GMM Estimation of the Spatial Autoregressive Model in a System of Interrelated Networks

Yan Bao

May, 2010

1 Introduction

In this paper, I extend the generalized method of moments framework based on linear and quadratic moment conditions for the estimation of the first-order spatial autoregressive model introduced by Lee (2001; 2007) to estimate the SAR model in a system of interrelated networks.

The SAR model in a system of interrelated networks can be used to describe a market situation with several chain stores competing against one another. Each chain is represented by a vector. The strategy of a store in the chain does not only involve coordination with the other stores in the same chain, but also competition against opponent stores in the other chains. Thus, we need to set up a SAR equation for each chain. These equations need to be estimated simultaneously because each equation contains spatial lag terms for every chain in the market.

This setting complicated the estimation of the model in two ways.

First, all the spatial lag terms have different coefficients. This makes it hard to implement the method of maximum likelihood even when the disturbances are normally distributed because the Jacobian transformation becomes very complicated. Ord’s (1975) device cannot be extended into this model.

Second, the disturbances are assumed to be correlated across equations. Because of this assumption, the estimation method developed by Lee and Liu (2010) cannot be directly applied in this model. Although this model can be rewritten into the form of a high order SAR model like in their paper, the heteroskedasticity and correlation affect the efficiency of their GMM estimator.
There has been some attempt at the estimation of this model in the literature. Kelejian and Prucha (2004) use 2SLS and 3SLS methods to estimate a very general SAR model with simultaneous equations. In addition, Tao (2005) studies the identification problem of a similar model and also applies the 2SLS method to school district data. The problem with these estimation methods, however, is that they do not have efficiency properties. On the other hand, the GMM approach of Lee yields efficient estimators. Therefore, we apply the GMM approach in this model in light of Lee’s work.

I consider a simpler two-equation SAR model in this paper. Similar to Lee’s approach, I set up linear and quadratic moment functions and derive the identification conditions for the parameters. Then I show the existence of the best GMME (BGMME) within the class of GMMEs based on linear and quadratic moment conditions. The proposed BGMME is asymptotically as efficient as the MLE when the disturbances are independently and normally distributed. It is not as efficient as the MLE when the disturbances are interrelated, but still more efficient than other linear estimators such as the 2SLSE.

By inspecting the first order conditions of the log likelihood function, I find that we need additional moments for the variance parameters in order to achieve efficiency. This finding is intuitive in that the variance parameters cannot be estimated efficiently in separation of the estimation of the main equations due to the complicated error structure. With the inclusion of the new moments, the resulting GMM estimator achieves the efficiency of the ML estimator.

The paper is organized as follows. In Section 2, I introduce a SAR model with two simultaneous equations and point out the difficulty in implementing the maximum likelihood estimation. Then I derive the identification condition of the model and propose a GMM estimation approach in Section 3. In Section 4 I investigate the consistency and asymptotic distribution of the GMMEs. In Section 5 I derive the best selection of moment functions and optimal IVs. The results are extended to multiple-equation systems in Section 7. Then I provide some Monte Carlo results for the comparison of finite sample properties of estimators in Section 6. Section 8 concludes.
We start from the basic MRSAR model introduced by Anselin (1988), which has the following form:

\[ Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n. \]

The interpretation here is that every individual is affected by its neighbors in the space, and the \((i,j)\)th element of \(W_n\) represents a measure of distance between each pair of \((y_i, y_j)\). Thus, by introducing the spatial lag term, \(W_n Y_n\), into the model, we can capture the neighboring effects in the coefficient \(\lambda\).

However, in market competitions between chain stores, we may face a more complicated situation. Specifically, we may consider supermarkets and gas stations. For example, the pricing strategy of a Shell gas station does not only depend on those of nearby Shell gas stations, but also depends on the strategies of Mobil, BP or other rival gas stations in the vicinity. Similarly, we may consider the interactions between supermarket branches in a MSA as a star network, but meanwhile the stores also engage in price and/or quantity competition with rival retailer chains, which means that there is correlation between different star networks.

Unlike in a simple social interaction model, we have to incorporate the correlation between networks into the model. Another distinctive feature of the model is that different networks have different spatial autoregressive coefficients. The mathematical form is the following. For simplicity, I only consider a market with two competitors in the model, which fits conveniently in the case of a duopoly.

\[
\begin{align*}
Y_{1,m} &= \lambda_{11} W_{11,m} Y_{1,m} + \lambda_{12} W_{12} Y_{2,n} + X_{1,m} \beta_1 + \epsilon_{1,m}, \\
Y_{2,n} &= \lambda_{22} W_{22,n} Y_{2,n} + \lambda_{21} W_{21} Y_{1,m} + X_{2,n} \beta_2 + \epsilon_{2,n}.
\end{align*}
\]

Here, \(Y_{1,m}\) represents the vector of prices or quantities of Wal-Mart stores and \(Y_{2,n}\) represents the vector of prices or quantities of K-Mart stores (or, in Ohio’s case, Meijer stores). \(W_{11,m}\), \(W_{22,n}\) are the weighting matrices for Wal-Mart and K-Mart stores respectively, designated in terms of distances between stores. Similarly, \(W_{12}\) and \(W_{21}\) denote the distances between rival stores. Unlike \(W_{11,m}\) and \(W_{22,n}\), they are \(m \times n\) and \(n \times m\) matrices rather than square matrices when \(m \neq n\).
Due to symmetry in this example, we have \( W_{12} = W_{21}' \). However, this condition can be relaxed in a more general setting. We will drop this assumption in the rest of the paper. In addition, \( X_{1,m} \) and \( X_{2,n} \) are exogenous variables which may be the same or contain different components.

There are three complications to this model. First, the exogenous variables also have different effects on \( Y_{1,m} \) and \( Y_{2,n} \). Secondly, we suspect that \( \lambda_{11} \neq \lambda_{22} \). This is because each competitor has its own cost structure, distribution channels and market strategies. It is not reasonable to assume that they share the same network structure. Finally, the coefficients for the cross terms \( W_{12}Y_{2,n} \) and \( W_{21}Y_{1,m} \) are also different than each other. It is actually an interesting empirical question to compare \( \lambda_{12} \) and \( \lambda_{21} \) and determine which retailer has a larger impact on its rival than the other way around. On the other hand, we can also assume that \( \lambda_{11} = \lambda_{22} \). It is very easy to extend the results in this case.

This model has many potential applications. Regarding the competition between chain stores, the model describe a Cournot competition if the \( Y \)'s represent quantities and a Bertrand competition if the \( Y \)'s represent prices. Furthermore, in the special case when \( m = n \), the model becomes a simultaneous equations system and can be used to describe the demand and supply system as an example.

We are tempted to follow the standards in the literature and assume that the disturbances of the vector \( \epsilon_{1,m} \) are i.i.d. \((0, \sigma^2_1)\) and the disturbances of the vector \( \epsilon_{2,n} \) are i.i.d. \((0, \sigma^2_2)\). However, due to the correlation between \( \epsilon_{1,m} \) and \( \epsilon_{2,n} \), we need a more general and complicated error structure. Thus, we let \( \epsilon_{1,m} = u_1t_m + v_{1,m} \) and \( \epsilon_{2,n} = u_2t_n + v_{2,n} \). Here, \( u_1 \) and \( u_2 \) are considered as group effects while \( v_{1,m} \) and \( v_{2,n} \) are simply white noises. The white noises \( v_{1,m} \) are assumed to be i.i.d. \((0, \sigma^2_{v_1})\) and \( v_{2,n} \) are assumed to be i.i.d. \((0, \sigma^2_{v_2})\), and they are not correlated with \( u_1 \), \( u_2 \) and the \( X \)'s. In this paper, we treat the group effects as random effects. I will leave the fixed effect model to future work. Thus, we assume that \( u_1 \) and \( u_2 \) are not correlated with the \( X \)'s. We also assume that they have zero mean, variances \( \sigma^2_{u_1} \) and \( \sigma^2_{u_2} \) and a covariance \( \sigma_{12} \). Therefore, the covariance matrix of the disturbances is

\[
\Sigma = \begin{pmatrix}
\sigma_{u_1}^2 t_m t_m' + \sigma_{v_1}^2 I_m & \sigma_{12} t_m t_n' \\
\sigma_{12} t_n t_m' & \sigma_{u_2}^2 t_n t_n' + \sigma_{v_2}^2 I_n
\end{pmatrix}.
\]

In practice, however, we cannot consistently estimate the variances of the group effects \( \sigma^2_{u_1} \) and
\( \sigma^2_{u2} \) when the number of groups is small. In this context, we may treat them as fixed effects and estimate the model with dummy variables. In this paper, we do not make specific assumption on the variance matrix but opt for a general structure of \( \Sigma \) with unknown parameters \( \eta_1, \ldots, \eta_t \).

Now we can write this model in the matrix form as follows.

\[
\begin{pmatrix}
Y_{1,m} \\
Y_{2,n}
\end{pmatrix} =
\begin{pmatrix}
\lambda_{11}W_{11,m} & \lambda_{12}W_{12} \\
\lambda_{21}W_{21} & \lambda_{22}W_{22,n}
\end{pmatrix}
\begin{pmatrix}
Y_{1,m} \\
Y_{2,n}
\end{pmatrix} +
\begin{pmatrix}
X_{1,m} & 0 \\
0 & X_{2,n}
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix} +
\begin{pmatrix}
\epsilon_{1,m} \\
\epsilon_{2,n}
\end{pmatrix}.
\]

This equation is not in the form of a spatial autoregressive model since the four weights matrices all have different coefficients. Thus, the standard estimation procedure is not applicable here.

However, we can rewrite the equation above as

\[
\begin{pmatrix}
Y_{1,m} \\
Y_{2,n}
\end{pmatrix} =
\lambda_{11}
\begin{pmatrix}
W_{11,m} & 0 \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
Y_{1,m} \\
Y_{2,n}
\end{pmatrix} +
\lambda_{12}
\begin{pmatrix}
0 & W_{12} \\
0 & 0
\end{pmatrix}
\begin{pmatrix}
Y_{1,m} \\
Y_{2,n}
\end{pmatrix} +
\lambda_{21}
\begin{pmatrix}
0 & 0 \\
W_{21} & 0
\end{pmatrix}
\begin{pmatrix}
Y_{1,m} \\
Y_{2,n}
\end{pmatrix} +
\lambda_{22}
\begin{pmatrix}
0 & 0 \\
0 & W_{22,n}
\end{pmatrix}
\begin{pmatrix}
Y_{1,m} \\
Y_{2,n}
\end{pmatrix} +
\begin{pmatrix}
X_{1,m} & 0 \\
0 & X_{2,n}
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix} +
\begin{pmatrix}
\epsilon_{1,m} \\
\epsilon_{2,n}
\end{pmatrix}.
\]

By changing notations, this model can be simplified into

\[
Y_N = \lambda_{11}W_1Y_N + \lambda_{12}W_2Y_N + \lambda_{21}W_3Y_N + \lambda_{22}W_4Y_N + X_N\beta + \epsilon_N.
\]

This equation is exactly in the form of a high order spatial autoregressive model. Lee and Liu (2010) show that this model can be estimated efficiently by the generalized method of moments. Thus, it may seem as if their estimation procedure can be directly used in this model. However, that is not the case.

The difficulty in using Lee and Liu’s method lies in the assumption of autocorrelation. While Lee and Liu assume that there is no correlation between the disturbances, we relaxed this assumption here by allowing for a more general error structure. Even if we let the correlations be zero, we still have to deal with the problem that the disturbances in the two equations have different variances whereas Lee and Liu assume homoskedasticity in their model.
Therefore, we have to handle the model as it is and find new ways to estimate it. Thus, we come back to the original equation:

\[ Y_N = W_N(\lambda)Y_N + X_N\beta + \epsilon_N, \]

where 

\[
Y_N = \begin{pmatrix}
Y_{1,m} \\
Y_{2,n}
\end{pmatrix},
\]

\[
W_N(\lambda) = \begin{pmatrix}
\lambda_{11}W_{11,m} & \lambda_{12}W_{12} \\
\lambda_{21}W_{21} & \lambda_{22}W_{22,n}
\end{pmatrix},
\]

\[
X_N = \begin{pmatrix}
X_{1,m} & 0 \\
0 & X_{2,n}
\end{pmatrix},
\]

\[
\beta = \begin{pmatrix}
\beta_1 \\
\beta_2
\end{pmatrix},
\]

\[
\epsilon_N = \begin{pmatrix}
\epsilon_{1,m} \\
\epsilon_{2,n}
\end{pmatrix}.
\]

Here, \( \text{var}(\epsilon_N) \equiv \Sigma \).

Let \( \lambda = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22})' \), \( \theta = (\lambda', \beta')' \) and \( \eta = (\eta_1, \ldots, \eta_t)' \). Denote \( \theta_0 \) and \( \eta_0 \) as the true parameters that generate the observed sample. Then denote \( S_N(\lambda) = I_N - W_N(\lambda) \) and \( S_N = S_N(\lambda_0) \). The structural equation implies the reduced form equation:

\[ Y_N = S_N^{-1}(X_N\beta_0 + \epsilon_N). \]

From this reduced form equation, we can see that the spatial autoregressive terms in the structural system are endogenous in general.

If \( \epsilon_N \) is normally distributed, the log likelihood function of this model is

\[
\ln L_N = -n \ln(2\pi) - \frac{1}{2} \ln |\Sigma| + \ln |S_N(\lambda)| - \frac{1}{2} (Y_N - W_N(\lambda)Y_N - X_N\beta)' \Sigma^{-1} (Y_N - W_N(\lambda)Y_N - X_N\beta).
\]

If we want to implement the maximum likelihood method to estimate the model, we need to put restrictions on the parameters to guarantee that the determinant of \( S_N(\lambda) \) is positive, i.e., \( |S_N(\lambda)| > 0 \). Let \( || \cdot || \) be any matrix norm. Then a necessary condition is

\[
\left| \begin{array}{cc}
\lambda_{11}W_{11,m} & \lambda_{12}W_{12} \\
\lambda_{21}W_{21} & \lambda_{22}W_{22,n}
\end{array} \right| < 1.
\]

When all the spatial weights matrices \( W_{ij} \) are row-normalized such that \( ||W_{ij}||_\infty = 1 \) for \( i = 1, 2 \) and \( j = 1, 2 \), a possible parameter space for \( \lambda \) can be those satisfying \( \lambda_{11} + \lambda_{12} < 1 \) and \( \lambda_{21} + \lambda_{22} < 1 \). If the weights matrices are not row-normalized, then the parameter space can be
such that $\lambda_{11} + \lambda_{12} < \left[ \max(||W_{11,m}||, ||W_{12}||) \right]^{-1}$ and $\lambda_{21} + \lambda_{22} < \left[ \max(||W_{21}||, ||W_{22,n}||) \right]^{-1}$ for some matrix norm. Then $S_N(\lambda)$ is invertible and $|S_N(\lambda)| > 0$ on this convex parameter space.

With the appropriate parameter space, it is still not an easy task to implement the maximum likelihood estimation in this model because of the complexity of $S_N(\lambda)$. The evaluation of the determinant of $S_N(\lambda)$ remains an issue because all the spatial weights matrices have different unknown coefficients, which makes $W_N(\lambda)$ undiagonalizable in general. Therefore, we need to consider other estimation methods which does not require the computation of the determinant of $S_N(\lambda)$. Following Lee and Liu (2010), the generalized method of moments seems like a viable option.

### 3 Identification and GMM Estimation

The regularity assumptions for GMM estimation specified in Lee (2007) need to be modified to fit the current model.

**Assumption 1.** The elements of $\epsilon_{1,m}$ and $\epsilon_{2,n}$ have zero mean and covariance matrix $\Sigma$. The fourth moment exists for both series.

**Assumption 2.** The elements of $X_{1,m}$ and $X_{2,n}$ are uniformly bounded constants. Moreover, $X_{1,m}$ and $X_{2,n}$ have full ranks $k_1$ and $k_2$, and $\lim_{m \to \infty} \frac{1}{m} X_{1,m}' X_{1,m}$ and $\lim_{n \to \infty} \frac{1}{n} X_{2,n}' X_{2,n}$ exist and are nonsingular.

**Assumption 3.** The spatial weights matrices $\{W_{ij}\}$ ($i = 1, 2, j = 1, 2$) and $\{S_N^{-1}\}$ are uniformly bounded in both row and column sums in absolute value.

Let $Q_N$ be a $(m + n) \times k_x$ matrix of IVs constructed as functions of $X_{1,m}$, $X_{2,n}$ and $W_{ij,n}$’s. $Q_N$ can be decomposed into $Q_{1,m}$ and $Q_{2,n}$ which correspond with the two equations respectively in the system, so $Q_N = (Q_{1,m}', Q_{2,n}')'$. Denote $\epsilon_N(\theta) = S_N(\lambda) Y_N - X_N \beta$ and $\epsilon_N = \epsilon_N(\theta_0)$. The moment functions correspond to the orthogonality conditions of $X_N$ and $\epsilon_N$ are $Q_N' \epsilon_N(\theta)$.

We can also use additional moment functions $\epsilon_N'(\theta) P_N \epsilon_N(\theta)$ suggested by Lee (2006). However, due to the general error structure in this model, we require the $P_N$’s to have the property $\text{tr}(\Sigma P_N) = 0$ rather than $\text{tr}(P_N) = 0$. We need to place the following restriction on these matrices.

**Assumption 4.** The matrices $P_N$’s with $\text{tr}(\Sigma P_N) = 0$ 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 ($j = 1, \cdots, r$) and IV matrix $Q_N$, the set of moment functions
for a vector
\[ g_N(\theta) = (Q_N, P'_{1N} \epsilon_N(\theta), \cdots, P'_{rN} \epsilon_N(\theta))' \epsilon_N(\theta), \]

which will be used for the estimation of \( \theta \).

First consider the identification of \( \theta_0 \) with these moments. At \( \theta_0 \), \( g_N(\theta_0) = (Q_N, P'_{1N} \epsilon_N, \cdots, P'_{rN} \epsilon_N)' \epsilon_N \) has a zero mean. At a feasible value of \( \theta \), let \( d_N(\theta) = S_N(\lambda)S_N^{-1}X_N \beta_0 - X_N \beta \).

As was defined, \( S_N(\lambda) = I_N - W_N(\lambda) \). Thus, \( S_N^{-1} = I_N + W_N(\lambda_0) + [W_N(\lambda_0)]^2 + \cdots \), and hence, \( S_N(\lambda)S_N^{-1} = I_N + [W_N(\lambda_0) - W_N(\lambda)]S_N^{-1} \).

Therefore, we have \( d_N(\theta) = [W_N(\lambda_0) - W_N(\lambda)]S_N^{-1}X_N \beta_0 + X_N(\beta_0 - \beta) \).

It follows that \( E(Q'_N \epsilon_N(\theta)) = Q'_N d_N(\theta) \) and

\[
E(\epsilon'_N(\theta)P_JN \epsilon_N(\theta)) = E[(d_N(\theta) + S_N(\lambda)S_N^{-1} \epsilon_N)'P_JN(d_N(\theta) + S_N(\lambda)S_N^{-1} \epsilon_N)]
\]
\[
= d'_N(\theta)P_JNd_N(\theta) + E[\epsilon'_N S_N^{-1} S'_N(\lambda)P_JN S_N(\lambda)S_N^{-1} \epsilon_N]
\]
\[
= d'_N(\theta)P_JNd_N(\theta) + tr[E(\epsilon'_N \epsilon_N)S_N^{-1} S'_N(\lambda)P_JN S_N(\lambda)S_N^{-1}]
\]
\[
= d'_N(\theta)P_JNd_N(\theta) + tr[\Sigma_0 S_N^{-1} S'_N(\lambda)P_JN S_N(\lambda)S_N^{-1}].
\]

Thus, we have
\[
E(g_N(\theta)) = \begin{pmatrix}
Q'_N d_N(\theta) \\
d'_N(\theta)P_N d_N(\theta) + tr[\Sigma_0 S_N^{-1} S'_N(\lambda)P_N S_N(\lambda)S_N^{-1}]
\end{pmatrix}.
\]
We can further analyze the first moment equation

\[ Q_N' d_N(\theta) = Q_N' [(W_N(\lambda_0) - W_N(\lambda))S_N^{-1}X_N\beta_0 + X_N(\beta_0 - \beta)] \]

\[ = \begin{pmatrix} Q_{1,m}' & Q_{2,n}' \end{pmatrix} \begin{pmatrix} (\lambda_{11,0} - \lambda_{11})W_{11,m} & (\lambda_{12,0} - \lambda_{12})W_{12} \\ (\lambda_{21,0} - \lambda_{21})W_{21} & (\lambda_{22,0} - \lambda_{22})W_{22,n} \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} X_N\beta_0 + Q_N'X_N(\beta_0 - \beta) \]

\[ = \begin{pmatrix} (\lambda_{11,0} - \lambda_{11})Q_{1,m}W'_{11,m} + (\lambda_{21,0} - \lambda_{21})Q_{2,n}W'_{21} \\ (\lambda_{12,0} - \lambda_{12})Q_{1,m}W'_{12} + (\lambda_{22,0} - \lambda_{22})Q_{2,n}W'_{22,n} \end{pmatrix}' \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} X_N\beta_0 + Q_N'X_N(\beta_0 - \beta) \]

\[ = Q_{1,m}'[(\lambda_{11,0} - \lambda_{11})(G_{11}^{11}X_{1,m}\beta_{1,0} + G_{11}^{12}X_{2,n}\beta_{2,0}) + (\lambda_{12,0} - \lambda_{12})(G_{12}^{11}X_{1,m}\beta_{1,0} + G_{12}^{12}X_{2,n}\beta_{2,0})] + Q_{2,n}'[(\lambda_{21,0} - \lambda_{21})(G_{21}^{11}X_{1,m}\beta_{1,0} + G_{21}^{12}X_{2,n}\beta_{2,0}) + (\lambda_{22,0} - \lambda_{22})(G_{22}^{11}X_{1,m}\beta_{1,0} + G_{22}^{12}X_{2,n}\beta_{2,0})] + Q_{1,m}'X_{1,m}(\beta_{1,0} - \beta_1) + Q_{2,n}'X_{2,n}(\beta_{2,0} - \beta_2). \]

Here, \( G_{ij}^{kl} \equiv W_{ij}S_{kl}^{ij} \) and \( S_{kl}^{ij} \) is the corresponding block in \( S_N^{-1} \).

Denote \( H_{11} = G_{11}^{11}X_{1,m}\beta_{1,0} + G_{11}^{12}X_{2,n}\beta_{2,0} \), \( H_{12} = G_{12}^{11}X_{1,m}\beta_{1,0} + G_{12}^{12}X_{2,n}\beta_{2,0} \), \( H_{21} = G_{21}^{11}X_{1,m}\beta_{1,0} + G_{21}^{12}X_{2,n}\beta_{2,0} \), and \( H_{22} = G_{22}^{11}X_{1,m}\beta_{1,0} + G_{22}^{12}X_{2,n}\beta_{2,0} \).

Thus, the moment equation \( Q_N' d_N(\theta) = 0 \) has the unique solution \((\lambda_0', \beta_0')\) if and only if the matrix \((Q_{1,n}'H_{11,n}, Q_{1,n}'H_{12,n}, Q_{2,n}'H_{21,n}, Q_{2,n}'H_{22,n}, Q_{1,n}'X_{1,n}, Q_{2,n}'X_{2,n})\) has a full column rank \((k_1 + k_2 + 4)\). Equivalently, we require \( Q_N'(H_N, X_N) \) to have a full rank \((k_1 + k_2 + 4)\). In this case, \( \lambda_0 \) and \( \beta_0 \) can be identified from this rank condition. A necessary condition is that the matrix \((H_N, X_N)\) has a full rank \((k_1 + k_2 + 4)\). Equivalently, we require both the matrices \((H_{11}, H_{12}, X_{1,m})\) and \((H_{21}, H_{22}, X_{2,n})\) to have full column ranks. Here, we define

\[ H_N \equiv (H_{1N}, H_{2N}, H_{3N}, H_{4N}) = \begin{pmatrix} H_{11} & H_{12} & 0 & 0 \\ 0 & 0 & H_{21} & H_{22} \end{pmatrix}. \]

On the other hand, if the rank condition is not satisfied, then we need the additional moments \((P_{1N}'\epsilon_N(\theta), \ldots, P_{nN}'\epsilon_N(\theta))'\epsilon_N(\theta)\) to identify \( \lambda_0 \) and \( \beta_0 \). Suppose that the matrix \((H_{11}, H_{12}, X_{1,m})\) has a column rank \((k_1 + 2 - p)\) and the matrix \((H_{21}, H_{22}, X_{2})\) has a column rank \((k_2 + 2 - q)\) with \( p, q \in \{1, 2\} \). As \( X_{1,m} \) and \( X_{2,n} \) both have full ranks, we can form a \((k_1 + 2 - p)\)-column basis and a
(k_2 + 2 - q)-column basis with X_{1,n}, X_{2,n} and other linearly independent columns. Then the other columns in these two matrices can all be expressed as linear combinations of these two bases.

Without loss of generality, we assume here that (H_{1,p+1}, \ldots, H_{12}, X_1) has full rank \((k_1 + 2 - p)\) and also that \((H_{2,q+1}, \ldots, H_{22}, X_2)\) has full rank \((k_2 + 2 - q)\). Then there exist constant vectors \(a_{1l}, a_{2l}\) and constants \(c_{1lj}, c_{2lj}\) such that \(H_{1l} = \sum_{j=p+1}^{2} H_{1j}c_{1lj} + X_1a_{1l}\) for \(l = 1, 2\) and \(H_{2l} = \sum_{j=q+1}^{2} H_{2j}c_{2lj} + X_2a_{2l}\) for \(l = 1, 2\). Hence, the linear moment equations \(Q_Nd_N(\theta) = 0\) are reduced to \(Q'_1\{\sum_{j=p+1}^{2} H_{1j}[\sum_{l=1}^{p} (\lambda_{1l,0} - \lambda_{1l})c_{1lj} + (\lambda_{1j,0} - \lambda_{1j})] + X_1[\sum_{l=1}^{p} (\lambda_{1l,0} - \lambda_{1l})a_{1l} + (\beta_{1,0} - \beta_1)]\} + Q'_2\{\sum_{j=q+1}^{2} H_{2j}[\sum_{l=1}^{q} (\lambda_{2l,0} - \lambda_{2l})c_{2lj} + (\lambda_{2j,0} - \lambda_{2j})] + X_2[\sum_{l=1}^{q} (\lambda_{2l,0} - \lambda_{2l})a_{2l} + (\beta_{2,0} - \beta_2)]\} = 0\), which have all their solutions satisfying

\[
\begin{align*}
\lambda_{1j} &= \lambda_{1j,0} + \sum_{l=1}^{p} (\lambda_{1l,0} - \lambda_{1l})c_{1lj}, & \beta_1 &= \sum_{l=1}^{p} (\lambda_{1l,0} - \lambda_{1l})a_{1l} + \beta_{1,0}, & & \text{for } j = p + 1, \ldots, 2, \\
\lambda_{2j} &= \lambda_{2j,0} + \sum_{l=1}^{q} (\lambda_{2l,0} - \lambda_{2l})c_{2lj}, & \beta_2 &= \sum_{l=1}^{q} (\lambda_{2l,0} - \lambda_{2l})a_{2l} + \beta_{2,0}, & & \text{for } j = q + 1, \ldots, 2.
\end{align*}
\]

From these conditions, we know that \(\beta_0\) and \(\lambda_0\) are identifiable as long as \(\lambda_{11}, \ldots, \lambda_{1p}\) and \(\lambda_{21}, \ldots, \lambda_{2q}\) are identified. In this situation, it is clear that \(d_N(\theta) = 0\), so the identification problem is reduced to \(tr[\Sigma_0 S_N^{-1}S_N'(\lambda)P_{jN}S_N(\lambda)S_N^{-1}] = 0\), for \(j = 1, \ldots, r\).

As we have already computed, \(S_N(\lambda)S_N^{-1} = I_N + [W_N(\lambda_0) - W_N(\lambda)]S_N^{-1}\). Let \(A^S = A + A'\) for any square matrix \(A\). Thus,

\[
tr[\Sigma_0 S_N^{r-1}S_N'(\lambda)P_{jN}S_N(\lambda)S_N^{-1}] = tr\{\Sigma_0[I_N + (W_N(\lambda_0) - W_N(\lambda))S_N^{-1}]P_{jN}[I_N + (W_N(\lambda_0) - W_N(\lambda))S_N^{-1}]\}
= tr(\Sigma_0 P_{jN}) + tr\{\Sigma_0[(W_N(\lambda_0) - W_N(\lambda))S_N^{-1}]P_{jN}(W_N(\lambda_0) - W_N(\lambda))S_N^{-1}\}
+ tr\{\Sigma_0[(W_N(\lambda_0) - W_N(\lambda))S_N^{-1}]P_{jN}[(W_N(\lambda_0) - W_N(\lambda))S_N^{-1}]\}.
\]

We will compute this expression by parts. First of all, we already assume that \(tr(\Sigma_0 P_{jN}) = 0\), so the first term disappears.
Next, we compute

\[
\Sigma_0[(W_N(\lambda_0) - W_N(\lambda)) S_N^{-1}]' P_{jN}^S
\]

\[
= \Sigma_0 \left[ \begin{pmatrix} (\lambda_{11,0} - \lambda_{11}) W_{11} & (\lambda_{12,0} - \lambda_{12}) W_{12} \\ (\lambda_{21,0} - \lambda_{21}) W_{21} & (\lambda_{22,0} - \lambda_{22}) W_{22} \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \right]' P_{jN}^S
\]

\[
= \Sigma_0 \begin{pmatrix} (\lambda_{11,0} - \lambda_{11}) G_{11}^{11} + (\lambda_{12,0} - \lambda_{12}) G_{12}^{12} & (\lambda_{11,0} - \lambda_{11}) G_{12}^{12} + (\lambda_{12,0} - \lambda_{12}) G_{12}^{12} \\ (\lambda_{21,0} - \lambda_{21}) G_{21}^{11} + (\lambda_{22,0} - \lambda_{22}) G_{22}^{12} & (\lambda_{21,0} - \lambda_{21}) G_{22}^{12} + (\lambda_{22,0} - \lambda_{22}) G_{22}^{12} \end{pmatrix}' P_{jN}^S
\]

Obviously, all the elements in the resulting matrix are linear terms regarding the \((\lambda_{ij,0} - \lambda_{ij})\)'s \((i, j \in \{1, 2\})\). Using the linear dependence relations and proper definitions of the \(\alpha_{1i,j}'\)s and \(\alpha_{2i,j}'\)s, we can rewrite the second term into \(tr\{\Sigma_0[(W_N(\lambda_0) - W_N(\lambda)) S_N^{-1}]' P_{jN}^S\} = \sum_{l=1}^p (\lambda_{1l,0} - \lambda_{1l}) \alpha_{1l,j} + \sum_{j=1}^q (\lambda_{2l,0} - \lambda_{2l}) \alpha_{2l,j,}\), for \(j = 1, \ldots, r\).

Lastly, we need to open up the third term:

\[
\Sigma_0[(W_N(\lambda_0) - W_N(\lambda)) S_N^{-1}]' P_{jN}[(W_N(\lambda_0) - W_N(\lambda)) S_N^{-1}]
\]

\[
= \begin{pmatrix} (\lambda_{11,0} - \lambda_{11}) G_{11}^{11} + (\lambda_{12,0} - \lambda_{12}) G_{12}^{12} & (\lambda_{11,0} - \lambda_{11}) G_{12}^{12} + (\lambda_{12,0} - \lambda_{12}) G_{12}^{12} \\ (\lambda_{21,0} - \lambda_{21}) G_{21}^{11} + (\lambda_{22,0} - \lambda_{22}) G_{22}^{12} & (\lambda_{21,0} - \lambda_{21}) G_{22}^{12} + (\lambda_{22,0} - \lambda_{22}) G_{22}^{12} \end{pmatrix}' P_{jN}.
\]

By taking the trace of the matrix above, the third term is essentially the sum of a number of quadratic terms with regards to the \((\lambda_{ij,0} - \lambda_{ij})\)'s. Once again, we can use the linear dependence relations and proper definitions of the \(\alpha_{1kl,j}'\)s, \(\alpha_{2kl,j}'\)s and \(\gamma_{lk,j}'\)s to rewrite the third term into

\[
\text{tr}\{\Sigma_0[(W_N(\lambda_0) - W_N(\lambda)) S_N^{-1}]' P_{jN}[(W_N(\lambda_0) - W_N(\lambda)) S_N^{-1}]\}
\]

\[
= \sum_{l=1}^p \sum_{k=1}^p (\lambda_{1l,0} - \lambda_{1l})(\lambda_{1k,0} - \lambda_{1k}) \alpha_{1l,k,j} + \sum_{l=1}^q \sum_{k=1}^q (\lambda_{2l,0} - \lambda_{2l})(\lambda_{2k,0} - \lambda_{2k}) \alpha_{2l,k,j}
\]

\[
+ \sum_{l=1}^p \sum_{k=1}^q (\lambda_{1l,0} - \lambda_{1l})(\lambda_{2k,0} - \lambda_{2k}) \gamma_{lk,j}, \quad \text{for } j = 1, \ldots, r.
\]
Combining all three terms together, we get the moment equation we need:

\[ tr[\Sigma_0 S_N'^{-1} S_N' (\lambda) P_{JN} S_N (\lambda) S_N'^{-1}] \]

\[ = \sum_{l=1}^{p} (\lambda_{1l,0} - \lambda_{1l}) \alpha_{1l,j} + \sum_{l=1}^{p} (\lambda_{2l,0} - \lambda_{2l}) \alpha_{2l,j} + \sum_{l=1}^{p} \sum_{k=1}^{q} (\lambda_{1l,0} - \lambda_{1l}) (\lambda_{1k,0} - \lambda_{1k}) \alpha_{1l,k,j} \]

\[ + \sum_{l=1}^{p} \sum_{k=1}^{q} (\lambda_{2l,0} - \lambda_{2l}) (\lambda_{2k,0} - \lambda_{2k}) \alpha_{2l,k,j} + \sum_{l=1}^{p} \sum_{k=1}^{q} (\lambda_{1l,0} - \lambda_{1l}) (\lambda_{2k,0} - \lambda_{2k}) \gamma_{l,k,j} = 0, \]

for \( j = 1, \ldots, r. \)

Apparently, \( \lambda_0 \) is a common solution of these \( m \) moment equations. Let \( \alpha_{1l} \) be the \( m \)-dimensional vector with \( \alpha_{1l,j} \) as its \( j \)th element. Also define \( \alpha_{2l}, \alpha_{1lr}, \alpha_{2lr} \) and \( \gamma_{lr} \) in a similar way. Therefore, the necessary and sufficient condition for this equation system to have a unique solution at \( \lambda_0 \) is that the vectors \( \alpha \)'s and \( \gamma \)'s do not have a linear combination with nonlinear coefficients in the form that

\[ \sum_{l=1}^{p} \delta_{1l} \alpha_{1l} + \sum_{l=1}^{q} \delta_{2l} \alpha_{2l} + \sum_{l=1}^{p} \sum_{r=1}^{q} \delta_{1l} \delta_{1r} \alpha_{1l,r,j} + \sum_{l=1}^{p} \sum_{r=1}^{q} \delta_{2l} \delta_{2r} \alpha_{2l,r,j} + \sum_{l=1}^{p} \sum_{r=1}^{q} \delta_{1l} \delta_{2r} \gamma_{l,r,j} = 0, \]

for some nonzero constants \( \delta_{11}, \ldots, \delta_{1p} \) and \( \delta_{21}, \ldots, \delta_{2q} \). This condition is necessary and sufficient for the identification of the model. The sufficient conditions for the identification of \( \theta_0 \) from the moment equations \( \lim_{N \to \infty} E(g_N(\theta)) = 0 \) are summarized in the following assumption.

**Assumption 5.** We must have that either (i) \( \lim_{N \to \infty} \frac{1}{N} Q_N' [H_N, X_N] \) has full rank \((k_1 + k_2 + 4)\), or (ii) \( \lim_{m \to \infty} \frac{1}{n} Q_N'[H_1,p+1, \ldots, H_12, X_1] \) has full rank \((k_1 + 2 - p)\) for some \( 0 \leq p \leq 2 \), \( \lim_{m \to \infty} \frac{1}{n} Q_N' [H_2,q+1, \ldots, H_22, X_2] \) has full rank \((k_2 + 2 - q)\) for some \( 0 \leq q \leq 2 \), and the vectors \( \alpha \)'s and \( \gamma \)'s do not have a linear combination with nonlinear coefficients in the form that \( \sum_{l=1}^{p} \delta_{1l} \alpha_{1l} + \sum_{l=1}^{q} \delta_{2l} \alpha_{2l} + \sum_{l=1}^{p} \sum_{r=1}^{q} \delta_{1l} \delta_{1r} \alpha_{1l,r,j} + \sum_{l=1}^{q} \sum_{r=1}^{q} \delta_{2l} \delta_{2r} \alpha_{2l,r,j} + \sum_{l=1}^{p} \sum_{r=1}^{q} \delta_{1l} \delta_{2r} \gamma_{l,r,j} = 0, \) for some nonzero constants \( \delta_{11}, \ldots, \delta_{1p} \) and \( \delta_{21}, \ldots, \delta_{2q} \).

Let \( \Omega_N = \text{var}(g_N(\theta_0)) \). The variance matrix \( \Omega_N \) involves variances and covariances of linear and quadratic forms of \( \epsilon_N \). Throughout the paper we assume normality of the disturbances. Then

\[ E(Q_N' \epsilon_N' \epsilon_N Q_N) = Q_N' \Sigma_0 Q_N. \]

Obviously, \( \epsilon_N' \equiv \Sigma_0^{-1/2} \epsilon_N \) follows a standard normal distribution. Thus, we can show that

\[ E(Q_N' \epsilon_N' P_N \epsilon_N) = E(Q_N' \Sigma_0^{1/2} \epsilon_N' \Sigma_0^{-1/2} P_N \Sigma_0^{1/2} \epsilon_N) = 0. \]

In addition,

\[ E(P_N \epsilon_N' P_N \epsilon_N) = E(\epsilon_N' \Sigma_0^{1/2} P_N \Sigma_0^{1/2} \epsilon_N' \Sigma_0^{-1/2} P_N \Sigma_0^{1/2} \epsilon_N) = tr(\Sigma_0^{1/2} P_N \Sigma_0^{1/2} P_N \Sigma_0^{1/2}) = \]

12
\( tr(\Sigma_0 P_N \Sigma_0 P^S_{1N}) \). It follows that

\[
\Omega_N = \begin{pmatrix}
Q_N' \Sigma_0 Q_N & 0 & \cdots & 0 \\
0 & tr(\Sigma_0 P_{1N} \Sigma_0 P^S_{1N}) & \cdots & tr(\Sigma_0 P_{rN} \Sigma_0 P^S_{rN}) \\
\vdots & \vdots & \ddots & \vdots \\
0 & tr(\Sigma_0 P_{rN} \Sigma_0 P^S_{rN}) & \cdots & tr(\Sigma_0 P_{rN} \Sigma_0 P^S_{rN})
\end{pmatrix} = \begin{pmatrix}
Q_N' \Sigma_0 Q_N & 0 \\
0 & \Delta_{rN}
\end{pmatrix},
\]

where \( \Delta_{rN} = \frac{1}{2} [vec(\Sigma_0^{1/2} P^S_{1N} \Sigma_0^{1/2}), \ldots, vec(\Sigma_0^{1/2} P^S_{rN} \Sigma_0^{1/2})]' [vec(\Sigma_0^{1/2} P^S_{1N} \Sigma_0^{1/2}), \ldots, vec(\Sigma_0^{1/2} P^S_{rN} \Sigma_0^{1/2})] \).

Then we impose the following conventional regularity condition on the limit of \( \frac{1}{N} \Omega_N \):

**Assumption 6.** The limit of \( \frac{1}{N} \Omega_N \) exists and is a nonsingular matrix.

Let \( g_N(\theta) \equiv (g_0(\theta)', g_1(\theta)', \ldots, g_m(\theta)')' \). Then we have

\[
\frac{\partial g_0(\theta)}{\partial \theta'} = -Q_N' \begin{pmatrix} W_{11}Y_1 & W_{12}Y_2 & 0 & 0 & X_1 & 0 \\
0 & 0 & W_{21}Y_1 & W_{22}Y_2 & 0 & X_2 \end{pmatrix}.
\]

Hence, the score is

\[
\frac{\partial E(g_0(\theta_0))}{\partial \theta'} = -Q_N' \begin{pmatrix} H_{11} & H_{12} & 0 & 0 & X_1 & 0 \\
0 & 0 & H_{21} & H_{22} & 0 & X_2 \end{pmatrix} = -Q_N' \begin{pmatrix} H_{1N} & H_{2N} & H_{3N} & H_{4N} & X_N \end{pmatrix}.
\]

In addition, we have

\[
\frac{\partial g_j(\theta)}{\partial \lambda_{11}} = \begin{pmatrix} (W_{11}Y_1)' & 0 \end{pmatrix} P^S_{jN} \epsilon_N(\theta),
\frac{\partial g_j(\theta)}{\partial \lambda_{12}} = \begin{pmatrix} (W_{12}Y_2)' & 0 \end{pmatrix} P^S_{jN} \epsilon_N(\theta),
\frac{\partial g_j(\theta)}{\partial \lambda_{21}} = \begin{pmatrix} 0 & (W_{21}Y_1)' \end{pmatrix} P^S_{jN} \epsilon_N(\theta),
\frac{\partial g_j(\theta)}{\partial \lambda_{22}} = \begin{pmatrix} 0 & (W_{22}Y_2)' \end{pmatrix} P^S_{jN} \epsilon_N(\theta),
\frac{\partial g_j(\theta)}{\partial \beta_1'} = \begin{pmatrix} X_1' & 0 \end{pmatrix} P^S_{jN} \epsilon_N(\theta),
\frac{\partial g_j(\theta)}{\partial \beta_2'} = \begin{pmatrix} 0 & X_2' \end{pmatrix} P^S_{jN} \epsilon_N(\theta),
\]
for \( j = 1, \ldots, m \).

First compute

\[
\frac{\partial E(g_j(\theta_0))}{\partial \lambda_{11}} = E\left[ (W_{11}Y_1)' \right] P_{jN}^S \epsilon_N(\theta_0) = \text{tr} \left[ P_{jN}^S E \left( \epsilon_N(\theta_0) \left( (W_{11}Y_1)' 0 \right) \right) \right]
\]

\[
= \text{tr} \left[ P_{jN}^S \left( \sigma_{1,0}^2 G_{11}' + \sigma_{1,0} G_{12}' 0 \right) \right]
\]

\[
= \text{tr} \left[ P_{jN}^S \Sigma_0 \left( G_{11}' 0 \right) \right]
\]

\[
= \text{tr}(P_{jN}^S \Sigma_0 G_{1N}') = \text{tr}(G_{1N} \Sigma_0 P_{jN}^S) = \text{tr}(\Sigma_0 P_{jN}^S G_{1N}).
\]

Here we define \( G_{1N} = \begin{pmatrix} G_{11} & G_{12} \\ 0 & G_{11} \end{pmatrix} \). Similarly, \( G_{2N} = \begin{pmatrix} G_{21} & G_{22} \\ 0 & 0 \end{pmatrix} \), \( G_{3N} = \begin{pmatrix} 0 & 0 \\ G_{21} & G_{22} \end{pmatrix} \).

It follows that

\[
D_N = \frac{\partial E(g_N(\theta_0))}{\partial \theta'} = \begin{pmatrix}
Q_N'H_{1N} & Q_N'H_{2N} & Q_N'H_{3N} & Q_N'H_{4N} & Q_N'X_N \\
\text{tr}(\Sigma_0 P_{1N}^S G_{1N}) & \text{tr}(\Sigma_0 P_{1N}^S G_{2N}) & \text{tr}(\Sigma_0 P_{1N}^S G_{3N}) & \text{tr}(\Sigma_0 P_{1N}^S G_{4N}) & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
\text{tr}(\Sigma_0 P_{rN}^S G_{1N}) & \text{tr}(\Sigma_0 P_{rN}^S G_{2N}) & \text{tr}(\Sigma_0 P_{rN}^S G_{3N}) & \text{tr}(\Sigma_0 P_{rN}^S G_{4N}) & 0
\end{pmatrix}
\]

4 Consistency and Asymptotic Distributions

The following proposition provides the asymptotic distribution of a GMME with a linear transformation of the moment equations, \( a_N g_N(\theta) \), where \( a_N \) is a matrix with a full row rank greater than or equal to \((k_1 + k_2 + 4)\). The \( a_N \) is assumed to converge to a constant full rank matrix \( a_0 \). As usual, we assume that the parameter space is a compact convex set including \( \theta_0 \) in its interior.

**Assumption 7.** \( \theta_0 \) is in the interior of the parameter space \( \Theta \), which a compact convex subset of \( R^{k_1+k_2+4} \).

**Proposition 1.** Under Assumptions 1-5, suppose that \( g_N(\theta) \) is the moment function such that \( \lim_{N \to \infty} a_N E(g_N(\theta)) = 0 \) has a unique root at \( \theta_0 \) in \( \Theta \). Then, the GMME \( \hat{\theta}_N \) derived from
\[
\min_{\theta \in \Theta} g_N'(\theta)a'_N a_N g_N(\theta) \text{ is a consistent estimator of } \theta_0, \text{ and } \sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{D} N(0, \Phi), \text{ where }
\]
\[
\Phi = \lim_{N \to \infty} \left[ \left( \frac{D'_N}{N} \right) a'_N a_N \left( \frac{D_N}{N} \right) \right]^{-1} \left( \frac{D'_N}{N} \right) a'_N a_N \left( \frac{\Omega_N}{N} \right) a'_N a_N \left( \frac{D_N}{N} \right) \left[ \left( \frac{D'_N}{N} \right) a'_N a_N \left( \frac{D_N}{N} \right) \right]^{-1},
\]
under the assumption that \( \lim_{N \to \infty} \frac{1}{N} a_N D_N \) exists and has the full rank \((k_1 + k_2 + 4)\).

From Proposition 1, the optimal choice of a weighting matrix \( a'_N a_N \) is \( \Omega_N^{-1} \) by the generalized Schwartz inequality. Since \( \Omega_N \) contains unknown parameters \( \sigma_{u_1}^2, \sigma_{u_2}^2, \sigma_{v_1}^2, \sigma_{v_2}^2 \), the optimal GMM objective function should be formulated with a consistent estimator of the covariance matrix. These parameters can be consistently estimated by using estimated residuals \( \epsilon_N \) from an initial consistent estimate of \( \theta_0 \). Then \( \Omega_N \) can be consistently estimated as \( \hat{\Omega}_N \).

The following proposition shows that the feasible optimal GMME with a consistently estimated \( \hat{\Omega}_N \) has the same limiting distribution as that of the optimal GMME based \( \Omega_N \). With the optimal GMM objective function, an overidentification test is available.

**Proposition 2.** Under Assumptions 1-6, suppose that \( (\hat{\Omega}_N/N)^{-1} - (\Omega_N/N)^{-1} = o_p(1) \), then the feasible optimal GMME \( \hat{\theta}_{f_0,N} \) derived from \( \min_{\theta \in \Theta} g_N'(\theta)a'_N a_N g_N(\theta) \) has the asymptotic distribution
\[
\sqrt{N}(\hat{\theta}_{f_0,N} - \theta_0) \xrightarrow{D} N(0, \lim_{n \to \infty} \frac{1}{N} D'_N \Omega_N^{-1} D_N^{-1}).
\]
Furthermore,
\[
g_N'(\hat{\theta}_N)\hat{\Omega}_N^{-1} g_N(\hat{\theta}_N) \xrightarrow{D} \chi^2((k_x + r) - (k_1 + k_2 + 4)).
\]

5 Efficiency and the BGMME

From the previous sections, we know that
\[
D'_N \Omega_N^{-1} D_N = \begin{pmatrix} A_N B_N^{-1} A_N & 0 \\ 0 & 0 \end{pmatrix} + C_N,
\]
where $A_N = \begin{pmatrix} tr(S_0^{1/2} P_1 S_0^{1/2} S_0^{1/2} G_1 \Sigma_0^{1/2}) & \cdots & tr(S_0^{1/2} P_r S_0^{1/2} S_0^{1/2} G_1 \Sigma_0^{1/2}) \\ \vdots & \ddots & \vdots \\ tr(S_0^{1/2} P_1 S_0^{1/2} S_0^{1/2} G_r \Sigma_0^{1/2}) & \cdots & tr(S_0^{1/2} P_r S_0^{1/2} S_0^{1/2} G_r \Sigma_0^{1/2}) \end{pmatrix}$,

$B_N = \begin{pmatrix} tr(S_0^{1/2} P_1 S_0^{1/2} S_0^{1/2} G_{1N} \Sigma_0^{1/2}) & \cdots & tr(S_0^{1/2} P_r S_0^{1/2} S_0^{1/2} G_{1N} \Sigma_0^{1/2}) \\ \vdots & \ddots & \vdots \\ tr(S_0^{1/2} P_1 S_0^{1/2} S_0^{1/2} G_{rN} \Sigma_0^{1/2}) & \cdots & tr(S_0^{1/2} P_r S_0^{1/2} S_0^{1/2} G_{rN} \Sigma_0^{1/2}) \end{pmatrix}$, and

$C_N = (H_N, X_N)' Q_N (Q_N' \Sigma_0 Q_N)^{-1} Q_N' (H_N, X_N)$.

From this asymptotic precision matrix, the generalized Schwartz inequality implies that the best selection of $Q_N$ will be the matrix $\Sigma_0^{-1}(H_N, X_N)$. As for the best selection of the $P_N$'s, we need to inspect $A_N B_N^{-1} A_N$.

We already know that we can rewrite $B_N$ as $B_N = \frac{1}{2} [vec(S_0^{1/2} P_1 S_0^{1/2}), \ldots, vec(S_0^{1/2} P_r S_0^{1/2})]'$ $\times$ $[vec(S_0^{1/2} P_1 S_0^{1/2}), \ldots, vec(S_0^{1/2} P_r S_0^{1/2})]$. On the other hand, we can compute

$$
tr(S_0^{1/2} P_j S_0^{1/2} S_0^{1/2} G_{kN} \Sigma_0^{1/2})
= \frac{1}{2} tr \left( S_0^{1/2} P_j S_0^{1/2} \left[ S_0^{1/2} \Sigma_0^{-1} G_{kN} \Sigma_0^{1/2} - \frac{tr(S_0^{1/2} \Sigma_0^{-1} G_{kN} \Sigma_0^{1/2})}{N} I_N \right] S \right)
= \frac{1}{2} vec' \left( \left[ S_0^{1/2} \Sigma_0^{-1} G_{kN} \Sigma_0^{1/2} - \frac{tr(G_{kN})}{N} I_N \right] S \right) vec(S_0^{1/2} P_j S_0^{1/2})
= \frac{1}{2} vec' \left( \left[ S_0^{1/2} \Sigma_0^{-1} \left( G_{kN} - \frac{tr(G_{kN})}{N} I_N \right) \Sigma_0^{1/2} \right] S \right) vec(S_0^{1/2} P_j S_0^{1/2})
$$

for $j = 1, \ldots, r$ and $k = 1, \ldots, 4$. Therefore, the generalized Schwartz inequality implies that

$$
A_N B_N^{-1} A_N \leq \frac{1}{2} [vec(S_0^{1/2} (\Sigma_0^{-1} \Gamma_1) S_0^{1/2}), \ldots, vec(S_0^{1/2} (\Sigma_0^{-1} \Gamma_4) S_0^{1/2})] \times
[vec(S_0^{1/2} (\Sigma_0^{-1} \Gamma_1) S_0^{1/2}), \ldots, vec(S_0^{1/2} (\Sigma_0^{-1} \Gamma_4) S_0^{1/2})]

= \begin{pmatrix} tr(S_0 (\Sigma_0^{-1} \Gamma_1) S G_1 N) & \cdots & tr(S_0 (\Sigma_0^{-1} \Gamma_1) S G_4 N) \\ \vdots & \ddots & \vdots \\ tr(S_0 (\Sigma_0^{-1} \Gamma_4) S G_1 N) & \cdots & tr(S_0 (\Sigma_0^{-1} \Gamma_4) S G_4 N) \end{pmatrix},
$$

where $\Gamma_{kN} = G_{kN} - (tr(G_{kN})/N) I_N$, for $k = 1, \ldots, 4$. Hence, the best selection of the $P_N$'s will be the matrices $\Sigma_0^{-1}[G_{kN} - (tr(G_{kN})/N) I_N]$ for $k = 1, \ldots, 4$. 

16
In practice, with initial consistent estimates of \( \lambda, \beta \) and the variances, we can get consistent estimates \( \hat{\Sigma}_N, \hat{H}_N \) and the \( \hat{G}_N \)'s. The following proposition summarizes the results above.

**Proposition 3.** Let \( Q^*_N = \Sigma_0^{-1}(H_N, X_N) \) and \( P^*_iN = \Sigma_0^{-1}[G_{K_N} - (tr(G_{K_N})/N)I_N] \) for \( i = 1, \ldots, 4 \). Under Assumptions 1-3, the consistent root \( \hat{\theta}_{b,N} \) derived from \( \min_{\theta \in \Theta} g_N^*(\theta)\Omega^{-1}_N g_N^*(\theta) \), where \( \Omega^*_N = \text{var}(g_N^*(\theta)) \) and \( g_N^*(\theta) = (Q^*_N, P^*_1N\epsilon_N(\theta), \ldots, P^*_4N\epsilon_N(\theta))'\epsilon_N(\theta) \), is the BGMME with the asymptotic distribution \( \sqrt{N}(\hat{\theta}_{b,N} - \theta_0) \overset{D}{\rightarrow} N(0, \Phi_b^{-1}) \) and

\[
\Phi_b = \lim_{N \rightarrow \infty} \frac{1}{N} \begin{pmatrix} \Phi_{11} & H_N'\Sigma_0^{-1}X_N \\ X_N'\Sigma_0^{-1}H_N & X_N'\Sigma_0^{-1}X_N \end{pmatrix},
\]

\[
\Phi_{11} = \begin{pmatrix} H_1^N\Sigma_0^{-1}H_1 + tr(\Sigma_0(\Sigma_0^{-1}G_{1N})^S G_{1N}) & \cdots & H_1^N\Sigma_0^{-1}H_4 + tr(\Sigma_0(\Sigma_0^{-1}G_{1N})^S G_{4N}) \\ \vdots & \ddots & \vdots \\ H_4^N\Sigma_0^{-1}H_1 + tr(\Sigma_0(\Sigma_0^{-1}G_{4N})^S G_{1N}) & \cdots & H_4^N\Sigma_0^{-1}H_4 + tr(\Sigma_0(\Sigma_0^{-1}G_{4N})^S G_{4N}) \end{pmatrix}.
\]

In order to study the efficiency of the BGMME, we write out the log likelihood function of the structural equation:

\[
\ln L_N = -n \ln(2\pi) - \frac{1}{2} \ln|\Sigma| + \ln|S_N(\lambda)| - \frac{1}{2}(Y_N - W_N(\lambda)Y_N - X_N\beta)'\Sigma^{-1}(Y_N - W_N(\lambda)Y_N - X_N\beta).
\]

The first order conditions are

\[
\frac{\partial \ln L_N}{\partial \lambda_{11}} = H_1^N\Sigma^{-1}\epsilon_N(\theta) + \epsilon_N'(\theta)\Sigma^{-1}G_{1N}\epsilon_N(\theta) - tr(G_{1N}) = 0,
\]

\[
\frac{\partial \ln L_N}{\partial \lambda_{12}} = H_2^N\Sigma^{-1}\epsilon_N(\theta) + \epsilon_N'(\theta)\Sigma^{-1}G_{2N}\epsilon_N(\theta) - tr(G_{2N}) = 0,
\]

\[
\frac{\partial \ln L_N}{\partial \lambda_{21}} = H_3^N\Sigma^{-1}\epsilon_N(\theta) + \epsilon_N'(\theta)\Sigma^{-1}G_{3N}\epsilon_N(\theta) - tr(G_{3N}) = 0,
\]

\[
\frac{\partial \ln L_N}{\partial \lambda_{22}} = H_4^N\Sigma^{-1}\epsilon_N(\theta) + \epsilon_N'(\theta)\Sigma^{-1}G_{4N}\epsilon_N(\theta) - tr(G_{4N}) = 0,
\]

\[
\frac{\partial \ln L_N}{\partial \beta} = X_N'\Sigma^{-1}\epsilon_N(\theta) = 0,
\]

and

\[
\frac{\partial \ln L_N}{\partial \eta_i} = -\frac{1}{2} tr \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \eta_i} \right) - \frac{1}{2} \epsilon_N'(\theta) \frac{\partial \Sigma^{-1}}{\partial \eta_i} \epsilon_N(\theta) = 0, \text{ for } i = 1, \ldots, t,
\]

\[17\]
where the \( \eta_i \) is the \( i \)th element in the vector \( \eta \), namely one of the variance parameters.

The similarity of the best GMM moments and the likelihood equations is revealing. However, the efficiency of the BGMME may be affected by the dependence between the disturbances because of the generalized form of the covariance matrix. Due to the complicated error structure, it is hard to solve the first order conditions with respect to the variance parameters. Thus, we cannot substitute the variance matrix in the first order conditions regarding \( \lambda \) and \( \beta \) and get the first order conditions for the concentrated likelihood function. This gives us some hint at the additional moments we may use to get efficient estimates.

First, we need to compare the variance matrices of the MLE and the BGMME. By inverting the Fisher Information and following Anselin and Bera (1998), we can get the asymptotic variance of the MLE \( (\hat{\theta}_{ml,N}, \hat{\eta}_{ml,N}) \):

\[
\text{AsyVar}(\hat{\theta}_{ml,N}, \hat{\eta}_{ml,N}) = \begin{pmatrix}
V_{11} & H_N' \Sigma^{-1} X_N & V_{13} \\
X_N' \Sigma^{-1} H_N & X_N' \Sigma^{-1} X_N & 0 \\
V_{13} & 0 & V_{33}
\end{pmatrix}^{-1},
\]

where \( V_{11} = \hat{H} + \hat{V}_{11} \) with \( \hat{H} \equiv \begin{pmatrix}
H_{1N}' \Sigma^{-1} H_{1N} & \cdots & H_{1N}' \Sigma^{-1} H_{4N} \\
\vdots & \ddots & \vdots \\
H_{4N}' \Sigma^{-1} H_{1N} & \cdots & H_{4N}' \Sigma^{-1} H_{4N}
\end{pmatrix} \) and

\[
\hat{V}_{11} \equiv \begin{pmatrix}
tr(\Sigma^{1/2} \Sigma^{-1} G_{1N} \Sigma^{1/2} \Sigma^{1/2} (\Sigma^{-1} G_{1N}) S \Sigma^{1/2}) & \cdots & tr(\Sigma^{1/2} \Sigma^{-1} G_{1N} \Sigma^{1/2} \Sigma^{1/2} (\Sigma^{-1} G_{4N}) S \Sigma^{1/2}) \\
\vdots & \ddots & \vdots \\
tr(\Sigma^{1/2} \Sigma^{-1} G_{4N} \Sigma^{1/2} \Sigma^{1/2} (\Sigma^{-1} G_{4N}) S \Sigma^{1/2}) & \cdots & tr(\Sigma^{1/2} \Sigma^{-1} G_{4N} \Sigma^{1/2} \Sigma^{1/2} (\Sigma^{-1} G_{4N}) S \Sigma^{1/2})
\end{pmatrix},
\]

\[
V_{13} = \begin{pmatrix}
-tr(\Sigma^{1/2} \Sigma^{-1} G_{1N} \Sigma^{1/2} \Sigma^{1/2} (\Sigma^{-1} \frac{\partial \Sigma^{-1}}{\partial \eta_1} \Sigma^{1/2})) & \cdots & -tr(\Sigma^{1/2} \Sigma^{-1} G_{1N} \Sigma^{1/2} \Sigma^{1/2} (\Sigma^{-1} \frac{\partial \Sigma^{-1}}{\partial \eta_1} \Sigma^{1/2})) \\
\vdots & \ddots & \vdots \\
-tr(\Sigma^{1/2} \Sigma^{-1} G_{4N} \Sigma^{1/2} \Sigma^{1/2} (\Sigma^{-1} \frac{\partial \Sigma^{-1}}{\partial \eta_1} \Sigma^{1/2})) & \cdots & -tr(\Sigma^{1/2} \Sigma^{-1} G_{4N} \Sigma^{1/2} \Sigma^{1/2} (\Sigma^{-1} \frac{\partial \Sigma^{-1}}{\partial \eta_1} \Sigma^{1/2}))
\end{pmatrix}
\]

\[
= -\frac{1}{4} \left[ \text{vec}(\Sigma_0^{1/2} (\Sigma^{-1}_0 G_{1N}) S \Sigma_0^{1/2}), \cdots, \text{vec}(\Sigma_0^{1/2} (\Sigma^{-1}_0 G_{4N}) S \Sigma_0^{1/2}) \right] \times
\]

\[
\left[ \text{vec}(\Sigma_0^{1/2} (\frac{\partial \Sigma^{-1}}{\partial \eta_1} S \Sigma_0^{1/2}), \cdots, \text{vec}(\Sigma_0^{1/2} (\frac{\partial \Sigma^{-1}}{\partial \eta_1} S \Sigma_0^{1/2})) \right]
\]

\[
= -\frac{1}{4} F' D,
\]
Hence, the asymptotic variance of the MLE would be the same as that of the BGMME with
is more efficient than any other estimator in the specified class such as the 2SLS estimator and the

Thus, the BGMME is not as efficient as the

From the inverse of a partitioned matrix, the asymptotic variance of the MLE \( \hat{\theta}_{ml,N} \) is

Here \( V_{13}V_{33}^{-1}V_{13}' = \frac{1}{N} F' D(D'D)^{-1} D' F \). If the disturbances are i.i.d., then \( V_{13}V_{33}^{-1}V_{13}' = \frac{1}{N} F' vec(I_N) vec'(I_N) F \).

Thus,

Hence, the asymptotic variance of the MLE would be the same as that of the BGMME with
\( \Phi_{11} = V_{11} - V_{13}V_{33}^{-1}V_{13}' \). However, when the disturbances are not i.i.d., these two matrices are not equal and there is no clear relationship between them. Thus, the BGMME is not as efficient as the MLE since the variance of the MLE attains the lower bound of Fisher Information. Nevertheless, it is more efficient than any other estimator in the specified class such as the 2SLS estimator and the
3SLS estimator. The following proposition summarizes these results and shows that the feasible BGMME has the same limiting distribution as the BGMME.

**Proposition 4.** Suppose \( \hat{\theta}_N \) and \( \hat{\eta}_N \) are consistent estimates of \( \theta_0 \) and \( \eta_0 \). Then the problem

\[
\min_{\theta \in \Theta} \bar{g}_N^* (\theta) \Omega_N^{-1} \bar{g}_N^* (\theta), \quad \text{where} \quad \bar{g}_N^* (\theta) = (\hat{Q}_N, \hat{P}_{1N} \epsilon_N (\theta), \ldots, \hat{P}_{4N} \epsilon_N (\theta))' \epsilon_N (\theta),
\]

yields the feasible BGMME \( \hat{\theta}_{fb,N} \) which has the same limiting distribution as that of \( \hat{\theta}_{b,N} \). This estimator is the best in the class, but generally not as efficient as the MLE. In the special case where the disturbances are i.i.d., the feasible BGMME is efficient.

This result is clearly different than what we see in the previous work of Lee and Liu. From the analysis of the first order conditions above, we know that the problem arises from the complicated structure of the variance matrix. In order to achieve the efficiency of the MLE, we need to estimate not only \( \lambda \) and \( \beta \), but also the variance parameters \( \eta \) simultaneously. Following this idea, we develop a new class of moment functions \( g_N (\theta) = (Q_N' \epsilon_N (\theta), \epsilon_N (\theta) P_{1N}' \epsilon_N (\theta) - \text{tr} (P_{1N} \Sigma (\theta)), \ldots, \epsilon_N (\theta) P_{rN}' \epsilon_N (\theta) - \text{tr} (P_{rN} \Sigma (\theta)))' \).

With this modification, we no longer require \( \text{tr} (P_{iN}) = 0 \) for \( i = 1, \ldots, N \). The identification conditions only need to be slightly modified due to the similar form of moment functions. Denote \( \vartheta = (\lambda_{11}, \lambda_{12}, \lambda_{21}, \lambda_{22}, \eta', \beta')' \). The score function \( \tilde{D}_N \) now becomes

\[
\frac{\partial E (g_N (\theta_0))}{\partial \vartheta} = \begin{pmatrix}
Q_N' H_{1N} & \cdots & Q_N' H_{4N} & 0 & \cdots & 0 & Q_N' X_N \\
\text{tr} (\Sigma_0 P_{1N}^S G_{1N}) & \cdots & \text{tr} (\Sigma_0 P_{1N}^S G_{4N}) & -\text{tr} (P_{1N} \frac{\partial \Sigma_0}{\partial \eta_0}) & \cdots & -\text{tr} (P_{1N} \frac{\partial \Sigma_0}{\partial \eta_0}) & 0 \\
\vdots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\text{tr} (\Sigma_0 P_{rN}^S G_{1N}) & \cdots & \text{tr} (\Sigma_0 P_{rN}^S G_{4N}) & -\text{tr} (P_{rN} \frac{\partial \Sigma_0}{\partial \eta_0}) & \cdots & -\text{tr} (P_{rN} \frac{\partial \Sigma_0}{\partial \eta_0}) & 0 \\
\end{pmatrix}
\]

\[
= \begin{pmatrix}
Q_N' H_{1N} & \cdots & Q_N' H_{4N} & 0 & \cdots & 0 & Q_N' X_N \\
\text{tr} (\Sigma_0 P_{1N}^S G_{1N}) & \cdots & \text{tr} (\Sigma_0 P_{1N}^S G_{4N}) & \text{tr} (P_{1N} \Sigma_0 \frac{\partial \Sigma_0^{-1}}{\partial \eta_0} \Sigma_0) & \cdots & \text{tr} (P_{1N} \Sigma_0 \frac{\partial \Sigma_0^{-1}}{\partial \eta_0} \Sigma_0) & 0 \\
\vdots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\
\text{tr} (\Sigma_0 P_{rN}^S G_{1N}) & \cdots & \text{tr} (\Sigma_0 P_{rN}^S G_{4N}) & \text{tr} (P_{rN} \Sigma_0 \frac{\partial \Sigma_0^{-1}}{\partial \eta_0} \Sigma_0) & \cdots & \text{tr} (P_{rN} \Sigma_0 \frac{\partial \Sigma_0^{-1}}{\partial \eta_0} \Sigma_0) & 0 \\
\end{pmatrix}.
\]
The variance matrix of the moment functions is slightly changed to  \( \tilde{\Omega}_N = \Omega_N^{1/2} Q_N \Omega_N^{-1/2} Q_N' \Omega_N^{1/2} \), where 

\[
\begin{pmatrix}
Q_N' \Sigma_0 Q_N & 0 & \ldots & 0 \\
0 & tr(\Sigma_0 P_{1N} \Sigma_0 P_{1N}' S_0) - tr(\Sigma_0 P_{1N}) tr(\Sigma_0 P_{1N}) & \ldots & tr(\Sigma_0 P_{1N} \Sigma_0 P_{1N}' S_0) - tr(\Sigma_0 P_{1N} \Sigma_0 P_{1N}' S_0)
\end{pmatrix}.
\]

Using similar steps in the previous case and exploiting the symmetry of the variance matrix, we can get the asymptotic precision matrix of the GMME for \( \lambda, \eta \) and \( \beta \):

\[
\lim_{N \to \infty} D_N \tilde{\Omega}_N^{-1} D_N = \begin{pmatrix}
\tilde{A}_N \tilde{B}_N^{-1} \tilde{A}_N & 0 \\
0 & 0
\end{pmatrix} + \tilde{C}_N,
\]

where \( \tilde{A}_N = (A_N, A^+) \) with

\[
A^+ = \frac{1}{2} \begin{pmatrix}
tr(\Sigma_0^{1/2} P_{1N} \Sigma_0^{1/2} S_0^{1/2}) & \ldots & tr(\Sigma_0^{1/2} P_{1N} \Sigma_0^{1/2} S_0^{1/2}) \\
\vdots & \ddots & \vdots \\
tr(\Sigma_0^{1/2} P_{1N} \Sigma_0^{1/2} S_0^{1/2}) & \ldots & tr(\Sigma_0^{1/2} P_{1N} \Sigma_0^{1/2} S_0^{1/2})
\end{pmatrix},
\]

\[
\tilde{B}_N = \begin{pmatrix}
tr(\Sigma_0^{1/2} P_{1N} \Sigma_0^{1/2} S_0^{1/2} P_{1N}' S_0^{1/2}) & \ldots & tr(\Sigma_0^{1/2} P_{1N} \Sigma_0^{1/2} S_0^{1/2} P_{1N}' S_0^{1/2}) \\
\vdots & \ddots & \vdots \\
tr(\Sigma_0^{1/2} P_{1N} \Sigma_0^{1/2} S_0^{1/2} P_{1N}' S_0^{1/2}) & \ldots & tr(\Sigma_0^{1/2} P_{1N} \Sigma_0^{1/2} S_0^{1/2} P_{1N}' S_0^{1/2})
\end{pmatrix},
\]

and \( \tilde{C}_N = (H_N, 0, X_N)' Q_N (Q_N' \Sigma_0^{-1} Q_N) (H_N, 0, X_N) \).

Again, by the generalized Schwartz inequality, the best selection of \( Q_N \) will be the matrix \( \Sigma_0^{-1}(H_N, X_N) \) just as before. As for the best selection of the \( P_N \)'s, it is not hard to figure out that we still need the matrices \( P_i^{*} = \Sigma_0^{-1} G_i \) for \( i = 1, \ldots, 4 \). In addition, we need some moments with \( P_j^{**} = \frac{\partial \Omega_j^{-1}}{\partial \theta_j} \) for \( j = 1, \ldots, t \). If we substitute these matrices in the asymptotic precision matrix, we can see that it is equal to the asymptotic precision matrix of MLE, which means that this GMME is efficient.

The following proposition summarizes the results above.

**Proposition 5.** Let \( Q_N^* = \Sigma_0^{-1}(H_N, X_N) \), \( P_i^{*} = \Sigma_0^{-1} G_i \) for \( i = 1, \ldots, 4 \) and \( P_j^{**} = \frac{\partial \Omega_j^{-1}}{\partial \theta_j} \) for \( j = 1, \ldots, t \). Under Assumptions 1-3, the consistent root \( \hat{\theta}_N \), derived from \( \min_{\theta \in \Theta} g_N^*(\theta) \Omega_N^{-1} g_N^*(\theta) \), where \( \Omega_N = \text{var}(g_N^*(\theta_0)) \) and \( g_N^*(\theta) = (Q_N^* \epsilon_N(\theta), \epsilon_N(\theta) P_{1N}^* \epsilon_N(\theta) - tr(P_{1N}^* \Sigma(\theta)), \ldots, \epsilon_N(\theta) P_{1N}^{**} \epsilon_N(\theta) - tr(P_{1N}^{**} \Sigma(\theta)), \ldots, \epsilon_N(\theta) P_{1N}^{**} \epsilon_N(\theta) - tr(P_{1N}^{**} \Sigma(\theta)))' \), is the best GMME for \( \lambda, \beta \) and \( \eta \). This GMME is efficient.
We will refer to this estimator as the EGMME in the Monte Carlo study. The asymptotic properties of the feasible EGMME are also easy to obtain.

6 Multiple Network Systems

So far we have only considered a system with two networks. In reality, however, we often need to deal with multiple networks. That is what we will briefly discuss in this section.

A spatial autoregressive model with \( R \) interrelated networks can be formulated as follows.

\[
Y_{1,n_1} = \lambda_{11}W_{11}Y_{1,n_1} + \lambda_{12}W_{12}Y_{2,n_2} + \cdots + \lambda_{1R}W_{1R}Y_{1,n_R} + X_{1,n_1}\beta_1 + \epsilon_{1,n_1},
\]
\[
Y_{2,n_2} = \lambda_{21}W_{21}Y_{1,n_1} + \lambda_{22}W_{22}Y_{2,n_2} + \cdots + \lambda_{2R}W_{2R}Y_{1,n_R} + X_{2,n_2}\beta_2 + \epsilon_{2,n_2},
\]
\[
\vdots
\]
\[
Y_{R,n_R} = \lambda_{R1}W_{R1}Y_{1,n_1} + \lambda_{R2}W_{R2}Y_{2,n_2} + \cdots + \lambda_{RR}W_{RR}Y_{1,n_R} + X_{R,n_R}\beta_1 + \epsilon_{R,n_R}.
\]

As we see in previous sections, the structure of the variance matrix does not affect the optimal selection of the moment functions. Therefore, we do not need to place any restriction on the variance matrix, but only assumes that it contains \( t \) unknown parameters \( \eta_1, \ldots, \eta_t \).

In practice, we can allow the disturbances to have a panel data-like structure if the number of networks is large. For example, we can assume that \( E(\epsilon_{r,n}, \epsilon'_{r,n}) = \sigma^2_u\epsilon_{n_r} + \sigma^2_vI_{n_r} \) for \( r = 1, \ldots, R \) and \( E(\epsilon_{r,n}, \epsilon'_{s,n}) = 0 \) if \( r \neq s \). These assumptions allow us to select the moment functions. Therefore, we do not need to place any restriction on the variance matrix, but only assumes that it contains \( t \) unknown parameters \( \eta_1, \ldots, \eta_t \).

We can conveniently extend the results in previous sections to the multiple network systems. Define all the \( G \) and \( H \) matrices in a similar fashion as in the previous sections. Then the EGMME can be obtained according to the following proposition.

**Proposition 6.** Let \( Q^*_N = \Sigma_0^{-1}(H_N, X_N) \), \( P^*_i = \Sigma_0^{-1}G_{iN} \) for \( i = 1, \ldots, R^2 \) and \( P^*_{jN} = \frac{\partial \Sigma_0^{-1}}{\partial \eta_j} \) for \( j = 1, \ldots, t \). Under Assumptions 1-3, the consistent root \( \hat{\theta}_{b,N} \) derived from \( g^*_{N}(\theta)\Omega_N^{-1}g^*_{N}(\theta) \), where \( \Omega_N = \text{var}(g^*_N(\theta_0)) \) and \( g^*_N(\theta) = (Q^*_N\epsilon_N(\theta), \epsilon'_N(\theta)P^*_{1N}\epsilon_N(\theta) - \text{tr}(P^*_{1N}\Sigma(\theta)), \cdots, \epsilon'_N(\theta)P^*_{R^2,2N}\epsilon_N(\theta) - \text{tr}(P^*_{R^2,2N}\Sigma(\theta)), \cdots, \epsilon'_N(\theta)P^*_{R^2,2N}\epsilon_N(\theta) - \text{tr}(P^*_{R^2,2N}\Sigma(\theta)))' \), is the best GMME for \( \lambda, \beta \) and \( \eta \). This GMME is efficient.

The asymptotic properties of the feasible EGMME are also easy to obtain.
7 Monte Carlo Study (in progress)

The model in the Monte Carlo study is specified as

\[ Y_{1,m} = \lambda_{11} W_{11} Y_{1,m} + \lambda_{12} W_{12} Y_{2,n} + X_{1,m} \beta_1 + \epsilon_{1,m}, \]
\[ Y_{2,n} = \lambda_{21} W_{21} Y_{1,m} + \lambda_{22} W_{22} Y_{2,n} + X_{2,n} \beta_2 + \epsilon_{2,n}, \]

where \( x_{1i} \) and \( x_{2i} \) are independently generated variables. \( x_{1i} \)'s are i.i.d. \( N(50, 1) \) and \( x_{2i} \)'s are i.i.d. \( N(50, 1.5) \). There is heteroskedasticity in the disturbances across the two networks: \( \epsilon_{1i} \)'s are i.i.d. \( N(0, 1) \) and \( \epsilon_{2i} \)'s are i.i.d. \( N(0, 2) \). For the ease of simulation we assume that the fixed effects are zero. The weights matrices \( W \)'s are randomly generated and row normalized. The sample size of the first network is either 100, 200 or 400 and the sample size of the second network is 80 percent of the first network. The estimation methods considered are the

1. 2SLS—the 2SLS method with IV’s \( W_{11} X_1, W_{12} X_1, W_{11} W_{12} X_2 \) for \( W_{11} Y_1, W_{12} Y_2, W_{11} W_{12} X_1, W_{12} W_{22} X_2 \) for \( W_{12} Y_2, W_{21} X_1, W_{21} W_{11} X_1, W_{21} W_{22} X_2 \) for \( W_{21} Y_1, W_{22} X_2, W_{22} W_{21} X_1, W_{22} W_{22} X_2 \) for \( W_{22} Y_2 \) and \( X_1, X_2 \);

2. BGMM—the best GMM approach by using \( Q_N = \hat{\Sigma}^{-1}(\hat{H}_N, X_N) \) for the linear moments and \( P_i = \hat{\Sigma}^{-1}[\hat{G}_N - (tr(\hat{G}_N)/N)I_N] \) for \( i = 1, \ldots, 4 \) for the quadratic moments with initial consistent estimates for \( \lambda, \beta \) and \( \eta \);

3. EGMM—the efficient GMM approach by using \( Q_N = \hat{\Sigma}^{-1}(\hat{H}_N, X_N) \) for the linear moments and \( P_i = \hat{\Sigma}^{-1}\hat{G}_N \) for \( i = 1, \ldots, 4 \) and \( P_j = \frac{\partial \hat{\Sigma}^{-1}}{\partial \eta_j} \) for \( j = 1, 2 \) for the quadratic moments with initial consistent estimates for \( \lambda, \beta \) and \( \eta \).

The number of repetitions is 300 in this Monte Carlo experiment. The regressors are randomly redrawn for each repetition. In each case, we report the bias, the standard deviation and the root mean squared error of the empirical distributions of the estimates. The following table reports the results where \( \lambda_{11,0} = 0.6, \lambda_{12,0} = -0.2, \lambda_{21,0} = -0.1, \lambda_{22,0} = 0.5, \beta_{1,0} = 1 \) and \( \beta_{2,0} = 0.8 \).
<table>
<thead>
<tr>
<th></th>
<th>2SLS</th>
<th>BGMM</th>
<th>EGMM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias(SD)[RMSE]</td>
<td>Bias(SD)[RMSE]</td>
<td>Bias(SD)[RMSE]</td>
</tr>
<tr>
<td>N=180</td>
<td>λ11  0.042(0.998)[0.999]</td>
<td>0.040(0.679)[0.681]</td>
<td>0.018(0.660)[0.660]</td>
</tr>
<tr>
<td></td>
<td>λ12  -0.167(1.552)[1.561]</td>
<td>-0.163(1.054)[1.066]</td>
<td>-0.128(1.024)[1.032]</td>
</tr>
<tr>
<td></td>
<td>λ21  0.311(1.658)[1.687]</td>
<td>0.123(1.041)[1.048]</td>
<td>0.126(0.939)[0.948]</td>
</tr>
<tr>
<td></td>
<td>λ22  -0.314(2.547)[2.566]</td>
<td>-0.025(1.593)[1.593]</td>
<td>-0.029(1.439)[1.440]</td>
</tr>
<tr>
<td></td>
<td>β1   0.001(0.107)[0.107]</td>
<td>0.001(0.106)[0.106]</td>
<td>0.001(0.106)[0.106]</td>
</tr>
<tr>
<td></td>
<td>β2   -0.014(0.135)[0.135]</td>
<td>-0.013(0.135)[0.135]</td>
<td>-0.012(0.134)[0.135]</td>
</tr>
<tr>
<td>N=360</td>
<td>λ11  -0.090(1.083)[1.087]</td>
<td>-0.007(0.590)[0.590]</td>
<td>0.043(0.587)[0.589]</td>
</tr>
<tr>
<td></td>
<td>λ12  0.044(1.681)[1.681]</td>
<td>-0.084(0.912)[0.916]</td>
<td>-0.162(0.908)[0.922]</td>
</tr>
<tr>
<td></td>
<td>λ21  0.317(1.685)[1.715]</td>
<td>0.205(0.959)[0.981]</td>
<td>0.164(0.856)[0.872]</td>
</tr>
<tr>
<td></td>
<td>λ22  -0.335(2.601)[2.622]</td>
<td>-0.159(1.479)[1.488]</td>
<td>-0.226(1.799)[1.813]</td>
</tr>
<tr>
<td></td>
<td>β1   -0.007(0.073)[0.073]</td>
<td>-0.006(0.072)[0.072]</td>
<td>-0.006(0.072)[0.072]</td>
</tr>
<tr>
<td></td>
<td>β2   -0.004(0.097)[0.097]</td>
<td>-0.004(0.097)[0.097]</td>
<td>-0.003(0.096)[0.096]</td>
</tr>
<tr>
<td>N=720</td>
<td>λ11  -0.118(3.016)[3.018]</td>
<td>-0.027(0.543)[0.544]</td>
<td>-0.031(0.504)[0.505]</td>
</tr>
<tr>
<td></td>
<td>λ12  0.090(4.667)[4.667]</td>
<td>-0.053(0.842)[0.844]</td>
<td>-0.046(0.778)[0.779]</td>
</tr>
<tr>
<td></td>
<td>λ21  0.218(3.637)[3.643]</td>
<td>0.092(0.950)[0.955]</td>
<td>0.072(0.886)[0.889]</td>
</tr>
<tr>
<td></td>
<td>λ22  -0.183(5.605)[5.608]</td>
<td>0.013(1.460)[1.460]</td>
<td>0.042(1.362)[1.363]</td>
</tr>
<tr>
<td></td>
<td>β1   -0.008(0.055)[0.056]</td>
<td>-0.007(0.052)[0.052]</td>
<td>-0.007(0.052)[0.052]</td>
</tr>
<tr>
<td></td>
<td>β2   -0.001(0.068)[0.068]</td>
<td>-0.001(0.065)[0.065]</td>
<td>-0.000(0.065)[0.065]</td>
</tr>
</tbody>
</table>

From this table, we have several findings. First of all, the bias and the standard deviation generally become smaller as the sample size grows larger for the BGMMs and the EGMMs. On the other hand, the 2SLSs have large biases and standard deviations, and thus provide poor performances. Also notice that the bias and the standard deviation for the $\beta$’s are very small in every case. This happens because $X_1$ and $X_2$ are exogenous regressors.

Secondly, the BGMM and EGMM approaches yield estimates with smaller bias and smaller standard deviation compared with the 2SLS approach for almost every variable and every sample
size. In the few situations where this is not the case, the BGMME and EGMME still have the smaller root mean squared errors. These results show that the GMM approaches almost always outperform the 2SLS approach in estimating the SAR model in a system of interrelated networks. These findings are in line with the conclusions in the previous sections.

Finally, when we compare the BGMME and the EGMME, it seems that the EGMME generally outperforms the BGMME. In most cases the EGMME has a smaller standard deviation than the BGMME. Once again, the Monte Carlo simulations support the conclusion that the EGMME is more efficient than the BGMME. This suggests that we estimate the variance parameters together with the $\lambda$’s and $\beta$’s when error term is not i.i.d.

## Conclusion

In this paper, I introduce a very general spatial autoregressive model with multiple equations. This model can be used to analyze a wide range of empirical problems which involve the interactions between individuals in the same network and across different networks.

Just like in the high order SAR models, we cannot estimate this model by maximum likelihood for practical reasons. While previous research uses the 2SLS method, it does not possess desirable efficiency properties. Therefore, this paper aims at developing GMM estimation approaches in this framework that have the same efficiency property as MLE.

I extend the GMM method in the previous literature for the high order SAR model. This approach improves upon 2SLS method by incorporating quadratic moment conditions. However, this GMME is generally not as efficient as the MLE in this framework due to the complex structure of the disturbances although it is still the best estimator in the defined class.

Because of the discrepancy I look for a new class of estimators by modifying the structure of the quadratic moments and including new moments for the variance parameters. The resulting new GMME are asymptotically as efficient as the MLE under normality.

I also present some evidence from Monte Carlo experiments that the GMMEs outperform the 2SLSE in finite sample. Furthermore, the new EGMME generally has smaller standard deviation than the BGMME. Therefore, the simulation results support the theoretical conclusions in this paper.
Appendix (in progress)

Proof of Proposition 1. This proof is similar to the one in Lee and Liu (2010). The difference between this model and a high order SAR model lies in the structure of the error term. While Lee and Liu assume that the disturbances are i.i.d., I do not place any structural assumption on the disturbances. However, dropping the i.i.d. assumption does not affect the validity of the original proof. This is because the variance parameters are all included in $\Omega_N$, which is the variance matrix of the moment functions. The proof only involves operations on $\Omega_N$ rather than the variance parameters. Therefore, Proposition 1 still holds when we have a generalized error structure.

Proof of Proposition 2. Once again, the proof is similar to the one in Lee and Liu (2010). Since the original proof only involves operations on $\Omega_N$, we can directly use it in the new context. The generalized Schwartz inequality implies that the optimal weighting matrix for $a'Na_N$ in Proposition 1 is $(\frac{1}{N} \Omega_N)^{-1}$. The rest of the proof follows.

References


