Abstract

We consider the estimation of the spatial auto regressive (SAR) model where the disturbances are homoskedastic but not necessarily normally distributed. For estimation of such a model, we provide an analytic procedure to find and construct best linear and quadratic moments for the GMM estimation of the model. We discuss possible implications of the additional moments on aspects of model structures beyond those best linear-quadratic moments for the estimation of the SAR model with normal disturbances and possibly unknown heteroskedasticity.

1. Introduction

The spatial autoregressive (SAR) model is a popular econometric model for modeling spatial interaction and dependence in regional economics and social interactions of labor economics. The spatial autoregressive model has often been estimated by the method of maximum likelihood (ML) (Ord 1975). Related econometric estimation methods are the method of 2SLS (Kelejian and Prucha 1999; Lee 2003) and the generalized method of moments (GMM) (Lee 2007).

The ML estimator would be consistent and asymptotic normal. When the disturbances are normally distributed, the MLE would be asymptotically efficient. The quasi-maximum likelihood (QML) estimator might not be asymptotically efficient when the disturbances of the SAR model were not known to be normal, but the likelihood were formulated as if the disturbances were normally distributed. The ML estimate (MLE) or the quasi-maximum likelihood estimate (QMLE) would be characterized by its scores, which are statistics in linear and quadratic form. This motivates the use of linear and quadratic moments for estimation for
the SAR model in Lee (2007). While linear moments correspond to IV estimation, quadratic moments capture spatial correlation of observed dependent variables. When disturbances of the SAR model are normally distributed, the best linear and quadratic moments which can give asymptotically efficient GMM estimate can be motivated by the score vector of the likelihood function. However, for a quasi-likelihood, the linear and quadratic moments motivated by the score vector from the quasi-likelihood would not characterize an asymptotic efficient estimator. So it remains a research issue in the search for possible best linear and quadratic moments so that the derived GMME can be the best within the class of GMMEs derived from the use of linear and quadratic moments. For the estimation of the SAR model, best linear and quadratic moments exist as claimed in Liu et al. (2010). However, derivations of those best moments were subject to trial and error by utilizing a characterization of redundant moments in Breusch et al. (1999). One might wonder whether there is an analytical procedure to derive and construct best moments. This paper provides an analytic procedure on the construction of the best linear and quadratic moments. One can see also those derived best linear and quadratic moments can provide useful statistics for testing important aspects of model specifications of the SAR model.

2. The SAR model, QML, GMM, linear-quadratic form, and martingale

A spatial autoregressive (SAR) model has the specification that

\[ Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n, \quad (1) \]

where \( Y_n \) is an \( n \)-dimensional vector of dependent variables, \( X_n \) is an \( n \times k \) matrix of exogenous variables of rank \( k \), \( W_n \) is an \( n \times n \) matrix of specified spatial weights matrix, and \( \epsilon_n = (\epsilon_{n1}, \cdots, \epsilon_{nn})' \) is an \( n \)-dimensional vector of disturbances with i.i.d. \( (0, \sigma^2) \) random variables with finite four plus moments for all units.

For estimation and inference, statistics of linear-quadratic form characterize testing, such as the Moran’s I test of spatial correlation, and estimation, such as the 2SLS, ML, GMM, and generalized empirical likelihood (GEL) (Jin and Lee 2019). For testing, Moran’s I test is widely used to test the existence of spatial correlation. In a linear regression setting, the Moran I statistic is based on checking whether \( \epsilon_{ni} \) is correlated with \( w_{n,i'}.\epsilon_n \) that represent neighbors’ outcomes of \( i \), where \( w_{n,i} \) is the \( i \)th row of \( W_n \). The statistic \( \frac{1}{\sqrt{n}} \epsilon_n' W_n \epsilon_n \) involves a quadratic form. In addition to Moran’s test, quadratic forms are also useful for various LM test statistics,
e.g., LM tests for spatial correlations in Anselin and Bera (1998) for linear SAR models, and for spatial Tobit models in Qu and Lee (2012). A survey on the use of linear-quadratic form for estimation of linear spatial econometric models is in Xu and Lee (2019).

Here we shall focus on the GMM estimation of the linear SAR model. The construction of possible best GMM estimation is motivated by the ML estimation of this SAR model under normal disturbances. In the event the disturbances of the SAR model were not normally distributed, we are searching for the best linear-quadratic moments for an improved GMME against the QMLE.

2.1 The QML estimation

For estimation, consider the QML of the SAR model. Assume that \( \epsilon_i \)'s are i.i.d. \((0, \sigma^2)\) in the linear SAR model (1). Let \( \theta = (\lambda, \beta)' \) be the vector of coefficients of this SAR model. The quasi-log-likelihood function of the SAR model is

\[
\ln L_n(\theta) = -\frac{n}{2} \ln(2\pi \sigma^2) - \frac{1}{2\sigma^2}[S_n(\lambda)Y_n - X_n\beta]'[S_n(\lambda)Y_n - X_n\beta] + \ln |I_n - \lambda W_n|,
\]

where \( S_n(\lambda) = I_n - \lambda W_n \). Denote \( \epsilon(\theta) \equiv S_n(\lambda)Y_n - X_n\beta \). The derivatives of \( \ln L_n(\theta) \) are

\[
\frac{\partial \ln L_n(\theta)}{\partial \beta} = \frac{1}{\sigma^2} X_n' \epsilon(\theta),
\]

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

gand

\[
\frac{\partial \ln L_n(\theta)}{\partial \lambda} = \frac{1}{\sigma^2} [W_n S_n(\lambda)^{-1} X_n \beta]' \epsilon(\theta) + \frac{1}{\sigma^2} \epsilon(\theta)' [W_n S_n(\lambda)^{-1}]' \epsilon(\theta) - \text{tr}[S_n(\lambda)^{-1} W_n].
\]

All these three derivatives are linear-quadratic forms of \( \epsilon(\theta) \). Linear-quadratic forms of random variables play a fundamental role in establishing the large sample properties of estimators for linear spatial econometric models. Thus, Kelejian and Prucha (2001) derive a central limit theorem (CLT) for linear-quadratic forms of independent random variables that are suitable to study linear spatial econometric models. In addition to QMLE of SAR models, linear-quadratic forms are also indispensable when we study 2SLS estimation (Kelejian and Prucha, 1998) and GMM estimation (Kelejian and Prucha, 1999; Lee, 2007), the estimation of SAR panel data models (Yu, de Jong and Lee, 2008; Lee and Yu, 2010; Kuersteiner and Prucha, 2013), and various test statistics for spatial panel data models in Baltagi et. al. (2007).
2.2 GMM and OGMM and martingale

For estimation of the SAR model (1), in addition to the linear moments \( Q'_n \varepsilon_n(\theta) \), where
\[
\varepsilon_n(\theta) = S_n(\lambda) Y_n - X_n \beta \text{ with } \theta = (\lambda, \beta)',
\]
based on the IV matrix \( Q_n \), other moment equations can also be constructed for estimation.

Now consider constant \( n \times n \) matrices \( P_{1n}, \ldots, P_{mn} \) of them each has zero trace. Denote
\[
P_{1n} = \{ P : P \text{ is an } n \times n \text{ matrix, } \text{tr}(P) = 0 \} \text{ the class of nonstochastic } n \times n \text{ matrices with zero traces. Thus those } P_{jn} \text{'s are taken from } P_{1n}.
\]
The moment functions \( (P_{jn} \varepsilon_n(\theta))' \varepsilon_n(\theta) \) can be used in addition to \( Q'_n \varepsilon_n(\theta) \). These moment functions form a vector
\[
g_n(\theta) = (\varepsilon_n'(\theta) P_{1n} \varepsilon_n(\theta), \ldots, \varepsilon_n'(\theta) P_{mn} \varepsilon_n(\theta), \varepsilon_n'(\theta) Q_n')
\]
for the estimation in the GMM framework.

For any constant \( n \times n \) matrix \( P_n \) with \( \text{tr}(P_n) = 0 \), \( E(\varepsilon_n' P_n \varepsilon_n) = \sigma^2 \text{tr}(P_n) = 0 \). Thus, \( P_n \varepsilon_n \) is uncorrelated with \( \varepsilon_n \), i.e., \( E((P_n \varepsilon_n)' \varepsilon_n) = 0 \). This shows, in particular, that at true \( \theta_0 \), \( E(g_n(\theta_0)) = 0 \). Hence, \( g_n(\theta) \) consists of valid moment equations for estimation.

There is an intuition on selecting those moments. As \( W_n Y_n = G_n X_n \beta_0 + G_n \varepsilon_n \), where \( G_n = W_n S_n^{-1}, S_n = S_n(\lambda_0), \) and \( G_n \varepsilon_n \) is correlated with the disturbance \( \varepsilon_n \) in the model \( Y_n = \lambda W_n Y_n + X_n \beta + \varepsilon_n \), hence any \( P_{jn} \varepsilon_n \), which is uncorrelated with \( \varepsilon_n \), can be used as IV for \( W_n Y_n \) as long as \( P_{jn} \varepsilon_n \) and \( G_n \varepsilon_n \) are correlated. Another motivation is from the score vectors of the quasi-likelihood (QL), where the scores (2)-(4) consist of linear and quadratic moments even though those IVs and quadratic matrices are also functions of unknown parameters.

Let \( \mu_3 \) and \( \mu_4 \) be respectively the third and fourth order moments of \( \varepsilon_{ni} \). For a square matrix \( A \), we denote \( A^* = A + A' \), and \( d_A \) is the (column) vector constructed with diagonal elements of \( A \).

First and second moments of quadratic forms

Let \( \varepsilon_{ni} \)'s in \( \varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})' \) are i.i.d. \((0, \sigma^2)\) and have finite fourth moment \( \mu_4 \).

Lemma. Let \( A_n = [a_{ij}] \) be an \( n \)-dimensional square matrix. Then
1) \( E(\varepsilon_n' A_n \varepsilon_n) = \sigma^2 \text{tr}(A_n) \),
2) \( E(\varepsilon_n' A_n \varepsilon_n)^2 = (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4[\text{tr}^2(A_n) + \text{tr}(A_n A_n') + \text{tr}(A_n^2)], \) and
3) \( \text{var}(\varepsilon_n' A_n \varepsilon_n) = (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4[\text{tr}(A_n A_n') + \text{tr}(A_n^2)] = (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \frac{\sigma^4}{2} [\text{tr}(A_n + A_n')^2] \).
In particular, for normal disturbances, as $\mu_4 - 3\sigma^4 = 0$, the above expressions can be simplified.

Then, for any $n$-dimensional vector $Q_n$ and $n \times n$ square matrices $A_n$ and $B_n$, we have

$$E(Q'_n \epsilon_n \cdot \epsilon_n' A_n \epsilon_n) = \mu_3 Q'_n d_{A_n}$$

and

$$E(\epsilon'_n A_n \epsilon_n \cdot \epsilon_n' B_n \epsilon_n) = (\mu_4 - 3\sigma^4) d_{A_n}^{\prime} d_{B_n} + \sigma^4 [\text{tr}(A_n)\text{tr}(B_n) + \text{tr}(A_n (B_n + B'_n))].$$

Furthermore,

$$\text{cov}(\epsilon'_n A_n \epsilon_n \cdot \epsilon_n' B_n \epsilon_n) = (\mu_4 - 3\sigma^4) d_{A_n}^{\prime} d_{B_n} + \sigma^4 \text{tr}((A_n + A'_n)(B_n + B'_n)).$$

**Theorem** (Kelejian and Prucha, 2001)

Suppose elements of $\epsilon_n = (\epsilon_{n1}, \cdots, \epsilon_{nn})'$ are independent and satisfy $E\epsilon_{ni} = 0$. $A_n = (a_{n,ij})$ is a nonstochastic symmetric matrix and its column sum is uniformly bounded. $b_n = (b_{n1}, \cdots, b_{nn})'$ is a nonstochastic vector satisfying $\sup_i \frac{1}{n} \sum_{i=1}^{n} |b_{ni}|^{2+\eta_1}$ for some $\eta_1 > 0$. Suppose $\sup_i E|\epsilon_{ni}|^{4+\eta_2} < \infty$ for some $\eta_2 > 0$ holds. Denote $Q_n \equiv \epsilon'_n A_n \epsilon_n + b'_n \epsilon_n$ and $\sigma_{Q_n} \equiv [\text{var}(Q_n)]^{1/2}$. If $\sigma_{Q_n}^2 / n \geq c$ for some $c > 0$ for all $n$, then

$$\frac{Q_n - EQ_n}{\text{var}^{1/2}(Q_n)} \overset{d}{\rightarrow} N(0, 1).$$

The above theorem can be generalized to a CLT with a multivariate linear-quadratic form. Let $m$ different linear-quadratic forms be defined as $Q_{r,n} = \epsilon'_n A_{r,n} \epsilon + b_{r,n} \epsilon_n, r = 1, \cdots, m$. Let $V_n = [Q_{1,n}, \cdots, Q_{m,n}]'$ and $\Sigma_n = \text{var}(V_n)$. Under similar conditions as those in the CLT on linear-quadratic form,

$$\Sigma_n^{-1/2}(V_n - EV_n) \overset{d}{\rightarrow} N(0, I_m).$$

With this theorem, Kelejian and Prucha (2010) show the consistency and asymptotic normality of a GMM estimator for a SAR equation with spatial autoregressive disturbances (SARAR) with unknown heteroskedastic disturbances.

5
The proofs of the CLT on linear-quadratic form in Kelejian and Prucha (2010) rely on a CLT of martingale difference arrays (MDA). A general linear-quadratic form of independent random variables $\epsilon_{ni}$ with zero mean and variance $\sigma_{ni}^2$ can first be written in a single summation:

$$Q_n - EQ_n \equiv \epsilon_n' A_n \epsilon + b_n' \epsilon_n - \text{tr}(\epsilon_n' A_n \epsilon)$$

$$= \sum_{i=1}^{n} b_{ni} \epsilon_{ni} + \sum_{i=1}^{n} a_{n,ii} \epsilon_{ni}^2 + 2 \sum_{i=1}^{n-1} \sum_{j=1}^{i} a_{n,ij} \epsilon_{ni} \epsilon_{nj} - \sum_{i=1}^{n} a_{n,ii} \sigma_{ni}^2 \equiv \sum_{i=1}^{n} Z_{ni},$$

where $Z_{ni} \equiv b_{ni} \epsilon_{ni} + a_{n,ii} \epsilon_{ni}^2 + 2 \sum_{j=1}^{i-1} a_{n,ij} \epsilon_{ni} \epsilon_{nj} - a_{n,ii} \sigma_{ni}^2$. Define the sigma field $\mathcal{F}_{ni} \equiv \sigma(\epsilon_{n1}, \ldots, \epsilon_{ni})$. Since $E\epsilon_{ni} = 0$, $E\epsilon_{ni}^2 = \sigma_{ni}^2$, and $\epsilon_{ni}$'s are independent, $E(Z_{ni}|\mathcal{F}_{ni,i-1}) = 0$. Hence, $\{Z_{ni}\}$ is a MDA. The MDA CLT for statistics of linear-quadratic forms provides an important tool for linear spatial models. This MDA CLT is useful for univariate linear spatial models. It can be extended to models with multivariate SAR equations, and panel data with both cross section and time dimensions.

### 2.3 Asymptotic distribution of GMM estimators

Let $a_n$ be a sequence of constants with a full row rank greater than or equal to the number of unknown parameters $(k + 1)$. The GMM estimation is $\min_{\theta} g_n'(\theta)a_n' a_n g_n(\theta)$.

Asymptotic property of our estimators can be shown under some regularity conditions on disturbances, regressors, spatial weights matrices and quadratic matrices. Specifically, for the results in this section, we assume that $v_{ni}$ are i.i.d. with zero mean, variance $\sigma^2$ and that a moment of order higher than the fourth exists ($\mu_3$ and $\mu_4$ denote, respectively, its third and four moments); 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; 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; the matrices $P_{jn}$ from $\mathcal{P}_1$ are uniformly bounded in both row and column sums in absolute value; and elements of $Q_n$ are uniformly bounded. The assumptions on the uniform boundedness of $W_n$ and related matrices guarantee stability of the SAR model as we have noted earlier.

Under those regular conditions, 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 $\theta_0$ (Lee 2007). For the asymptotic distribution of $\hat{\theta}_n$, one investigates
the Taylor expansion of \( \frac{\partial g'_n(\hat{\theta}_n)}{\partial \theta} a'_n a_n g_n(\hat{\theta}_n) = 0 \) at \( \theta_0 \),

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) = -\left[ \frac{1}{n} \frac{\partial g'_n(\hat{\theta}_n)}{\partial \theta} a'_n a_n \right] \frac{1}{n} \frac{\partial g_n(\bar{\theta}_n)}{\partial \theta} \frac{1}{\sqrt{n}} a_n g_n(\theta_0).
\]

The asymptotic distribution of a GMM estimator will involve the variance matrix of linear-quadratic moments of the disturbance vector \( \epsilon_n \). For any square \( n \times n \) matrix \( A \), \( d_A = (a_{11}, \ldots, a_{nn})' \) denotes the column vector formed with the diagonal elements of \( A \). As shown earlier, \( E(\epsilon'_n P_n \epsilon_n) = \mu_3' Q'_n d_P \) and \( E(\epsilon'_n P_{jn} \epsilon_n) = (\mu_4 - 3\sigma_0^4) d'_P d_P + \sigma_0^4 \text{tr}(P_{jn} P_{tn}) \). It follows that \( \text{var}(g_n(\theta_0)) = \Omega_n \), where

\[
\Omega_n = \begin{pmatrix}
(\mu_4 - 3\sigma_0^4) \omega_{nm} \omega_{nm} & \mu_3 \omega_{nm} Q_n \\
\mu_3 Q'_n \omega_{nm} & 0
\end{pmatrix} + \Omega^0_n,
\]

with \( \omega_{nm} = (d_{P_{kn}}, \ldots, d_{P_{mn}})' \) and

\[
\Omega^0_n = \sigma_0^4 \begin{pmatrix}
\text{tr}(P_{1n} P_{1n}^s) & \ldots & \text{tr}(P_{1n} P_{mn}^s) & 0 \\
\vdots & \ddots & \vdots & \vdots \\
\text{tr}(P_{mn} P_{1n}^s) & \ldots & \text{tr}(P_{mn} P_{mn}^s) & 0 \\
0 & \ldots & 0 & \frac{1}{\sigma_0^4} Q'_n Q_n
\end{pmatrix} = \sigma_0^4 \begin{pmatrix}
\Delta_{mn} & 0 \\
0 & \frac{1}{\sigma_0^4} Q'_n Q_n
\end{pmatrix},
\]

where \( \Delta_{mn} = [\text{vec}(P_{1n}), \ldots, \text{vec}(P_{mn})]'[\text{vec}(P_{1n}), \ldots, \text{vec}(P_{mn})] \), because, for any conformable matrices \( A \) and \( B \), \( \text{tr}(AB) = \text{vec}(A') \text{vec}(B) \). When \( \epsilon_n \) is normally distributed, \( \Omega_n \) is simplified to \( \Omega^0_n \) because \( \mu_3 = 0 \) and \( \mu_4 = 3\sigma_0^4 \).

As

\[
\frac{\partial g_n(\theta)}{\partial \theta} = \begin{pmatrix}
\epsilon'_n(\theta) P_{1n}^s \\
\vdots \\
\epsilon'_n(\theta) P_{mn}^s \\
Q'_n
\end{pmatrix},
\]

\[
\frac{\partial g'_n(\theta)}{\partial \theta} = -\begin{pmatrix}
(\mu_4 - 3\sigma_0^4) Q'_n \omega_{nm} & \mu_3 Q'_n \omega_{nm} & 0 \\
\mu_3 Q'_n \omega_{nm} & 0 & \mu_3 Q'_n \omega_{nm} \\
0 & \mu_3 Q'_n \omega_{nm} & 0 \\
0 & 0 & \mu_3 Q'_n \omega_{nm}
\end{pmatrix},
\]

it follows \( \frac{1}{n} \frac{\partial g_n(\hat{\theta}_n)}{\partial \theta} = -\frac{1}{n} D_n + o_P(1) \), where

\[
D_n = \begin{pmatrix}
\sigma_0^2 \text{tr}(P_{1n} G_n) & 0 \\
\vdots & \ddots & \vdots \\
\sigma_0^2 \text{tr}(P_{mn} G_n) & 0 \\
Q'_n G_n X_n \beta_0 & Q'_n X_n
\end{pmatrix}.
\]

By the CLT for a linear-quadratic function,

\[
\frac{1}{\sqrt{n}} a_n g_n(\theta) = \frac{1}{\sqrt{n}} \left[ \epsilon'_n \left( \sum_{j=1}^m a_{nj} P_{jn} \right) \epsilon_n + a_{nx} Q'_n \epsilon_n \right] \xrightarrow{D} N(0, \lim_{n \to \infty} \frac{1}{n} a_n \Omega_n a_n' a_n). \]
Consequently,
\[ \sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \Sigma), \]
where
\[ \Sigma = \lim_{n \to \infty} \left[ \left( \frac{1}{n} D_n \right)^\prime a_n(\frac{1}{n} D_n) \right]^{-1} \left( \frac{1}{n} D_n \right)^\prime a_n(\frac{1}{n} \Omega_n) a_n \left( \frac{1}{n} D_n \right) \left[ \left( \frac{1}{n} D_n \right)^\prime a_n(\frac{1}{n} D_n) \right]^{-1}. \]

As in the usual case, an optimal selection of \( a_n \) will be \( a_n = \Omega_n^{-1/2} \). With an initial consistent estimated \( \hat{\Omega}_n \), we have a feasible optimum GMME (OGMME). 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 \( P_{jn} \)'s has the asymptotic distribution
\[ \sqrt{n}(\hat{\theta}_{o,n} - \theta_0) \xrightarrow{D} N(0, (\lim_{n \to \infty} \frac{1}{n} D_n^\prime \Omega_n^{-1} D_n)^{-1}). \]

The preceding theory on GMME is presented with arbitrary IVs \( Q_n \) and quadratic matrices \( P_{jn} \)'s with zero traces. The remaining issue is how to select those IV and quadratic matrices. In particular, we are interested in the selection of possible best matrices so that the corresponding OGMME can be asymptotically the best among GMMEs derived from those linear-quadratic moments.

3. The search for linear-quadratic moments for GMM estimation

We are considering estimation issues for the spatial autoregressive (SAR) model \( Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n \) in (1). Let \( \theta = (\lambda, \beta)^\prime \) be the vector of coefficients of this SAR model. The disturbances \( \epsilon_n \) are distributional fee, i.e., not assumed to be normally distributed but are homoskedastic with variance \( \sigma^2 \). We are interested in the GMM estimation based on linear and quadratic moments. We would like to derive analytically the best linear and quadratic moments for estimation. The existence of best linear and quadratic moments are shown in Lee (2007) for the model with normal disturbances, and their existence has been confirmed in Liu et al.(2010) even the disturbances are non-normal. However, the construction of the best linear and quadratic moments in Liu et al.(2010) is not analytic except the case with third moments of disturbances being zero. The existence and possible best linear and quadratic moments under non-normal disturbances and with nonzero third moment are confirmed in Liu et al.(2010), but their derivations are trial by error as their constructions are based on the characterization of redundant moments in Breusch et al.,(1999), i.e., any additional linear-quadratic moments would be redundant given a set of possible linear and quadratic moments. The later set of linear and quadratic moments would be a set of best moments.
The redundant moments issue studied in Breusch et al. (1999) is based on the comparison of limiting variances of GMME based on a whole set of moments and that of a proper subset of moments. Suppose that the whole set of moments \( g_n(\theta) \) can be divided into two subsets, namely, \( g_n(\theta) = (g_{n1}(\theta), g_{n2}(\theta))' \) where the subset \( g_{n1}(\theta) \) can identity \( \theta_0 \). We might estimate \( \theta_0 \) by the OGMM with \( g_n \) as well as only by \( g_{n1} \). The OGMM min\( \theta \) \( g_n'(\theta)\Omega^{-1}g_n(\theta) \), where \( \Omega \) is the (limiting) variance of \( \sqrt{n}g_n(\theta_0) \), gives the OGMME \( \hat{\theta} \). On the other hand, the OGMME based on \( g_{n1} \) is \( \hat{\theta}_1 \) from min\( \theta \) \( g_{n1}'(\theta)\Omega_{11}^{-1}g_{n1}(\theta) \), where \( \Omega_{11} \) is the (limiting) variance of \( \sqrt{n}g_{n1}(\theta_0) \).

The limiting variances of \( \hat{\theta} \) and \( \hat{\theta}_1 \) will be respectively \( (D'\Omega^{-1}D)^{-1} \) and \( (D_1'\Omega_{11}^{-1}D_1)^{-1} \), where

\[
\text{plim}_{n \to \infty} \frac{\partial g_n^*(\theta_0)}{\partial \theta} = D = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix} = \text{plim}_{n \to \infty} \left( \begin{pmatrix} \frac{\partial g_{n1}^*(\theta_0)}{\partial \theta} \\ \frac{\partial g_{n2}^*(\theta_0)}{\partial \theta} \end{pmatrix} \right).
\]

By the quadratic partition (see, e.g., Ruud 2000), one has

\[
D'\Omega^{-1}D = D_1'\Omega_{11}^{-1}D_1 + (D_2 - \Omega_{21}\Omega_{11}^{-1}D_1)'(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}(D_2 - \Omega_{21}\Omega_{11}^{-1}D_1).
\]

Therefore, the redundancy of \( g_{n2} \) given \( g_{n1} \) is characterized by \( (D_2 - \Omega_{21}\Omega_{11}^{-1}D_1)'(\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1}(D_2 - \Omega_{21}\Omega_{11}^{-1}D_1) = 0 \), and hence, equivalently

\[
(D_2 - \Omega_{21}\Omega_{11}^{-1}D_1) = 0.
\]

Thus, in the search for the best possible linear-quadratic moments based on this characterization, one looks for a specific set of linear-quadratic moments \( g_{n1} \) with their specific \( \Omega_{11} \) such that any additional linear-quadratic moments \( g_{n2} \) with their corresponding covariance matrix \( \Omega_{21} \) with \( g_{n1} \) satisfying the redundancy condition \( (D_2 - \Omega_{21}\Omega_{11}^{-1}D_1) = 0 \). Apparently, such a search of the best would be trial by error.

The complication for the search for the best is due to the variance \( \Omega \) of any set of linear-quadratic moments in the presence of the excess kurtosis \( (\mu_3 - 3\sigma_0^2) \) and \( \mu_3 \) not being zero. For the relatively special case with the third moment \( \mu_3 = 0 \), an analytical derivation of the best linear and quadratic moments is possible as shown in Liu at al. (2010). The idea is to explore the generalized Schwartz inequality to obtain a possible sharp lower bound of variance matrices with the hope of possible construction of relevant linear-quadratic moments with which their GMME variance might attain the lower bound. However, for the general situation with possible non-zero third moment, the analysis were not available. In this paper, eventually
we are able to derive analytically and constructively the best linear and quadratic moments for the estimation of the SAR model in the general case.

We consider the following general framework of GMM estimation. A GMM with $m$ quadratic matrices $P_{jn}$, $j = 1, \ldots, m$ with zero diagonal, and $d$ diagonal matrices $A_{jn}$, $j = 1, \ldots, d$ with zero trace, and IV matrix $Q_n$ with a finite number of IV variables. The linear-quadratic moment vector can be

$$g_n(\theta) = (\epsilon'_n(\theta)P_{1n}\epsilon_n(\theta), \ldots, \epsilon'_n(\theta)P_{mn}\epsilon_n(\theta), \epsilon'_n(\theta)A_{1n}\epsilon_n(\theta), \ldots, \epsilon'_n(\theta)A_{dn}\epsilon_n(\theta), \epsilon'_n(\theta)Q_n)'$$

where $\epsilon_n(\theta) = (I_n - \lambda W_n)Y_n - X_n\beta$. At the true parameter vector $\theta_0$, $\epsilon_n = \epsilon_n(\theta_0)$. For the statistics with a quadratic form, $E(\epsilon'_nP_{jn}\epsilon_n) = \sigma_0^2tr(P_{jn}) = 0$ as $P_{jn}$ has a zero diagonal. For the diagonal matrix with its trace zero, $E(\epsilon'_nA_{jn}\epsilon_n) = \sigma_0^2tr(A_{jn}) = 0$ because $A_{jn}$ has the trace on its diagonal being zero. Indeed, we might start with quadratic matrices with trace zero to begin with. However, it is desirable for technical reasons to consider quadratic matrices with zero diagonal and also diagonal matrices with zero trace. Later on, we may see the advantage of using those matrices.

At the true parameter vector $\theta_0$,

$$\text{var}(g_n(\theta_0)) = \tilde{\Omega}_n = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & (\mu_{40} - 3\sigma_0^4)\omega_{nd}\omega_{nd} & \mu_{30}\omega_{nd}'Q_n & 0 \\ 0 & \mu_{30}Q_n\omega_{nd} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_0^4}Q_n'Q_n \end{pmatrix} + \begin{pmatrix} \Delta_{mn} & 0 & 0 \\ 0 & \Delta_{dn} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where

$$\tilde{\Delta}_{mn} = [\text{vec}(P^n_{1n}), \ldots, \text{vec}(P^n_{mn})]'[\text{vec}(P_{1n}), \ldots, \text{vec}(P_{mn})],$$

$$\omega_{nd} = (d_{A_{1n}}, \ldots, d_{A_{dn}}),$$

and

$$\Delta_{dn} = [\text{vec}(A^n_{1n}), \ldots, \text{vec}(A^n_{dn})]'[\text{vec}(A_{1n}), \ldots, \text{vec}(A_{dn})].$$

For the SAR model, let $G_n = W_nS^{-1}_n$. The variance-covariance matrix of a finite number of linear and quadratic moments with zero trace has been obtained in, e.g., Lee (2007), and shown earlier. The preceding quadratic moments with their quadratic matrices having zero diagonal and additional diagonal matrices with zero trace are specific quadratic moments. The reason for the use of quadratic moments with zero diagonal is that those moments at the true parameters are uncorrelated with linear IV moments as well as quadratic statistics.
with diagonal matrices with zero trace. For the optimum GMM estimator of $\theta_0$ with those moments, its asymptotic variance is $(\bar{D}_n'\bar{\Omega}_n^{-1}\bar{D}_n)^{-1}$, where

$$\bar{D}_n = \begin{pmatrix}
\sigma_0^2 tr(P_{1n}^s G_n) & 0 \\
\vdots & \vdots \\
\sigma_0^2 tr(P_{mn}^s G_n) & 0 \\
\sigma_0^2 tr(A_{1n}^s G_n) & 0 \\
\vdots & \vdots \\
\sigma_0^2 tr(A_{dn}^s G_n) & 0 \\
Q_n'(G_nX_n\beta_0') & Q_n'X_n
\end{pmatrix} = \begin{pmatrix}
\sigma_0^2 [vec(P_{1n}^s), \ldots, vec(P_{mn}^s)]'vec(G_n) & 0 \\
\sigma_0^2 [vec(A_{1n}^s), \ldots, vec(A_{dn}^s)]'vec(G_n) & 0 \\
Q_n'(G_nX_n\beta_0') & Q_n'X_n
\end{pmatrix}.$$  

Let $B_n = [tr(P_{1n}^s G_n), \ldots, tr(P_{mn}^s G_n)]' = [vec(P_{1n}^s), \ldots, vec(P_{mn}^s)]'vec(G_n)$, and

$C_n = [tr(A_{1n}^s G_n), \ldots, tr(A_{dn}^s G_n)]' = [vec(A_{1n}^s), \ldots, vec(A_{dn}^s)]'vec(G_n)$.

Then $\bar{D}_n = \begin{pmatrix}
\sigma_0^2 B_n & 0 \\
\sigma_0^2 C_n & 0 \\
Q_n'(G_nX_n\beta_0') & Q_n'X_n
\end{pmatrix}$.

With these $\bar{D}_n$ and $\bar{\Omega}_n$, it follows that

$$\bar{D}_n'\bar{\Omega}_n^{-1}\bar{D}_n = \begin{pmatrix}
B_n'\bar{\Delta}^{-1}_{mn}B_n & 0 \\
0 & 0
\end{pmatrix} + D_n'\Omega_n^{-1}D_n,$$

where

$$D_n = \begin{pmatrix}
\sigma_0^2 C_n \\
Q_n'(G_nX_n\beta_0') \\
Q_n'X_n
\end{pmatrix}$$

and

$$\Omega_n = \begin{pmatrix}
(\mu_4 - 3\sigma_0^4)/\omega_n'\omega_n & \mu_3\omega_n'Q_n & 0 \\
\mu_3\omega_n'\omega_n & \mu_3Q_n'\omega_n & 0 \\
0 & 0 & \sigma_0^4 Q_n'Q_n
\end{pmatrix}.$$  

From this result, we see that the best selection of $P'_{jn}$s would be separately from those selections of $A_{jn}$ and $Q_n$ because the optimization of $P'_{jn}$’s would be just for $B_n'\bar{\Delta}^{-1}_{mn}B_n$ but the optimal selection of $A_{jn}$’s and $Q_n$ would be from the optimization of $D_n'\Omega_n^{-1}D_n$.

In the GMM estimation of SAR models in the literature, one usually starts with a finite number of quadratic moments with quadratic matrices having zero trace and with a finite number of linear (IV) moments. The preceding GMM estimation with its empirical quadratic moments having zero diagonal and additional diagonal matrices with zero trace are uncorrelated with linear IV statistics as well as those quadratic statistics with diagonal matrices with zero trace. Those properties can render the analytical derivation of the best moments easier to
deal with. This GMM setting would not lose its generality as a quadratic moment with a zero trace quadratic matrix can be spitted out to two quadratic moments with one quadratic matrix having zero diagonal and the other being a diagonal matrix with zero trace. The optimum pooling of these two quadratic moments for estimation will result in GMM estimates asymptotically efficient relative to those of the single quadratic moment with zero trace quadratic matrix.

3.1 Best selection of quadratic matrices with zero diagonal

The selection of best quadratic matrix $P_n$ with zero diagonal is simple as it follows directly from the Schwartz inequality as in Lee (2007). The $B_n$ can be rewritten as

$$B_n = \begin{pmatrix} \text{tr}(P_{s1n}^s G_n) \\ \vdots \\ \text{tr}(P_{mn}^s G_n) \end{pmatrix} = \begin{pmatrix} \text{tr}(P_{s1n}^s (G_n - \text{Diag}(G_n))) \\ \vdots \\ \text{tr}(P_{mn}^s (G_n - \text{Diag}(G_n))) \end{pmatrix},$$

where $\text{Diag}(G_n)$ denotes a diagonal matrix formed by the diagonal elements of $G_n$. This follows because $P_{jn}$’s have zero diagonals. The variance component $\bar{\Delta}_{mn}$ can be rewritten in a symmetric form with

$$\bar{\Delta}_{mn} = \begin{pmatrix} \text{tr}(P_{s1n}^s P_{1n}) & \cdots & \text{tr}(P_{s1n}^s P_{mn}) \\ \vdots & \ddots & \vdots \\ \text{tr}(P_{mn}^s P_{1n}) & \cdots & \text{tr}(P_{mn}^s P_{mn}) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \text{tr}(P_{s1n}^s P_{1n}) & \cdots & \text{tr}(P_{s1n}^s P_{mn}) \\ \vdots & \ddots & \vdots \\ \text{tr}(P_{mn}^s P_{1n}) & \cdots & \text{tr}(P_{mn}^s P_{mn}) \end{pmatrix},$$

because $\text{tr}(P_{jn}^s P_{kn}^t) = \text{tr}(P_{jn}^t P_{kn}^s)$. Hence, by the Schwartz inequality

$$B_n^{-1} \bar{\Delta}_{mn}^{-1} B_n = 2 \text{vec}(G_n - \text{Diag}(G_n))' [\text{vec}(P_{s1n}^s), \cdots, \text{vec}(P_{mn}^s)]$$

$$\cdot (\text{vec}(P_{s1n}^s), \cdots, \text{vec}(P_{mn}^s))' [\text{vec}(P_{s1n}^s), \cdots, \text{vec}(P_{mn}^s)]^{-1}$$

$$\cdot [\text{vec}(P_{s1n}^s), \cdots, \text{vec}(P_{mn}^s)]' \text{vec}(G_n - \text{Diag}(G_n))$$

$$\leq 2 \text{vec}(G_n - \text{Diag}(G_n))' \text{vec}(G_n - \text{Diag}(G_n)).$$

The equality can be attained with the single optimal quadratic matrix $P_n^* = G_n - \text{Diag}(G_n)$, which has a zero diagonal.

3.2 Variance bound of GMME with linear-quadratic moments with zero trace diagonal matrices
The selection of possible optimal diagonal matrix \( A_n \) and IV matrix is more demanding for the general case without normality.

A well known formulae for ‘quadratic partition’ is the one below (find in e.g., Ruud 2000). Let \( z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \) be a partitioned vector and \( \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \) be conformable partition of a positive definite matrix with the same dimension as \( z \). Then

\[
\begin{align*}
z' \Omega^{-1} z &= (z_1', z_2') \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\
&= (z_1 - \Omega_{12} \Omega_{22}^{-1} z_2)'(\Omega_{11} - \Omega_{12} \Omega_{22}^{-1} \Omega_{21})^{-1} (z_1 - \Omega_{12} \Omega_{22}^{-1} z_2) + z_2' \Omega_{22}^{-1} z_2.
\end{align*}
\]

It follows that, by the quadratic partition,

\[
D_n' \Omega_n^{-1} D_n = \left( \frac{\sigma_0^2 C_n'}{\sigma_0} \right) - \frac{\mu_3}{\sigma_0^3} \left( \begin{pmatrix} G_n X_n \beta \\ X_n' \end{pmatrix} \right) Q_n (Q_n' Q_n)^{-1} Q_n' \omega_{nd}
\]

\[
\cdot H_n^{-1} \left( \frac{(\sigma_0^2 C_n, 0)}{\sigma_0} \right) - \frac{\mu_3}{\sigma_0^3} \omega_{nd} Q_n (Q_n' Q_n)^{-1} Q_n' (G_n X_n \beta, X_n)
\]

\[
+ \frac{1}{\sigma_0^2} (G_n X_n \beta, X_n)' Q_n (Q_n' Q_n)^{-1} Q_n' (G_n X_n \beta, X_n),
\]

where

\[
H_n = (\mu_4 - 3\sigma_0^4) \omega_{nd} \omega_{nd} + \sigma_0^4 \Delta_{dn} - \left( \frac{\mu_3^2}{\sigma_0^3} \right) \omega_{nd} Q_n (Q_n' Q_n)^{-1} Q_n' \omega_{nd}.
\]

For a diagonal matrix \( A_n \) with trace zero, it can be represented as \( A_n = \text{Diag}(a_n) \), where \( a_n = (a_{n1}, \ldots, a_{nn})' \) with \( \sum_{i=1}^n a_{ni} = 0 \). We have for this case that \( tr(A_n^2) = a_n' a_n \), which is \( n \) times the empirical variance of \( a_n \). Then, for a single diagonal moment case with a diagonal matrix \( A_n \), it corresponds to \( d = 1 \). For this single moment case, \( \omega_{nd} = a_n \) and \( \Delta_{dn} = \text{vec}'(A_n')\text{vec}(A_n) = tr(A_n' A_n) = 2a_n' a_n \). Hence, \( (\mu_4 - 3\sigma_0^4) \omega_{nd} \omega_{nd} + \sigma_0^4 \Delta_{dn} = (\mu_4 - \sigma_0^4) a_n' a_n = \sigma_0^4 (\eta_4 - 1) a_n' a_n \), where \( \eta_4 = \mu_4/\sigma_0^4 \). Therefore,

\[
\Omega_n = \sigma_0^4 \left( \frac{(\eta_4 - 1)a_n' a_n}{\mu_3 \frac{Q_n'}{\sigma_0} a_n} \frac{\mu_3^2 Q_n}{\sigma_0^3} a_n \right) \quad \text{and} \quad D_n = \begin{pmatrix} 2\sigma_0^2 a_n' dG_n & 0 \\ Q_n' (G_n X_n \beta) & Q_n' X_n \end{pmatrix},
\]

and

\[
H_n = \sigma_0^4 a_n' [(\eta_4 - 1)I_n - \left( \frac{\mu_3}{\sigma_0^3} \right)^2 Q_n (Q_n' Q_n)^{-1} Q_n'] a_n
\]

\[
= \sigma_0^4 a_n' [(\eta_4 - 1)(I_n - Q_n (Q_n' Q_n)^{-1} Q_n') + (\eta_4 - 1 - \left( \frac{\mu_3}{\sigma_0^3} \right)^2) Q_n (Q_n' Q_n)^{-1} Q_n'] a_n
\]

\[
= \sigma_0^4 a_n' h_n h_n a_n,
\]

where

\[
h_n = (\eta_4 - 1)^{\frac{1}{2}} (I_n - Q_n (Q_n' Q_n)^{-1} Q_n') + (\eta_4 - 1 - \left( \frac{\mu_3}{\sigma_0^3} \right)^2)^{\frac{1}{2}} Q_n (Q_n' Q_n)^{-1} Q_n'.
\]
Note that \((\eta_4 - 1)\) and \((\eta_4 - 1 - \left(\frac{\mu_3}{\sigma^3}\right)^2)\) are positive; therefore \(h_n\) is invertible and its inverse is simply

\[
h_n^{-1} = (\eta_4 - 1)^{-\frac{1}{2}}(I_n - Q_n(Q_n'Q_n)^{-1}Q_n') + (\eta_4 - 1 - \left(\frac{\mu_3}{\sigma^3}\right)^2)^{-\frac{1}{2}}Q_n(Q_n'Q_n)^{-1}Q_n'.
\]

Furthermore, because \(\text{tr}(A_n^aG_n) = \text{tr}(A_n^a\tilde{G}_n)\) with \(\tilde{G}_n = G_n - \frac{\text{tr}(G_n)}{n}I_n\),

\[
X_n'Q_n(Q_n'Q_n)^{-1}Q'_nd_{A_n} = X_n'Q_n(Q_n'Q_n)^{-1}Q'_nh_n^{-1}h_na_n
\]

and

\[
\sigma^2\text{tr}(A_n^aG_n) - \frac{\mu_3}{\sigma^2}(G_nX_n\beta)'Q_n(Q_n'Q_n)^{-1}Q'_n d_{A_n}
\]

\[
= 2\sigma^2d_{\tilde{G}_n}a_n - \frac{\mu_3}{\sigma^2}(G_nX_n\beta)'Q_n(Q_n'Q_n)^{-1}Q'_n a_n
\]

\[
= k_nh_n a_n
\]

because \(A_n\) has a zero trace, where

\[
k_n = \left(2\sigma^2d_{\tilde{G}_n}^2 - \frac{\mu_3}{\sigma^2}(G_nX_n\beta)'Q_n(Q_n'Q_n)^{-1}Q'_n\right)h_n^{-1}
\]

\[
= 2\sigma^2d_{\tilde{G}_n}h_n^{-1} - \frac{\mu_3}{\sigma^2}(\eta_4 - 1 - \left(\frac{\mu_3}{\sigma^3}\right)^2)^{-\frac{1}{2}}(G_nX_n\beta)'Q_n(Q_n'Q_n)^{-1}Q'_n
\]

\[
= 2\sigma^2d_{\tilde{G}_n}h_n^{-1} - \frac{\mu_3}{\sigma^2}(\eta_4 - 1 - \left(\frac{\mu_3}{\sigma^3}\right)^2)^{-\frac{1}{2}}(G_nX_n\beta)'Q_n(Q_n'Q_n)^{-1}Q'_n,
\]

with

\[
d_{\tilde{G}_n}h_n^{-1} = [(\eta_4 - 1)^{-\frac{1}{2}}d_{\tilde{G}_n}(I_n - Q_n(Q_n'Q_n)^{-1}Q_n') + (\eta_4 - 1 - \left(\frac{\mu_3}{\sigma^3}\right)^2)^{-\frac{1}{2}}d_{\tilde{G}_n}Q_n(Q_n'Q_n)^{-1}Q_n']
\]

It follows that

\[
D_n'Q_n^{-1}D_n = \frac{1}{\sigma^4} \begin{pmatrix} k_n \\ (h_n a_n) (h_n a_n)' \\ k_n \\ (\sigma^2)^{-1} (G_nX_n\beta, X_n)'Q_n(Q_n'Q_n)^{-1}Q_n (G_nX_n\beta, X_n) \end{pmatrix}
\]

\[
\cdot \begin{pmatrix} \sigma^2(X_n'Q_n(Q_n'Q_n)^{-1}Q_n')^{-1}Q_n' \\ \sigma^2(X_n'Q_n(Q_n'Q_n)^{-1}Q_n')^{-1}Q_n' \\ \sigma^2(X_n'Q_n(Q_n'Q_n)^{-1}Q_n')^{-1}Q_n' \end{pmatrix}
\]

We note that some terms in (5) can be replaced appropriately under the design that the constant vector \(l_n\) is a component in \(Q_n\). These follow because \(Q_n(Q_n'Q_n)^{-1}Q_n'l_n = l_n\) and, hence,

\[
(h_n a_n)' l_n = a_n' h_n' l_n = (\eta_4 - 1 - \left(\frac{\mu_3}{\sigma^3}\right)^2)^{-\frac{1}{2}} a_n' Q_n(Q_n'Q_n)^{-1}Q_n' l_n = (\eta_4 - 1 - \left(\frac{\mu_3}{\sigma^3}\right)^2)^{-\frac{1}{2}} a_n' l_n = 0,
\]

14
as \( \sum_{i=1}^{n} a_{ni} = 0 \). Therefore,
\[
(h_n a_n)' Q_n (Q_n' Q_n)^{-1} Q_n' X_n = (h_n a_n)' Q_n (Q_n' Q_n)^{-1} Q_n' \tilde{X}_n, \tag{6}
\]
and
\[
(h_n a_n)' Q_n (Q_n' Q_n)^{-1} Q_n' (G_n X_n \beta) = (h_n a_n)' Q_n (Q_n' Q_n)^{-1} Q_n' G_n \tilde{X}_n \beta, \tag{7}
\]
where each column of \( \tilde{X}_n \) is the corresponding regressors in \( X_n \) deviated from its sample mean, and so is \( \tilde{c}_n \) for a column vector \( c_n \).

In order to obtain an upper bound with the Schwartz inequality, we would face a difficulty due to the term \( d' \tilde{G}_n (I_n - Q_n (Q_n' Q_n)^{-1} Q_n' \) ). However, if \( d \tilde{G}_n \) is already a component in \( Q_n \), then \( d' \tilde{G}_n (I_n - Q_n (Q_n' Q_n)^{-1} Q_n' \) ) = 0. So a strategy for the searching for the best linear-quadratic moment might start with \( d \tilde{G}_n \) as a component in \( Q_n \) (or more general, \( d \tilde{G}_n \) can be a linear combination of columns of \( Q_n \)). While one uses IVs \( Q_n \) for estimation, one may extend \( Q_n \) to include \( d \tilde{G}_n \) as an additional IV. With the extended list of IVs, the asymptotic variance of the corresponding GMM estimate would not become larger but might be improved. Assuming \( d \tilde{G}_n \) is a component of \( Q_n \), then \( (I_n - Q_n (Q_n' Q_n)^{-1} Q_n' d \tilde{G}_n = 0 \). Hence
\[
d' \tilde{G}_n h_n^{-1} = (\eta_4 - 1 - (\frac{\mu_3}{\sigma^3})^2)^{-\frac{1}{2}} d' \tilde{G}_n,
\]
and
\[
k_n = (\eta_4 - 1 - (\frac{\mu_3}{\sigma^3})^2)^{-\frac{1}{2}} \left[ 2 \sigma^2 d' \tilde{G}_n - \frac{\mu_3}{\sigma^2} G_n X_n \beta \right] Q_n (Q_n' Q_n)^{-1} Q_n'.
\]
With (7), it follows that \( (h_n a_n)' k_n' = (h_n a_n)' \tilde{k}_n' \), where
\[
\tilde{k}_n = (\eta_4 - 1 - (\frac{\mu_3}{\sigma^3})^2)^{-\frac{1}{2}} \left[ 2 \sigma^2 d' \tilde{G}_n - \frac{\mu_3}{\sigma^2} G_n X_n \beta \right] Q_n (Q_n' Q_n)^{-1} Q_n'.
\]
Thus, under such circumstances, (5) can be written as
\[
D_n' \Omega_n^{-1} D_n = \frac{1}{\sigma^4} \left( (-\frac{\mu_3}{\sigma^2}) (\eta_4 - 1 - (\frac{\mu_3}{\sigma^3})^2)^{-\frac{1}{2}} \tilde{k}_n' Q_n (Q_n' Q_n)^{-1} Q_n' \right) \cdot (h_n a_n)' (h_n a_n)' (h_n a_n)' \cdot \left( (-\frac{\mu_3}{\sigma^2}) (\eta_4 - 1 - (\frac{\mu_3}{\sigma^3})^2)^{-\frac{1}{2}} \tilde{k}_n' Q_n (Q_n' Q_n)^{-1} Q_n' \right)'
+ \frac{1}{\sigma^2} (G_n X_n \beta, X_n)' Q_n (Q_n' Q_n)^{-1} Q_n' (G_n X_n \beta, X_n).
\]
It follows from the Schwartz inequality that
\[
D_n' \Omega_n^{-1} D_n \leq \frac{1}{\sigma^4} \left( (\frac{\mu_3}{\sigma^2}) (\eta_4 - 1 - (\frac{\mu_3}{\sigma^2})^2)^{-\frac{1}{2}} \tilde{k}_n Q_n (Q_n' Q_n)^{-1} Q_n' \right) \\
\cdot \left( (\frac{\mu_3}{\sigma^2})(\eta_4 - 1 - (\frac{\mu_3}{\sigma^2})^2)^{-\frac{1}{2}} \tilde{k}_n Q_n (Q_n' Q_n)^{-1} Q_n' \right)' \\
+ \frac{1}{\sigma^2} (G_n X_n \beta, X_n)' Q_n (Q_n' Q_n)^{-1} Q_n' (G_n X_n \beta, X_n) \\
= \frac{1}{\sigma^4} (\eta_4 - 1 - (\frac{\mu_3}{\sigma^2})^2)^{-\frac{1}{2}} \left( (2\sigma^2 d_{G_n}, 0) - \frac{\mu_3}{\sigma^2} (G_n \tilde{X}_n \beta, \tilde{X}_n) \right)' Q_n (Q_n' Q_n)^{-1} Q_n' \\
\cdot \left( (2\sigma^2 d_{G_n}, 0) - \frac{\mu_3}{\sigma^2} (G_n \tilde{X}_n \beta, \tilde{X}_n) \right) \\
+ \frac{1}{\sigma^2} (G_n X_n \beta, X_n)' Q_n (Q_n' Q_n)^{-1} Q_n' (G_n X_n \beta, X_n) \\
\leq \frac{1}{\sigma^4} (\eta_4 - 1 - (\frac{\mu_3}{\sigma^2})^2)^{-\frac{1}{2}} ( (2\sigma^2 d_{G_n}, 0) - \frac{\mu_3}{\sigma^2} (G_n \tilde{X}_n \beta, \tilde{X}_n) )' \\
\cdot \left( (2\sigma^2 d_{G_n}, 0) - \frac{\mu_3}{\sigma^2} (G_n \tilde{X}_n \beta, \tilde{X}_n) \right) + \frac{1}{\sigma^2} (G_n X_n \beta, X_n)'(G_n X_n \beta, X_n).
\]

The previous deviation is simplified with $d = 1$ but it can be applied to the general case with a finite number $d$ quadratic moments with diagonal matrices with zero trace. With a finite $d$, the $a_n$ would simply be a matrix of dimension $n \times d$ in the preceding analysis. Thus the upper bound of (8) provides an upper precision bound of those GMMEs.

The subsequent section provides the optimum design on $a_n$ and $Q_n$ to attain the upper precision bound.

### 3.3 Designs for optimal $a_n$ and $Q_n$:

We see that in the preceding derivation of the upper precision bound, the two projectors $Q_n (Q_n' Q_n)^{-1} Q_n'$ and $(h_n a_n)[(h_n a_n)'(h_n a_n)]^{-1}(h_n a_n)'$ need to coordinate with each other. As $a_n$ has its column sums to be zero but columns of $Q_n$ might not, it is desirable to extend $Q_n$ to include $l_n$ as an column (i.e., an IV) to fulfill the need that columns of $a_n$ can be formed as linear combinations of columns of $Q_n$ so that $a_n$ lies in the column space of $Q_n$.

With such features for $a_n$ and $Q_n$, we see $Q_n (Q_n' Q_n)^{-1} Q_n' a_n = a_n$ and, in consequence, $(I_n - Q_n (Q_n' Q_n)^{-1} Q_n') a_n = 0$ and $h_n a_n = (\eta_4 - 1 - (\frac{\mu_3}{\sigma^2})^2)^{\frac{1}{2}} a_n$. It follows then that
\[
k_n(h_n a_n) = [2\sigma^2 d_{G_n} h_n^{-1} - \frac{\mu_3}{\sigma^2} (\eta_4 - 1 - (\frac{\mu_3}{\sigma^2})^2)^{-\frac{1}{2}} (G_n X_n \beta)' Q_n (Q_n' Q_n)^{-1} Q_n' h_n a_n \\
= 2\sigma^2 d_{G_n} a_n - \frac{\mu_3}{\sigma^2} (G_n X_n \beta)' Q_n (Q_n' Q_n)^{-1} Q_n' a_n \\
= [2\sigma^2 d_{G_n} - \frac{\mu_3}{\sigma^2} (G_n X_n \beta)'] a_n \\
= [2\sigma^2 d_{G_n} - \frac{\mu_3}{\sigma^2} (G_n X_n \beta)] a_n.
\]
The preceding considerations motivate the following best selection of linear and quadratic moments.

1) The optimum \( a_n^* \) is \( a_n^* = \left[ d_{\tilde{G}_n}, G_n \bar{X}_n \beta, \bar{X}_n, \ldots, \bar{X}_{nk} \right] \), where \( \bar{X}_n = [l_n, X_{n2}, \ldots, X_{nk}] \).

2) The optimum \( Q_n \) can be \( Q_n^* = [G_n X_n \beta, X_n, d_{G_n}] \).

With those \( a_n^* \) and \( Q_n^* \), we have \( Q_n^*(Q_n^* Q_n^*)_1 Q_n^* a_n^* = a_n^* \), because \( d_{\tilde{G}_n} = d_{G_n} - \frac{tr G_n}{n} l_n \) and \( G_n \bar{X}_n \beta = G_n X_n \beta - \frac{1}{n} (l_n^* G_n X_n \beta) l_n \). With the corresponding optimum \( \tilde{h}_n \)

\[
h_n^* = (\eta_4 - 1)^{\frac{1}{2}} (I_n - Q_n^*(Q_n^* Q_n^*)_1 Q_n^*) + (\eta_4 - 1) - \left( \frac{\mu_3}{\sigma^3} \right)^2 \frac{1}{2} Q_n^*(Q_n^* Q_n^*)_1 Q_n^*,
\]

we have

\[
(h_n^* a_n^*)^\prime (h_n^* a_n^*)^{-1} (h_n^* a_n^*)^\prime = a_n^* (a_n^* a_n^*)^{-1} a_n^*.
\]

Furthermore, as

\[
k_n^* = (\eta_4 - 1 - \left( \frac{\mu_3}{\sigma^3} \right)^2)^{-\frac{1}{2}} [2 \sigma^2 d_{G_n}^\prime - \frac{\mu_3}{\sigma^3} (G_n \bar{X}_n \beta)^\prime Q_n^*(Q_n^* Q_n^*)_1 Q_n^*],
\]

we have

\[
(h_n^* a_n^*)^\prime (h_n^* a_n^*)^{-1} (h_n^* a_n^*)^\prime k_n^* = k_n^*,
\]

and

\[
(h_n^* a_n^*)^\prime (h_n^* a_n^*)^{-1} (h_n^* a_n^*)^\prime Q_n^* (Q_n^* Q_n^*)_1 Q_n^* \bar{X}_n = a_n^* (a_n^* a_n^*)^{-1} a_n^* \bar{X}_n
\]

\[
= \bar{X}_n = Q_n^* (Q_n^* Q_n^*)_1 Q_n^* \bar{X}_n.
\]

Also, \( Q_n^*(Q_n^* Q_n^*)_1 Q_n^* (G_n X_n \beta, X_n) = (G_n X_n \beta, X_n) \). Thus, we conclude that those \( a_n^* \) and \( Q_n^* \) attain the upper bound of (8) and therefore, they provide the asymptotic best quadratic moments and IV matrices.

### 3.4 Best GMM with linear-quadratic moments, and martingale differences

In conclusion, the set of best linear-quadratic moments is

\[
g_n^*(\theta) = [\epsilon_n^\prime (\theta) (G_n - \text{Diag}(G_n)) \epsilon_n (\theta), \quad \epsilon_n^\prime (\theta) \left( \text{Diag}(G_n) - \frac{tr(G_n)}{n} I_n \right) \epsilon_n (\theta),
\]

\[
\epsilon_n^\prime (\theta) \left( \text{Diag}(G_n X_n \beta) - \frac{l_n^* G_n X_n \beta}{n} I_n \right) \epsilon_n (\theta),
\]

\[
\epsilon_n^\prime (\theta) \left( \text{Diag}(X_{n2}) - \frac{l_n^* X_{n2}}{n} I_n \right) \epsilon_n (\theta), \quad \ldots, \quad \epsilon_n^\prime (\theta) \left( \text{Diag}(X_{nk}) - \frac{l_n^* X_{nk}}{n} I_n \right) \epsilon_n (\theta),
\]

\[
\epsilon_n^\prime (\theta) G_n X_n \beta, \quad \epsilon_n^\prime (\theta) X_n, \quad \epsilon_n^\prime (\theta) d_{G_n}]'.
\]

(9)
For the set of feasible linear-quadratic moments, the parameter $\lambda$ in $G_n$ and $\beta$ should be replaced by initial consistent estimates.

At the true parameter vector $\theta_0$,

$$
\text{var}(g_n^*(\theta_0)) = \Omega_n^* = \begin{pmatrix}
0 & 0 & 0 \\
0 & (\mu_{40} - 3\sigma_0^4)\omega_{nd}^2\omega_{nd} & \mu_{30}\omega_{nd}^2 Q_n^* \\
0 & \mu_{30}Q_n^* & \omega_{nd}
\end{pmatrix} + \sigma_0^4 \begin{pmatrix}
\Delta_n & 0 & 0 \\
0 & \Delta_{dn} & 0 \\
0 & 0 & \frac{1}{\sigma_0}Q_n^*Q_n^*
\end{pmatrix},
$$

where $\Delta_n = \frac{1}{2} \text{tr}[(G_n - \text{Diag}(G_n))^s(G_n - \text{Diag}(G_n))^s]$, $\omega_{nd} = [d_{G_n}, G_nX_n\beta, \tilde{X}_{n2}, \ldots, \tilde{X}_{nk}]$, and

$$
\Delta_{dn} = 2 \begin{pmatrix}
\tilde{d}_{G_n}'d_{G_n} & \tilde{d}_{G_n}'(G_nX_n\beta), & \ldots, & \tilde{d}_{G_n}'(G_nX_n\gamma) \\
(d_nX_n\gamma)'d_{G_n}, & (d_nX_n\gamma)'(G_nX_n\gamma) & \ldots & (d_nX_n\gamma)'(G_nX_n\gamma) \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{X}_{nk}'d_{G_n}, & \tilde{X}_{nk}'(G_nX_n\gamma) & \ldots & \tilde{X}_{nk}'(G_nX_n\gamma)
\end{pmatrix},
$$

because

$$
\text{tr}([\text{Diag}(G_n) - \frac{\text{tr}(G_n)}{n}I_n]^s[\text{Diag}(G_n) - \frac{\text{tr}(G_n)}{n}I_n]) = 2\tilde{d}_{G_n}'\tilde{d}_{G_n},
$$

and

$$
\text{tr}([\text{Diag}(G_n) - \frac{\text{tr}(G_n)}{n}I_n]^s[\text{Diag}(G_nX_n\beta) - \frac{\text{tr}(G_nX_n\beta)}{n}I_n]) = 2\tilde{l}_{G_n}'\tilde{l}_{G_n}(G_nX_n\beta)
$$

etc. With an initial consistent estimated $\tilde{\Omega}_n^*$, the best GMME of $\theta$ can be derived as the argument of

$$
\min_{\hat{\theta}} g_n^*(\hat{\theta})\tilde{\Omega}_n^* g_n^*(\hat{\theta}).
$$

In this approach, there is a need to estimate the third and fourth moments of $\epsilon_{ni}$. On the other hand, one may avoid such estimates by exploring the estimation of $\tilde{\Omega}_n^*$ by using martingale structure of a linear quadratic form.

For asymptotic distributions of statistics with linear-quadratic forms, martingale CLT provides an essential tool. The martingale property of those statistics may also provide additional uses. Here one may have a simple estimate of $\tilde{\Omega}_n^*$ based on martingale representation of statistics with linear-quadratic form. A general statistic $\Xi_n$ in a linear-quadratic form is

$$
\Xi_n = \epsilon_n'P_n\epsilon_n - \sigma_0^2\text{tr}(P_n) + b_n'\epsilon_n,
$$

where $P_n = [p_{ni,j}]$ is an $n \times n$ nonstochastic matrix, $b_n = [b_{ni}]$ is an $n \times 1$ nonstochastic vector, and elements of $\epsilon_n = [\epsilon_{ni}]$ are mutually independent with a
homogenous variance $\sigma_0^2$. We can rewrite $\Xi_n$ as a sum of martingale differences. Specifically $\Xi_n = \sum_{i=1}^n \xi_{ni}$, where

$$\xi_{ni} = p_{nii}(\epsilon_{ni}^2 - \sigma_0^2) + \epsilon_{ni} \sum_{j=1}^{i-1} (p_{nij} + p_{nji})\epsilon_{nj} + b_{ni}\epsilon_{ni}.$$  

Consider the $\sigma$-fields $F_{n0} = \{\emptyset, \Omega\}$, $F_{ni} = \sigma(\epsilon_{n1}, \ldots, \epsilon_{ni})$, $1 \leq i \leq n$. As $F_{n,i-1} \subset F_{ni}$ and $E(\xi_{ni}|F_{n,i-1}) = 0$, $\{\xi_{ni}, F_{ni}, 1 \leq i \leq n, n \geq 1\}$ forms a martingale difference array. Thus, $\xi_{ni}$'s are uncorrelated and the variance of $\Xi_n$ is simply

$$\text{var}(\Xi_n) = \sum_{i=1}^n E(\xi_{ni}\xi_{ni}') = E(\varphi_n'\varphi_n),$$

where $\varphi_n = (\xi_{n1}, \ldots, \xi_{nn})'$. Under regularity conditions, $\text{var}(\Xi_n)$ can be estimated by the outer product

$$\hat{\varphi}_n'\hat{\varphi}_n = \sum_{i=1}^n \hat{\xi}_{ni}\hat{\xi}_{ni}',$$

where $\hat{\xi}_{ni}$'s are components of an estimated $\hat{\varphi}_n$, such that $\frac{1}{n}\hat{\varphi}_n'\hat{\varphi}_n = \frac{1}{n}\text{var}(\Xi_n) + o_p(1)$. The advantage of this variance estimator is that no analytical form of the variance is needed.

For the best linear quadratic moments in (9), $g^*_n(\theta) = \frac{1}{n}\sum_{i=1}^n \xi^*_ni(\theta)$, where

$$\xi^*_ni(\theta) = \left(\epsilon_{ni}(\theta)\sum_{j=1}^{i-1} (G_{n,ij}(\theta) + G_{n,ji})\epsilon_{nj}(\theta), \tilde{G}_{n,ii}(\epsilon_{ni}^2(\theta) - \sigma_0^2), (G_n\tilde{X}_n\beta)i(\epsilon_{ni}^2(\theta) - \sigma_0^2), (X_n\epsilon_{ni}(\theta), G_{n,ii}\epsilon_{ni}(\theta))'\right).$$

Note that here, we have to adjust terms involving $\epsilon_{ni}^2$ by its variance $\sigma_0^2$ so that the resulted $(\epsilon_{ni}^2, \sigma_0^2)$ has zero mean so it turns $\xi^*_ni$ at the true parameter vector into a martingale difference array. With initial consistent estimates of $\theta_0$ and $\sigma_0^2$, all the unknown parameters in $\xi^*_ni(\theta)$ can be replaced by the initial consistent estimates and its estimate will be denoted by $\hat{\xi}^*_ni$, so $\hat{\Omega}_n^*$ can be consistently estimated by $\frac{1}{n}\sum_{i=1}^n \hat{\xi}^*_ni\hat{\xi}^*_ni'$. A feasible OGMM estimation can then be

$$\min_{\theta} g^*_n(\theta)(\frac{1}{n}\sum_{i=1}^n \hat{\xi}^*_ni\hat{\xi}^*_ni')^{-1} g^*_n(\theta).$$

In this estimation, the total number of unknown parameters $\lambda$ and $\beta$ is of dimension $(k+1)$, but the total number of best moments has $(k+2)$ linear moments, and $(k+2)$ number
of quadratic moments. So there is a \((k + 3)\) number of over-identified moments for estimation. Let \(\hat{\theta}_n\) be the OGMM estimate. Then

\[
n g_n^* (\hat{\theta}_n) = \frac{1}{n} \sum_{i=1}^{n} \xi_{ni}^* \xi_{ni}^{*'} - 1 g_n^* (\hat{\theta}_n) \xrightarrow{d} \chi^2 (k + 3),
\]

which is a goodness-of-fit measure, known as the J-statistic.

When the disturbance vector \(\epsilon_n\) in the SAR model is normally \(N(0, \sigma_0^2 I_n)\) distributed, or more generally the disturbances have zero third moment and zero excess kurtosis, Lee (2007) has shown that the set of best linear-quadratic moments is

\[
g_{n1}^*(\theta) = \epsilon_n^'(\theta)(G_n - \text{Diag}(G_n))\epsilon_n(\theta), \quad \epsilon_n^'(\theta)(\text{Diag}(G_n) - \frac{\text{tr}(G_n)}{n} I_n)\epsilon_n(\theta),
\]

\[
\epsilon_n^'(\theta)G_nX_n\beta, \quad \epsilon_n^'(\theta)X_n^',
\]

which follows from the scores of the likelihood function under normality.

So in terms of asymptotic efficiency, the asymptotic variance of the OGMME of \(\theta\) derived from \(g_{n1}^*(\theta)\) will be the same as that using the larger set of moments \(g_n^*(\theta)\). Therefore, when \(\epsilon_{ni}\)’s are homoskedastic and normally distributed, the remaining moments

\[
g_{n2}^*(\theta) = [\epsilon_n^'(\theta)(\text{Diag}(G_n)X_n\beta) - \frac{l_n'G_nX_n\beta}{n} I_n]\epsilon_n(\theta),
\]

\[
\epsilon_n^'(\theta)(\text{Diag}(X_{n2}) - \frac{l_n'X_{n2}}{n} I_n)\epsilon_n(\theta), \ldots, \epsilon_n^'(\theta)(\text{Diag}(X_{nk}) - \frac{l_n'X_{nk}}{n} I_n)\epsilon_n(\theta),
\]

\[
\epsilon_n^'(\theta)d_{G_n}'
\]

would be redundant (Breusch et al. 1999). The redundancy of \(g_{n2}^*\) given \(g_{n1}^*\) is characterized by the limiting variance of the \(\sqrt{n}\) normalized GMME based on \(g_{n1}^*\) being equal to that based on \(g_{n}^*\). As \(g_{n2}^*\) can provide extra identification of \(\theta_0\), even \(g_{n2}^*\) given \(g_{n1}^*\) is redundant, the J statistic of over-identification based on the whole \(g_{n}^*\) moments would still be asymptotically chi-square distributed with the same degrees of freedom.

4. The best linear-quadratic moments for specification tests

The best linear-quadratic moments derived seem to have various implications on some aspects of the SAR model specification.

4.1 Inconsistency of QML and GMM under ignored heteroskedasticity in disturbances
For the estimation of the SAR model $Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + \epsilon_n$ in (1) if the elements $\epsilon_{ni}$’s of the $n$-dimensional vector $\epsilon_n$ are independent with mean 0 but heteroskedastic variances $\sigma_{ni}^2$, $i = 1, \cdots, n$. For a regression model, it is known that the usual 2SLS estimate of regression coefficients can be consistent even variances were heteroskedastic. For the SAR model, it is apparent that the linear moments $X_n' \epsilon_n$, $(G_n X_n \beta_0)' \epsilon_n$ and $d'_{G_n} \epsilon_n$ are orthogonality conditions for IV estimation for the SAR model as $X_n$, $G_n X_n \beta_0$ and even $G_n$ are exogenous variables. Those linear moments would hold for estimation even the disturbances $\epsilon_{ni}$’s are heteroskedastic.

The quadratic moment $\epsilon'_n (G_n - \text{Diag}(G_n)) \epsilon_n$ will also be robust under heteroskedasticity because

$$E(\epsilon'_n (G_n - \text{Diag}(G_n)) \epsilon_n) = \text{tr}((G_n - \text{Diag}(G_n)) \Sigma_n) = \text{tr}((\text{Diag}(G_n) \Sigma_n) - \text{Diag}(G_n) \Sigma_n) = 0,$$

where $\Sigma_n = \text{Diag}(\sigma_{n1}^2, \cdots, \sigma_{nn}^2)$. However, the remaining quadratic moments with diagonal quadratic matrices could be problematic.

For the ML (or QML) of the SAR model, however, without taking into account of heteroskedastic variances, the MLE could be inconsistent under some circumstances as shown below (Lin and Lee 2010). By ignoring the heteroskedastic disturbances but regarding them with a homoskastic variance, the log quasi-likelihood would be

$$\ln L_n(\delta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |S_n(\lambda)| - \frac{1}{2\sigma^2} \epsilon'_n(\theta) \epsilon_n(\theta).$$

Given $\lambda$, the MLE of $\beta$ is $\hat{\beta}_n(\lambda) = (X_n' X_n)^{-1} X_n' S_n(\lambda) Y_n$, and the MLE of $\sigma^2$ is

$$\hat{\sigma}_n^2(\lambda) = \frac{1}{n} [S_n(\lambda) Y_n - X_n' \hat{\beta}_n(\lambda)] [S_n(\lambda) Y_n - X_n' \hat{\beta}_n(\lambda)]' = \frac{1}{n} Y_n' S_n(\lambda) M_n S_n(\lambda) Y_n,$$

where $M_n = I_n - X_n (X_n' X_n)^{-1} X_n'$. The concentrated log likelihood function of $\lambda$ is

$$\ln L_n(\lambda) = -\frac{n}{2} (\ln(2\pi) + 1) - \frac{n}{2} \ln \hat{\sigma}_n^2(\lambda) + \ln |S_n(\lambda)|.$$

The first order condition for the concentrated log likelihood function is

$$\frac{\partial \ln L_n(\lambda)}{\partial \lambda} = \frac{1}{\hat{\sigma}_n^2(\lambda)} Y_n' Y_n' M_n S_n(\lambda) Y_n - tr(W_n S_n^{-1}(\lambda)).$$

For consistency of the MLE $\hat{\lambda}_n$, the necessary condition is $\lim_{n \to \infty} \frac{1}{n} \frac{\partial \ln L_n(\lambda_n)}{\partial \lambda} = 0$. But with heteroskedastic disturbances, this condition would not necessarily be satisfied. In consequently,
the consistency of the MLE might not be guaranteed. In the presence of heteroskedasticity, at the true $\lambda_0$,

$$\hat{\sigma}_n^2(\lambda_0) = \frac{1}{n}[S_n'Y_n - X_n'\hat{\beta}_n(\lambda_0)][S_n'Y_n - X_n'\hat{\beta}_n(\lambda_0)]$$

$$= \frac{1}{n}\epsilon_n'M_n\epsilon_n = \frac{1}{n}\epsilon_n'\epsilon_n + o_p(1) = \frac{1}{n}\sum_{i=1}^{n}\sigma_{ni}^2 + o_p(1).$$

So, $\hat{\sigma}_n^2(\lambda_0)$ and the average $\bar{\sigma}_n^2$ of $\sigma_{ni}^2$ are asymptotically equivalent. Then, at $\lambda_0$,

$$\frac{1}{n}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = \frac{1}{n}[\frac{1}{\bar{\sigma}_n^2(\lambda_0)}Y_n'W_n'\epsilon_n - tr(W_nS_n^{-1})]\$$

$$= \frac{1}{n}\epsilon_n'G_n'\epsilon_n + \frac{1}{n}(X_n\beta_0)'G_n'M_n\epsilon_n - \frac{1}{n}tr(G_n)$$

$$= \frac{\sum_{i=1}^{n}G_{n,ii}\sigma_{ni}^2}{\sum_{i=1}^{n}\sigma_{ni}^2} - G_n + o_p(1)$$

$$= \frac{1}{n}\sum_{i=1}^{n}[G_{n,ii} - \bar{G}_n](\sigma_{ni}^2 - \bar{\sigma}_n^2) + o_p(1)$$

$$= \frac{cov(G_{n,ii},\sigma_{ni}^2)}{\bar{\sigma}_n^2} + o_p(1),$$

where $\bar{G}_n = \frac{1}{n}tr(G_n) = \frac{1}{n}\sum_{i=1}^{n}G_{n,ii}$ and $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_{ni}^2$ are means. Therefore, the limit of $\frac{1}{n}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda}$ will be zero if and only if the (empirical) covariance between the diagonal elements of the matrix $G_n$, namely $G_{n,ii}$, $i = 1, \ldots, n$, and the individual variances $\sigma_{ni}^2$, $i = 1, \ldots, n$, is zero in the limit. In the heteroskedastic case, this condition of zero covariance would occur for some $W_n$; for example, the case with all the diagonal elements of the matrix $G_n$ are equal. Constant diagonal elements in $G_n$ for some special cases may hold such as a “circular” world where the units are arranged on a circle such that the last unit $y_n$ has neighbors $y_1$ and $y_{n-1}$, $y_1$ has neighbors $y_2$ and $y_n$, and so forth. If we assign equal weight to each neighbor of the same unit, the diagonal elements of the resulting $G_n$ matrix will be constant. The units in a “circular” world can have more neighbors, as long as each unit has the same numbers of neighbors and with half of the neighbors lead and the rest half lag, the diagonal elements of the $G_n$ matrix will be the same. In general, if some spatial units do not possess some similar neighboring structure, one would not expect that the diagonal of $G_n$ would be a constant.

Following the inconsistency of the MLE of $\lambda_0$, a consequence is the inconsistency of the MLE of $\beta_0$ because

$$\hat{\beta}_n(\lambda) = \beta_0 + (\lambda_0 - \lambda)(X_n'X_n)^{-1}X_n'G_nX_n\beta_0 + o_p(1),$$

22
which will not converge to $\beta_0$ in the limit if $\hat{\lambda}$ is not consistent. Thus, besides the computational burden it entails, the MLE for the SAR model with unknown heteroskedasticity could be inconsistent due to the inconsistence of $\hat{\lambda}$.

4.2 Specification test for inconsistency of QML

The above consideration suggests an useful test statistic for testing consistency of the QMLE as if disturbances are homoscedastic while the true disturbances might be heteroskedastic. The test statistic can be based on testing the validity of the moment

$$
\epsilon_n'(\hat{\theta})\text{Diag}[G_n(\hat{\theta}) - \frac{1}{n}\text{tr}(G_n(\hat{\theta}))I_n]\epsilon_n(\hat{\theta}),
$$

where $\hat{\theta}$ is a consistent estimate of $\theta_0$ under unknown heteroskedasticity. Alternatively, a test on the validity of this moment can be based on the difference of optimized distance functions with or without this moment condition (see Ruud (2000)). This moment test is to test whether $G_{nii}$ would be correlated with $\sigma_{ni}^2$ or not. If it is accepted, then it has the implication that the conventional QMLE of the SAR model would be consistent. However, the disturbances might still be correlated. So the variance matrix of the QMLE under such a circumstance would take a ‘sandwich’ form and one has to be careful in its evaluation.

4.3 Specification test for unknown heteroskedasticity

It can be possible that variances, $\sigma_{ni}^2$, of disturbances of the SAR model are heteroscedastic due to $\sigma_{ni}^2$ depending on some exogenous regressors. For example, $x_{ni}$ would be correlated with $\sigma_{ni}^2$ under the possibility that $\sigma_{ni}^2$ would depend on $x_{ni}$. One may then suggest a test based on testing whether individual varying regressors $x_{ni,2}, \cdots, x_{ni,k}$ of $X_n$ are correlated with $\sigma_{ni}^2$ or not. Let $X_{nl}$ be the $l$th column of $X_n$ for $l = 2, \cdots, k$, and $A_{nl} = \text{Diag}((I_n - \frac{1}{n}l_n'l_n)X_{nl})$. Note that $(I_n - \frac{1}{n}l_n'l_n)$ is the deviation from mean operator, so $\tilde{X}_{nl} = (I_n - \frac{1}{n}l_n'l_n)X_{nl}$ is the vector of regressor $x_{ni,l}$ deviation from its sample mean. The test of $\sigma_{ni}^2$ being heteroskedastic can be based on the empirical moments

$$
\epsilon_n'(\hat{\theta})A_{nl}\epsilon_n(\hat{\theta}), \quad l = 2, \cdots, k.
$$

The theoretical average moment would have

$$
\frac{1}{n}E(\epsilon_n'A_{nl}\epsilon_n) = \frac{1}{n}\text{tr}(A_n\Sigma_n) = \frac{1}{n}\sum_{i=1}^{n}a_{ni}\sigma_{ni}^2 = \frac{1}{n}\sum_{i=1}^{n}a_{ni}(\sigma_{ni}^2 - \bar{\sigma}_n^2) = \text{cov}(\sigma_{ni}^2, a_{ni});
$$
thus when \( a_{ni} \) is an \( x_{ni,l} \), we are testing the covariance of \( \sigma^2_{ni} \) with \( x_{ni,l} \).

5. Concluding Remark

While in this paper we have focused our attention on the search of best linear and quadratic moments for the GMM estimation of the SAR with a single spatial lag and i.i.d. disturbances, we expect that the search can be extended to more complicated and richer SAR models with spatial errors. Indeed, we expect that we might construct best linear and quadratic moments for the estimation of higher order SAR models with spatial lags and higher order spatial errors as those in Liu and Lee (2010).

References


