound effects on technical efficiency estimation.

Keywords: autocorrelation, Moran I, productivity, stochastic frontier, spatial dependence.

* Corresponding author. Department of Economics, University of Arizona, Tucson, AZ 85721-0108, USA. Voice: (520) 621-6230, Fax: (520) 521-8450. Email: whorrace@u.arizona.edu.
1. Introduction

Over the last few years there have been several papers on the subject of the spatial analysis of economic, cross sectional data. Theoretical or empirical spatial issues have been addressed in Case (1991), Conley (1999), Delong and Summers (1991), Dubin (1988), Fishback, Horrace and Kantor (1999), Kelejian and Robinson (1993), Moulton (1990), Quah (1992) and Topa (1996). These cross sectional specifications address the important economic phenomena of geographical proximity, infrastructure effects or spillovers, to name a few. Recently, Kelejian and Prucha (1999) considered generalized moments estimation of a certain class of these models which permits autocorrelation of the disturbances across cross sectional units. This “cross sectional autocorrelation” can be likened to the usual autocorrelation theory of the time series literature, however in this case estimation hinges on the ex ante specification of a “spatial weighting matrix” in the cross sectional dimension of the error process. The form of the weighting matrix is at the discretion of the analyst, but very often can be based on some underlying economic, geographic, or meteorological theory. The specification of the spatial weighting matrix may seem like a fairly strong parametric assumption to impose on the model, however the assumption is testable and is the price paid for lack of a time dimension in the data. “Of course, if panel data are available one can consider, e.g., a seemingly unrelated regression model, or an error component model to permit for cross sectional correlation, and estimate the cross sectional correlations via the time dimension of the panel if the time dimension is large”.

1 Unfortunately, the usual panel data case is when the cross sectional dimension is large and the time dimension is small (fixed), so consistent estimation of the cross sectional correlations is typically not justified.

This paper is concerned with generalized moments estimation of this class of models under the usual panel data conditions. Specifically, we generalize the Kelejian and Prucha (1999) estimator to the usual panel data case. It is important to stress that the panel data theory presented is for the case where \( T \) is fixed, consequently the current discussion also hinges on the ex ante specification of a spatial weighting matrix. Once we allow the time dimension to become large, the specification of the weighting matrix
becomes unnecessary as the estimation techniques presented herein, become empirically inferior to more standard approaches that hinge on $T$ asymptotics.

This paper also presents an application of these spatial panel techniques to the efficiency measurement problem of the well-developed stochastic frontier literature, which attempts to model a common production function for a sample of firms based on observables (inputs and outputs) and an unobservable, one-sided component viewed as technical or production inefficiency. Cross sectional estimation of these models is usually traced back to Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), while panel estimation is due to Schmidt and Sickles (1984). Our concern is, of course, the panel specification, and we select a panel of 171 Indonesian rice farms observed over 6 periods for our example. Output is rice, and inputs are things like seed, fertilizer and land acreage. The time dimension of the data is small, so consistent estimation of cross sectional correlations in the error process may not be justified. Consequently, we specify an spatial weighting scheme in the error process which allows for spillovers across farms based on geographic proximity, altitude and weather conditions. The results indicate that spatial correlations may exist in the data and have ramifications for the estimation of the production function and the estimation of farm-level technical efficiency.

The paper is organized as follows. The next section presents an unrestricted, general panel data model that is not identified and fully- and partially-restricted models that are identified. Section 3 discusses feasible estimation of these models. Section 4 presents the Indonesian rice farm example. Section 5 summarizes and concludes.

2. A Panel Model with Spatial Disturbances

Consider the standard fixed effect (FE) model:

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it}; \quad i = 1, ..., N; \quad t = 1, ..., T,$$

where $\beta$ is $(k \times 1)$ and $x_{it}$ is $(1 \times k)$. Here we assume that $T$ is fixed, so we cannot rely on $T$-asymptotics. Collecting $i$ the model becomes

$$y_t = \alpha + x_\beta + u_t, \quad t = 1, ..., T,$$

(1)
where $\alpha' = [\alpha_1, ..., \alpha_N]$ and $x_t$ is $(N \times k)$. Now suppose that the error term is spatially lagged such that

$$u_t = \rho_t M_t u_t + \varepsilon_t; \quad t = 1, ..., T,$$

where $\rho_t$ is a scalar, spatial autoregressive parameter and $M_t$ is a $(N \times N)$ spatial weighting matrix of known constants, which captures the spatial correlations across cross sectional units. (In the sequel we allow for time invariant spatial parameter and weighting matrix.) Elements of $M_t$ are $m_{ijt}$ and are chosen based on some geographic or economic proximity measure such as contiguity or physical, economic or climatic distances or dissimilarities. For example, in section 4, we select $m_{ijt}$ to be the inverse of the physical distance $(1/km)$ between unit $i$ and unit $j$ in time period $t$.

Assumption 1: The elements of $\varepsilon_t$ are independently and identically distributed with zero-mean and finite variance $\sigma^2_t$, the fourth moment of $\varepsilon_t$ is finite, and $\varepsilon_t$ is independent of $\varepsilon_s$, $\forall t \neq s$.

Assumption 2: All diagonal elements of $M_t$ are zero. The matrix $(I_N - \rho_t M_t)$ is non-singular. $|\rho_t| < 1$.

Collecting $t$,

$$y = \iota_T \otimes \alpha + x \beta + u, \quad u = \rho^* M^* u + \varepsilon,$$

where $\iota_T$ is a $T$ dimensional column vector of ones, $x$ is $(TN \times k)$ and

$$M^* = \begin{bmatrix} M_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & M_T \end{bmatrix}, \quad \rho^* = \begin{bmatrix} \rho_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \rho_T \end{bmatrix}.$$

Notice that
so the disturbance is heteroskedastic. Define \( \Phi_t = (I_N - \rho_t M_t)/\sigma_t \), then we can pre-multiply the model in equations (1) and (2) to get,

\[
E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma_t^2 I_N & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_t^2 I_N \end{bmatrix},
\]

(4) \( y_t^* = \alpha_t^* + x_t^* \beta + \varepsilon_t^* \),

where \( y_t^* = \Phi_y y_t, x_t^* = \Phi_x x_t, \alpha_t^* = \Phi_\alpha \alpha \) and \( \varepsilon_t^* = \Phi_\nu \nu_t = \varepsilon_t / \sigma_t \). Collecting \( t \)

(5) \( y^* = \alpha^* + x^* \beta + \varepsilon^* \),

where \( \alpha^* = [\alpha_1^* \ldots \alpha_T^*] \), a \( TN \) dimensional column vector. Notice that equation (5) possess a “well-behaved” disturbance. That is, \( E(\varepsilon^*) = 0 \) and \( E(\varepsilon^* \varepsilon^*') = I_{TN} \).

Unfortunately this system consists of \( TN \) observations and \( TN + k \) parameters, so estimates are not identified. The unobserved heterogeneity term varies across cross sectional units and the time series; what identifies the parameters in the usual fixed effect specification is the fact that unobserved heterogeneity is time-invariant. Therefore, we need to impose some additional structure on equation (5) to identify estimates.

### 2.1 A Fully Restricted Specification.

One obvious restriction would be to assume that some subset of the weighting matrices, autoregressive parameters and variance parameters are equal. As an extreme case we could assume that \( M_1 = \ldots = M_T = M, \rho_1 = \ldots = \rho_T = \rho \), \( \sigma_1^2 = \ldots = \sigma_T^2 = \sigma^2 \), implying \( \Phi_1 = \ldots = \Phi_T = \Phi \). Then \( \alpha_t^* = \Phi \alpha^* \) in equation (4), \( \alpha^* = \nu_T \otimes \Phi \alpha \) in equation (5), and the system consists of \( N + k \) parameters which are identified by the \( TN \) observations. Of course, the error term \( \varepsilon \) of equation (3) is no longer heteroskedastic; it has variance matrix \( E(\varepsilon \varepsilon') = \sigma^2 I_{TN} \), so \( \Phi \) need not be a function of \( \sigma \) for efficiency. Fixed effect estimation of equation (5) under this full restriction, will then be efficient for \( \alpha^* \) and \( \beta \), if \( \rho \) and \( \sigma^2 \) are known, and if the restriction is true. It is also consistent for fixed \( T \) as
Additionally an estimate of \( \alpha \) can be recovered by transforming the estimate of \( \alpha^* \) with \( \Phi \). Of course \( \rho \) and \( \sigma^2 \) are not known, so the challenge is to consistently estimate them, so that equation (5) can be feasibly estimated; this is undertaken in section 3.

2.2 A Partially Restricted Specification.

As another example of a reasonable restriction on the parameters of the model, briefly consider the empirical example of the sequel. We observe \( N = 171 \) Indonesian rice farms over \( T = 6 \) periods. Periods 1, 3 and 5 are “wet or rainy seasons” and periods 2, 4 and 6 are “dry seasons”. It may be reasonable to suspect that \( \rho_1 = \rho_3 = \rho_5 = \rho_w \) (wet) and \( \rho_2 = \rho_4 = \rho_6 = \rho_d \) (dry), similarly for \( M_t, \sigma^2_t \) and \( \Phi_t \). (This may be true on the island of Java, since during the rainy season many roads in the low country are impassable, and hence spillovers based on infrastructure are potentially diminished.) Then \( \alpha^* = [\Phi_w \Phi_d \Phi_w \Phi_d \Phi_w \Phi_d] \Theta \alpha \) in equation (5), a \( TN \) dimensional column vector, consisting of only \( 2N \) different parameters. The system in (5) then consists of \( 2N + k \) parameters and can effectively be treated as \( 2 \times 171 = 342 \) farms observed over \( 6/2 = 3 \) periods, so fixed effect estimation of equation (5) is identified, since it has been assumed that realizations of the error \( \varepsilon_t \) are independent across both \( t \) and \( i \). Of course, there will be an efficiency loss in the estimate of \( \alpha^* \), since the time series dimension has been effectively cut in half from 6 to 3, but the slope parameter \( \beta \) will still be efficient (and consistent in \( N \)) since it is still based on the same number of observations, \( TN \). Again the challenge is estimation of \( \rho_w, \rho_d, \sigma^2_w \) and \( \sigma^2_d \), which is taken up in the following section.

3. Feasible Estimation

Kelejian and Prucha (1999) develop a moments estimator of the parameters \( \rho_t \) and \( \sigma^2_t \) in the cross sectional setting (\( T = 1 \)). We now generalize their results for the case where \( \rho_t \) and \( \sigma^2_t \) are different across \( t \). Using their notation, let \( \tilde{u}_t \) be a predictor of \( u_t \) from the fixed effect (or within) regression implied by equation (1), ignoring equation (2). That is,

\[ 2 \text{ We present no proofs of our results, because they are all straight-forward extensions of Kelejian and Prucha’s proofs. Proofs are available from the second author upon request.} \]
\tilde{u}_t \text{ converges in distribution to the random variable } u_t. \text{ Additionally, let } \tilde{u}_t = M_t \tilde{u}_t, \\
\tilde{\tilde{u}}_t = M_t \tilde{u}_t, \quad \tilde{\varepsilon}_t = M_t \varepsilon_t, \text{ and } \tilde{\tilde{\varepsilon}}_t = M_t \tilde{\varepsilon}_t. \text{ Consider the following } 3T \text{ moment conditions implied by equations (1) and (2) and assumptions 1 and 2.}

\begin{align*}
E[N^{-1} \varepsilon_t \varepsilon_t'] &= \sigma_t^2, \\
E[N^{-1} \varepsilon_t' \varepsilon_t] &= \sigma_t^2 N^{-1} \text{tr}(M_t', M_t), \\
E[N^{-1} \varepsilon_t \varepsilon_t] &= 0,
\end{align*}
\quad t = 1, ..., T.

Noting that \( \varepsilon_t = (I_N - \rho_t M_t) u_t \), these moment conditions imply the following system of \( 3T \) equations,

\begin{align*}
\Gamma_t [\rho_t, \sigma_t^2] - \gamma_t &= 0,
\end{align*}
where,

\begin{align*}
\Gamma_t &= \begin{bmatrix}
\frac{1}{N} E(u_t', \varepsilon_t) & \frac{1}{N} E(\varepsilon_t', \varepsilon_t) & 1 \\
\frac{1}{N} E(\varepsilon_t', \varepsilon_t) & \frac{1}{N} E(\varepsilon_t', \varepsilon_t) & \frac{1}{N} \text{tr}(M_t', M_t) \\
\frac{1}{N} E(u_t', \tilde{u}_t + \tilde{\varepsilon}_t, \tilde{\varepsilon}_t') & \frac{1}{N} E(\tilde{\varepsilon}_t', \tilde{u}_t') & 0
\end{bmatrix},
\gamma_t = \begin{bmatrix}
\frac{1}{N} E(u_t', u_t) \\
\frac{1}{N} E(\tilde{\varepsilon}_t', \tilde{\varepsilon}_t) \\
\frac{1}{N} E(u_t', \tilde{u}_t)
\end{bmatrix},
\end{align*}
\quad t = 1, ..., T.

The sample analogs based on \( \tilde{u}_t \) are

\begin{align*}
G_t [\rho_t, \sigma_t^2] - g_t &= v_t [\rho_t, \sigma_t^2],
\end{align*}
where,

\begin{align*}
G_t &= \begin{bmatrix}
\frac{1}{N} \tilde{u}_t' \tilde{u}_t & \frac{1}{N} \tilde{u}_t' \tilde{\varepsilon}_t & 1 \\
\frac{1}{N} \tilde{\varepsilon}_t' \tilde{u}_t & \frac{1}{N} \tilde{\varepsilon}_t' \tilde{\varepsilon}_t & \frac{1}{N} \text{tr}(M_t', M_t) \\
\frac{1}{N} \tilde{\varepsilon}_t' \tilde{\varepsilon}_t & \frac{1}{N} \tilde{\varepsilon}_t' \tilde{\varepsilon}_t & 0
\end{bmatrix},
\quad g_t = \begin{bmatrix}
\frac{1}{N} \tilde{u}_t' \tilde{u}_t \\
\frac{1}{N} \tilde{u}_t' \tilde{\varepsilon}_t \\
\frac{1}{N} \tilde{\varepsilon}_t' \tilde{u}_t
\end{bmatrix},
\quad t = 1, ..., T.
\end{align*}

Here \( v_t \) is the usual error associated with a sample of statistical realizations (i.e. it will ultimately be squared, summed, then minimized by selecting parameters optimally). The system consists of \( 3T \) equations and \( 3T \) unknowns, but the system is actually \( T \) separate subsystems of 3 equations and 3 unknowns. If these \( T \) subsystems satisfy Assumptions 1 and 2 above and Assumptions 3, 4 and 5 of Kelejian and Prucha.
(1999), then Theorem 1 of Kelejian and Prucha (1999) is applicable to the individual subsystems. That is, $\hat{\rho}_t$ and $\hat{\sigma}_t^2$ that solve the non-linear optimization

$$(\hat{\rho}_t, \hat{\sigma}_t^2) = \arg \min_{\rho, \sigma} \{V_t(r, s^2) \mid V_t(r, s^2) : s^2 \geq 0\}$$

are consistent for $\rho_t$ and $\sigma_t^2$ as $N \to \infty$. For a proof see Kelejian and Prucha (1999). Let

$$\hat{\Phi}_t = (I_N - \hat{\rho}_t M_t) / \sqrt{\sigma_t^2}.$$ 

We can substitute $\hat{\Phi}_t$ for $\Phi_t$ in equation (5), but again the system will be over-parameterized and will require some restrictions for parameter estimates to be identified. Let us called the $\hat{\rho}_t$ and $\hat{\sigma}_t^2$ unrestricted estimates. We now consider feasible estimation of the fully restricted and partially restricted models discussed in the last section.

### 3.1 Feasible Estimation of the Fully Restricted System.

If we can assume that $M_1 = \ldots = M_T = M$, $\rho_1 = \ldots = \rho_T = \rho$, $\sigma_1^2 = \ldots = \sigma_T^2 = \sigma^2$ as before, then we can impose the assumption $M_1 = \ldots = M_T = M$ in equation (6) and estimate $\hat{\rho}_t$ and $\hat{\sigma}_t^2$, $t = 1, \ldots, T$ as above. Then average estimates of $\rho$ and $\sigma^2$ are

$$\hat{\rho} = T^{-1} \sum_t \hat{\rho}_t \quad \text{and} \quad \hat{\sigma}^2 = T^{-1} \sum_t \hat{\sigma}_t^2.$$ 

We shall call these estimates the fully restricted average estimates. The estimates will be consistent as $N \to \infty$, so long as the restriction is true. These are essentially two-stage estimates, where in the first stage unrestricted estimates are calculated ($\hat{\rho}_t$ and $\hat{\sigma}_t^2, t = 1, \ldots, T)$, and the restriction is imposed in the second stage of averaging over $t$. Since the estimates are based on the unrestricted estimates they do not exploit all the information in

---

3 Assumptions 3, 4 and 5 of Kelejian and Prucha are contained in the Appendix.

4 The fact that we estimate $\rho_t, t = 1, \ldots, T$, implies a test of the hypothesis $\rho_1 = \ldots = \rho_T = \rho$. We are not aware of any such test, nor are we aware of a standard error calculation for the estimate of $\rho_t$. Of course, the standard error could be boot-strapped. In the sequel we use the Moran I test to test the significance of the over-all weighting scheme.
the data set simultaneously. That is, each \( \hat{\rho} \) and \( \hat{\sigma}^2 \) is calculated from one of \( T \) separate sub-samples of the data. These estimates imply

\[
\hat{\Phi} = (I_N - \hat{\rho}M),
\]

which can be substituted into equation (5). Then fixed effect estimation of equation (5) with \( \alpha^* = \mathcal{I}_T \otimes \hat{\Phi} \alpha \) is consistent for \( \alpha^* \) and \( \beta \).

If we can a) impose the restriction, b) estimate the parameters in a single step and c) do so such that the data is not divided into \( T \) sub-samples, then the resulting parameter estimates should be more efficient than the average fully restricted estimates. One such estimate is based on the moment conditions:

\[
E[(TN)^{-1}e'e] = \sigma^2, \quad E[(TN)^{-1}e'M] = \sigma^2 (N)^{-1}tr(M'M),
\]

where \( e = M\varepsilon \), and \( \bar{e} = M\bar{e} \). Letting \( \tilde{u} \) be a predictor of \( u \) from the fixed effect (or within) regression implied by equation (3), \( \tilde{u} = M\tilde{u} \) and \( \bar{u} = M\bar{u} \), equation (6) becomes

\[
(7) \quad G[\rho, \rho^2, \sigma^2] - g = v(\rho, \sigma^2),
\]

where,

\[
G = \begin{bmatrix}
\frac{2}{TN} \tilde{u}'\tilde{u} & \frac{1}{TN} \tilde{u}'\bar{u} & 1 \\
\frac{1}{TN} \tilde{u}'\tilde{u} & \frac{1}{TN} \bar{u}'\bar{u} & \frac{1}{N} tr(M'M) \\
\frac{1}{TN} \tilde{u}'\tilde{u} + \bar{u}'\bar{u} & \frac{1}{TN} \bar{u}'\bar{u} & 0
\end{bmatrix}, \quad g = \begin{bmatrix}
\frac{1}{TN} \tilde{u}'\tilde{u} \\
\frac{1}{TN} \bar{u}'\bar{u} \\
\frac{1}{N} \bar{u}'\tilde{u}
\end{bmatrix}.
\]

The system consists of 3 equations and 3 unknowns and is exactly the Kelejian and Prucha (1999) result. Then estimates \( \tilde{\rho} \) and \( \tilde{\sigma}^2 \) follow from

\[
(\tilde{\rho}, \tilde{\sigma}^2) = \underset{r, s^2}{\arg \min} \{v(r, s^2) : v(r, s^2), s^2 \geq 0\}
\]

We shall call these the fully restricted moment estimates (to differentiate them from the fully restricted average estimates). The efficiency gain over the estimates \( \hat{\rho} \) and the \( \hat{\sigma}^2 \)

\[\text{Notice that the middle moment condition contains } N^{-1} \text{ and not } (TN)^{-1}, \text{ since it is based on } M \text{ and not } M^*\text{.} \]
hinges on the fact that equation (7) exploits the information contained in \(TN\) observations and imposes a hypothetically correct restriction, while equation (6) exploits that contained in \(N\) observations over \(t = 1, \ldots, T\), and no restriction. Again, \(\bar{\rho}\) and \(\bar{\sigma}^2\) imply \(\tilde{\Phi}\), which can be inserted in equation (5); then fixed effect estimation of equation (5) with \(\alpha^* = t_I \otimes \tilde{\Phi} \alpha\) is unbiased for \(\alpha^*\) and \(\beta\). Unbiased estimation of \(\alpha\) follows by transforming the estimate of \(\alpha^*\) by \(\tilde{\Phi}\).

3.2 Feasible Estimation of the Partially Restricted System.

For our 171 Indonesian rice farms observed over 6 periods, if we can assume that \(M_1 = \ldots = M_6 = M, \rho_1 = \rho_3 = \rho_5 = \rho_W, \rho_2 = \rho_4 = \rho_6 = \rho_D, \sigma_1^2 = \sigma_3^2 = \sigma_5^2 = \sigma_W^2, \sigma_2^2 = \sigma_4^2 = \sigma_6^2 = \sigma_D^2,\) then we can impose the assumption \(M_1 = \ldots = M_6 = M\) in equation (6) and estimate \(\hat{\rho}_t\) and \(\hat{\sigma}^2_t, t = 1, \ldots, 6\) as above. Then consistent estimates of \(\rho_W, \rho_D, \sigma_W^2\) and \(\sigma_D^2\) are the partially restricted average estimates.

\[
\hat{\rho}_W = \frac{1}{3}(\hat{\rho}_1 + \hat{\rho}_3 + \hat{\rho}_5), \quad \hat{\rho}_D = \frac{1}{3}(\hat{\rho}_2 + \hat{\rho}_4 + \hat{\rho}_6)
\]

and

\[
\hat{\sigma}_W^2 = \frac{1}{3}(\hat{\sigma}_1^2 + \hat{\sigma}_3^2 + \hat{\sigma}_5^2), \quad \hat{\sigma}_D^2 = \frac{1}{3}(\hat{\sigma}_2^2 + \hat{\sigma}_4^2 + \hat{\sigma}_6^2).
\]

Again, these are two-stage estimates, which imply

\[
\tilde{\Phi}_W = (I_N - \hat{\rho}_WM) / \sqrt{\hat{\sigma}_W^2}, \quad \tilde{\Phi}_D = (I_N - \hat{\rho}_DM) / \sqrt{\hat{\sigma}_D^2},
\]

which can be substituted into equation (5). Then fixed effect estimation of equation (5) with \(\alpha^* = [\tilde{\Phi}_W \tilde{\Phi}_D \tilde{\Phi}_W \tilde{\Phi}_D \tilde{\Phi}_W \tilde{\Phi}_D] \otimes \alpha\) is consistent for \(\alpha^*\) and \(\beta\).

Define \(\varepsilon_W' = [\varepsilon_1' \varepsilon_3' \varepsilon_5']\), \(\varepsilon_D' = [\varepsilon_2' \varepsilon_4' \varepsilon_6']\), \(\tilde{\alpha}_{W'} = [\tilde{\alpha}_1' \tilde{\alpha}_3' \tilde{\alpha}_5']\) and \(\tilde{\alpha}_{D'} = [\tilde{\alpha}_2' \tilde{\alpha}_4' \tilde{\alpha}_6']\). Additionally, let \(\tilde{\alpha}_{j} = M\tilde{\alpha}_{j}, \tilde{\alpha}_{j} = M\tilde{\alpha}_{j}, \varepsilon_j = M\varepsilon_j, \varepsilon_j = M\varepsilon_j, j = W, D,\)

It follows analogously that the single stage estimates are:

\[
(\tilde{\rho}_j, \tilde{\sigma}_j^2) = \arg \min_{r,s^2} \{v_j(r, s^2) v_j(r, s^2) : s^2 \geq 0\}, \ j = W, D,
\]
where,

\[ v_j(\rho_j, \sigma^2_j) = G_j[\rho_j, \rho_j, \sigma^2_j] - g_j, j = W, D, \]

and where \( G_j \) and \( g_j \) are \( G_t \) and \( g_t \) of equation (6), but with \( j \) substituted for \( t \) and \( 3N \) substituted for \( N \). Call these estimates the *partially restricted moment estimates*. The \( \tilde{\rho}_j \)
and \( \tilde{\sigma}^2_j \) imply \( \tilde{\Phi}_j \) for wet and dry seasons, and fixed effect estimation of equation (5) is again consistent for \( \alpha^* \) and \( \beta \).

So to summarize, the unrestricted estimation procedure yields \( \hat{\rho}_t \) and \( \hat{\sigma}^2_t \), but the models slope parameters are not identified. These estimates imply fully restricted average estimates (\( \hat{\rho} \) and \( \hat{\sigma}^2 \)) or partially restricted average estimates (\( \hat{\rho}_j \) and \( \hat{\sigma}^2_j \)). These are two-stage estimates. Fully restricted moment estimates (\( \tilde{\rho} \) and \( \tilde{\sigma}^2 \)) and partially restricted moments estimates (\( \tilde{\rho}_j \) and \( \tilde{\sigma}^2_j \)) are single stage estimates. We now illustrate these estimates in an empirical example.

4. Application to Indonesian Rice Farms

We now apply the estimators to a balanced panel of Indonesian rice farms. For the panel specification of a stochastic frontier model, \( y \) is the natural logarithm of output (\( \ln(\text{rice}) \)), \( x \) is a vector of inputs (e.g. seed and fertilizer) and \( \alpha_i \) embodies farm-level technical inefficiency. In order to perform the spatial analysis we first specify the spatial weighting matrix, \( M_t \) for the error process, which captures productivity spillovers across farms. The following section considers geographical and climate characteristics of the island of West Java, which motivate the construction of three different weighting matrixes used in subsequent analyses.

4.1 Geographical and Climatic Characteristics of West Java.

In 1977 the Indonesian Ministry of Agriculture began to survey 171 rice farms concerning farming practices over 6 (3 wet and 3 dry) growing seasons. The farms were

---

6 Data set was previously analyzed by e.g. Erwidodo (1990), Lee (1991), Lee and Schmidt (1993), Horrace and Schmidt (1996, 1999). For detailed discussion of the data see Erwidodo (1990).
selected from 6 villages located in the production area of Cimanuk River Basin in West Java. Of the six villages included in the sample, two are located on the north coast of the island in an area with average altitudes of 10-15 meters above sea level; these are classified as *lowland* villages. Another three villages are situated in a *highland* area (600-1100 meters) in the central part of West Java. The last village with average altitude of 375 meters is classified as a *midland* village (for lack of better terminology). The infrastructure in the Cimanuk River Basin is fairly heterogeneous. Some of the villages (in both high and lowland areas) lack reliable transportation systems with local roads being almost impassable in the wet (rainy) season. Other villages, located in close proximity to province capital cities, are highly accessible along paved, all-weather roads.

Across the island of West Java average weather conditions are also highly variable. Annual temperatures average in the range of 72-84°F (22-29°C) with average humidity of 75%. However, the north coastal plains are usually hotter in dry season 94°F (34°C) and are more humid then the rest of the island. Highlands of the island receive an average annual rainfall of 156 inches (4000 mm) while lowlands of the north coast receive less than one forth of that amount - 35 inches (900 mm).

Based on these facts we constructed and performed our analysis using three different weighting matrices $M_{1t}, M_{2t}$, and $M_{3t}$. The first one, $M_{1t}$, was based on the *inverse of geographical distance* between individual farms. We used geographical coordinates of the villages to determine physical distances between producing units. Distances between individual villages are between 31 and 91 km. The individual distances between farms within the same village is unavailable and is therefore arbitrarily chosen to be 10 km. The $M_{1t}$ weighting matrix then consists of the inverse values of these distances. That is, $m_{ijt}$ equals the inverse of the distance between farms $i$ and $j$. In the second weighting matrix we employed an intra-village contiguity scheme. For $M_{2t}$, we

---

7 Per Schmidt and Sickle (1984) a measure of technical efficiency for farm $i$ is given by $\exp(\alpha_i - \max_j \alpha_j)$.
8 The survey ended in 1983, so the infrastructure description may be different from the current state.
9 Cliff and Ord (1973,1981) first measured potential interactions between spatial units using a combination of distance measures and relative length of the common border (contiguity). Since there is no measure of contiguity available in our case we use physical distance only as a proxy for interdependence between spatial units.
10 Moran (1948) and Geary (1954) advanced initial measures of spatial dependence (spatial correlation) that were based on the notion of binary contiguity between spatial units. Underlying spatial structure is expressed by 0-1 values: if spatial units have common border (are contiguous) value of 1 is assigned.
let $m_{ijt}$ equal 1 if farms $i$ and $j$ are in the same village, equal 0 otherwise. That is the weighting scheme is based on common villages. The last weighting scheme, $M_{3n}$, reflects the geographic and climate conditions of villages at different altitudes. The highland villages receive substantially more precipitation than low land villages and are characterized by lower average temperature than lowland villages. Moreover, certain villages are inaccessible during the rainy season. Therefore in the last weighting matrix, $M_{3n}$, we assigned values based on differences in altitudes as follows.

$$
\begin{align*}
    m_{ijt} &= 1.00 \text{ if farms } i \text{ and } j \text{ are in the same village} \\
    m_{ijt} &= 0.75 \text{ if farms } i \text{ and } j \text{ are in different villages but at same altitudes.} \\
    m_{ijt} &= 0.50 \text{ if farms } i \text{ is highland and farm } j \text{ is in midland.} \\
    m_{ijt} &= 0.50 \text{ if farms } i \text{ is midland and farm } j \text{ is in lowland.} \\
    m_{ijt} &= 0.00 \text{ if farms } i \text{ is highland and farm } j \text{ is in lowland.}
\end{align*}
$$

For subsequent computational simplification and as a standard practice in forming weighting matrices we normalize each weighting matrix, so the elements of each row sum to one. Additionally, all the weighting schemes are time invariant, so the $t$ subscript can be dropped.

### 4.2 Spatial Analysis of Indonesian Rice Farms.

We first estimate the standard fixed effect model of stochastic production frontier described by (1) and (2). Inputs to the production of rice included in the data set are seed (kg), urea (kg), trisodium phosphate (TSP) (kg), labor (labor-hours) and land (hectares). Output is measured in kilograms of rice. The data also include dummy variables. $DP$ equals 1 if pesticides were used and 0 otherwise. $DV1$ equals 1 if high yield varieties of rice were planted and $DV2$ equals 1 if mixed varieties were planted; the omitted category represents that traditional varieties were planted. $DSS$ equals 1 if it was a wet season, zero otherwise. Results are contained in column I of Table 1 and are based on the restriction that $\rho_1 = \ldots = \rho_6 = 0$. These results are identical to those contained in Horrace and Schmidt (1996).

Before embarking on a spatial analysis, we use the residuals from the standard fixed effect estimation to determine whether or not spatial dependence (based on each of
our three weighting schemes) exists in the data. As before, let the usual fixed effect residuals in period \( t \) be \( \tilde{u}_t \). We employ the most widely used test for spatial dependence, called the Moran I statistic (see e.g. Anselin (1988)). (To preclude confusion with the symbol for the identity matrix we adopt the script \( \vartheta \).) The \( \vartheta \) statistic for period \( t \) is:

\[
\vartheta_t = \frac{N}{S} \{ [\tilde{u}_t'M\tilde{u}_t] / \tilde{u}_t'\tilde{u}_t \}
\]

where \( N \) is the number of farms, \( S \) is the sum of all elements in weighting matrix \( M \) (\( S \) equals \( N \) if \( M \) is normalized so that sum of row elements equals one.) The null hypothesis for this test is “absence of spatial dependence”. Notice that we have dropped the \( t \) subscript on the weighting matrix \( M \), because our empirical analysis assumes time invariance for spatial dependence. As shown by Cliff and Ord (1972, 1973, 1981) the asymptotic distribution for the Moran statistic is standard normal, if \( \vartheta \) is transformed in the usual manner:

\[
z_t = \{ \vartheta_t - E[ \vartheta_t ] \} / \sqrt{V[\vartheta_t]}
\]

where \( E[ \vartheta_t ] \) is the mean, and \( V[\vartheta_t] \) is the variance of Moran \( \vartheta \) statistics in period \( t \), derived under the null of no spatial dependence. In the general case of a non-normalized weighting matrix these can be expressed in the form:

\[
E[\vartheta_t] = \frac{(N/S) tr(PM)}{(N-k)}
\]

\[
V[\vartheta_t] = \frac{(N/S)^2 \{ tr(PMPM') + tr(PM)^2 + [tr(PM)]^2 (N-k)(N-k+2) - \{E[\vartheta_t]\}^2 \}}{N-k+2}
\]

where \( P \) is the projection matrix \( I_N - x(x'x)^{-1}x' \), and \( x \) is the demeaned exogenous variables from the standard model. The test was conducted for each weighting scheme \( (M1, M2, M3) \) in each time period \( t = 1, \ldots, 6 \). The \( z_t \) -scores for weighting scheme \( M1 \) are contained in the top row \( (z_t) \) of Table 2 and range from 6.0702 in period \( t=2 \) to 26.4159 in period \( t=4 \). It is therefore safe to conclude that at the 95% confidence level, we reject the hypothesis of no spatial dependence based on weighting scheme \( M1 \). Test results for weighting schemes \( M2 \) and \( M3 \) were similar and are contained in the top rows \( (z_t) \) of Table 4 and Table 6, respectively.
Based on these test results, each of our proposed weighting schemes appears justified. Consequently, we estimated the unrestricted spatial autoregressive parameters and error variances for each period for each schemes, using equation (6). Notice that the autoregressive and variance parameter estimates are identified for each period, even though the parameters $\alpha^*$ and $\beta$ in equation (5) are not. Estimation results are contained in Tables 2, 4 and 6 for schemes $M1$, $M2$ and $M3$, respectively. Note that for all weighting schemes, the $\rho$-parameter tends to be larger in period 1 than in period 2, larger in period 3 than period 4, and larger in period 5 than in period 6. These differences correspond to differences in wet seasons ($t = 1, 3, 5$) and dry seasons ($t = 2, 4, 6$). All autoregressive parameters are positive, and in only a single case does it exceed unity (scheme $M3$, period 3).

To identify parameter estimates for $\alpha^*$ and $\beta$ in equation (5) we feasibly estimated the fully and partially-restricted systems described in sections 3.1 and 3.2, respectively. The fully restricted system, $\rho_1 = \ldots = \rho_6 = \rho$, was estimated using both the average autoregressive parameter, $\hat{\rho}$, and the moments autoregressive parameter, $\tilde{\rho}$, for each weighting scheme. Estimation results for $\hat{\rho} = 0.7248$ and for $\tilde{\rho} = 1.0557$ using weighting scheme $M1$ are contained in Table 1, columns II and III, respectively. There is little difference in the slope parameter estimates based on $\hat{\rho}$ or $\tilde{\rho}$ or the standard FE model of column I. This is not surprising, given that ignoring the spatial dependence manifests itself as an efficiency loss in the slope parameter estimates (not a bias). Indeed, the most noticeable differences in the estimates of columns I, II and III are in the standard error estimates, with Columns II and III being generally more efficient that column I, the standard model. The sign of the coefficient on the pesticide variable (DP) changes from positive to negative when we include spatial effects, however it is always insignificant. The difference in magnitudes of $\hat{\rho}$ and $\tilde{\rho}$ is troublesome. Perhaps this difference indicates that the restriction $\rho_1 = \ldots = \rho_6 = \rho$, does not hold. We did not attempt to test this, however it would be possible if the variance matrix of the $\rho_i$ were estimable; this is currently under investigation by the authors. The results of the fully restricted model under weighting schemes $M2$ and $M3$ are contained in columns II and III in Tables 3 and 5, respectively. The results are similar to the $M1$ case: slope coefficients do not change
much, standard error estimates decrease, and there is a large difference between the two estimates of $\rho$.

Feasible estimation of the partially-restricted system follows the same pattern, except that instead of only one correlation coefficient fixed for all time periods now we estimate and utilize two correlation coefficients – one for wet and one for dry season. We calculate the *average parameter estimates* ($\hat{\rho}_w$, $\hat{\rho}_D$) and *the moments estimates* ($\tilde{\rho}_w$, $\tilde{\rho}_D$) for weighting scheme. Fixed effect estimation results for ($\hat{\rho}_w$, $\hat{\rho}_D$) and for ($\tilde{\rho}_w$, $\tilde{\rho}_D$), based on weighting scheme $M1$, are contained in Table 1, columns IV and V, respectively. The differences between the average and moments parameter estimates are much less pronounced than the fully restricted case (compare estimates $\hat{\rho}_w = 0.7584$ to $\tilde{\rho}_w = 0.8218$, estimates $\hat{\rho}_D = 0.6914$ to $\tilde{\rho}_D = 0.7476$, and estimates $\hat{\rho} = 0.7248$ to $\tilde{\rho} = 1.0557$). One might conclude that the partially restricted model seems to fit the data better, however this is not formally tested. Again, the standard errors of the slope parameter estimates are smaller for the partially restricted model than for the standard model (column I). Of course the coefficient on the season variable (DSS) is not identified, since it is effectively time invariant now that the data set have been dichotomized into “wet” and “dry” sub-samples. The coefficients on the partially restricted system are generally higher than those of the fully restricted system (columns II and III) and the standard model (column I). As in fully restricted system, the coefficient on the pesticide variable (DP) is negative and insignificant. Even though it is insignificant, this is troubling, since economic theory usually dictates increasing marginal products. However, one could argue that too much pesticide might have a negative effect on output. Alternatively, one could argue that we have not adequately controlled for the interaction between pesticides (DP), output (y) and weather (DSS, $\rho_w$ and $\rho_D$). Perhaps, pesticide use is higher during the wet season (more water, more insects), and our simple dummy variable for pesticide does not adequately capture a more complex relationship. Nonetheless, the results are compelling and the coefficient is insignificant. Estimation results for weighting schemes M2 and M3 are similarly presented in columns IV and V of
Table 3 and Table 5, respectively. For both the “common village” and “altitude difference” specifications, the coefficient of pesticides (DP) is negative but insignificant. Therefore these weighting schemes may be more reasonable than scheme \( M1 \) for this particular sample.

### 4.3 Technical Efficiency Rankings.

Stochastic frontier models are often concerned with estimating firm level technical inefficiency and, in particular, determining the relative magnitudes of the resulting inefficiency measures, using a rank or order statistic. In the next analysis we demonstrate how the various weighting schemes effect the technical efficiency rankings of the farms. Specifically, for each weighting scheme we estimated and ranked the technical efficiencies, \( \exp(\alpha_i - \max_j \alpha_j) \), for each farm. This was performed for the standard fixed effect model (corresponding to column I of Table 1) and for the fully restricted moments estimator (corresponding to column III of Tables 1, 3, and 5).\(^{12}\) The idea was to see how the rankings differed between the standard model and the spatial model for each of the three weighting schemes. Order statistics for each model are contained in Table 7. The first three columns of the table are results for the standard fixed effect model. Since there are 171 farms we only report results for the four farms with the highest technical efficiency, the four farms with the median technical efficiency, and the four farms with the lowest technical efficiency. Column 1 contains the farm number, column 2 contains the ordered estimates of farm-level technical efficiency, and column 3 contains the ordinal rankings for the standard fixed effect model (numbered 1 to 171). To see the effects of the spatial dependence on technical efficiency estimation, we also report the ordinal rankings for the same 12 farms for the fully restricted spatial model under weighting schemes \( M1, M2 \) and \( M3 \) in columns 4, 5 and 6, respectively.\(^{13}\)

While there are some changes across weighting schemes in the rank ordering of the most-

---

\(^{11}\) Note that even though the time dimension have effectively been cut in half by this dichotomy, the estimates of the slope parameters are still based on the entire sample \( TN \) after the observables have been demeaned based on whether they are “dry” or “wet”.

\(^{12}\) We did not consider comparing the rankings of the partially-restricted system, since the \( \alpha \) estimates in are based on only 3 observations and in the other cases they are based on 6 observations.

\(^{13}\) To save space we do not report the actual technical efficiency estimates for the spatial models only the efficiency ranking.
and least-efficient farms, these are minor. For instance in the standard model farm 152 had a technical efficiency rank of 4, but it has a rank of 6 under weighting schemes $M1$ and $M2$ and a rank of 4 under weighting scheme $M3$. Notice that the ranking of the most efficient farm (farm 164) is always 1 and that of the least efficient farm (farm 45) is always 171. Most of the largest differences in ranking appear in the median farms. For example, farm 166 has a standard fixed effect ranking of 85 but spatial rankings of 166, 166 and 108 for $M1$, $M2$ and $M3$, respectively. These are large changes in the ranking, that could only be detected with a spatial analysis.

To further summarize the changes in the efficiency ranking in Table 7, we calculate Spearman’s rho ($r_s$) for each weighting scheme, using the standard fixed effect model as the baseline. Spearman’s rho is a standard measure of rank correlation between two rank statistics given by

$$r_s = 1 - \frac{6 \sum \delta_i^2}{N(N^2 - N)},$$

where $\delta_i$ is the difference in the rankings for the $i^{th}$ farm. For example when comparing the rank statistic for the standard model and the $M1$ model in Table 7, $\delta_{164} = 0$ and $\delta_{15} = 24$. Here we always compare the rankings of the $M1$, $M2$ and $M3$ models to the standard model ranking. It is true that $r_s \in [-1, 1]$, $r_s = 1$ when the two rank statistics are identical and $r_s = -1$ when the rank statistics are completely reversed (i.e. as we move from one order statistic to the other, the most efficient farm becomes the least efficient, the second most efficient farm become the second-least efficient....). Spearman statistics are contained in the last row of Table 7 and are on the order of 0.8 for each of the three weighting schemes. We can interpret this result as saying that 80% of the rank statistic is preserved when we use a spatial weighting specification over the standard fixed effect specification. One could interpret this as a profound change in the technical efficiency rankings.

5. Conclusions

This paper has presented a straight-forward generalization of the cross sectional model of Kelejian and Prucha (1999). However, the implications of the results transcend
the spatial econometrics literature. Because economic agents and entities have finite lives, one cannot always rely on large $T$ in economic panel data sets. There can be no denying that most panel data sets (with the exception of perhaps microeconomic financial data) have large $N$ and small $T$. Moreover, the costliness of empirical research necessarily impedes long, protracted data collection exercises; it is just cheaper to collect data sets with large $N$ and small $T$. Additionally, if $T$ is somewhat large, the usually time-invariant unobserved heterogeneity models (e.g. fixed effect) may not be applicable, since it is widely held that heterogeneity may change in long-run dynamic economic systems (particularly when it is viewed as technical inefficiency). The result is that consistency arguments usually must hinge on $N \rightarrow \infty$. This is fine for estimating conditional means (the model’s slope parameters). However, any second moment parameters (such as the elements of $M$) that embody cross sectional dependence cannot be consistently estimated in the sense that they will necessarily rely on $T \rightarrow \infty$. This is unfortunate, because there can also be no denying that cross sectional dependencies do exist in economic field data.

When faced with this dilemma, empirical economic researchers have two recourses, 1) collect more data or 2) impose more structure on the model and hope that the structure will be testable. Given the aforementioned arguments against large $T$, it would seem that we are faced with the alternative of imposing more structure on our models. The question then becomes, “what structure is reasonable”? Spatial weighting schemes based on some geographic or economic proximity measures seem to be a reasonable and natural approach. The theoretical economic literature is rife with arguments for economic spillovers, and spatial analysis provides a means to make these spillovers explicit. Moreover, the structures are testable (as demonstrated) with the Moran I statistic. Therefore, if we must make assumptions about the second moments of our data, spatial weighting schemes may be less ad hoc than they first appear.

We have presented two special cases of a more general, unidentified model: the fully-restricted case and the partially-restricted case. Clearly, the possibilities for the partially restricted case are limitless. Our partial restriction based on wet and dry seasons was obviously data driven, so alternative restrictions for different types of data sets are possible. For example, infrastructure changes, catastrophic events or political dynamics
may imply a different set of restrictions on the spatial correlation parameterization. Tests of these partial restrictions should be a high priority in subsequent research.

Notice that, dynamic spatial dependence in the second moment of our estimators has implications for dynamics in the first moment. The original model in equation (1) has a time invariant \( \alpha \) parameter, but the transformed model had time varying \( \alpha^* \). It is this loss of time-invariance that makes the general model not-identified, and forces us to impose some restrictions on the dynamics of the spatial dependence. This could be important. Most panel data models that attempt to make the \( \alpha \) term dynamic, do so by imposing structure on the first moments of the models. For instance, several papers in the stochastic frontier literature impose special structure on the conditional first moment of \( \alpha \). For example: \( \alpha_t = \mu + \delta t + \gamma t^2 \).

The models presented here create dynamic \( \alpha \) through second moment conditions on \( u_t \). What the implications of this difference will be for the dynamic panel data literature is unknown, but it is interesting to point this difference this out. Additionally, spatial dependence may be a way to indirectly incorporate time-invariant regressors into a fixed effect model. For example, Horrace and Schmidt (1996) analyze the same data and incorporate five dummy variables for the six villages into a GLS or random effects specification, but they are forced to leave these dummy variable out of the fixed effect specification, because they are time-invariant at the farm level. In the application presented here, village effects are incorporated into second moment of the residual in the form of distances between villages in weighting matrix \( M1 \) and in the form of contiguity or “common village” in matrix \( M2 \). While there are commonly employed techniques for incorporating time-invariant regressors into a fixed effect model, the research presented here provides analysts with an alternative means.

Our empirical example demonstrates the utility of this research; there are direct applications of this procedure to the stochastic frontier literature. We have illustrated the dangers associated with ignoring spatial dependence: a) a potential efficiency loss and b) the possibility of incorrect assessment of the technical efficiency rank statistics implied by the model. However, this modeling approach has much broader empirical

---

14 For examples see Cornwell, Schmidt and Sickles (1990), Lee and Schmidt (1993), Battese and Coelli (1992) and Kumbhakar (1990).
implications. Any fixed effect model for a panel of data with fixed $T$ could benefit from this type of analysis tool, if there are compelling reasons to believe that spatial dependence exists and if there are additional data (such as geographic or economic proximity measures) that motivate selection of a reasonable spatial weighting scheme.

---

15 See Hausman and Taylor (1981) for a technique for time-invariant regressors.
References


Table 1. Weighting Scheme $M1$ – Inverse of Distance

<table>
<thead>
<tr>
<th></th>
<th>Standard FE Model</th>
<th>Fully Restricted Average</th>
<th>Fully Restricted Moment</th>
<th>Partially Restricted Average</th>
<th>Partially Restricted Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I I I I I I</td>
<td>I I I I I I</td>
<td>I I I I I I</td>
<td>I I I I I I</td>
<td>I I I I I I</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>-</td>
<td>0.7248</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>-</td>
<td>-</td>
<td>1.0557</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_w$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.7584</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_D$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.6914</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_w$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.8218</td>
</tr>
<tr>
<td>$\hat{\rho}_D$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.7476</td>
</tr>
</tbody>
</table>

Seed 0.1208 0.1038 0.0998 0.1292 0.1248
(0.030) (0.025) (0.024) (0.024) (0.024)

Urea 0.0918 0.0894 0.0901 0.1405 0.1440
(0.021) (0.018) (0.017) (0.015) (0.015)

TSP 0.0892 0.0353 0.0244 0.0340 0.0307
(0.013) (0.012) (0.012) (0.011) (0.011)

Labor 0.2431 0.2366 0.2379 0.2254 0.2204
(0.032) (0.029) (0.028) (0.026) (0.026)

Land 0.4521 0.4879 0.4931 0.5046 0.5141
(0.035) (0.031) (0.030) (0.027) (0.027)

DP 0.0338* -0.0178* -0.0298* -0.0224* -0.0212*
(0.032) (0.028) (0.028) (0.025) (0.025)

DV1 0.1788 0.1084 0.0935 0.1250 0.1320
(0.041) (0.038) (0.038) (0.034) (0.035)

DV2 0.1754 0.1060 0.0952 0.0917 0.0947
(0.057) (0.049) (0.048) (0.048) (0.048)

DSS 0.0533 0.0759* 0.1062* - -
(0.022) (0.063) (0.302) - -

R-sq 0.910228 0.9246 0.9271 0.9190 0.9177

Note: Numbers in parenthesis are standard errors. All estimates are significant at least at 5% significance level except those marked with an asterisk.

Table 2. Unrestricted Estimates, Weighting Scheme $M1$.

<table>
<thead>
<tr>
<th>Time period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_i$</td>
<td>0.6217</td>
<td>0.5217</td>
<td>0.8694</td>
<td>0.8375</td>
<td>0.7705</td>
<td>0.7101</td>
</tr>
<tr>
<td>$\sigma_i^2$</td>
<td>0.0405</td>
<td>0.0828</td>
<td>0.0782</td>
<td>0.0713</td>
<td>0.0485</td>
<td>0.0658</td>
</tr>
</tbody>
</table>
Table 3. Weighting Scheme $M2$ – Common Villages

<table>
<thead>
<tr>
<th></th>
<th>Standard FE Model</th>
<th>Fully Restricted Average</th>
<th>Fully Restricted Moment</th>
<th>Partially Restricted Average</th>
<th>Partially Restricted Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>-</td>
<td>0.6604</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>-</td>
<td>-</td>
<td>0.9882</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_w$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.6811</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_D$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.6398</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_{wD}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.7388</td>
</tr>
<tr>
<td>$\hat{\rho}_{D}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.6999</td>
</tr>
<tr>
<td>Seed</td>
<td>0.1208</td>
<td>0.1035</td>
<td>0.0996</td>
<td>0.1255</td>
<td>0.1248</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.025)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Urea</td>
<td>0.0918</td>
<td>0.0909</td>
<td>0.0901</td>
<td>0.1435</td>
<td>0.1446</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.018)</td>
<td>(0.017)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>TSP</td>
<td>0.0892</td>
<td>0.0356</td>
<td>0.0239</td>
<td>0.0326</td>
<td>0.0301</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Labor</td>
<td>0.2431</td>
<td>0.2385</td>
<td>0.2376</td>
<td>0.2201</td>
<td>0.2198</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.029)</td>
<td>(0.028)</td>
<td>(0.026)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Land</td>
<td>0.4521</td>
<td>0.4855</td>
<td>0.4934</td>
<td>0.5131</td>
<td>0.5148</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.031)</td>
<td>(0.030)</td>
<td>(0.028)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>DP</td>
<td>0.0338*</td>
<td>-0.0189*</td>
<td>-0.0306*</td>
<td>-0.0208*</td>
<td>-0.0219*</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.025)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>DV1</td>
<td>0.1788</td>
<td>0.1116</td>
<td>0.0928</td>
<td>0.1335</td>
<td>0.1326</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.038)</td>
<td>(0.038)</td>
<td>(0.034)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>DV2</td>
<td>0.1754</td>
<td>0.1080</td>
<td>0.0947</td>
<td>0.0970</td>
<td>0.0961</td>
</tr>
<tr>
<td></td>
<td>(0.057)</td>
<td>(0.049)</td>
<td>(0.048)</td>
<td>(0.049)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>DSS</td>
<td>0.0533</td>
<td>0.0789*</td>
<td>0.0844*</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.051)</td>
<td>(1.424)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

R-sq | 0.910228 | 0.9240 | 0.9271 | 0.9171 | 0.9174

Note: Numbers in parenthesis are standard errors. All estimates are significant at least at 5% significance level except those marked with an asterisk.

Table 4. Unrestricted Estimates, Weighting Scheme $M2$.

<table>
<thead>
<tr>
<th>Time period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_i$</td>
<td>7.4954</td>
<td>5.0873</td>
<td>23.5416</td>
<td>23.5057</td>
<td>13.6585</td>
<td>11.1836</td>
</tr>
<tr>
<td>$\hat{\rho}_z$</td>
<td>0.5682</td>
<td>0.4803</td>
<td>0.7875</td>
<td>0.7889</td>
<td>0.6875</td>
<td>0.6501</td>
</tr>
<tr>
<td>$\sigma_i^2$</td>
<td>0.0407</td>
<td>0.0842</td>
<td>0.0774</td>
<td>0.0718</td>
<td>0.0481</td>
<td>0.0661</td>
</tr>
</tbody>
</table>
### Table 5. Weighting Scheme M3 – Altitude Differences

<table>
<thead>
<tr>
<th></th>
<th>Standard FE Model</th>
<th>Fully Restricted Average</th>
<th>Fully Restricted Moment</th>
<th>Partially Restricted Average</th>
<th>Partially Restricted Moment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>$r_{\hat{\rho}}$</td>
<td>0.8648</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}$</td>
<td>$r_{\hat{\rho}}$</td>
<td>-</td>
<td>1.0673</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_w$</td>
<td>$r_{\hat{\rho}_w}$</td>
<td>-</td>
<td>-</td>
<td>0.9130</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_D$</td>
<td>$r_{\hat{\rho}_D}$</td>
<td>-</td>
<td>-</td>
<td>0.8166</td>
<td>-</td>
</tr>
<tr>
<td>$\hat{\rho}_w$</td>
<td>$r_{\hat{\rho}_w}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.9929</td>
</tr>
<tr>
<td>$\hat{\rho}_D$</td>
<td>$r_{\hat{\rho}_D}$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.8446</td>
</tr>
</tbody>
</table>

Seed       0.1208 0.1114 0.1095 0.1430 0.1500  
(0.030) (0.025) (0.025) (0.023) (0.023)  
Urea       0.0918 0.0943 0.0963 0.1377 0.1342  
(0.021) (0.018) (0.018) (0.015) (0.015)  
TSP        0.0892 0.0325 0.0283 0.0330 0.0341  
(0.013) (0.012) (0.012) (0.011) (0.011)  
Labor      0.2431 0.2489 0.2522 0.2362 0.2407  
(0.032) (0.028) (0.028) (0.026) (0.026)  
Land       0.4521 0.4794 0.4800 0.4902 0.4803  
(0.035) (0.031) (0.030) (0.028) (0.027)  
DP         0.0338* -0.0185* -0.0256* 0.0033* -0.0057*  
(0.032) (0.028) (0.028) (0.023) (0.023)  
DV1        0.1788 0.1216 0.1203 0.1631 0.1610  
(0.041) (0.037) (0.036) (0.034) (0.035)  
DV2        0.1754 0.1304 0.1309 0.1205 0.1168  
(0.057) (0.049) (0.048) (0.050) (0.048)  
DSS        0.0533 0.0804* 0.0683* - -  
(0.022) (0.128) (0.254) - -  

R-sq        0.910228 0.9280 0.9289 0.9180 0.9196  

Note: Numbers in parenthesis are standard errors. All estimates are significant at least at 5% significance level except those marked with an asterisk.

### Table 6. Unrestricted Estimates, Weighting Scheme M3

<table>
<thead>
<tr>
<th>Time period</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$z_t$</td>
<td>14.0899</td>
<td>15.9659</td>
<td>57.4523</td>
<td>54.5444</td>
<td>24.8664</td>
<td>26.3768</td>
</tr>
<tr>
<td>$\hat{\rho}_t$</td>
<td>0.7510</td>
<td>0.7129</td>
<td>1.0384</td>
<td>0.9044</td>
<td>0.9497</td>
<td>0.8325</td>
</tr>
<tr>
<td>$\sigma_t^2$</td>
<td>0.0415</td>
<td>0.0811</td>
<td>0.0765</td>
<td>0.0764</td>
<td>0.0501</td>
<td>0.0661</td>
</tr>
</tbody>
</table>
### Table 7. Orders Statistics, Various Models

<table>
<thead>
<tr>
<th>Farm #</th>
<th>Standard FE Efficiency</th>
<th>Standard FE Model</th>
<th>Weight Scheme M1</th>
<th>Weight Scheme M2</th>
<th>Weight Scheme M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>164</td>
<td>100%</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>118</td>
<td>93.23%</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>163</td>
<td>93.03%</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>152</td>
<td>89.93%</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>55.62%</td>
<td>84</td>
<td>106</td>
<td>106</td>
<td>114</td>
</tr>
<tr>
<td>166</td>
<td>55.47%</td>
<td>85</td>
<td>116</td>
<td>116</td>
<td>108</td>
</tr>
<tr>
<td>15</td>
<td>55.40%</td>
<td>86</td>
<td>62</td>
<td>62</td>
<td>72</td>
</tr>
<tr>
<td>40</td>
<td>55.35%</td>
<td>87</td>
<td>54</td>
<td>54</td>
<td>64</td>
</tr>
<tr>
<td>86</td>
<td>39.80%</td>
<td>168</td>
<td>165</td>
<td>165</td>
<td>166</td>
</tr>
<tr>
<td>143</td>
<td>38.37%</td>
<td>169</td>
<td>169</td>
<td>169</td>
<td>170</td>
</tr>
<tr>
<td>117</td>
<td>37.90%</td>
<td>170</td>
<td>168</td>
<td>168</td>
<td>168</td>
</tr>
<tr>
<td>45</td>
<td>36.55%</td>
<td>171</td>
<td>171</td>
<td>171</td>
<td>171</td>
</tr>
</tbody>
</table>

$r_5$: 1.0000 0.8027 0.8095 0.8674

### Appendix

Assumptions 3, 4, and 5 from Kelejian and Prucha (1999). Let $P(p_t) = (I_N - \rho M_t)^{-1}$ with typical element $p_{ij}(\rho_t)$.

**Assumption 3:**

(i) The sums $\sum |m_{ij}|$ and $\sum |m_{ji}|$ are bounded by say, $c_m < \infty$ for all $1 \leq i, j \leq N, N \geq 1$. (ii) The sums $\sum |p_{ij}(\rho_t)|$ and $\sum |p_{ji}(\rho_t)|$ are bounded by say, $c_p < \infty$ for all $1 \leq i, j \leq N, N \geq 1, |\rho_t| < 1$.

**Assumption 4:** Let $\tilde{u}_{it}$ be the $i^{th}$ element of $\tilde{u}_t$. There exists finite dimensional random vectors $d_{it}$ and $\Delta_t$ such that $|\tilde{u}_{it} - u_{it}| \leq ||d_{it}|| ||\Delta_t||$ with $N^{-1}\sum ||d_{it}||^{2+\delta} = O_p(1)$ for some $\delta > 0$ and $N^{\frac{1}{2}}\sum ||\Delta_t|| = O_p(1)$.

**Assumption 5:** The smallest eigen value of $\Gamma_t^\top \Gamma_t$ is bounded away from zero.