Lecture 1: Introduction to Spatial Autoregressive (SAR) Models

1. Some SAR models:

Some specified models for cross sectional data which may capture possible spatial interactions across spatial units.

The following models are generalizations of autoregressive processes and autoregression models in time series.

1.1) **(First order) spatial autoregressive (SAR) process:**

$$y_i = \lambda w_{i,n} Y_n + \epsilon_i, \qquad i = 1, \cdots, n,$$

where $Y_n = (y_1, \cdots, y_n)'$ is the column vector of dependent variables, $w_{i,n}$ is a n-dimensional row vector of constants, and ϵ_i's are i.i.d.$(0, \sigma^2)$. In the vector/matrix form,

$$Y_n = \lambda W_n Y_n + \mathcal{E}_n.$$

The term $W_n Y_n$ has been termed 'spatial lag'. This is supposed to be an equilibrium model. Under the assumption that $S_n(\lambda) = I_n - \lambda W_n$ is nonsingular, one has

$$Y_n = S_n^{-1}(\lambda)\mathcal{E}_n.$$

This process model may not be not useful alone in empirical econometrics but it is used to model possible spatial correlations in disturbances of a regression equation. The regression model with SAR disturbance U_n is specified as

$$Y_n = X_n\beta + U_n, \qquad U_n = \rho W_n U_n + \mathcal{E}_n,$$

where $\mathcal{E}_n$ has zero mean and variance $\sigma^2 I_n$. In this regression model, the disturbances ϵ_i's in $\mathcal{E}_n$ are correlated across units according to a SAR process. The variance matrix of U_n is $\sigma_0^2 S_n^{-1}(\rho_0) S_n^{'-1}(\rho_0)$. As the diagonal elements of $S_n^{-1}(\rho_0) S_n^{'-1}(\rho_0)$ may not be zero, there are correlations of disturbances u_i's across units.

1.2) **Mixed regressive, spatial autoregressive model (MRSAR):**

This model generalizes the SAR process by incorporating exogenous variables x_i in the SAR process. It has also simply been called the spatial autoregressive model. In vector/matrix form,

$$Y_n = \lambda W_n Y_n + X_n\beta + \mathcal{E}_n,$$

where $\mathcal{E}_n$ is $(0, \sigma^2 I_n)$. This model has the feature of a simultaneous equation model and its reduced form is

$$Y_n = S_n^{-1}(\lambda) X_n \beta + S_n^{-1}(\lambda) \mathcal{E}_n.$$

1.3) Some Intuitions on Spatial Weights Matrix W_n

The matrix W_n is called the spatial weights matrix. The value $w_{n,ij}$ of the jth element of $w_{i,n}$ represents the link (or distance) between the neighbor j to the spatial unit i. Usually, the diagonal of W_n is specified to be zero, i.e., $w_{n,ii} = 0$ for all i, because $\lambda w_{i,n} Y_n$ represents the effect of other spatial units on the spatial unit i.

In some empirical applications, it is a common practice to have W_n having a zero diagonal and being row-normalized such that the summation of elements in each row of W_n is a unity. In some applications, the ith row $w_{i,n}$ of W_n may be constructed as $w_{i,n} = (d_{i1}, d_{i2}, \ldots, d_{in}) / \sum_{j=1}^n d_{ij}$, where $d_{ij} \geq 0$, represents a function of the spatial distance of the ith and jth units, for example, the inverse of the distance of the i and jth units, in some (characteristic) space. The weighting operation may be interpreted as an average of neighboring values. However, in some cases, one may argue that row-normalization for weights matrix may not be meaningful (e.g., Bell and Bochstael (2000) RESTAT for real estate problems with microlevel data).

When neighbors are defined as adjacent ones for each unit, the correlation is local in the sense that correlations across units will be stronger for neighbors but will be weak for units far away. This becomes clear with an expansion of $S_n^{-1}(\rho_0)$. Suppose that $\| \rho W_n \| \leq 1$ for some matrix norm $\| \dot{} \|$, then

$$S_n^{-1}(\rho) = I_n + \sum_{i=1}^{\infty} \rho^i W_n^i,$$

(see, e.g., Horn and Johnson 1985). As $\| \sum_{i=m}^{\infty} \rho^i W_n^i \| \leq |\rho W_n|^i \| S_n^{-1}(\rho) \|$, in particular, if W_n is row normalized, $\| \sum_{i=m}^{\infty} \rho^i W_n^i \|_\infty \leq | \sum_{i=m}^{\infty} |\rho|^i = \frac{|\rho|^m}{1-|\rho|}$, will become small when m becomes larger. The U_n can be represented as

$$U_n = \mathcal{E}_n + \rho W_n \mathcal{E}_n + \rho^2 W_n^2 \mathcal{E}_n + \cdots,$$

where ρW_n may represent the influence of neighbors on each unit, $\rho^2 W_n^2$ represents the second layer neighborhood influence, etc.

In the social science (or social interactions) literature, $W_n S_n^{-1}(\rho)$ is a vector of *measures of centrality*, which summaries the position of each spatial unit (in a network).

In conventional spatial weights matrices, neighboring units are defined by only a few adjacent ones. However, there are cases where 'neighbors' may consist of many units. An example is a social interactions model, where 'neighbors' refer to individuals in a same group. The latter may be regarded as models with large group interactions (a kind of models with social interactions). For models with a large number of interactions for each unit, the spatial weights matrix W_n will associate with the sample size. In a model, suppose there are R groups and there are m individuals in each group. The sample size will be $n = mR$. In a model without special network (e.g., frindship) information, one may assume that each individual in a group is given equal weight. In that case, $W_n = I_R \otimes B_m$ where $B_m = (l_m l_m' - I_m)/(m-1)$, $\otimes$ is the Kronecker product, and l_m is the m-dimensional vector of ones. Both m and R can be large in a sample. In which case, asymptotic analysis for large sample will be relevant with both R and m go to infinity. Of course, in the general case, the number of members in each district may be large but have different sizes. This model has many interesting applications in the study of social interactions issues.

1.4) Other generalizations of those models can have high order spatial lags and/or SAR disturbances.

A more rich SAR model may combine the MRSAR equation with SAR distrubances

$$Y_n = \lambda W_n Y_n + X_n \beta + U_n, \quad U_n = \rho M_n U_n + \mathcal{E}_n, \tag{1.5}$$

where W_n and M_n are spatial weights matrices, which may or may not be identical.

Further extension of a SAR model may allow high-order spatial lags as in

$$Y_n = \sum_{j=1}^{p} \lambda_j W_{jn} Y_n + X_n \beta + \mathcal{E}_n,$$

where W_{jn}'s are p distinct spatial weights matrices.

2. Some matrix algebra

2.1) Vector and Matrix Norms

Definition. Let V be a vector space. A function $\|\cdot\|: V \longrightarrow R$ is a *vector norm* if for any $x, y \in V$,

(1) nonnegative, $\| x \| \geq 0$,

(1a) positive, $\| x \|= 0$ if and only if $x = 0$,

(2) homogeneous, $\| cx \|= |c| \| x \|$ for any scalar c,

(3) triangle inequality, $\| x + y \| \leq \| x \| + \| y \|$.

Examples:

a) The Euclidean norm (or l_2 norm) is

$$\| x \|_2= (|x_1|^2 + \cdots + |x_n|^2)^{1/2}.$$

b) The sum norm (or l_1 norm) is

$$\| x \|_1= |x_1| + \cdots + |x_n|.$$

c) The max norm (or l_∞ norm) is

$$\| x \|_\infty= \max\{|x_1|, \cdots, |x_n|\}.$$

Definition. A function $\| \cdot \|$ of a square matrix is a *matrix norm* if, for any square matrices A and B, it satisfies the following axioms:

(1) $\| A \| \geq 0$,

(1a) $\| A \|= 0$ if and only if $A = 0$,

(2) $\| cA \|= |c| \| A \|$ for any scalar c,

(3) $\| A + B \| \leq \| A \| + \| B \|$,

(4) submultiplicative, $\| AB \| \leq \| A \| \cdot \| B \|$.

Associated with each vector norm is a natural matrix norm induced by the vector norm.

Let $\| \cdot \|$ be a vector norm. Define $\| A \|= \max_{\|x\|=1} \| Ax \|= \max_x \frac{\|Ax\|}{\|x\|}$. This function $\| \cdot \|$ for A is a matrix norm and it has the properties that $\| Ax \| \leq \| A \| \cdot \| x \|$ and $\| I \|= 1$.

Examples:

a) The *maximum column sum matrix norm* $\| \cdot \|_1$ is

$$\| A \|_1= \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|,$$

which is a matrix norm induced by the l_1 vector norm.

b) The *maximum row sum matrix norm* $\| \cdot \|_\infty$ is

$$\| A \|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|,$$

which is a matrix norm induced by the l_∞ vector norm.

One important application of matrix norms is in giving bounds for the eigenvalues of a matrix.

Definition: The *spectral radius* $\rho(A)$ of a matrix A is

$$\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

Let λ be an eigenvalue of A and x be the corresponding eigenvector, i.e., $Ax = \lambda x$. Then,

$$|\lambda| \cdot \| x \| = \| \lambda x \| = \| Ax \| \le \| A \| \cdot \| x \|,$$

and, therefore, $|\lambda| \le \| A \|$. That is, $\rho(A) \le \| A \|$.

Another useful application of the matrix norm is that, if $\| \cdot \|$ is a matrix norm, and if $\| A \| < 1$, then $I - A$ is invertible and $(I - A)^{-1} = \sum_{j=0}^{\infty} A^j$.

When all of the elements of W_n are non-negative and its rows are normalized to one, $\| W_n \|_\infty = 1$. For this case, when $|\lambda| < 1$,

$$(I_n - \lambda W_n)^{-1} = \sum_{j=0}^{\infty} (\lambda W_n)^j = I_n + \lambda W_n + \lambda^2 W_n^2 + \cdots.$$

2.2) Useful (important) regularity conditions on W_n and associated matrix:

Definition: Let $\{A_n\}$ be a sequence of square matrices, where A_n is of dimension n.

(i) $\{A_n\}$ is said to be *uniformly bounded in row sums* if $\{\| A_n \|_1\}$ is a bounded sequence;

(ii) $\{A_n\}$ is said to be *uniformly bounded in column sums* if $\{\| A_n \|_\infty\}$ is a bounded sequence.

A useful property is that if $\{A_n\}$ and $\{B_n\}$ are uniformly bounded in row sums (column sums), then $\{A_n B_n\}$ is also uniformly bounded in row sums (column sums). This follows from the submultiplicative property of matrix norms. For example, $\| A_n B_n \|_\infty \le \| A_n \|_\infty \| B_n \|_\infty \le c^2$ when $\| A_n \|_\infty \le c$ and $\| B_n \|_\infty \le c$.

An Important Assumption: The spatial weights matrices $\{W_n\}$ and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums.

Equivalently, this assumes that $\{\| W_n \|_1\}$ and $\{\| W_n \|_\infty\}$ are bounded sequences. Similarly, the maximum row and column sum norms of S_n^{-1} are bounded.

Rule out unit root process:

The assumption that S_n^{-1} are uniformly bounded in both row and column sums rules out the unit root case (in time series). Consider the unit root process that $y_t = y_{t-1} + \epsilon_t$, $t = 2, \cdots, n$ and $y_1 = \epsilon_1$, i.e.,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{pmatrix}.$$

It implies that

$$y_t = \sum_{j=1}^{t} \epsilon_j.$$

That is, S_n^{-1} is a lower diagonal matrix with all nonzero elements being unity. The sum of its first column is n, and the sum of its last row is also n. For the unit root process, S_n^{-1} are neither bounded in row sums nor column sums.

3. Estimation Methods

Estimation methods, which have been considered in the existing literature, are mainly the ML (or QML) method, the 2SLS (or IV) method, and the generalized method of moments (GMM). The QML method has usually good finite sample properties relative to the other methods for the estimation of SAR models with the first order of spatial lag. However, the ML method is not computationally attractive for models with more than one single spatial lag. For the higher spatial lags model, the IV and GMM methods are feasible. With properly designed moment equations, the best GMM estimator exists and can be asymptotically efficient as the ML estimate under normal disturbances (Lee 2007).

The most popular and traditional estimation method is the ML under the assumption that $\mathcal{E}_n$ is $N(0, \sigma^2 I_n)$. Other estimation methods are the method of moments (MOM), GMM and the 2SLS. The latter is applicable only to the MRSAR model.

3.1) **ML** for the SAR process,

$$\ln L_n(\lambda, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}Y_n' S_n'(\lambda)S_n(\lambda)Y_n,$$

where $S_n(\lambda) = I_n - \lambda W_n$.

For the regression model with SAR disturbances, the log likelihood function is

$$\ln L_n(\lambda, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}(Y_n - X_n\beta)' S_n'(\lambda)S_n(\lambda)(Y_n - X_n\beta).$$

The log likelihood function for the MRSAR model is

$$\ln L_n(\lambda, \beta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}(Y_n S_n(\lambda) - X_n\beta)'(S_n(\lambda)Y_n - X_n\beta).$$

The likelihood function involves the computation of the determinant of $S_n(\lambda)$, which is a function of the unknown parameter λ, and may have a large dimension n.

A computationally tractable method is due to Ord 1975, where W_n is a row-normalized weights matrix with $W_n = D_n W_n^*$, where W_n^* is a symmetric matrix and $D_n = Diag\{\sum_{j=1}^n w_{n,ij}^*\}^{-1}$. Note that, in this case, W_n is not a symmetric matrix. However, the eigenvalues of W_n are still all real. This is so, because

$$\begin{aligned}
|W_n - \nu I_n| &= |D_n W_n^* - \nu I_n| = |D_n W_n^* D_n^{1/2} D_n^{-1/2} - \nu D_n^{1/2} D_n^{-1/2}| \\
&= |D_n^{\frac{1}{2}}| \cdot |D_n^{\frac{1}{2}} W_n^* D_n^{\frac{1}{2}} - \nu I_n| \cdot |D_n^{-\frac{1}{2}}| = |D_n^{1/2} W_n^* D_n^{1/2} - \nu I_n|,
\end{aligned}$$

the eigenvalues of W_n are the same as those of $D_n^{1/2} W_n^* D_n^{1/2}$, which is a symmetric matrix. As the eigenvalues of a symmetric matrix are real, the eigenvalues of W_n are real. Let μ_i's be the eigenvalues of W_n, which are the same as those of $D_n^{1/2} W_n^* D_n^{1/2}$. Let Γ be the orthogonal matrix such that $D_n^{1/2} W_n^* D_n^{1/2} = \Gamma \mathrm{Diag}\{\mu_i\}\Gamma'$. The above relations show also that

$$\begin{aligned}
|I_n - \lambda W_n| &= |I_n - \lambda D_n W_n^*| = |I_n - \lambda D_n^{1/2} W_n^* D_n^{1/2}| = |I_n - \lambda \Gamma \mathrm{Diag}\{\mu_i\}\Gamma'| \\
&= |I_n - \lambda \mathrm{Diag}\{\mu_i\}| = \prod_{i=1}^n (1 - \lambda\mu_i).
\end{aligned}$$

Thus, $|I_n - \lambda W_n|$ can be easily updated during iterations within a maximization subroutine, as μ_i's need be computed only once.

When W_n is a spare matrix, specific numerical methods for handling spare matrices are useful.

Another computational tractable method is the characteristic polynomial. The determinant $|W_n - \mu I_n|$ is a polynomial in μ and is called the *characteristic polynomial* of W_n. (note: the zeros of the characteristic polynomial are the eigenvalues of W_n). Thus,

$$|I_n - \lambda W_n| = a_n\lambda^n + \cdots + a_1\lambda + a_0,$$

the constants a's depend only on W_n. So the a's can be computed once during the maximization algorithm.

3.2) **2SLS Estimation** for the MRSAR model

For the MRSAR $Y_n = \lambda W_n Y_n + X_n\beta + \mathcal{E}_n$, the spatial lag $W_n Y_n$ can be correlated with the disturbance vector $\mathcal{E}_n$. So, in general, OLS may not be a consistent estimator. However, there is a class of spatial W_n (with large group interaction) that the OLS estimator can be consistent (Lee 2002).

To avoid the bias due to the correlation of $W_n Y_n$ with $\mathcal{E}_n$, Kelejian and Prucha (1998) has suggested the use of instrumental variables (IVs). Let Q_n be a matrix of instrumental variables. Denote $Z_n = (W_n Y_n, X_n)$ and $\theta = (\lambda, \beta')'$. The MRSAR equation can be rewritten as $Y_n = Z_n\theta + \mathcal{E}_n$. The 2SLS estimator of θ with Q_n is

$$\hat{\theta}_{2sl,n} = [Z_n' Q_n (Q_n' Q_n)^{-1} Q_n' Z_n]^{-1} Z_n' Q_n (Q_n' Q_n)^{-1} Q_n' Y_n.$$

It can be shown that the asymptotic distribution of $\hat{\theta}_{2sl,n}$ follows from

$$\sqrt{n}(\hat{\theta}_{2sl,n} - \theta_0) \xrightarrow{d} N(0, \sigma_0^2 \lim_{n\to\infty} \{\frac{1}{n}(G_n X_n \beta_0, X_n)' Q_n (Q_n' Q_n)^{-1} Q_n' (G_n X_n \beta_0, X_n)\}^{-1}),$$

where $G_n = W_n S_n^{-1}$, under the assumption that the limiting matrix $\frac{1}{n}(G_n X_n \beta_0, X_n)$ has the full column rank $(k+1)$ where k is the dimension of β or, equivalently, the number of columns of X_n. In practice, Kelejian and Prucha (1998) suggest the use of linearly independent variables in $(X_n, W_n X_n)$ for the construction of Q_n.

By the Schwartz inequality, the optimum IV matrix is $(G_n X_n \beta_0, X_n)$.

This 2SLS method can not be used for the estimation of the (pure) SAR process. The SAR process is a special case of the MRSAR model with $\beta_0 = 0$. However, when $\beta_0 = 0$, $G_n X_n \beta_0 = 0$

and, hence, $(G_nX_n\beta_0, X_n) = (0, X_n)$ would have rank k but not full rank $(k+1)$. Intuitively, when $\beta_0 = 0$, X_n does not appear in the model and there is no other IV available.

3.3) **Method of Moments** (MOM) for SAR process:

Kelejian and Prucha (1999) has suggested a MOM estimation:

$$\min_{\theta} g_n'(\theta)g_n(\theta).$$

The moment equations are based on three moments:

$$E(\mathcal{E}_n'\mathcal{E}_n) = n\sigma^2, \quad E(\mathcal{E}_n'W_n'W_n\mathcal{E}_n) = \sigma^2 tr(W_n'W_n), \quad \text{and} \quad E(\mathcal{E}_n'W_n\mathcal{E}_n) = 0.$$

These correspond to

$$g_n(\theta) = (Y_n'S_n'(\lambda)S_n(\lambda)Y_n - n\sigma^2,\ Y_n'S_n'(\lambda)W_n'W_nS_n(\lambda)Y_n - \sigma^2 tr(W_n'W_n),\ Y_n'S_n'(\lambda)W_nS_n(\lambda)Y_n)'$$

for the SAR process. For the regression model with SAR disturbances, Y_n shall be replaced by least squares residuals.

Remark: In Kelejian and Prucha (1999), they have shown consistency of the MOM estimator only but not asymptotic normality of the estimator.

3.4) **Moments for GMM estimation**

For the MRSAR model, in addition to the moments based on the IV matrix Q_n, other moment equations can also be constructed for estimation.

Let Q_n be an $n\times k_x$ IV matrix constructed as functions of X_n and W_n. Let $\epsilon_n(\theta) = S_n(\lambda)Y_n - X_n\beta$ for any possible value θ. The k_x moment functions correspond to the orthogonality conditions are $Q_n'\epsilon_n(\theta)$.

Now consider a finite number, say m, of $n \times n$ constant matrices $P_{1n}, \cdots, P_{mn}$ of which each has a zero diagonal. The moment functions $(P_{jn}\epsilon_n(\theta))'\epsilon_n(\theta)$ can be used in addition to $Q_n'\epsilon_n(\theta)$. These moment functions form a vector

$$g_n(\theta) = (P_{1n}\epsilon_n(\theta), \cdots, P_{mn}\epsilon_n(\theta), Q_n)'\epsilon_n(\theta) = \begin{pmatrix} \epsilon_n'(\theta)P_{1n}\epsilon_n(\theta) \\ \vdots \\ \epsilon_n'(\theta)P_{mn}\epsilon_n(\theta) \\ Q_n'\epsilon_n(\theta) \end{pmatrix}$$

in the estimation in the GMM framework.

Proposition. For any constant $n \times n$ matrix P_n with $tr(P_n) = 0$, $P_n \epsilon_n$ is uncorrelated with ϵ_n, i.e., $E((P_n \epsilon_n)' \epsilon_n) = 0$.

Proof: $E((P_n\epsilon_n)'\epsilon_n) = E(\epsilon_n' P_n' \epsilon_n) = E(\epsilon_n' P_n \epsilon_n) = \sigma_0^2 tr(P_n) = 0$.

This shows, in particular, that $E(g_n(\theta_0)) = 0$. Thus, $g_n(\theta)$ are valid moment equations for estimation.

INTUITION:

As $W_n Y_n = G_n X_n \beta_0 + G_n \epsilon_n$, where $G_n = W_n S_n^{-1}$ and $S_n = S_n(\lambda_0)$, and $G_n \epsilon_n$ is correlated with the disturbance ϵ_n in the model,

$$Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n,$$

any $P_{jn}\epsilon_n$, which is uncorrelated with ϵ_n, can be used as IV for $W_n Y_n$ as long as $P_{jn}\epsilon_n$ and $G_n \epsilon_n$ are correlated.

4. Basic Asymptotic Foundation

Some laws of large numbers and central limit theorem

Lemma. *Suppose that $\{A_n\}$ are uniformly bounded in both row and column sums. Then, $E(\epsilon_n' A_n \epsilon_n) = O(n)$ and $\text{var}(\epsilon_n' A_n \epsilon_n) = O(n)$. Furthermore, $\epsilon_n' A_n \epsilon_n = O_P(n)$ and, $\frac{1}{n}\epsilon_n' A_n \epsilon_n - \frac{1}{n}E(\epsilon_n' A_n \epsilon_n) = o_P(1)$.*

Lemma. *Suppose that A_n is a $n \times n$ matrix with its column sums being uniformly bounded, elements of the $n \times k$ matrix C_n are uniformly bounded, and $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. with zero mean, finite variance σ^2 and $E|\epsilon|^3 < \infty$. Then, $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n = O_P(1)$ and $\frac{1}{n} C_n' A_n \epsilon_n = o_p(1)$. Furthermore, if the limit of $\frac{1}{n} C_n' A_n A_n' C_n$ exists and is positive definite, then*

$$\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n \xrightarrow{D} N(0, \sigma^2 \lim_{n\to\infty} \frac{1}{n} C_n' A_n A_n' C_n).$$

This lemma can be shown with the Liapounov central limit theorem.

Theorem: (Liapounov's theorem) Let $\{X_n\}$ be a sequence of independent random variables. Let $E(X_n) = \mu_n$, $E(X_n - \mu_n)^2 = \sigma_n^2 \neq 0$. Denote $C_n = (\sum_{i=1}^n \sigma_i^2)^{1/2}$. If $\frac{1}{C_n^{2+\delta}} \sum_{k=1}^n E|X_k - \mu_k|^{2+\delta} \to 0$ for a positive $\delta > 0$, then $\frac{\sum_{i=1}^n (X_i - \mu_i)}{C_n} \xrightarrow{d} N(0,1)$.

Lemma. *Suppose that A_n is a constant $n \times n$ matrix uniformly bounded in both row and column sums. Let c_n be a column vector of constants. If $\frac{1}{n}c_n'c_n = o(1)$, then $\frac{1}{\sqrt{n}}c_n'A_n\epsilon_n = o_P(1)$. On the other hand, if $\frac{1}{n}c_n'c_n = O(1)$, then $\frac{1}{\sqrt{n}}c_n'A_n\epsilon_n = O_P(1)$.*

Proof: The results follow from Chebyshev's inequality by investigating $var(\sqrt{\frac{1}{n}}c_n'A_n\epsilon_n) = \sigma_0^2\frac{1}{n}c_n'A_nA_n'c_n$. Let Λ_n be the diagonal matrix of eigenvalues of A_nA_n' and Γ_n be the orthonormal matrix of eigenvectors. As eigenvalues in absolute values are bounded by any norm of the matrix, eigenvalues in Λ_n in absolute value are uniformly bounded because $\parallel A_n \parallel_\infty$ (and $\parallel A_n \parallel_1$) are uniformly bounded. Hence, $\frac{1}{n}c_n'A_nA_n'c_n \leq \frac{1}{n}c_n'\Gamma_n\Gamma_n'c_n \cdot |\lambda_{n,max}| = \frac{1}{n}c_n'c_n|\lambda_{n,max}| = o(1)$, where $\lambda_{n,max}$ is the eigenvalue of A_nA_n' with the largest absolute value.

When $\frac{1}{n}c_n'c_n = O(1)$, $\frac{1}{n}c_n'A_nA_n'c_n \leq \frac{1}{n}c_n'c_n|\lambda_{n,max}| = O(1)$. In this case, $\mathrm{var}(\sqrt{\frac{1}{n}}c_n'A_n\epsilon_n) = \sigma^2\frac{1}{n}c_n'A_nA_n'c_n = O(1)$. Therefore, $\sqrt{\frac{1}{n}}c_n'A_n\epsilon_n = O_P(1)$. Q.E.D.

A Martingale Central Limit Theorem:

Definition: Let $\{z_t\}$ be a sequence of random scalars, and let $\{\mathcal{J}_t\}$ be a sequence of σ-fields such that $\mathcal{J}_{t-1} \subseteq \mathcal{J}_t$ for all t. If z_t is measurable with respect to $\mathcal{J}_t$, then $\{z_t, \mathcal{J}_t\}$ is called an *adapted stochastic sequence.*

The σ-field $\mathcal{J}_t$ can be thought of as the σ-algebra generated by the entire current and past history of z_t and some other random variables.

Definition: Let $\{y_t, \mathcal{J}_t\}$ be an adapted stochastic sequence. Then $\{y_t, \mathcal{J}_t\}$ is a *martingale difference sequence* if and only if $E(y_t|\mathcal{J}_{t-1}) = 0$, for all $t \geq 2$.

Theorem (A Martingale Central Limit Theorem):

Let $\{\xi_{Tt}\}$ be a martingale-difference array. If as $T \to \infty$,

(i) $\sum_{t=1}^T v_{Tt} \xrightarrow{p} \sigma^2$, where $\sigma^2 > 0$ and $v_{Tt} = E(\xi_{Tt}^2|\mathcal{J}_{T,t-1})$ denotes the conditional variances,

(ii) for any $\epsilon > 0$, $\sum_{t=1}^T E(\xi_{Tt}^2 I(|\xi_{Tt}| > \epsilon)\big|\mathcal{J}_{T,t-1})$ converges in probability to zero (a Lindeberg condition),

then $\xi_{T1} + \cdots + \xi_{TT} \xrightarrow{D} N(0, \sigma^2)$.

A CLT for quadratic-linear function of ϵ_n:

Suppose that $\{A_n\}$ is a sequence of symmetric $n \times n$ matrices with row and column sums uniformly bounded and b_n is a n-dimensional column vector satisfying $\sup_n \frac{1}{n}\sum_{i=1}^n |b_{ni}|^{2+\eta} < \infty$

for some $\eta > 0$. The $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. random variables with zero mean and finite variance σ^2, and its moment $E(|\epsilon|^{4+\delta})$ for some $\delta > 0$ exists. Let $\sigma^2_{Q_n}$ be the variance of Q_n where $Q_n = \epsilon'_n A_n \epsilon_n + b'_n \epsilon_n - \sigma^2 tr(A_n)$. Assume that the variance $\sigma^2_{Q_n}$ is bounded away from zero at the rate n. Then, $\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0,1)$.

This theorem has been proved by Kelejian and Prucha (2001) with a martingale central limit theorem.

5. The Cliff and Ord (1973) Test for Spatial Correlation

Consider the model

$$Y = X\beta + U, \quad U = \rho W U + \epsilon,$$

where Y is an n-dimensional column vector of the dependent variable, X is an $n \times k$ matrix of regressors, U is an $n \times 1$ vector of disturbances, W is a $n \times n$ spatial weights matrix with a zero diagonal, and ϵ is $N(0, \sigma^2 I_n)$. One may want to test $H_0 : \rho = 0$, i.e., the test of spatial correlation in U.

The log likelihood of this model is

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln|I - \rho W| - \frac{1}{2\sigma^2}(y - X\beta)'(I - \rho W)'(I - \rho W)(y - X\beta).$$

The first-order conditions are

$$\frac{\partial \ln L}{\partial \beta'} = \frac{1}{\sigma^2} X'(I - \rho W)'(I - \rho W)(y - X\beta),$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)'(I - \rho W)'(I - \rho W)(y - X\beta),$$

$$\frac{\partial \ln L}{\partial \rho} = -tr(W_n(I - \rho W)^{-1}) + \frac{1}{2\sigma^2}(y - X\beta)'[(I - \rho W)'W + W'(I - \rho W)](y - X\beta),$$

where $\frac{\partial \ln|I - \rho W|}{\partial \rho} = -tr(W(I - \rho W)^{-1})$ by using the matrix differentiation formulae $\frac{d}{d\xi}\ln|A| = tr(A^{-1}\frac{dA}{d\xi})$ (see, e.g., Amemiya (1985) Advanced Econometrics, p461).

The second order derivatives with respect to ρ have

$$\frac{\partial^2 \ln L}{\partial \beta' \partial \rho} = -\frac{1}{\sigma^2} X'[W'(I - \rho W) + (I - \rho W)'W](y - X\beta),$$

$$\frac{\partial^2 \ln L}{\partial \sigma^2 \partial \rho} = -\frac{1}{2\sigma^4}(y - X\beta)'[W'(I - \rho W) + (I - \rho W)'W](y - X\beta),$$

and

$$\frac{\partial^2 \ln L}{\partial \rho^2} = -tr([W(I-\rho W)^{-1}]^2) - \frac{1}{\sigma^2}(y - X\beta)'W'W(y - X\beta),$$

where $\frac{\partial (I-\rho W)^{-1}}{\partial \rho} = (I-\rho W)^{-1}W(I-\rho W)^{-1}$, by using the matrix differentiation formulae $\frac{d}{d\xi}A^{-1} = -A^{-1}\frac{dA}{d\xi}A^{-1}$. We see that $E_{H_0}(\frac{\partial^2 \ln L}{\partial \beta \partial \rho}) = 0$, $E_{H_0}(\frac{\partial^2 \ln L}{\partial \sigma \partial \rho}) = 0$ because $E(\epsilon' W \epsilon) = \sigma^2 tr(W) = 0$ as W has a zero diagonal. Furthermore,

$$E_{H_0}(\frac{\partial^2 \ln L}{\partial \rho^2}) = -tr(W^2 + W'W).$$

Because of the block diagonal structure of the information matrix $E_{H_0}(\frac{\partial^2 \ln L}{\partial \theta \partial \theta'})$, an efficient score test statistic for testing H_0 can have a simple expression. The efficient score test is also known as a Lagrange multiplier (LM) test.

The efficient score test (Rao) statistic in its general form can be

$$\frac{\partial \ln L(\bar{\theta})}{\partial \theta'}[-E(\frac{\partial^2 \ln L(\bar{\theta})}{\partial \theta \partial \theta'})]^{-1}\frac{\partial \ln L(\bar{\theta})}{\partial \theta}$$

where $\bar{\theta}$ is the constrained MLE of θ under H_0.

Let $\hat{\beta}$ and $\hat{\sigma}^2$ be the MLE under $H_0 : \rho = 0$, i.e., $\hat{\beta}$ is the usual OLS estimate and $\hat{\sigma}^2 = \frac{1}{n}e'e$, where $e = y - X\hat{\beta}$ the least squares residual, is the average sum of squared residuals. Then $\bar{\theta} = (\hat{\beta}', \hat{\sigma}^2, 0)'$ and

$$\frac{\partial \ln L(\bar{\theta})}{\partial \rho} = -tr(W) + \frac{1}{2\hat{\sigma}^2}(y - X\hat{\beta})'(W + W')(y - X\hat{\beta}) = \frac{1}{2\hat{\sigma}^2}(y - X\hat{\beta})'(W + W')(y - X\hat{\beta}),$$

because $tr(W) = 0$. Hence, the LM test statistic is

$$ne'(W + W')e/(2e'e)/(tr(W^2 + W'W))^{\frac{1}{2}},$$

which shall be asymptotically normal $N(0, 1)$.

The test statistic given by Moran (1950) and Cliff and Ord (1973) is

$$I = e'We/e'e,$$

which is the the LM statistic with an unscaled denominator. The LM interpretation of the Cliff and Ord (1973) test of spatial correlation is due to Burridge (1980), *J.R.Statist.Soc.*B, vol 42, pp.107-108.

Burridge (1980) has also pointed out the LM test for a moving average alternative, i.e., $U = \epsilon + \rho W \epsilon$, is exactly the same as that given above.

Note that for the regression model with SAR disturbances, $y = X\beta + u$ with $u = \rho W u + \epsilon$, $\frac{1}{\sqrt{n}} u' W u$ and $\frac{1}{\sqrt{n}} \hat{u}' W \hat{u}$, where $\hat{u}$ is the least squares residual vector, can be asymptotically equivalent(under both H_0 and H_1, i.e., $\frac{1}{\sqrt{n}} \hat{u}' W \hat{u} = \frac{1}{\sqrt{n}} u' W u + o_p(1)$. This can be shown as follows. As $\hat{u} = [I - X(X'X)^{-1}X']u$,

$$\begin{aligned}
\hat{u}' W \hat{u} &= u'[I - X(X'X)^{-1}X']W[I - X(X'X)^{-1}X']u \\
&= u'Wu - u'(W + W')X(X'X)^{-1}X'u + u'X(X'X)^{-1}X'WX(X'X)^{-1}X'u.
\end{aligned}$$

Because $u = (I - \rho W)^{-1}\epsilon$, under the assumptions that elements of X are uniformly bounded and $(I - \rho W)^{-1}$ are uniformly bounded in both row and column sums, $\frac{1}{\sqrt{n}} X' u = \frac{1}{\sqrt{n}} X'(I - \rho W)^{-1}\epsilon = O_p(1)$ and $\frac{1}{\sqrt{n}} X' W u = \frac{1}{\sqrt{n}} X' W (I - \rho W)^{-1}\epsilon = O_p(1)$ by the CLT for the linear form. These imply, in particular, $\frac{1}{n} X' u = o_p(1)$ and $\frac{1}{n} X' W u = o_p(1)$. Hence,

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \hat{u}' W \hat{u} \\
&= \frac{u'Wu}{\sqrt{n}} - \frac{u'X}{n}\left(\frac{X'X}{n}\right)^{-1} \frac{1}{\sqrt{n}} X'(W + W')u + \frac{u'X}{n}\left(\frac{X'X}{n}\right)^{-1} \frac{X'WX}{n}\left(\frac{X'X}{n}\right)^{-1} \frac{X'u}{\sqrt{n}} \\
&= \frac{1}{\sqrt{n}} u'Wu + o_P(1).
\end{aligned}$$

Consequently, the Cliff-Ord test using the least squares residual is asymptotically $\chi^2(1)$ under H_0.

6. More on GMM and 2SLS Estimation of MRSAR Models

6.1) IVs and Quadratic Moments

The MRSAR model is

$$Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n,$$

where X_n is a $n \times k$ dimensional matrix of nonstochastic exogenous variables, W_n is a spatial weights matrix of known constants with a zero diagonal.

We assume that

Assumption 1. *The ϵ_{ni}'s are i.i.d. with zero mean, variance σ^2 and that a moment of order higher than the fourth exists.*

Assumption 2. *The elements of X_n are uniformly bounded constants, X_n has the full rank k, and $\lim_{n \to \infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular.*

Assumption 3. *The spatial weights matrices $\{W_n\}$ and $\{(I_n - \lambda W_n)^{-1}\}$ at $\lambda = \lambda_0$ are uniformly bounded in both row and column sums in absolute value.*

Let Q_n be an $n \times k_x$ matrix of IVs constructed as functions of X_n and W_n in a 2SLS approach. Denote $\epsilon_n(\theta) = (I_n - \lambda W_n)Y_n - X_n\beta$ for any possible value θ. Let $\mathcal{P}_{1n} = \{P : P \text{ is } n \times n \text{ matrix }, tr(P) = 0\}$ be the class of constant $n \times n$ matrices which have a zero trace. A subclass $\mathcal{P}_{2n}$ of $\mathcal{P}_{1n}$ is $\mathcal{P}_{2n} = \{P : P \in \mathcal{P}_{1n}, diag(P) = 0\}$.

Assumption 4. *The matrices P_{jn}s from $\mathcal{P}_{1n}$ are uniformly bounded in both row and column sums in absolute value, and elements of Q_n are uniformly bounded.*

With the selected matrices P_{jn}s and IV matrices Q_n, the set of moment functions forms a vector

$$g_n(\theta) = (P_{1n}\epsilon_n(\theta), \cdots, P_{mn}\epsilon_n(\theta), Q_n)'\epsilon_n(\theta) = (\epsilon_n'(\theta)P_{1n}\epsilon_n(\theta), \cdots, \epsilon_n'(\theta)P_{mn}\epsilon_n(\theta), \epsilon_n'(\theta)Q_n)'$$

for the GMM estimation.

At θ_0, $g_n(\theta_0) = (\epsilon_n'P_{1n}\epsilon_n, \cdots, \epsilon_n'P_{mn}\epsilon_n, \epsilon_n'Q_n)'$. It follows that $E(g_n(\theta_0)) = 0$ because $E(Q_n'\epsilon_n) = Q_n'E(\epsilon_n) = 0$ and $E(\epsilon_n'P_{jn}\epsilon_n) = \sigma_0^2 tr(P_{jn}) = 0$ for $j = 1, \cdots, m$. The intuition is as follows:

(1) The IV variables in Q_n, can be used as IV variables for W_nY_n and X_n.

(2) The $P_{jn}\epsilon_n$ is uncorrelated with ϵ_n. As $W_nY_n = G_n(X_n\beta_0 + \epsilon_n)$, $P_{jn}\epsilon_n$ can be used as an IV for W_nY_n as long as $P_{jn}\epsilon_n$ and $G_n\epsilon_n$ are correlated, where $G_n = W_nS_n^{-1}$.

6.2) Identification

In the GMM framework, the identification condition requires the unique solution of the limiting equations, $\lim_{n\to\infty} \frac{1}{n}E(g_n(\theta)) = 0$ at θ_0 (Hansen 1982). The moment equations corresponding to Q_n are

$$\lim_{n\to\infty} \frac{1}{n}Q_n'd_n(\theta) = \lim_{n\to\infty} \frac{1}{n}(Q_n'X_n, Q_n'G_nX_n\beta_0)((\beta_0 - \beta)', \lambda_0 - \lambda)' = 0.$$

They will have a unique solution at θ_0 if $(Q_n'X_n, Q_n'G_nX_n\beta_0)$ has a full column rank, i.e., rank $(k+1)$, for large enough n.

This sufficient rank condition implies the necessary rank condition that $(X_n, G_nX_n\beta_0)$ has a full column rank $(k+1)$ and that Q_n has a rank at least $(k+1)$, for large enough n. The sufficient rank condition requires Q_n to be correlated with W_nY_n in the limit as n goes to infinity. This is so because $E(Q_n' W_n Y_n) = Q_n' G_n X_n \beta_0$. Under the sufficient rank condition, θ_0 can thus be identified via $\lim_{n\to\infty} \frac{1}{n} Q_n' d_n(\theta) = 0$.

The necessary full rank condition of $(X_n, G_n X_n \beta_0)$ for large n is possible only if the set consisting of $G_n X_n \beta_0$ and X_n is not asymptotically linearly dependent. This rank condition would not hold, in particular, if β_0 were zero. There are other cases when this dependence can occur (see, e.g., Kelejian and Prucha 1998).

As X_n has rank k but $(X_n, G_n X_n \beta_0)$ does not have a full rank $(k+1)$, the corresponding moment equations $Q_n' d_n(\theta) = 0$ will have many solutions but the solutions are all described by the relation that $\beta = \beta_0 + (\lambda_0 - \lambda)c$ as long as $Q_n' X_n$ has a full rank k. Under this scenario, β_0 can be identified only if λ_0 is identifiable. The identification of λ_0 will rely on the remaining quadratic moment equations. In this case,

$$Y_n = \lambda_0 (G_n X_n \beta_0) + X_n \beta_0 + S_n^{-1} \epsilon_n = X_n(\beta_0 + \lambda_0 c) + u_n,$$

where $u_n = S_n^{-1}\epsilon_n$. The relationship $u_n = \lambda_0 W_n u_n + \epsilon_n$ is a SAR process. The identification of λ_0 thus comes from the SAR process u_n.

The following assumption summarizes some sufficient conditions for the identification of θ_0 from the moment equations $\lim_{n\to\infty} \frac{1}{n} E(g_n(\theta)) = 0$.

Assumption 5. *Either (i)* $\lim_{n\to\infty} \frac{1}{n} Q_n'(G_n X_n \beta_0, X_n)$ *has the full rank* $(k+1)$*, or (ii)* $\lim_{n\to\infty} \frac{1}{n} Q_n' X_n$ *has the full rank* k*,* $\lim_{n\to\infty} \frac{1}{n} tr(P_{jn}^s G_n) \neq 0$ *for some* j*, and* $\lim_{n\to\infty} \frac{1}{n}[tr(P_{1n}^s G_n), \cdots, tr(P_{mn}^s G_n)]'$ *is linearly independent of*

$$\lim_{n\to\infty} \frac{1}{n}[tr(G_n' P_{1n} G_n), \cdots, tr(G_n' P_{mn} G_n)]'.$$

5.3) Optimum GMM

The variance matrix of these moments functions involves variances and covariances of linear and quadratic forms of ϵ_n. For any square $n \times n$ matrix A, let $vec_D(A) = (a_{11}, \cdots, a_{nn})'$ denote the column vector formed with the diagonal elements of A. Then,

$$E(Q_n' \epsilon_n \cdot \epsilon_n' P_n \epsilon_n) = Q_n' \sum_{i=1}^{n} \sum_{j=1}^{n} p_{n,ij} E(\epsilon_n \epsilon_{ni} \epsilon_{nj}) = Q_n' vec_D(P_n) \mu_3$$

and

$$E(\epsilon_n' P_{jn}\epsilon_n \cdot \epsilon_n' P_{ln}\epsilon_n) = (\mu_4 - 3\sigma_0^4) vec_D'(P_{jn}) vec_D(P_{ln}) + \sigma_0^4 tr(P_{jn}P_{ln}^s).$$

It follows that $var(g_n(\theta_0)) = \Omega_n$ where

$$\Omega_n = \begin{pmatrix} (\mu_4 - 3\sigma_0^4)\omega_{nm}'\omega_{nm} & \mu_3\omega_{nm}' Q_n \\ \mu_3 Q_n'\omega_{nm} & 0 \end{pmatrix} + V_n,$$

with $\omega_{nm} = [vec_D(P_{1n}), \cdots, vec_D(P_{mn})]$ and

$$V_n = \sigma_0^4 \begin{pmatrix} tr(P_{1n}P_{1n}^s) & \cdots & tr(P_{1n}P_{mn}^s) & 0 \\ \vdots & & \vdots & \vdots \\ tr(P_{mn}P_{1n}^s) & \cdots & tr(P_{mn}P_{mn}^s) & 0 \\ 0 & \cdots & 0 & \frac{1}{\sigma_0^2}Q_n'Q_n \end{pmatrix} = \sigma_0^4 \begin{pmatrix} \Delta_{mn} & 0 \\ 0 & \frac{1}{\sigma_0^2}Q_n'Q_n \end{pmatrix},$$

where $\Delta_{mn} = [vec(P_{1n}'), \cdots, vec(P_{mn}')]'[vec(P_{1n}^s), \cdots, vec(P_{mn}^s)]$, by $tr(AB) = vec(A')'vec(B)$ for any conformable matrices A and B.

When ϵ_n is normally distributed, Ω_n is simplified to V_n because $\mu_3 = 0$ and $\mu_4 = 3\sigma_0^4$. If P_{jn}'s are from $\mathcal{P}_{2n}$, $\Omega_n = V_n$ also because $\omega_{nm} = 0$. In general, Ω_n is nonsingular if and only if both $(vec(P_{1n}), \cdots, vec(P_{mn}))$ and Q_n have full column ranks. This is so, because Ω_n would be singular if and only if the moments in $g_n(\theta_0)$ are functionally dependent, equivalently, if and only if $\sum_{j=1}^m a_j P_{jn} = 0$ and $Q_n b = 0$ for some constant vector $(a_1, \cdots, a_m, b') \neq 0$.

Let a_n converge to a constant full rank matrix a_0.

Proposition 1. *Under Assumptions 1-5, suppose that P_{jn} for $j = 1, \cdots, m$, are from $\mathcal{P}_{1n}$ and Q_n is a $n \times k_x$ IV matrix so that $a_0 \lim_{n\to\infty} \frac{1}{n}E(g_n(\theta)) = 0$ has a unique root at θ_0 in Θ. Then, the GMME $\hat{\theta}_n$ derived from $\min_{\theta\in\Theta} g_n'(\theta)a_n'a_n g_n(\theta)$ is a consistent estimator of θ_0, and $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \Sigma)$, where*

$$\Sigma = \lim_{n\to\infty} \left[(\frac{1}{n}D_n')a_n'a_n(\frac{1}{n}D_n)\right]^{-1} (\frac{1}{n}D_n')a_n'a_n(\frac{1}{n}\Omega_n)a_n'a_n(\frac{1}{n}D_n) \left[(\frac{1}{n}D_n')a_n'a_n(\frac{1}{n}D_n)\right]^{-1},$$

and

$$D_n = \begin{pmatrix} \sigma_0^2 tr(P_{1n}^s G_n) & \cdots & \sigma_0^2 tr(P_{mn}^s G_n) & (G_n X_n \beta_0)' Q_n \\ 0 & \cdots & 0 & X_n' Q_n \end{pmatrix}',$$

under the assumption that $\lim_{n\to\infty} \frac{1}{n} a_n D_n$ exists and has the full rank $(k+1)$.

Proposition 2. *Under Assumptions 1-6, suppose that $(\frac{\hat{\Omega}_n}{n})^{-1} - (\frac{\Omega_n}{n})^{-1} = o_P(1)$, then the feasible OGMME $\hat{\theta}_{o,n}$ derived from $\min_{\theta\in\Theta} g_n'(\theta)\hat{\Omega}_n^{-1} g_n(\theta)$ based on $g_n(\theta)$ with P_{jn}'s from $\mathcal{P}_{1n}$ has* *the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_{o,n} - \theta_0) \xrightarrow{D} N(0, (\lim_{n\to\infty} \frac{1}{n} D_n' \Omega_n^{-1} D_n)^{-1}).$$

Furthermore,

$$g_n'(\hat{\theta}_{o,n})\hat{\Omega}_n^{-1} g_n(\hat{\theta}_{o,n}) \xrightarrow{D} \chi^2((m + k_x) - (k+1)).$$

5.4) Efficiency and the BGMME

The optimal GMME $\hat{\theta}_{o,n}$ can be compared with the 2SLSE. With Q_n as the IV matrix, the 2SLSE of θ_0 is

$$\hat{\theta}_{2sl,n} = [Z_n' Q_n (Q_n' Q_n)^{-1} Q_n' Z_n]^{-1} Z_n' Q_n (Q_n' Q_n)^{-1} Q_n' Y_n, \tag{4.1}$$

where $Z_n = (W_n Y_n, X_n)$. The asymptotic distribution of $\hat{\theta}_{2sl,n}$ is

$$\begin{aligned} &\sqrt{n}(\hat{\theta}_{2sl,n} - \theta_0) \xrightarrow{D} \\ &N(0, \sigma_0^2 \lim_{n\to\infty} \{\frac{1}{n}(G_n X_n \beta_0, X_n)' Q_n (Q_n' Q_n)^{-1} Q_n' (G_n X_n \beta_0, X_n)\}^{-1}), \end{aligned} \tag{4.2}$$

under the assumptions that $\lim_{n\to\infty} \frac{1}{n} Q_n' Q_n$ is nonsingular and $\lim_{n\to\infty} \frac{1}{n} Q_n (G_n X_n \beta_0, X_n)$ has the full column rank $(k+1)$ (Kelejian and Prucha 1998). Because the 2SLSE can be a special case of the GMM estimation, by Proposition 2, $\hat{\theta}_{o,n}$ shall be efficient relative to $\hat{\theta}_{2sl,n}$. The best Q_n is $Q_n^* = (G_n X_n \beta_0, X_n)$ by the Schwartz inequality.

The remaining issue is on the best selection of P_{jn}'s. When the disturbance ϵ_n is normally distributed or P_{jn}'s are from $\mathcal{P}_{2n}$, with Q_n^* for Q_n,

$$D_n' \Omega_n^{-1} D_n = \begin{pmatrix} C_{mn} \Delta_{mn}^{-1} C_{mn}' & 0 \\ 0 & 0 \end{pmatrix} + \frac{1}{\sigma_0^2} (G_n X_n \beta_0, X_n)' (G_n X_n \beta_0, X_n),$$

where $C_{mn} = [tr(P_{1n}^s G_n), \cdots, tr(P_{mn}^s G_n)]$. Note that, because $tr(P_{jn} P_{ln}^s) = \frac{1}{2} tr(P_{jn}^s P_{ln}^s)$, Δ_{mn} can be rewritten as

$$\Delta_{mn} = \frac{1}{2} \begin{pmatrix} tr(P_{1n}^s P_{1n}^s) & \cdots & tr(P_{1n}^s P_{mn}^s) \\ \vdots & & \vdots \\ tr(P_{mn}^s P_{1n}^s) & \cdots & tr(P_{mn}^s P_{mn}^s) \end{pmatrix} = \frac{1}{2} [vec(P_{1n}^s) \cdots vec(P_{mn}^s)]' [vec(P_{1n}^s) \cdots vec(P_{mn}^s)].$$

(i) When P_{jn}s are from $\mathcal{P}_{2n}$,

$$\begin{aligned} tr(P_{jn}^s G_n) &= \frac{1}{2} tr\Big(P_{jn}^s [G_n - Diag(G_n)]^s\Big) \\ &= \frac{1}{2} vec'\Big([G_n - Diag(G_n)]^s\Big) vec(P_{jn}^s) \end{aligned}$$

for $j = 1, \cdots, m$, in C_{mn}, where $Diag(A)$ denotes the diagonal matrix formed by the diagonal elements of a square matrix A. Therefore, the generalized Schwartz inequality implies that

$$\begin{aligned} C_{mn}\Delta_{mn}^{-1}C_{mn}' &\leq \frac{1}{2}vec'\Big([G_n - Diag(G_n)]^s\Big) \cdot vec\Big([G_n - Diag(G_n)]^s\Big) \\ &= tr\Big([G_n - Diag(G_n)]^s G_n\Big). \end{aligned}$$

Thus, in the subclass $\mathcal{P}_{2n}$, $[G_n - Diag(G_n)]$ and together with $[G_n X_n \beta_0, X_n]$ provide the set of best IV functions.

(ii) For the case where ϵ_n is $N(0, \sigma_0^2 I_n)$, because, for any $P_{jn} \in \mathcal{P}_{1n}$,

$$tr(P_{jn}^s G_n) = \frac{1}{2}vec'\Big([G_n - \frac{tr(G_n)}{n}I_n]^s\Big)vec(P_{jn}^s),$$

for $j = 1, \cdots, m$, the generalized Schwartz inequality implies that

$$C_{mn}\Delta_{mn}^{-1}C_{mn}' \leq tr\Big([G_n - \frac{tr(G_n)}{n}I_n]^s G_n\Big).$$

Hence, in the broader class $\mathcal{P}_{1n}$, $[G_n - \frac{tr(G_n)}{n}I_n]$ and $[G_n X_n \beta_0, X_n]$ provide the best set of IV functions.

Proposition 3. *Under Assumptions 1-3, suppose that $\hat{\lambda}_n$ is a $\sqrt{n}$-consistent estimate of λ_0, $\hat{\beta}_n$ is a consistent estimate of β_0, and $\hat{\sigma}_n^2$ is a consistent estimate of σ_0^2.*

Within the class of GMME's derived with $\mathcal{P}_{2n}$, the BGMME $\hat{\theta}_{2b,n}$ has the limiting distribution that $\sqrt{n}(\hat{\theta}_{2b,n} - \theta_0) \xrightarrow{D} N(0, \Sigma_{2b}^{-1})$ where

$$\Sigma_{2b} = \lim_{n\to\infty}\frac{1}{n}\begin{pmatrix} tr[(G_n - Diag(G_n))^s G_n] + \frac{1}{\sigma_0^2}(G_n X_n \beta_0)'(G_n X_n \beta_0) & \frac{1}{\sigma_0^2}(G_n X_n \beta_0)' X_n \\ \frac{1}{\sigma_0^2}X_n'(G_n X_n \beta_0) & \frac{1}{\sigma_0^2}X_n' X_n \end{pmatrix},$$

which is assumed to exist.

In the event that $\epsilon_n \sim N(0, \sigma_0^2 I_n)$, within the broader class of GMME's derived with $\mathcal{P}_{1n}$, the BGMME $\hat{\theta}_{1b,n}$ has the limiting distribution that $\sqrt{n}(\hat{\theta}_{1b,n} - \theta_0) \xrightarrow{D} N(0, \Sigma_{1b}^{-1})$ where

$$\Sigma_{1b} = \lim_{n\to\infty}\frac{1}{n}\begin{pmatrix} tr[(G_n - \frac{tr(G_n)}{n}I_n)^s G_n] + \frac{1}{\sigma_0^2}(G_n X_n \beta_0)'(G_n X_n \beta_0) & \frac{1}{\sigma_0^2}(G_n X_n \beta_0)' X_n \\ \frac{1}{\sigma_0^2}X_n'(G_n X_n \beta_0) & \frac{1}{\sigma_0^2}X_n' X_n \end{pmatrix},$$

which is assumed to exist.

5.5) Links between BGMME and MLE

When ϵ_n is $N(0, \sigma_0^2 I_n)$, the model can be estimated by the ML method. The log likelihood function of the MRSAR model via its reduced form equation is

$$\begin{aligned}\ln L_n = & -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|(I_n - \lambda W_n)| \\ & - \frac{1}{2\sigma^2}[Y_n - (I_n - \lambda W_n)^{-1}X_n\beta]'(I_n - \lambda W_n')(I_n - \lambda W_n)[Y_n - (I_n - \lambda W_n)^{-1}X_n\beta].\end{aligned} \tag{4.6}$$

The asymptotic variance of the MLE $(\hat{\theta}_{ml,n}, \hat{\sigma}^2_{ml,n})$ is

$$\begin{aligned}&\text{AsyVar}(\hat{\theta}_{ml,n}, \hat{\sigma}^2_{ml,n}) \\ &= \begin{pmatrix} tr(G_n^2) + tr(G_n'G_n) + \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'(G_nX_n\beta_0) & \frac{1}{\sigma_0^2}(X_n'G_nX_n\beta_0)' & \frac{tr(G_n)}{\sigma_0^2} \\ \frac{1}{\sigma_0^2}X_n'G_nX_n\beta_0 & \frac{1}{\sigma_0^2}X_n'X_n & 0 \\ \frac{tr(G_n)}{\sigma_0^2} & 0 & \frac{n}{2\sigma_0^4}\end{pmatrix}^{-1}\end{aligned}$$

(see, e.g., Anselin and Bera (1998), p.256). From the inverse of a partitioned matrix, the asymptotic variance of the MLE $\hat{\theta}_{ml,n}$ is

$$\begin{aligned}&AsyVar(\hat{\theta}_{ml,n}) \\ &= \begin{pmatrix} tr(G_n^2) + tr(G_n'G_n) + \frac{1}{\sigma_0^2}(G_nX_n\beta_0)'(G_nX_n\beta_0) - \frac{2}{n}tr^2(G_n) & \frac{1}{\sigma_0^2}(X_n'G_nX_n\beta_0)' \\ \frac{1}{\sigma_0^2}X_n'G_nX_n\beta_0 & \frac{1}{\sigma_0^2}X_n'X_n\end{pmatrix}^{-1}.\end{aligned}$$

As $tr(G_n^2) + tr(G_n'G_n) - \frac{2}{n}tr^2(G_n) = tr((G_n - \frac{tr(G_n)}{n}I_n)^s G_n)$, the GMME $\hat{\theta}_{1b,n}$ has the same limiting distribution as the MLE of θ_0 from Proposition 3.

There is an intuition on the best GMM approach compared with the maximum likelihood one. The derivatives of the log likelihood in (4.6) are

$$\frac{\partial \ln L_n}{\partial \beta} = \frac{1}{\sigma^2}X_n'\epsilon_n(\theta),$$

$$\frac{\partial \ln L_n}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\epsilon_n'(\theta)\epsilon_n(\theta),$$

and

$$\begin{aligned}\frac{\partial \ln L_n}{\partial \lambda} = & -tr(W_n(I_n - \lambda W_n)^{-1}) + \frac{1}{\sigma^2}[W_n(I_n - \lambda W_n)^{-1}X_n\beta]'\epsilon_n(\theta) \\ & + \frac{1}{\sigma^2}\epsilon_n'(\theta)[W_n(I_n - \lambda W_n)^{-1}]'\epsilon_n(\theta).\end{aligned}$$

The equation $\frac{\partial \ln L_n}{\partial \sigma^2} = 0$ implies that the MLE is $\hat{\sigma}_n^2(\theta) = \frac{1}{n}\epsilon_n'(\theta)\epsilon_n(\theta)$ for a given value θ. By substituting $\hat{\sigma}_n^2(\theta)$ into the remaining likelihood equations, the MLE $\hat{\theta}_{ml,n}$ will be characterized by the moment equations: $X_n'\epsilon_n(\theta) = 0$, and

$$[W_n(I_n - \lambda W_n)^{-1}X_n\beta]'\epsilon_n(\theta) + \epsilon_n'(\theta)\left[W_n(I_n - \lambda W_n)^{-1} - \frac{1}{n}tr(W_n(I_n - \lambda W_n)^{-1})\right]\epsilon_n(\theta) = 0.$$

The similarity of the best GMM moments and the above likelihood equations is revealing. The best GMM approach has the linear and quadratic moments of $\epsilon_n(\theta)$ in its formation and uses consistently estimated matrices in its linear and quadratic form.

5.6) More recent results

1) generalization to the estimation of higher order autoregressive models

2) robust GMM estimation in the presence of unknown heteroskedasticity

3) improve GMM estimators when disturbances are non-normal.

Lecture 2: Econometric Models with Social Interactions

We consider interactions of individuals in a group setting.

2.1) Linear in the Mean (Expectation) Model and the Reflection Problem (Manski)

Endogenous effects, wherein the propensity of an individual to behave in some way varies with the behaviour of the group;

Exogenous (contextual) effects, wherein the propensity of an individual to behave in some way varies with the exogenous characteristics of the group, and

Correlated effects, wherein individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environment.

The Manski's linear model specification:

$$y_{ri} = \lambda_0 E(y_{rj}) + x_{ri,1}\beta_{10} + E(x_{rj,2})\beta_{20} + \alpha_r + \epsilon_{ri}$$

where $E(y_{rj})$ with the coefficient λ_0 captures the endogenous interactions, $E(x_{rj,2})$ with β_{20} captures the exogenous interactions, and α_r represents the correlation effect. The $E(y_r)$ is assumed to solve the "social equilibrium" equation,

$$E(y_r) = \lambda_0 E(y_r) + E(x_{r,1})\beta_{10} + E(x_{r,2})\beta_{20} + \alpha_r.$$

Provided that $\lambda_0 \neq 1$, this equation has a unique solution, which is

$$E(y_r) = E(x_{r1})\frac{\beta_{10}}{1-\lambda_0} + E(x_{r,2})\frac{\beta_{20}}{1-\lambda_0} + \frac{\alpha_r}{1-\lambda_0}.$$

The case where $x_{r1} = x_{r2} (= x_r)$, or more general, $x_{r,1}$ is a subset of $x_{r,2}$, will have an underidentification problem.

When $x_{r,1} = x_{r,2} = x_r$, it follows $E(y_r) = E(x_r)\frac{\beta_{10}+\beta_{20}}{(1-\lambda_0)} + \frac{\alpha_r}{1-\lambda_0}$ and that the reduced form of this model is

$$y_{ri} = E(x_r)[\beta_{20} + \frac{\lambda_0}{1-\lambda_0}(\beta_{10}+\beta_{20})] + x_{ri}\beta_{10} + \frac{\alpha_r}{1-\lambda_0} + \epsilon_{ri}.$$

The β_{10} and the composite parameter vecotr $\beta_{20} + \frac{\lambda_0}{1-\lambda_0}(\beta_{10}+\beta_{20})$ can be identified, but λ_0 and β_{20} can not be separately identified. This is so, because $E(y_r)$ is linearly dependent on $E(x_r)$ and the constant intercept term α_r — the "reflection problem".

From the implied equation, β_{10} is apparently identifiable. For the identification of λ_0 and β_{20}, a sufficent identification condition is that some relevant variables are in $x_{ri,1}$, but they are not in $x_{ri,2}$ (under the random components specification for the overall disturbance). This follows, because, with the relevant variables in $x_{ri,1}$ but excluded in $x_{ri,2}$, it allows us to identify λ_0 from the composite parameter vector. In that case, the other coefficients λ and β_2 are also identifiable.

2.2) A SAR Model with Social Interactions

Instead of rational expectation formulation, an alternative model is to specify direct interactions of individuals in a group setting. This model will be useful in particular for cases where each group has a small or moderate number of members. Suppose there are m_r members in the group r, for each unit i in the group r,

$$y_{ri} = \lambda_0(\frac{1}{m_r - 1} \sum_{j=1, j\neq i}^{m_r} y_{rj}) + x_{ri,1}\beta_{10} + (\frac{1}{m_r - 1} \sum_{j=1, j\neq i}^{m_r} x_{rj,2})\beta_{20} + \alpha_r + \epsilon_{ri},$$

with $i = 1, \cdots, m_r$ and $r = 1, \cdots, R$, where y_{ri} is the ith individual in the rth group, $x_{ri,1}$ and $x_{ri,2}$ are, respectively, k_1 and k_2-dimensional row vectors of exogenous variables, and ϵ_{ri}'s are i.i.d. $(0, \sigma_0^2)$.

It is revealing to decompose this equation into two parts. Because of the group structure, the model is equivalent to the following two equations:

$$(1 - \lambda_0)\bar{y}_r = \bar{x}_{r1}\beta_{10} + \bar{x}_{r2}\beta_{20} + \alpha_r + \bar{\epsilon}_r, \quad r = 1, \cdots, R,$$

and

$$(1 + \frac{\lambda_0}{m_r - 1})(y_{ri} - \bar{y}_r) = (x_{ri,1} - \bar{x}_{r1})\beta_{10} - \frac{1}{m_r - 1}(x_{ri,2} - \bar{x}_{r2})\beta_{20} + (\epsilon_{ri} - \bar{\epsilon}_r),$$

for $i = 1, \cdots, m_r$; $r = 1, \cdots, R$, where $\bar{y}_r = \frac{1}{m_r}\sum_{i=1}^{m_r} y_{ri}$, $\bar{x}_{r1} = \frac{1}{m_r}\sum_{i=1}^{m_r} x_{ri,1}$ and $\bar{x}_{r2} = \frac{1}{m_r}\sum_{i=1}^{m_r} x_{ri,2}$ are means for the r-th group. The first equation may be called a 'between' equation and that in the second one is a 'within' equation as they have similarity with those of a panel data regression model (Hsiao 1986). The possible effects due to interactions are revealing in the reduced-form between and within equations:

$$\bar{y}_r = \bar{x}_{r1}\frac{\beta_{10}}{(1 - \lambda_0)} + \bar{x}_{r2}\frac{\beta_{20}}{(1 - \lambda_0)} + \frac{\alpha_r}{(1 - \lambda_0)} + \frac{\bar{\epsilon}_r}{(1 - \lambda_0)}, \quad r = 1, \cdots, R,$$

and

$$(y_{ri} - \bar{y}_r) = (x_{ri,1} - \bar{x}_{r1})\frac{(m_r - 1)\beta_{10}}{(m_r - 1 + \lambda_0)} - (x_{ri,2} - \bar{x}_{r2})\frac{\beta_{20}}{(m_r - 1 + \lambda_0)}$$
$$+ \frac{(m_r - 1)}{(m_r - 1 + \lambda_0)}(\epsilon_{ri} - \bar{\epsilon}_r),$$

with $i = 1, \cdots, m_r;\ r = 1, \cdots, R.$

Identification of the within equation

When all groups have the same number of members, i.e., m_r is a constant, say m, for all r, the effect λ can not be identifiable from the within equation. This is apparent as only the functions $\frac{(m-1)\beta_{10}}{(m-1+\lambda_0)}$, $\frac{\beta_{20}}{(m-1+\lambda_0)}$, and $\frac{(m-1)\sigma_0^2}{(m-1+\lambda_0)}$ may ever be identifiable from the within equation.

Estimation of the within equation

Conditional likelihood and CMLE:

$$\ln L_{w,n}(\theta) = c + \sum_{r=1}^{R}(m_r - 1)\ln(m_r - 1 + \lambda) - \frac{(n - R)}{2}\ln \sigma^2$$
$$- \frac{1}{2\sigma^2}\sum_{r=1}^{R}[\frac{1}{c_r(\lambda)}Y_r - Z_r\delta_m]' J_r [\frac{1}{c_r(\lambda)}Y_r - Z_r\delta_m],$$

where $c_r(\lambda) = (\frac{m_r-1}{m_r-1+\lambda})$, $z_{ri} = (x_{ri,1}, -\frac{m}{m_r-1}x_{ri,2})$, $m = \frac{1}{R}\sum_{r=1}^{R} m_r$; $\delta_m = (\beta_1, \frac{\beta_2}{m})$.

CMLE of β and σ^2:

$$\hat{\beta}_n(\lambda) = \begin{pmatrix} I_{k_1} & 0 \\ 0 & mI_{k_2} \end{pmatrix} \left(\sum_{r=1}^{R} Z_r' J_r Z_r\right)^{-1} \sum_{r=1}^{R} Z_r' J_r Y_r \frac{1}{c_r(\lambda)},$$

and

$$\hat{\sigma}_n^2(\lambda) = \frac{1}{n - R}\Big\{ \sum_{r=1}^{R} \frac{1}{c_r^2(\lambda)} Y_r' J_r Y_r$$
$$- \sum_{r=1}^{R} \frac{1}{c_r(\lambda)} Y_r' J_r Z_r \left(\sum_{r=1}^{R} Z_r' J_r Z_r\right)^{-1} \sum_{r=1}^{R} Z_r' J_r Y_r \frac{1}{c_r(\lambda)}\Big\}.$$

Concentrated log likelihood at λ:

$$\ln L_{c,n}(\lambda) = c_1 + \sum_{r=1}^{R}(m_r - 1)\ln(m_r - 1 + \lambda) - \frac{(n - R)}{2}\ln \hat{\sigma}_n^2(\lambda).$$

Instrumental Variables Estimation

$$y_{ri} - \bar{y}_r = -\lambda_0 \frac{(y_{ri} - \bar{y}_r)}{m_r - 1} + (x_{ri,1} - \bar{x}_{r,1})\beta_{10} - \frac{(x_{ri,2} - \bar{x}_{r,2})}{m_r - 1}\beta_{20} + (\epsilon_{ri} - \bar{\epsilon}_r).$$

The IV estimator of $\theta_0 = (\lambda_0, \beta_{10}', \beta_{20}')'$ is

$$\hat{\theta}_{n,IV} = \left[\sum_{r=1}^{R} (\frac{Q_r}{m_r - 1}, X_{r,1}, -\frac{X_{r,2}}{m_r - 1})' J_r (-\frac{Y_r}{m_r - 1}, X_{r,1}, -\frac{X_{r,2}}{m_r - 1}) \right]^{-1}$$
$$\cdot \sum_{r=1}^{R} (\frac{Q_r}{m_r - 1}, X_{r,1}, -\frac{X_{r,2}}{m_r - 1})' J_r Y_r.$$

2.3) Expectation Model (self-influence excluded)

The social interactions model under consideration is

$$y_{ri} = \lambda_0 (\frac{1}{m_r - 1} \sum_{j=1,\neq i}^{m_r} E(y_{rj} | \mathcal{J}_r)) + x_{ri,1}\beta_{10}$$
$$+ (\frac{1}{m_r - 1} \sum_{j=1,\neq i}^{n} x_{rj,2})\beta_{20} + \alpha_r + \epsilon_{ri},$$

where $\mathcal{J}_r$ is the information set of the rth group, which includes all observed x's and the group fixed effect α_r. This model assumes that each invidual knows all the exogenous characteristics of all other members in the group but does not know their actual outcomes. Because he does not know what the outcomes of others will be, he formulates his expectation about their possible (rational expectation) outcomes.

The group means of the expected outcomes are

$$\frac{1}{m_r} \sum_{i=1}^{m_r} E(y_{ri} | \mathcal{J}_r) = \frac{1}{1 - \lambda_0} (\frac{1}{m_r} \sum_{i=1}^{m_r} x_{ri,1}\beta_{10} + \frac{1}{m_r} \sum_{i=1}^{m_r} x_{ri,2}\beta_{20} + \alpha_r),$$

for $r = 1, \cdots, R$, which are linear functions of group means of exogenous variables. The model implies the between equation representation

$$\bar{y}_r = \frac{1}{(1 - \lambda_0)} (\bar{x}_{r,1}\beta_{10} + \bar{x}_{r,2}\beta_{20}) + v_r + \bar{\epsilon}_r,$$

and the within equation representation

$$y_{ri} - \bar{y}_r = \frac{m_r - 1}{m_r - 1 + \lambda_0} (x_{ri,1} - \bar{x}_{r,1})\beta_{10} - \frac{1}{m_r - 1 + \lambda_0} (x_{ri,2} - \bar{x}_{r,2})\beta_{20}$$
$$+ (\epsilon_{ri} - \bar{\epsilon}_r).$$

From these equations, we see no implication on the possible correlation of disturbances due to endogenous interactions, while the mean components are exactly the same as the simultaneous interaction model. For this model, there is an underidentification when $\beta_{20} + \lambda_0\beta_{10} = 0$ with $x_{ri,1} = x_{ri,2} (= x_{ri})$. These features imply $\frac{(m_r-1)\beta_{10}-\beta_{20}}{m_r-1+\lambda_0} = \beta_{10}$ and, hence,

$$y_{ri} - \bar{y}_r = (x_{ri} - \bar{x}_r)\beta_{10} + (\epsilon_{ri} - \bar{\epsilon}_r).$$

We see that, in this case, λ_0 and β_{20} do not appear in the equation and they can not be identified from the within equation. In the simultaneous equation interactions model, the λ_0 can be identified from the reduced form disturbances of the within equation, even that constraint on β_{10}, β_{20} and λ_0 occurs.

Lee (2007) suggests the estimation of the within equation by the conditional ML method. In addition, the IV method is also feasible, and optimal IV's are also available.

We note that if the model includes self-influence as in

$$y_{ri} = \lambda_0(\frac{1}{m_r}\sum_{j=1}^{m_r} E(y_{rj}|\mathcal{J}_r)) + x_{ri,1}\beta_{10} + (\frac{1}{m_r}\sum_{j=1}^{m_r} x_{rj,2})\beta_{20} + \alpha_r + \epsilon_{ri},$$

the social effects in the model would also not be identifiable when $x_{ri,1} = x_{ri,2}$. For this model, the group means of the expected outcomes are

$$\frac{1}{m_r}\sum_{i=1}^{m_r} E(y_{ri}|\mathcal{J}_r) = \frac{1}{1-\lambda_0}(\frac{1}{m_r}\sum_{i=1}^{m_r} x_{ri,1}\beta_{10} + \frac{1}{m_r}\sum_{i=1}^{m_r} x_{ri,2}\beta_{20} + \alpha_r),$$

which is a linear function of group means of exogenous variables.

2.4) A Network Model with Social Interactions

Network Model: rth group

$$Y_r = \lambda W_r Y_r + X_r\beta_1 + W_r X_r \beta_2 + l_{m_r}\alpha_r + u_r,$$

and

$$u_r = \rho M_r u_r + \epsilon_r,$$

where $\epsilon_r = (\epsilon_{1r}, \cdots, \epsilon_{m_r r})'$, ϵ_{ir} i.i.d. $(0, \sigma^2)$.

R # of groups

m_r # of members in r group

W_r, M_r exogenous network (social) matrices

$\lambda W_r Y_r$ endogenous effect

$W_r X_r \beta_2$ exogenous effect

α_r group (fixed) effect; $l_{m_r} = (1, \cdots, 1)'$

$\rho M_r u_r$ correlation effects

Lin (2005, 2006) — AddHealth Data;

The Add Health Survey:

Students in grades 7-12; 132 schools

Wave I in school survey — 90,118 students;

Friendship network — name up to 5 male and 5 female friends

Special features (in this paper):

Group structure (R can be large – incidental parameters);

W_r and M_r – row normalized, i.e., $W_r l_{m_r} = l_{m_r}$.

Bonacich measure of centrality of members in a group:

$$(I_{m_r} - \lambda_0 W_r)^{-1} W_r l_{m_r} = (1 - \lambda_0) l_{m_r},$$

i.e., all members in the same group has the same measure of centrality.

Related works:

Lin (2005, 2006): Eliminate α_r by $(I_n - W_n)$ and estimation by IV method.

Calvo-Armengol, Patacchini and Zenou (2006) — social network in education;

Bramoulle, Djebbari and Fortin (2006) — identification.

This study:

- allow 'correlated effects';
- more on identification issues;
- more efficient estimation methods;
- MC study on sensitivity of estimated effects by omission of a certain effects;
- empirical applications with AddHealth data.

3.5) Estimation

Elimination of group fixed effects:

1) Quasi-difference

$$R_r Y_r = \lambda_0 R_r W_r Y_r + R_r Z_r \beta_0 + (1-\rho_0) l_{m_r} \alpha_{r0} + \epsilon_r,$$

where $Z_r = (X_r, W_r X_r)$ and $R_r = I_{m_r} - \rho_0 M_r$.

2) Eliminate group fixed effect

- deviation from group mean
- and eliminate linear dependence on disturbances

$$R_r^* Y_r^* = \lambda_0 R_r^* W_r^* Y_r^* + R_r^* Z_r^* \beta_0 + \epsilon_r^*,$$

where

i) $J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l'_{m_r} = F_r F'_r$ – an eigenvalue-eigenvector decomposition; $[F_r, \frac{1}{\sqrt{m_r}} l_{m_r}]$ an orthonormal matrix,

ii) $R_r^* = F'_r R_r F_r$; $W_r^* = F'_r W_r F_r$

iii) $Y_r^* = F'_r Y_r$; $Z_r^* = F'_r Z_r$;

iv) using $W_r = W_r(F_r F'_r + \frac{1}{m_r} l_{m_r} l'_{m_r})$, $W_r l_{m_r} = l_{m_r}$, and $F'_r l_{m_r} = 0$, one has

$$F'_r W_r = F'_r W_r F_r F'_r.$$

Similarly for M_r;

v) ϵ_r^* has zero mean and the variance matrix $\sigma_0^2 I_{m_r-1}$.

—- spatial AR model.

Estimation: quasi-ML

$$\ln L_n(\theta) = -\frac{(n-R)}{2} \ln(2\pi\sigma^2) + \sum_{r=1}^{R} \ln \frac{|S_r(\lambda)|}{1-\lambda} + \sum_{r=1}^{R} \ln \frac{|R_r(\rho)|}{1-\rho}$$
$$- \frac{1}{2\sigma^2} \sum_{r=1}^{R} [R_r(\rho)(S_r(\lambda)Y_r - Z_r\beta)]' J_r [R_r(\rho)(S_r(\lambda)Y_r - Z_r\beta)].$$

Concentrated log likelihood function

$$
\begin{aligned}
\ln L_n(\lambda,\rho) = & -\frac{(n-R)}{2}(\ln(2\pi)+1) - \frac{(n-R)}{2}\ln\hat{\sigma}_n^2(\lambda,\rho) \\
& + \ln|S_n(\lambda)| + \ln|R_n(\rho)| - R\ln[(1-\lambda)(1-\rho)].
\end{aligned}
$$

Another (feasible) estimation method: 2SLS or G2SLS

apply IVs

$$
F_r' R_r Y_r = \lambda_0 F_r' R_r W_r Y_r + F_r' R_r Z_r \beta_0 + \epsilon_r^*. \tag{**}
$$

(stacked up all the groups together – all sample observations)

i) Estimate $F_r' Y_r = \lambda_0 F_r' W_r Y_r + F_r' Z_r \beta_0 + F_r' u_r$ by 2sls with some IV matrix Q_{1r} to get an initial (consistent) estimate of λ and β;

ii) use the estimated residuals $F_r' u_r (= u_r^*)$ to estimate ρ; via (Kelejian-Prucha 1999, MOM)

$$
(I_{m_r-1} - \rho_0 M_r^*) u_r^* = \epsilon_r^*;
$$

iii) estimate R_r;

iv) then feasible 2SLS for (**).

- The best IV matrix is $F_r' R_r [W_r (I_{m_r} - \lambda_0 W_r)^{-1} Z_r \beta_0, Z_r]$.
- Under normal disturbances, the (transformed) ML is asymptotically more efficient.