Lecture 6: Spatial Linear Regression Model with Unknown Heteroskedasticity

Zhenlin Yang

School of Economics, Singapore Management University
zlyang@smu.edu.sg
http://www.mysmu.edu.sg/faculty/zlyang/

ECON747 Spatial Econometric Models and Methods, Term I 2024-25

6.1. Introduction

This lecture presents estimation and inference methods for spatial linear regression (SLR) model with errors being independent but not identically distributed (inid). In particular, in the SLR model with spatial lag (SL) and spatial error (SE), studied in Lecture 2:

$$Y_n = \lambda W_{1n} Y_n + X_n \beta + u_n, \quad u_n = \rho W_{2n} u_n + \epsilon_n, \tag{6.1}$$

the variances of ϵ_{ni} are allowed to be different for different spatial units, leading to the so-called cross-sectional heteroskedasticity (CH).

- The CH can be a function (of known form) of, e.g., some regressor values and some additional parameters.

  See e.g., Breusch & Pagan (1979) and Baltagi, Pirotte & Yang (2021).

- It can be of an unknown form. This lecture concerns the unknown CH and assumes that $\text{Var}(\epsilon_{ni}) = \sigma_0^2 h_{ni}$, where $h_{ni} > 0$ and $\frac{1}{n}\sum_{i=1}^{n} h_{ni} = 1$.

  See, e.g., Baltagi & Yang (2013b) and Liu & Yang (2015).

While **CH** is common in regular cross-section studies, it may be more so for a spatial econometrics model due to aggregation, clustering, etc.

- Anselin (1988b) identifies that heteroskedasticity can broadly occur due to "idiosyncrasies in model specification and affect the statistical validity of the estimated model", e.g., the misspecification of model that feeds to the disturbance term or more naturally the presence of peer interactions.
- Data related heteroskedasticity may also occur, e.g., if the model deals with a mix of aggregate and non aggregate data, the aggregation may cause errors to be heteroskedastic. See, e.g., Glaeser et al. (1996), LeSage & Pace (2009), Lin & Lee (2010), Kelejian & Prucha (2010), for more discussions.
- As such, the assumption of homoskedastic disturbances is likely to be invalid in a spatial context in general. However, as Liu & Yang (2015) comment, much of the present spatial econometrics literature has focused on estimators developed under the assumption that the errors are homoskedastic.
- This is in a clear contrast to the standard cross-section econometrics literature where the use of CH-robust estimators is a standard practice.

In the presence of unknown CH, Lin & Lee (2010) show that the QMLE of SL model can be inconsistent as a 'necessary' condition for consistency can be violated, and thus propose robust GMM estimators for the model.

In this lecture,

1. we first show that this condition may hold in certain situations and when it does the regular QML estimator can still be consistent.
2. In cases where this condition is violated, we propose modified QML (MQML) estimation methods, robust against unknown CH.
3. In both cases, asymptotic distributions of the estimators are derived, and methods for estimating robust variances are given.
4. The proposed MQML methods are extended to SLE model. GMM-type methods for SL and SLE models are presented.
5. We then present LM-type tests for the existence of spatial effects in the SLR models, robust against unknown CH.
6. Some Monte Carlo results are presented to show the 'necessity' of using CH-robust methods when unknown CH is present.

To ease the exposition, we study in detail the robustness of QML estimator of the spatial lag (SL) model presented in Lecture 2:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + \epsilon_n. \tag{6.2}$$

The Gaussian loglikelihood of $\theta = (\beta', \sigma^2, \lambda)'$ **as if** $\epsilon_{ni} \overset{iid}{\sim} N(0, \sigma_0^2)$ is

$$\ell_n(\theta) = -\tfrac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\lambda)| - \tfrac{1}{2\sigma^2}\|\epsilon_n(\beta, \lambda)\|^2, \tag{6.3}$$

where $\epsilon_n(\beta, \lambda) = A_n(\lambda)Y_n - X_n\beta$ and $A_n(\lambda) = I_n - \lambda W_n$.

Maximizing $\ell_n(\theta)$ gives the QMLE $\hat{\theta}_n$ of θ_0. It has been shown: $\hat{\theta}_n$ is consistent and $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma)$ as long as $\epsilon_{ni} \overset{iid}{\sim} (0, \sigma_0^2)$.

However, some important issues need to be further considered:

(*i*) conditions under which $\hat{\theta}_n$ remains consistent when $\epsilon_{ni} \overset{inid}{\sim} (0, \sigma_0^2 h_{ni})$,

(*ii*) methods to modify the QML estimation method so that the modified QMLE becomes generally consistent under unknown CH, and

(*iii*) methods for CH-robust estimation of the variance of modified QMLE.

We now address issue (i). Issues (ii) and (iii) are addressed in Sec. 6.3.

It is well-known that the regular QMLE of the usual linear regression model, developed under homoskedastic errors, is still consistent when the errors are in fact heteroskedastic. However, for correct inferences the standard error of the estimator has to be adjusted to account for this unknown CH (White, 1980).

Suppose now we have an SL model (6.2) with $\epsilon_{ni} \overset{inid}{\sim} (0, \sigma_0^2 h_{ni})$, $h_{ni} > 0$ and $\frac{1}{n}\sum_{i=1}^n h_{ni} = 1$. Consider the quasi score function derived from (6.3),

$$S_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial \theta} = \begin{cases} \frac{1}{\sigma^2} X_n' \epsilon_n(\beta, \lambda), \\ \frac{1}{2\sigma^4}[\epsilon_n'(\beta,\lambda)\epsilon_n(\beta,\lambda) - n\sigma^2], \\ \frac{1}{\sigma^2} Y_n' W_n' \epsilon_n(\beta,\lambda) - \mathrm{tr}[F_n(\lambda)], \end{cases} \tag{6.4}$$

where $F_n(\lambda) = W_n A^{-1}(\lambda)$.

Note: σ^2 is the average of Var$(\epsilon_{n,i})$. Under homoskedasticity, $h_{ni} = 1, \forall i$. This parameterization, a nonparametric version of Breusch and Pagan (1979), is useful as it allows the estimation of the average scale parameter.

For an extremum estimator, such as the QMLE $\hat{\theta}_n$ given above, to be consistent, it is necessary that

$$\text{plim}_{n\to\infty} \frac{1}{n} S_n(\theta_0) = 0 \ \text{ at the true parameter } \theta_0$$

(Amemiya, 1985). This is always the case for the β and σ^2 components of $\psi_n(\theta_0)$ whether or not the errors are homoskedastic. However, it may not be the case for the λ-component of $S_n(\theta_0)$.

Let $\mathbf{h}_n = (h_{n,1}, \ldots, h_{n,n})'$, $\mathbf{f}_n = \mathtt{diagv}(F_n)$, and $\text{Cov}(\mathbf{h}_n, \mathbf{f}_n)$ denote the sample covariance between the two vectors. Let $\bar{f}_n = \frac{1}{n}\sum_{i=1}^n f_{n,i}$ and $H_n = \mathtt{diag}(h_n)$. We have, similarly to Lin and Lee (2010),

$$\begin{aligned} \frac{1}{n}\frac{\partial}{\partial\lambda}\ell_n(\theta_0) &= \frac{1}{n}\text{tr}(H_n F_n - F_n) + o_p(1) \\ &= \frac{1}{n}\sum_{i=1}^n (h_{n,i} - 1)(f_{n,i} - \bar{f}_n) + o_p(1) \qquad (6.5) \\ &= \text{Cov}(\mathbf{h}_n, \mathbf{f}_n) + o_p(1). \end{aligned}$$

Therefore, for $\hat{\theta}_n$ to be consistent, it is necessary that $\text{Cov}(\mathbf{h}_n, \mathbf{f}_n) \to 0$; in other words, if $\lim_{n\to\infty} \text{Cov}(\mathbf{h}_n, \mathbf{f}_n) \neq 0$, $\hat{\theta}_n$ cannot be consistent.

- Lin and Lee (2010) noted that this condition is satisfied if almost all the diagonal elements of the matrix F_n are equal.
- Liu & Yang (2015) argued: by Cauchy-Schwartz inequality, this condition is satisfied if $\text{Var}(f_n) \to 0$, which boils down to $\text{Var}(k_n) \to 0$, where k_n is the vector of number of neighbours for each unit.
- This is because (i) $F_n = W_n + \lambda W_n^2 + \lambda^2 W_n^3 + \ldots$, if $|\lambda| < 1$ and $w_{n,ij} < 1$, and (ii) the diagonal elements of W_n^r, $r \geq 2$ inversely relate to k_n, see Anselin (2003). In fact, when W_n is row-normalized and symmetric, $\text{diag}(W_n^2) = \{k_{n,i}^{-1}\}$.
- $\text{Var}(k_n) = o(1)$ can be seen to be true for many popular spatial layouts such as Rook, Queen, group interactions such that variation in group sizes becomes small when n gets large, etc, see Yang (2010).
- Furthermore, if CH occurs due to reasons unrelated to the number of neighbours, e.g., the nature of the exogenous regressors X_n, then the required condition will still be satisfied. These suggest that the regular QMLE of the SL model can still be consistent under CH.

Recall: a quantity defined at the true parameter is represented with a suppressed variable notation, e.g., $A_n \equiv A_n(\lambda_0)$ and $F_n \equiv F_n(\lambda_0)$.

Assumption 1: *The true λ_0 is in the interior of a compact parameter set Λ.*

Assumption 2: $\epsilon_n \sim (0, \sigma_0^2 H_n)$*, where* $H_n = \text{diag}(h_{n,1}, \dots, h_{n,n})$*,* $h_{n,i} > 0, \forall i$*, and* $\frac{1}{n}\sum_{i=1}^n h_{n,i} = 1$*.* $E|\epsilon_{n,i}|^{4+\delta} < c$ *for some* $\delta > 0$ *and constant c for all n and i.*

Assumption 3: *The elements of X_n are uniformly bounded for all n, X_n has the full rank k, and* $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ *exists and is nonsingular.*

Assumption 4: *The spatial weights matrix W_n is uniformly bounded in absolute value in both row and column sums and its diagonal elements are zero.*

Assumption 5: *The matrix A_n is non-singular and A_n^{-1} is uniformly bounded in absolute value in both row and column sums. Further, $A_n^{-1}(\lambda)$ is uniformly bounded in either row or column sums, uniformly in $\lambda \in \Lambda$.*

Assumption 6: *The limit* $\lim_{n\to\infty} \frac{1}{n}(F_n X_n \beta_0)' M_n (F_n X_n \beta_0) = k$*, where either* $k > 0$*, or* $k = 0$ *but* $\lim_{n\to\infty} \frac{1}{n} \ln|\sigma_0^2 A_n^{-1} A_n^{'-1}| - \frac{1}{n} \ln|\sigma_n^2(\lambda) A_n^{-1}(\lambda) A_n^{'-1}(\lambda)| \neq 0$*, whenever* $\lambda \neq \lambda_0$*, where* $\sigma_n^2(\lambda) = \frac{1}{n}\sigma_0^2 \text{tr}(H_n A_n^{'-1} A_n^{'-1}(\lambda) A_n^{-1}(\lambda) A_n^{-1})$*.*

Parameter space: Λ must be such that $A_n(\lambda)$ is non-singular $\forall \lambda \in \Lambda$.

Since $|A_n(\lambda)| = \prod_{i=1}^{n}(1 - \lambda\omega_i)$, where ω_i are the eigenvalues of W_n (Ord, 1975),

- if the eigenvalues of W_n are all real, then $\Lambda = (\omega_{\min}^{-1}, \ \omega_{\max}^{-1})$, where $\omega_{\min}$ and $\omega_{\max}$ are, respectively, the smallest and the largest eigenvalues of W_n.
- If W_n is row normalized, then $\omega_{\max} = 1$ and $\omega_{\min}^{-1} < -1$, and the parameter space becomes $\Lambda = (w_{\min}^{-1}, 1)$ (Anselin, 1988b, p. 78-79).
- In general, the eigenvalues of W_n may not be all real as W_n can be asymmetric. LeSage and Pace (2009, p. 88-89) argue that only the purely real eigenvalues can affect the singularity of $A_n(\lambda)$. Consequently,
- the interval of λ that guarantees non-singular $A_n(\lambda)$ is $\Lambda = (w_s^{-1}, 1)$ where w_s is the most negative real eigenvalue of W_n.
- Kelejian and Prucha (2010) suggest Λ be $(-\tau_n^{-1}, \tau_n^{-1})$ where τ_n is the spectral radius of W_n, or $(-1, 1)$ after normalization.

Assumptions 2-5 are standard for the SL model (Lin & Lee, 2010; Kelejian & Prucha, 1999). Assumption 6 is the heteroskedastic version of the identification condition by Lee (2004) for the homoskedastic SL model.

From the quasi score function $S_n(\theta)$ given in (6.4), define

$$\Gamma_n = -\mathrm{E}[\tfrac{\partial}{\partial\theta'}S_n(\theta_0)] \text{ and } \Omega_n = \mathrm{E}[S_n(\theta_0)S_n'(\theta_0)],$$

w.r.t. Assumption 2. Let $\mathbf{q}_n = \mathrm{diagv}(F_n'F_n))$. We have the following results.

Theorem 6.1: *Under Assumptions 1-6, Cov*$(\mathbf{f}_n, \mathbf{h}_n) = o(1)$ *and Cov*$(\mathbf{q}_n, \mathbf{h}_n) = o(1)$*, we have as* $n \to \infty$, $\hat{\theta}_n \xrightarrow{p} \theta_0$*; under Assumptions 1-6 and Cov*$(\mathbf{f}_n, \mathbf{h}_n) = o(n^{-1/2})$*, we have as* $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} n\Gamma_n^{-1}\Omega_n\Gamma_n^{-1}\big), \qquad (6.6)$$

where the limits of $\frac{1}{n}\Gamma_n$ *and* $\frac{1}{n}\Omega_n$ *are assumed to exist and* $\frac{1}{n}\Gamma_n$ *is assumed to be nonsingular for large enough n.*

- Γ_n can be estimated by plug-in estimator or by sample analogue.
- However, plug-in estimation of Ω_n runs into a problem, as its σ_0^2-σ_0^2 component, $\frac{1}{4n\sigma_0^4}\sum_{i=1}^n(\kappa_{ni} + 2h_{ni}^2)$, cannot be consistently estimated. The method of M.D. decomposition does not work either.

Remark 6.1. Under unknown CH, joint inference for β, λ and σ^2 does not seem possible, but joint inference for $\vartheta = (\beta', \lambda)'$ can be made. This is important as ϑ is the parameter set that is of main interest.

Remark 6.2. This approach is in line with the GMM approach, where it is typical that the estimation proceeds without σ^2 (e.g., Lin and Lee, 2010).

To do so, we concentrate out σ^2 from (6.4) and work with the concentrated QS (CQS) function of ϑ:

$$S_n^c(\vartheta) = \begin{cases} \tilde{\sigma}_n^{-2}(\vartheta) X_n' \epsilon_n(\vartheta), \\ \tilde{\sigma}_n^{-2}(\vartheta) Y_n' W_n' \epsilon_n(\vartheta) - \mathrm{tr}[F_n(\lambda)], \end{cases} \tag{6.7}$$

where $\tilde{\sigma}_n^2(\vartheta) = \frac{1}{n} \epsilon_n'(\vartheta) \epsilon_n(\vartheta)$, the constrained estimate of σ_0^2 given ϑ.

When $\mathrm{Cov}(\mathbf{f}_n, \mathbf{h}_n) = o(n^{-\frac{1}{2}})$, $\mathrm{plim} \frac{1}{n} S_n^c(\vartheta) = 0$. Theorem 6.1 implies:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} n \Gamma_n^{c-1} \Omega_n^c \Gamma_n^{c-1}\big), \tag{6.8}$$

where $\Gamma_n^c = -\mathrm{E}[\frac{\partial}{\partial \vartheta'} S_n^c(\vartheta_0)]$ and $\Omega_n^c = \mathrm{E}[S_n^c(\vartheta_0) S_n^c(\vartheta_0)']$.

For joint inference for ϑ under CH, Γ_n^c is estimated by its sample analogue,

$$\widehat{\Gamma}_n^c = -\tfrac{\partial}{\partial \vartheta'} S_n^c(\vartheta_0)|_{\vartheta_0 = \hat{\vartheta}_n}. \tag{6.9}$$

Ω_n is estimated by an OPMD estimator, based on the asymptotic M.D. representation (noting that $\frac{1}{\sqrt{n}} \sum_{i=1}^n f_{n,ii}(e_{ni}^2 - 1) = o_p(1)$):

$$\frac{1}{\sqrt{n}} S_n^c(\vartheta_0) \overset{a}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n s_{ni}^c(\vartheta_0) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} z_{ni} e_{ni}; \\ e_{ni}\mu_{ni} + e_{ni}\xi_{ni} \end{pmatrix},$$

where, z_{ni}', e_{ni}, μ_{ni}, and ξ_{ni} are, respectively, the rows of

- $z_n \equiv z_n(\vartheta_0) = \tilde{\sigma}_n^{-1}(\vartheta_0) X_n$, $\quad e_n \equiv e_n(\vartheta_0) = \tilde{\sigma}_n^{-1}(\vartheta_0)\epsilon_n$,
- $\mu_n \equiv \mu_n(\vartheta_0) = \tilde{\sigma}_n^{-1}(\vartheta_0) F_n X_n \beta_0$, $\quad \xi_n \equiv \xi_n(\vartheta_0) = (F_n^{u\prime} + F_n^l) e_n$,

with F_n^u and F_n^l being the strict upper and lower triangular matrices of F_n.

Under the conditions in Theorem 6.1, it is easy to see that $\{s_{ni}^c(\vartheta_0)\}$ are asymptotically uncorrelated with means zero. Thus, $\Omega_n^c \overset{a}{=} \mathrm{Var}[S_n^c(\vartheta_0)] \overset{a}{=} \sum_{i=1}^n \mathrm{Var}[s_{ni}^c(\vartheta_0)]$, and a consistent estimate of Ω_n^c is

$$\widehat{\Omega}_n^c = \sum_{i=1}^n s_{ni}^c(\hat{\vartheta}_n) s_{ni}^c(\hat{\vartheta}_n)'. \tag{6.10}$$

In QML estimation, the use of Gaussian likelihood renders the QS functions being of a linear-quadratic form: $\epsilon_n' A_n \epsilon_n + b_n' \epsilon_n$.

- If $\epsilon_n \sim (0, \sigma_0^2 I_n)$, then $\mathrm{E}(\epsilon_n' A_n \epsilon_n + b_n' \epsilon_n) = \sigma_0^2 \mathrm{tr}(A_n)$. Thus, a valid moment condition would be $\epsilon_n' A_n \epsilon_n + b_n' \epsilon_n - \sigma_0^2 \mathrm{tr}(A_n)$;
- If $\epsilon_n \sim (0, \sigma_0^2 H_n)$, then $\mathrm{E}(\epsilon_n' A_n \epsilon_n + b_n' \epsilon_n) = \sigma_0^2 \mathrm{tr}(A_n H_n)$. But, the similar quantity $\epsilon_n' A_n \epsilon_n + b_n' \epsilon_n - \sigma_0^2 \mathrm{tr}(A_n H_n)$ would not give a valid moment condition due to the presence of the unknown H_n,
- **unless** $\mathrm{tr}(A_n H_n = 0)$, which is the case if $\mathrm{diag}(A_n) = 0$. This presents the key idea in developing H_n-robust estimation method.

In this spirit, simply modifying the λ-component of $S_n(\theta)$ given in (6.4) as:

$$\sigma_0^{-2}[Y_n' W_n' \epsilon_n - \epsilon_n' \mathrm{diag}(F_n) \epsilon_n].$$

We immediately see: $\mathrm{plim} \frac{1}{n\sigma_0^2}[Y_n' W_n' \epsilon_n - \epsilon_n' \mathrm{diag}(F_n) \epsilon_n] = 0$, in light of (6.5).

- The β-component of $S_n(\theta_0)$ is robust against unknown CH.
- Therefore, a set of modified QS functions for ϑ robust against unknown CH would be: $\{\sigma_0^{-2} X_n' \epsilon_n(\vartheta_0);\ \sigma_0^{-2}[Y_n' W_n' \epsilon_n - \epsilon_n' \mathrm{diag}(F_n)\epsilon_n]\}$.
- Obviously, σ_0^{-2} does not play a role here.
- As indicated in Remark 6.1, in the presence of unknown CH, joint inference can only be done for ϑ.

A set of modified quasi score (MQS) of $\vartheta = (\beta', \lambda)'$ is thus,

$$S_n^{\circ}(\vartheta) = \begin{cases} X_n' \epsilon_n(\vartheta), \\ Y_n' W_n' \epsilon_n(\vartheta) - \epsilon_n'(\vartheta)\mathrm{diag}(F_n(\lambda))\epsilon_n(\vartheta). \end{cases} \quad (6.11)$$

Solving $S_n^{\circ}(\vartheta) = 0$ gives the modified QML (MQML) estimator $\hat{\vartheta}_n^{\circ}$ of ϑ_0, fully robust against unknown CH.

It is easy to see that $\mathrm{E}[S_n^\circ(\vartheta_0)] = 0$ and $\mathrm{plim}\frac{1}{n}S_n^\circ(\vartheta_0 = 0$. This paves the way for $\hat{\vartheta}_n^\circ$ to achieve the regular asymptotic properties.

Theorem 6.2. Under Assumptions 1-6, we have as $n \to \infty$, $\hat{\vartheta}_n^\circ \xrightarrow{p} \vartheta_0$ and

$$\sqrt{n}(\hat{\vartheta}_n^\circ - \vartheta_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} n\Gamma_n^{\circ-1}\Omega_n^\circ\Gamma_n^{\circ-1}\big). \tag{6.12}$$

where $\Gamma_n^\circ = -\mathrm{E}[\frac{\partial}{\partial\vartheta'}S_n^\circ(\vartheta_0)]$ and $\Omega_n^\circ = \mathrm{E}[S_n^\circ(\vartheta_0)S_n^\circ(\vartheta_0)']$.

For statistical inference, Γ_n° and Ω_n° are estimated by

- $\widehat{\Gamma}_n^\circ = -\frac{\partial}{\partial\vartheta'}S_n^\circ(\vartheta_0)|_{\vartheta_0=\hat{\vartheta}_n^\circ}$,
- $\widehat{\Omega}_n^\circ = \sum_{i=1}^n s_{ni}^\circ(\hat{\vartheta}_n)s_{ni}^\circ(\hat{\vartheta}_n)'$,

where $s_{ni}^\circ(\vartheta_0) = (x_{ni}'\epsilon_{ni},\ \epsilon_{ni}\mu_{ni} + \epsilon_{ni}\xi_{ni})'$, with x_{ni}' being the rows of X_n, and μ_{ni} and ξ_n the elements of $\mu_n = F_nX_n\beta_0$ and $\xi_n = (F_n^{u\prime} + F_n^l)\epsilon_n$.

$\frac{\partial}{\partial\vartheta'}S_n^\circ(\vartheta) =???$; $\Omega_n^\circ =???$ Does plug-in estimation for Ω_n° work?

The modified QMLE $\hat{\lambda}_n^\circ$ does not taken into account the variability from the estimation of β, and thus it may not perform well in finite sample. Similar phenomenon holds for the QMLE $\hat{\lambda}_n$ when it is robust against CH.

Liu and Yang (2015) propose to work on the concentrated QS function of λ, obtained by substituting $\tilde{\beta}_n(\lambda)$ and $\tilde{\sigma}_n^2(\lambda)$ into the last component of (6.4) for β and σ^2 and then dividing by n:

$$\tilde{\psi}_n(\lambda) = \frac{Y_n' A_n'(\lambda) M_n [F_n(\lambda) - \frac{1}{n}\mathrm{tr}(F_n(\lambda)) I_n] A_n(\lambda) Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n}. \tag{6.13}$$

The average concentrated score $\tilde{\psi}_n(\lambda)$ contains the variability coming from estimating β and σ^2, but to account for it modifications are needed.

Clearly, the QMLE $\hat{\lambda}_n = \arg\{\tilde{\psi}_n(\lambda) = 0\}$, which may not be robust against unknown CH when the conditions in Theorem 6.1 are violated.

Our idea is to modify the numerator of (6.13) so that its expectation at the true parameter λ_0 is zero even under unknown CH. This is achieved by replacing $F_n(\lambda) - \frac{1}{n}\text{tr}(F_n(\lambda))I_n$ by,

$$F_n^{\circ}(\lambda) = F_n(\lambda) - \text{diag}(M_n)^{-1}\text{diag}(M_n F_n(\lambda)). \tag{6.14}$$

This gives a modified concentrated score function,

$$\tilde{\psi}_n^*(\lambda) = \frac{Y_n' A_n'(\lambda) M_n F_n^{\circ}(\lambda) A_n(\lambda) Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n}, \tag{6.15}$$

and hence a modified QML estimator of λ_0 as,

$$\hat{\lambda}_n^* = \arg\{\tilde{\psi}_n^*(\lambda) = 0\}. \tag{6.16}$$

The resulting CH-robust estimators of β and σ^2 are, respectively,

$$\hat{\beta}_n^* = \tilde{\beta}_n(\hat{\lambda}_n^*) \quad \text{and} \quad \hat{\sigma}_n^{*2} = \tilde{\sigma}_n^2(\hat{\lambda}_n^*),$$

as the estimating functions (first two components of $S_n(\theta)$ given in (6.4)) leading to $\tilde{\beta}_n(\lambda)$ and $\tilde{\sigma}_n^2(\lambda)$ (also in (2.18)) are robust to unknown CH.

- Making the expectation of an estimating function to be zero leads potentially to a finite sample bias corrected estimation.
- This is in line with Baltagi and Yang (2013a,b) in constructing standardized or heteroskedasticity-robust LM tests.
- See also Kelejian and Prucha (2001, 2010) and Lin and Lee (2010) for useful methods on LQ forms of heteroskedastic random vectors.

To ensure that the modified estimation function given in (6.15) uniquely identifies λ_0, Assumption 6 needs to be modified as follows.

Assumption 6*: Let $\Omega_n(\lambda) = A_n'(\lambda)[F_n(\lambda) - \mathrm{diag}(F_n(\lambda))]A_n(\lambda)$.
$\lim_{n\to\infty} \frac{1}{n}[\beta_0' X_n' A_n'^{-1}\Omega_n(\lambda)A_n^{-1}X_n\beta_0 + \sigma_0^2 \mathrm{tr}(H_n A_n'^{-1}\Omega_n(\lambda)A_n^{-1})] \neq 0$, $\forall \lambda \neq \lambda_0$.

The CLT for LQ forms of Kelejian and Prucha (2001) allows for CH and can be used to prove the asymptotic normality of $\hat{\lambda}_n^*$. First,

$$\sqrt{n}\tilde{\psi}_n^* \equiv \sqrt{n}\tilde{\psi}_n^*(\lambda_0) = \frac{1}{\sqrt{n}\sigma_0^2}\left(\epsilon_n' B_n^\circ \epsilon_n + c_n'\epsilon_n\right) + o_p(1), \qquad (6.17)$$

where $B_n^\circ = M_n F_n^\circ$ and $c_n = M_n F_n^\circ X_n \beta_0$, because $\hat{\sigma}_n^{-2}(\lambda_0) = \sigma_0^{-2} + o_p(1)$.

It is important to note that $\mathrm{diag}(B_n^\circ) = 0$, arising from the way that $F_n^\circ(\lambda)$ is defined. This greatly simplifies the expressions and their subsequent estimation in Liu an Yang (2015, Theorems 2 & 3).

Based on this and the asymptotic representation for $\sqrt{n}\tilde{\psi}_n^*$ given in (6.17),

$$\begin{aligned}\mathrm{Var}(\sqrt{n}\tilde{\psi}_n^*) &= \tfrac{1}{n}\mathrm{tr}[H_n B_n(H_n B_n + H_n B_n')] + \tfrac{1}{n\sigma_0^2} c_n' H_n c_n + o(1), \\ &\equiv \Omega_{n\lambda} + o(1). \qquad (6.18)\end{aligned}$$

Theorem 6.3: *Under Assumptions 1-5 and 6*, the modified QMLE $\hat{\lambda}_n^*$ is consistent and asymptotically normal, i.e., as $n \to \infty$, $\hat{\lambda}_n^* \xrightarrow{p} \lambda_0$, and*

$$\sqrt{n}(\hat{\lambda}_n^* - \lambda_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} \Gamma_{n\lambda}^{-2}\Omega_{n\lambda}\big),$$

where $\Gamma_{n\lambda} = \frac{1}{n}\mathrm{tr}[H_n(G_n^\circ F_n + F_n^{\circ\prime} F_n - \dot{F}_n^\circ)] + \frac{1}{n\sigma_0^2} c_n'\eta_n,\ \eta_n = F_n X_n \beta_0,$ *and* $\dot{F}_n^\circ = \frac{d}{d\lambda} F_n^\circ = F_n^2 - \mathrm{diag}(M_n)^{-1}\mathrm{diag}(M_n F_n^2)$.

Now consider the modified QMLEs $\hat{\beta}_n^*$ and $\hat{\sigma}_n^{*2}$ of β_0 and σ_0^2 defined below (6.16). Using the relation $A_n(\tilde{\lambda}_n) = A_n - (\tilde{\lambda}_n - \lambda_0)W_n$, we can write,

$$\hat{\beta}_n = \tilde{\beta}_n(\lambda_0) - (\tilde{\lambda}_n - \lambda_0)(X_n'X_n)^{-1}X_n'F_nA_nY_n, \text{ and}$$
$$\hat{\sigma}_n^2 = \tilde{\sigma}_n^2(\lambda_0) - 2(\tilde{\lambda}_n - \lambda_0)\tfrac{1}{n}Y_n'W_n'M_nA_nY_n + (\tilde{\lambda}_n - \lambda_0)^2\tfrac{1}{n}Y_n'W_n'M_nW_nY_n.$$

The asymptotic properties of $\hat{\beta}_n^*$ and $\hat{\sigma}_n^{*2}$ are summarized below.

Corollary 6.1: *Under the assumptions of Theorem 6.3, we have, as $n \to \infty$, $\hat{\beta}_n^* \xrightarrow{p} \beta_0$ and $\hat{\sigma}_n^{*2} \xrightarrow{p} \sigma_0^2$, and further,*

$$\sqrt{n}(\hat{\beta}_n^* - \beta_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty}(X_n'X_n)^{-1}X_n'\Sigma_{n\beta}X_n(X_n'X_n)^{-1}\big],$$

where $\Sigma_{n\beta} = n\sigma_0^2 H_n + \Omega_{n\lambda}\eta_n\eta_n' - 2\Gamma_{n\lambda}^{-1}H_nc_n\eta_n'$.

For applications of Theorem 6.3 and Corollary 6.1, $\Gamma_{n\lambda}$, $\Omega_{n\lambda}$ and $\Sigma_{n\beta}$ can all by estimated by plug-in method, i.e., plugging in $(\hat{\beta}_n^*, \hat{\sigma}_n^{*2}, \hat{\lambda}_n^*)$ for $(\beta_0, \sigma_0^2, \lambda_0)$, and $\hat{H}_n = \mathrm{diag}(\hat{\sigma}_n^{*-2}\hat{\epsilon}_n^2)$ for H_n, where $\hat{\epsilon}_n = A_n(\hat{\lambda}_n^*)Y_n - X_n\hat{\beta}_n^*$.

The modified QML estimation methods can be easily extended to suit for more general models (spatial or non-spatial) with few concentrated score elements that need to be modified to account for the unknown CH.

One of the most needed extensions would be to the SLE model, of which QML estimation is considered in Lecture 2:

$$Y_n = \lambda W_{1n} Y_n + X_n \beta + u_n, \quad u_n = \rho W_{2n} u_n + \epsilon_n. \tag{6.19}$$

The quasi Gaussian loglikelihood function of $\theta = (\beta', \sigma^2, \delta')'$ with $\delta = (\lambda, \rho)'$, **as if** $\epsilon_n \sim N(0, \sigma_0^2 I_n)$, takes the form:

$$\ell_n(\theta) = -\frac{n}{2} \log(2\pi\sigma^2) + \log|A_n(\lambda)| + \log|B_n(\rho)| - \frac{1}{2\sigma^2} \|\epsilon_n(\beta, \delta)\|^2, \tag{6.20}$$

where $A_n(\lambda) = I_n - \lambda W_{1n}$, $B_n(\rho) = I_n - \rho W_{2n}$, $\epsilon_n(\beta, \delta) = \mathbb{Y}_n(\delta) - \mathbb{X}_n(\rho)\beta$, $\mathbb{Y}_n(\delta) = B_n(\rho) A_n(\lambda) Y_n$, and $\mathbb{X}_n(\rho) = B_n(\rho) X_n$.

The QMLE $\hat{\theta}_n$ of θ maximizes $\ell_n(\theta)$.

As in Sec. 6.2., under an extended set of conditions, the QMLE $\hat{\theta}_n$ can be robust to unknown CH, and when it does, an extended set of inference methods can be developed in a similar way.

We will skip such a detail, and concentrated on the modified QML estimators that are fully robust to CH.

Recall the QS function given in Lecture 2, written in slightly different way:

$$S_{\text{SLE}}(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbb{X}_n'(\rho)\epsilon_n(\beta,\delta), \\ \frac{1}{2\sigma^4}\epsilon_n'(\beta,\lambda)\epsilon_n(\beta,\delta) - \frac{n}{2\sigma^2}, \\ \frac{1}{\sigma^2}\epsilon_n'(\beta,\delta)B_n(\rho)W_{1n}Y_n - \text{tr}[F_n(\lambda)], \\ \frac{1}{\sigma^2}\epsilon_n'(\beta,\delta)G_n(\rho)\epsilon_n(\beta,\delta) - \text{tr}[G_n(\rho)], \end{cases} \tag{6.21}$$

where $F_n(\lambda) = W_{1n}A_n^{-1}(\lambda)$, and $G_n(\rho) = W_{2n}B_n^{-1}(\rho)$.

As in (6.11) for SL model, the modified quasi score (MQS) function of $\vartheta = (\beta', \delta')'$ has the following form:

$$S^{\circ}_{\mathrm{SLE}}(\vartheta) = \begin{cases} \mathbb{X}'_n(\rho)\epsilon_n(\vartheta), \\ \epsilon'_n(\vartheta)B_n(\rho)W_{1n}Y_n - \epsilon'_n(\vartheta)\mathrm{diag}(\bar{F}_n(\lambda))\epsilon_n(\vartheta), \\ \epsilon'_n(\vartheta)G_n\epsilon_n(\vartheta) - \epsilon'_n(\vartheta)\mathrm{diag}(G_n(\lambda))\epsilon_n(\vartheta). \end{cases} \tag{6.22}$$

where $\bar{F}_n(\delta) = B_n(\rho)F_n(\lambda)B_n^{-1}(\rho)$.

- Solving $S^{\circ}_n(\vartheta) = 0$ gives the modified QML (MQML) estimators $\hat{\vartheta}^{\circ}_n$ of ϑ_0, and $\hat{\sigma}^{\circ 2}_n = \frac{1}{n}\epsilon'_n(\hat{\vartheta}^{\circ}_n)\epsilon_n(\hat{\vartheta}^{\circ}_n)$ of σ^2_0, fully robust against unknown CH.
- Inference methods can be developed in a similar manner as for the SL model, in particular along the line of Theorem 6.2.

Under an extended set of assumptions of Theorem 6.2, we have as $n \to \infty$, $\hat{\sigma}^{\circ 2}_n \xrightarrow{p} \sigma^2_0$, $\hat{\vartheta}^{\circ}_n \xrightarrow{p} \vartheta_0$, and

$$\sqrt{n}(\hat{\vartheta}^{\circ}_n - \vartheta_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} n\Gamma^{\circ -1}_{\mathrm{SLE}}\Omega^{\circ}_{\mathrm{SLE}}\Gamma^{\circ -1}_{\mathrm{SLE}}\big). \tag{6.23}$$

where $\Gamma^{\circ}_{\mathrm{SLE}} = -\mathrm{E}[\frac{\partial}{\partial\vartheta'}S^{\circ}_{\mathrm{SLE}}(\vartheta_0)]$ and $\Omega^{\circ}_{\mathrm{SLE}} = \mathrm{E}[S^{\circ}_{\mathrm{SLE}}(\vartheta_0)S^{\circ}_{\mathrm{SLE}}(\vartheta_0)']$.

Recall from (2.31) the CQS function of δ upon dividing by n:

$$\psi^c_{\mathtt{SLE}}(\delta) = \begin{cases} -\frac{1}{n}\mathrm{tr}(F_n(\lambda)) + \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\bar{F}_n(\delta)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)M_n(\rho)\mathbb{Y}_n(\delta)}, \\ -\frac{1}{n}\mathrm{tr}(G_n(\rho)) + \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\bar{G}_n(\rho)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\lambda)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \end{cases} \tag{6.24}$$

where $\bar{G}_n(\rho) = G_n(\rho)\mathbb{M}_n(\rho)$, and $\mathbb{M}_n(\rho) = I_n - \mathbb{X}_n(\rho)[\mathbb{X}_n'(\rho)\mathbb{X}_n(\rho)]^{-1}\mathbb{X}_n'(\rho)$.

Using similar arguments as given in Section 6.3, we have, after some algebraic manipulations, the following modified CQS function,

$$\tilde{\psi}^*_{\mathtt{SLE}}(\delta) = \begin{cases} \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\bar{F}^\circ_n(\delta)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \\ \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\bar{G}^\circ_n(\rho)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \end{cases} \tag{6.25}$$

where $\bar{F}^\circ_n(\delta) = \bar{F}_n(\delta) - \mathrm{diag}(\mathbb{M}_n(\rho))^{-1}\mathrm{diag}[\mathbb{M}_n(\rho)\bar{F}_n(\delta)]$, and
$\bar{G}^\circ_n(\rho) = \bar{G}_n(\rho) - \mathrm{diag}(\mathbb{M}_n(\rho))^{-1}\mathrm{diag}[\mathbb{M}_n(\rho)\bar{G}_n(\rho)]$.

- The modified QMLE of δ is defined as $\hat{\delta}_n^* = \arg\{\tilde{\psi}^*_{\mathrm{SLE}}(\delta) = 0\}$.
- The modified QMLEs of β and σ^2 are $\hat{\beta}_n^* \equiv \tilde{\beta}_n(\hat{\delta}_n^*)$ and $\hat{\sigma}_n^{*2} \equiv \tilde{\sigma}_n^{*2}(\hat{\delta}_n^*)$, $\tilde{\beta}_n(\delta)$ and $\tilde{\sigma}_n^2(\delta)$: the constrained QMLEs given in (2.28) and (2.29).
- The MQMLE $\hat{\delta}_n^*$ potentially improves over $\hat{\delta}_n$ and $\hat{\delta}_n^\circ$, and
- over the three-step estimator of Kelejian and Prucha (2010).

It would be interesting to give a more detailed theoretical and empirical (Monte Carlo) study on the topic:

"Heteroskedasticity Robust Estimation of Spatial Linear Regression"

Theorem 6.4: *Under Assumptions 1-3 and extended Assumptions 4, 5, and 6* to SLE model, the modified QMLE* $\hat{\delta}_n^*$ *is consistent and asymptotically normal, i.e., as* $n \to \infty$, $\hat{\delta}_n^* \xrightarrow{p} \lambda_0$, *and*

$$\sqrt{n}(\hat{\delta}_n^* - \delta_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} \Gamma_{n\delta}^{-1}\Omega_{n\delta}\Gamma_{n\delta}^{-1}\big),$$

where $\Gamma_{n\delta} = -\mathrm{E}\big[\frac{\partial}{\partial\delta'}\tilde{\psi}^*_{\mathrm{SLE}}(\delta_0)\big]$ *and* $\Omega_{n\delta} = \mathrm{Var}[\sqrt{n}\tilde{\psi}^*_{\mathrm{SLE}}(\delta_0)]$.

Inferences on δ require consistent estimators of $\Gamma_{n\lambda}$ and $\Omega_{n\delta}$.

- The negative Hessian, $-\frac{\partial}{\partial\delta'}\tilde{\psi}^*_{\mathtt{SLE}}(\delta)$, can easily be derived, which provides a consistent estimator of $\Gamma_{n\lambda}$.
- It is easy to see that $\mathrm{diag}[\mathbb{M}_n(\rho)\bar{F}^\circ_n(\delta)] = 0$ and $\mathrm{diag}[\mathbb{M}_n(\rho)\bar{G}^\circ_n(\rho)] = 0$.
- These greatly facilitate the estimation of $\Omega_{n\delta}$, as it involves only θ_0 and H_n asymptotically, not the 3rd and 4th moments of the errors.

The above idea is made clear by the following asymptotic representation:

$$\sqrt{n}\tilde{\psi}^*_n = \begin{cases} \frac{1}{\sqrt{n}\sigma_0^2}\left(\epsilon'_n C_n \epsilon_n + c'_n \epsilon_n\right) + o_p(1), \\ \frac{1}{\sqrt{n}\sigma_0^2}\left(\epsilon'_n D_n \epsilon_n + d'_n \epsilon_n\right) + o_p(1), \end{cases} \tag{6.26}$$

where $C_n = \mathbb{M}_n\bar{F}^\circ_n$ and $c_n = \mathbb{M}_n\bar{F}^\circ_n\mathbb{X}_n\beta_0$, and D_n and d_n are defined similarly with $\bar{F}^\circ_n$ being replaced by $\bar{G}^\circ_n$.

With (6.26), a first-order approximation to $\Omega_{n\delta}$ is derived giving a plug-in estimator of $\Omega_{n\delta}$, or an asymptotic M.D. representation for $\sqrt{n}\tilde{\psi}^*_n$ is given leading to an OPMD estimator.

With the asymptotic distribution of $\hat{\delta}_n^*$, one can easily derive the asymptotic distribution for $\hat{\beta}_n^*$.

Corollary 6.2: *Under the assumptions of Theorem 6.4, we have, as* $n \to \infty$, $\hat{\beta}_n^* \xrightarrow{p} \beta_0$ *and* $\hat{\sigma}_n^{*2} \xrightarrow{p} \sigma_0^2$, *and further,*

$$\sqrt{n}(\hat{\beta}_n^* - \beta_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty}(\mathbb{X}_n'\mathbb{X}_n)^{-1}\mathbb{X}_n'\mathbb{A}_n\mathbb{X}_n\mathbb{X}_n'\mathbb{X}_n)^{-1}\big),$$

where $\mathbb{A}_n = n\sigma_0^2 H_n + \Omega_{n\delta,11}\eta_n\eta_n' + 2\sqrt{n}H_n(c_n, d_n)\Gamma_{n\delta}^{-1}(\eta_n,\ 0_n)'$, $\Omega_{n\delta,11}$ *is the top-left corner of* $\Omega_{n\delta}$, *and* $\eta_n = B_n F_n X_n \beta_0$.

With the methods of estimation for $\Omega_{n\delta}$ and $\Gamma_{n\delta}$ discussed around (6.26), estimating H_n by $\hat{H}_n = \mathrm{diag}(\hat{\epsilon}_n^{*2})$, where $\hat{\epsilon}_n^* = \epsilon_n(\hat{\beta}_n^*, \hat{\delta}_n^*)$ are the estimate residuals, inference about β are made.

6.5. GMM-Type Estimation of SLR Models under unknown CH

GMM-type methods have been proposed for the estimation of spatial linear regression models under homoskedasticity or heteroskedasticity.

- Kelejian & Prucha (1999) proposed a generalized moment (GM) estimator for homoskedastic SE model.
- Lee (2007a) proposed GMM methods for homoskedastic SL model.
- Lee (2007b) proposed modified GMM estimator for homoskedastic SL model, which reduces the joint maximization to the maximization w.r.t. SL parameter only.
- Lee & Liu (2010) extended the GMM method of Lee (2007a) to higher-order SLR models.
- Lin & Lee (2010) proposed a GMM method for estimating the SL model with unknown heteroskedasticity.
- Kelejian & Prucha (2010) proposed a three-step estimator for the SLE model with unknown heteroskedasticity.

Recall: the SL model $Y_n = \lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n$ in (6.2), its reduced form $Y_n = A_n^{-1} X_n\beta_0 + A_n^{-1}\epsilon_n$, $A_n(\lambda) = (I_n - \lambda W_n)$, and $F_n(\lambda) = W_n A_n^{-1}(\lambda)$.

- If $\|\lambda W_n\| < 1$ where $\|\cdot\|$ is a matrix norm, we have

$$(I_n - \lambda W_n)^{-1} = I_n + \lambda W_n + \lambda^2 W_n^2 + \cdots,$$

- from which the endogenous spatial lag term $W_n Y_n$ can be written as

$$W_n Y_n = W_n X_n\beta_0 + \lambda_0 W_n^2 X_n\beta_0 + \lambda_0^2 W_n^3 X_n\beta_0 + \cdots + F_n\epsilon_n. \qquad (6.27)$$

- The $n \times k^*$ matrix Q_n, constructed from X_n, $W_n X_n$, $W_n^2 X_n$, etc., forms the IV's for the deterministic part of $W_n Y_n$: $\mathrm{E}(W_n Y_n | X_n)$.
  Clearly, Q_n is correlated with $W_n Y_n$ but uncorrelated with ϵ_n;

- To increase efficiency of estimation, additional IV's for the stochastic part of $W_n Y_n$: $F_n\epsilon_n$, are needed. A valid candidate would be $P_n\epsilon_n$
  IF $\mathrm{Corr}(P_n\epsilon_n, F_n\epsilon_n) \neq 0$ but $\mathrm{Corr}(P_n\epsilon_n, \epsilon_n) = 0$!

The condition $\mathrm{Corr}(P_n\epsilon_n, \epsilon_n) = 0$ is true when $\epsilon_n \sim (0, \sigma_0^2 I_n)$ and $\mathrm{tr}(P_n) = 0$. This form the base of GMM estimation of Lee (2007a).

When $\epsilon_n \sim (0, \sigma_0^2 H_n)$ and $H_n \neq I_n$, the condition is true when $\mathrm{diag}(P_n) = 0$. This forms the base for CH-robust GMM estimation of Lin and Lee (2010).

Assume (i) the elements of $\mathbb{Q}_n$ are uniformly bounded, and (ii) the matrices $P_{jn}, j = 1, \ldots, m$, are such that $\mathrm{diag}(P_{jn}) = 0$, and are uniformly bounded in both row and column sum norms.

The set of moment functions for GMM estimation of $\vartheta = (\beta', \sigma^2)'$ is

$$\begin{aligned} \mathbf{g}_n(\vartheta) &= \big(\mathbb{Q}_n,\ P_{1n}\epsilon_n(\vartheta), \ldots, P_{mn}\epsilon_n(\vartheta)\big)'\epsilon_n(\vartheta) \\ &= \big(\epsilon_n'(\vartheta)\mathbb{Q}_n,\ \epsilon_n'(\vartheta)P_{1n}\epsilon_n(\vartheta), \ldots, \epsilon_n'(\vartheta)P_{mn}\epsilon_n(\vartheta)\big)', \end{aligned} \tag{6.28}$$

which is CH-robust as $\mathrm{E}[\mathbf{g}_n(\vartheta_0)] = 0$ under unknown heteroskedasticity.

The matrices P_{jn} can be W_n, $W_n^2 - \mathrm{diag}(W_n^2)$, $F_n - \mathrm{diag}(F_n)$, etc.

The CH-robust **GMM estimator** of ϑ_0 minimizes $\mathbf{g}_n'(\vartheta)\Omega_n\mathbf{g}_n(\vartheta)$, i.e.,

$$\tilde{\vartheta}_n = \text{argmin } \mathbf{g}_n'(\vartheta)\Omega_n\mathbf{g}_n(\vartheta),$$

for a **chosen** GMM weight matrix Ω_n or its feasible version.

Under regularity conditions (Lin and Lee, 2010), $\tilde{\vartheta}_n$ is asymptotically normal with mean zero and asymptotic variance (AVar):

$$\text{AVar}(\tilde{\vartheta}_n) = (\Sigma_n'\Omega_n\Sigma_n)^{-1}(\Sigma_n'\Omega_n\Gamma_n\Omega_n\Sigma_n)(\Sigma_n'\Omega_n\Sigma_n)^{-1},$$

where $\Sigma_n = -\text{E}[\frac{\partial}{\partial\vartheta'}\mathbf{g}_n(\vartheta_0)]$ and $\Gamma_n = \text{Var}[\mathbf{g}_n(\vartheta_0)]$, with

$$\Sigma_n = \begin{pmatrix} \mathbb{Q}_n'X_n & \mathbb{Q}_n'F_nX_n\beta_0 \\ 0 & \text{tr}(\mathbf{H}_nP_{1n}^sF_n) \\ \vdots & \vdots \\ 0 & \text{tr}(\mathbf{H}_nP_{mn}^sF_n) \end{pmatrix}, \quad \text{and} \qquad (6.29)$$

$$\Gamma_n = \begin{pmatrix} \mathbb{Q}_n' \mathbf{H}_n \mathbb{Q}_n & 0 & 0 & \cdots \\ 0 & \mathrm{tr}[\mathbf{H}_n P_{1n} (\mathbf{H}_n P_{1n})^s] & \mathrm{tr}[\mathbf{H}_n P_{1n} (\mathbf{H}_n P_{2n})^s] & \cdots \\ 0 & \mathrm{tr}[\mathbf{H}_n P_{2n} (\mathbf{H}_n P_{1n})^s] & \mathrm{tr}[\mathbf{H}_n P_{2n} (\mathbf{H}_n P_{2n})^s] & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 0 & \mathrm{tr}[\mathbf{H}_n P_{mn} (\mathbf{H}_n P_{1n})^s] & \mathrm{tr}[\mathbf{H}_n P_{mn} (\mathbf{H}_n P_{2n})^s] & \cdots \end{pmatrix}, \quad (6.30)$$

where $P_{kn}^s = P_{kn}' + P_{kn}, k = 1, \cdots, m$, and $\mathbf{H}_n = \sigma_0^2 H_n$.

- In practical applications, Σ_n and Γ_n are estimated by replacing ϑ_0 by $\hat{\vartheta}_n$ and $\mathbf{H}_n$ by $\widehat{\mathbf{H}}_n = \mathrm{diag}(\hat{\epsilon}_{n1}^2, \ldots, \hat{\epsilon}_{nn}^2)$ and $\{\hat{\epsilon}_{ni}\}$ are the residuals of the model with ϑ_0 being estimated by $\hat{\vartheta}_n$.
- When $\mathbb{Q}_n$ and P_{jn} contain unknown parameters, e.g., $\mathbb{Q}_n = (X_n, \ F_n X_n \beta_0)$ and $P_{jn} = F_n - \mathrm{diag}(F_n)$, their consistently estimated versions are used.
- "Optimal" RGMM uses Γ_n^{-1} as the weighting matrix. However,
- with unknown CH, the best choices of $\mathbb{Q}_n$ and P_{jn} are unavailable.

Kelejian & Prucha (2010) proposed a three-step estimation of the SLE model with unknown heteroskedasticity. First, write the model as

$$Y_n = Z_n\vartheta + u_n \tag{6.31}$$

$$u_n = \rho W_{2n} u_n + \epsilon_n \tag{6.32}$$

where $Z_n = (X_n, W_{1n}Y_n)$, and $\vartheta = (\beta', \lambda)'$. Assume $\epsilon_n \sim (0, \mathbf{H}_n)$.

The three-step estimation is summarized as follows.

1. Run a 2SLS on Model (6.31) using instruments Q_n to give $\tilde{\vartheta}_{\text{2SLS}}$.
2. Compute the 2SLS residuals $\tilde{u}_{2SLS} = Y_n - Z_n\tilde{\vartheta}_{\text{2SLS}}$. Using $\tilde{u}_{2SLS}$ for u_n in Model (6.32), perform a GM estimation of ρ to give $\tilde{\rho}_{\text{GM}}$, as described in Kelejian & Prucha (2010, Sec. 3.1).
3. Transform (6.31) as $B_n(\rho)Y_n = B_n(\rho)Z_n\vartheta + \epsilon_n$. Run another 2SLS on this transformed model after replacing ρ by $\tilde{\rho}_{\text{GM}}$.

Liu & Yang (2015) conduct Monte Carlo experiments to compare the finite sample performance of available estimators of SL model. They conclude:

(i) MQMLE of λ performs well in all cases considered, and it generally outperforms all other estimators in terms of bias and rmse. Further, in cases where QMLE is consistent, MQMLE can be significantly less biased than QMLE, and is as efficient as QMLE.

(ii) RGMME and ORGMME of λ perform reasonably well when $\beta = (3, 1, 1)'$, but deteriorates significantly when $\beta = (.3, .1, .1)'$ and in this case GMME and 2SLSE can be very erratic. In contrast, MQMLE is much less affected by the magnitude of β, and is less biased and more efficient than RGMME and ORGMME more significantly when $\beta = (.3, .1, .1)'$.

(iii) Root estimator (Jin & Lee, 2012) of λ performs equally well as MQMLE when $|\lambda|$ is not big and *n* is not small, but otherwise tends to give imaginary roots.

(iv) The GMM-type estimators can perform quite differently when the errors are normal as opposed to non-normal errors, especially when $\beta = (.3, .1, .1)'$. It is interesting to note that RGMME often outperforms the ORGMME.

(v) The OPG-based estimate of the robust standard errors of MQMLE of λ performs well in general with their values very close to their Monte Carlo counter parts.

(vi) Finally, the relative performance of various estimators of β is much less contrasting than that of various estimators of λ, although it can be seen that MQMLE of β is slightly more efficient RGMME and ORGMME.

Liu & Yang (2015) also compare the proposed MQMLE with the three-step estimator of Kelejian and Pruch (2010), and conclude:

- the modified QMLE has an excellent finite sample performance, and
- it outperforms the three-step estimator of Kelejian and Prucha (2010) from a combined consideration in terms of bias, consistency and efficiency.

Some results from Liu & Yang (2015) are reproduced in Table 6.1 (SL model) and Table 6.2 (SLE model) for illustration.

Table 6.1: Empirical Mean(rmse)[sd]{$\hat{sd}$} of Estimators of λ for SL Model
Case I of Inconsistent QMLE: Circular Neighbours (REG-1)

λ_0	n	QMLE	MQMLE	RGMM	ORGMM
DGP 2: $\beta_0 = (.3, .1, .1)'$					
.50	100	.416(.147)[.121]	.482(.123)[.121]{.119}	.475(.138)[.136]	.592(.342)[.329]
	250	.438(.101)[.080]	.490(.081)[.080]{.079}	.487(.090)[.089]	.528(.157)[.154]
	500	.448(.074)[.053]	.496(.053)[.053]{.052}	.494(.054)[.053]	.511(.068)[.067]
	1000	.452(.061)[.038]	.499(.038)[.038]{.037}	.498(.038)[.038]	.508(.047)[.047]
.25	100	.184(.152)[.137]	.236(.154)[.154]{.157}	.224(.165)[.163]	.304(.305)[.301]
	250	.203(.100)[.088]	.242(.097)[.097]{.091}	.236(.099)[.098]	.271(.149)[.147]
	500	.211(.073)[.062]	.246(.067)[.067]{.066}	.243(.068)[.068]	.264(.109)[.109]
	1000	.217(.055)[.044]	.250(.048)[.048]{.047}	.249(.048)[.048]	.258(.058)[.058]
.00	100	-.040(.144)[.139]	-.021(.171)[.169]{.164}	-.039(.180)[.176]	.014(.262)[.262]
	250	-.016(.091)[.089]	-.010(.107)[.107]{.104}	-.016(.109)[.108]	.008(.134)[.134]
	500	-.007(.063)[.063]	-.003(.075)[.075]{.074}	-.006(.075)[.075]	.008(.090)[.090]
	1000	-.003(.046)[.046]	-.001(.054)[.054]{.053}	-.003(.054)[.054]	.006(.066)[.066]
-.25	100	-.232(.133)[.131]	-.259(.169)[.169]{.159}	-.281(.180)[.177]	-.254(.266)[.266]
	250	-.216(.090)[.083]	-.254(.106)[.106]{.107}	-.262(.108)[.107]	-.249(.138)[.138]
	500	-.210(.073)[.061]	-.251(.077)[.077]{.077}	-.255(.077)[.077]	-.246(.088)[.088]
	1000	-.207(.063)[.046]	-.249(.057)[.057]{.055}	-.251(.057)[.057]	-.247(.067)[.067]
-.50	100	-.424(.148)[.127]	-.503(.163)[.163]{.160}	-.535(.191)[.187]	-.549(.246)[.241]
	250	-.410(.123)[.084]	-.499(.105)[.105]{.099}	-.507(.106)[.105]	-.513(.151)[.151]
	500	-.409(.108)[.058]	-.500(.071)[.071]{.072}	-.504(.071)[.071]	-.507(.086)[.086]
	1000	-.409(.100)[.041]	-.503(.050)[.050]{.051}	-.506(.051)[.050]	-.509(.063)[.062]

Table 6.1: Cont'd

λ_0	n	QMLE	MQMLE	RGMM	ORGMM
			DGP 3: $\beta_0 = (.3, .1, .1)'$		
.50	100	.416(.147)[.120]	.480(.118)[.116]{.099}	.473(.130)[.128]	.652(.453)[.426]
	250	.439(.096)[.074]	.490(.071)[.070]{.065}	.486(.073)[.071]	.572(.247)[.236]
	500	.449(.074)[.054]	.497(.050)[.050]{.048}	.495(.051)[.051]	.547(.189)[.184]
	1000	.453(.060)[.037]	.498(.034)[.034]{.035}	.497(.035)[.034]	.523(.104)[.101]
.25	100	.174(.153)[.133]	.224(.147)[.144]{.137}	.212(.156)[.152]	.335(.387)[.378]
	250	.210(.089)[.080]	.249(.087)[.087]{.083}	.243(.087)[.087]	.310(.245)[.237]
	500	.211(.072)[.061]	.244(.065)[.065]{.061}	.242(.066)[.065]	.283(.198)[.195]
	1000	.214(.057)[.044]	.247(.046)[.046]{.044}	.246(.047)[.046]	.266(.116)[.115]
.00	100	-.027(.135)[.133]	-.008(.161)[.160]{.153}	-.026(.172)[.170]	.077(.422)[.414]
	250	-.014(.087)[.086]	-.006(.103)[.103]{.099}	-.013(.105)[.104]	.052(.263)[.258]
	500	-.008(.059)[.058]	-.004(.070)[.070]{.069}	-.008(.071)[.070]	.026(.151)[.149]
	1000	-.003(.042)[.042]	-.001(.050)[.050]{.050}	-.003(.050)[.050]	.025(.116)[.114]
-.25	100	-.234(.131)[.130]	-.262(.172)[.172]{.179}	-.288(.184)[.180]	-.238(.295)[.295]
	250	-.218(.090)[.084]	-.254(.105)[.105]{.099}	-.262(.107)[.106]	-.223(.239)[.238]
	500	-.213(.073)[.063]	-.252(.076)[.076]{.071}	-.256(.077)[.076]	-.233(.161)[.160]
	1000	-.208(.062)[.046]	-.250(.055)[.055]{.053}	-.252(.055)[.055]	-.238(.128)[.127]
-.50	100	-.418(.151)[.127]	-.495(.158)[.158]{.151}	-.526(.178)[.176]	-.544(.304)[.301]
	250	-.411(.126)[.089]	-.503(.105)[.105]{.099}	-.511(.105)[.104]	-.508(.199)[.198]
	500	-.408(.113)[.066]	-.500(.073)[.073]{.069}	-.504(.072)[.072]	-.501(.156)[.156]
	1000	-.403(.109)[.049]	-.496(.051)[.051]{.049}	-.498(.051)[.051]	-.502(.129)[.129]

Table 6.2: Empirical Mean(rmse)[sd] of Estimators of λ and ρ for SARAR(1,1) Model

Case I of Inconsistent QMLEs: Circular Neighbours (REG-1)

Par	QMLE-λ	MQMLE-λ	KP-λ	QMLE-ρ	MQMLE-ρ	KP-ρ
			DGP 1: $\beta_0 = (3, 1, 1)'$			
1-1	.470(.141)[.138]	.472(.197)[.195]	.578(.219)[.204]	.409(.195)[.172]	.446(.237)[.231]	.335(.341)[.299]
	.484(.080)[.078]	.482(.118)[.117]	.528(.109)[.105]	.445(.116)[.102]	.488(.140)[.139]	.479(.180)[.179]
	.487(.065)[.064]	.489(.097)[.097]	.515(.093)[.092]	.454(.088)[.075]	.491(.110)[.109]	.512(.156)[.156]
	.490(.043)[.042]	.495(.060)[.059]	.505(.057)[.057]	.458(.066)[.051]	.497(.070)[.070]	.533(.103)[.097]
1-2	.372(.173)[.116]	.418(.233)[.218]	.494(.143)[.143]	-.307(.249)[.158]	-.505(.252)[.239]	-.507(.244)[.244]
	.411(.109)[.063]	.488(.095)[.094]	.501(.072)[.072]	-.324(.202)[.100]	-.502(.153)[.153]	-.492(.150)[.150]
	.400(.112)[.050]	.498(.071)[.071]	.498(.060)[.060]	-.305(.208)[.072]	-.504(.126)[.125]	-.476(.121)[.119]
	.421(.084)[.030]	.502(.047)[.047]	.499(.035)[.035]	-.321(.186)[.051]	-.506(.109)[.108]	-.470(.083)[.078]
2-1	.280(.144)[.141]	.250(.200)[.200]	.333(.239)[.224]	.374(.208)[.165]	.441(.225)[.217]	.358(.307)[.272]
	.292(.095)[.086]	.253(.128)[.127]	.297(.133)[.124]	.399(.140)[.097]	.470(.135)[.131]	.464(.176)[.172]
	.293(.080)[.067]	.252(.106)[.106]	.276(.105)[.101]	.408(.119)[.075]	.491(.109)[.107]	.499(.146)[.146]
	.287(.057)[.043]	.250(.064)[.064]	.259(.064)[.064]	.421(.093)[.049]	.494(.065)[.065]	.524(.092)[.089]
2-2	.113(.189)[.130]	.233(.188)[.163]	.235(.186)[.186]	-.330(.231)[.156]	-.582(.269)[.249]	-.507(.259)[.259]
	.156(.120)[.074]	.239(.131)[.131]	.248(.092)[.092]	-.337(.188)[.095]	-.503(.209)[.209]	-.484(.151)[.150]
	.140(.125)[.059]	.248(.099)[.099]	.247(.079)[.079]	-.319(.193)[.069]	-.510(.115)[.114]	-.484(.117)[.116]
	.164(.093)[.036]	.250(.052)[.052]	.250(.045)[.045]	-.332(.175)[.047]	-.501(.102)[.101]	-.475(.080)[.076]

Note: (i) Par = i-j, where '$i = 1, 2, 3, 4, 5$' represents '$\lambda = .5, .25$'; '$j = 1, 2$' represents '$\rho = .5, -.5$'.

(ii) Under each Par setting, $n = 100, 250, 500, 1000$, corresponding to the four rows.

(iii) KP denotes Kelejian and Prucha's (2010) three-step estimator.

6.6. CH-Robust LM-Type Tests for Spatial Effects

As argued in Lecture 3, for testing the existence of spatial effects in a linear regression model, the LM test is preferred as it requires only the estimation of null model, in particular when it is an OLS regression.

- However, the regular LM tests for spatial effects may perform poorly in finite samples, and may not be robust against nonnormality.
- A simple way to improve the regular LM tests is **standardization** (see Baltagi and Yang, 2013a, and the references therein).
- Furthermore, LM or SLM tests may not be robust against unknown CH. An outer-product-of-martingale-difference (OPMD) version of the SLM test can be obtained through a martingale differences (M.D.) representation, making it robust against unknown heteroskedasticity.
- The latter property is extremely useful in the context of spatial models where unknown heteroskedasticity may be a standard feature (see Baltagi and Yang 2013b, and the references therein).

Consider the three tests associated with three models:

- $H_0^{\mathrm{SL}}: \lambda = 0$, in the SL model;
- $H_0^{\mathrm{SE}}: \rho = 0$, in the SE model;
- $H_0^{\mathrm{SLE}}: \lambda = \rho = 0$, in the SLE model.

Various tests for these three hypotheses and others have been presented in Lecture 3, and extensive discussions are given therein. However, the studies in Lecture 3 are limited to the case of homoskedasticity.

In this section, we introduce LM-type tests that are robust against both unknown heteroskedasticity and nonnormality.

The corresponding LM tests, due to by Anselin (1988a,b) and Burridge (1980), which are presented in Lecture 3 separately for each model, are summarized below using unified notation:

$$\mathrm{LM}^{\mathtt{FI}}_{\mathtt{SL}} = \frac{\tilde{\epsilon}'_n W_{1n} Y_n}{\tilde{\sigma}^2_n \, (\tilde{D}_n + T_{1n})^{\frac{1}{2}}}, \tag{6.33}$$

$$\mathrm{LM}^{\mathtt{FI}}_{\mathtt{SE}} = \frac{\tilde{\epsilon}'_n W_{2n} \tilde{\epsilon}_n}{\tilde{\sigma}^2_n \, T^{\frac{1}{2}}_{2n}}, \tag{6.34}$$

$$\mathrm{LM}^{\mathtt{FI}}_{\mathtt{SLE}} = \frac{1}{\tilde{\sigma}^4_n} \begin{pmatrix} \tilde{\epsilon}'_n W_{1n} Y_n \\ \tilde{\epsilon}'_n W_{2n} \tilde{\epsilon}_n \end{pmatrix}' \begin{pmatrix} T_{1n} + \tilde{D}_n & T_{3n} \\ T_{3n} & T_{2n} \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\epsilon}'_n W_{1n} Y_n \\ \tilde{\epsilon}'_n W_{2n} \tilde{\epsilon}_n \end{pmatrix}, \tag{6.35}$$

- $\tilde{\varepsilon}_n$ are the OLS residuals from regressing Y_n on X_n,
- $\tilde{\beta}_n$ and $\tilde{\sigma}^2_n$ are the OLS estimators of β and σ^2, respectively,
- $T_{rn} = \mathrm{tr}[(W_{rn} + W'_{rn}) W_{rn}], r = 1, 2,$
- $T_{3n} = \mathrm{tr}[(W_{2n} + W'_{2n}) W_{1n}],$
- $\tilde{D}_n = \tilde{\sigma}_n^{-2} (W_n X_n \tilde{\beta}_n)' M_n W_n X_n \tilde{\beta}_n, \quad M_n = I_n - X_n (X'_n X_n)^{-1} X'_n.$

Born and Breitung (2011) proposed OPMD variants of the above LM tests:

$$\mathrm{LM}_{\mathrm{SL}}^{\mathrm{MD}} = \frac{\tilde{\epsilon}_n' W_{1n} Y_n}{(\tilde{\epsilon}_n^{2\,\prime}\, \tilde{\xi}_{1n}^2)^{\frac{1}{2}}}, \tag{6.36}$$

$$\mathrm{LM}_{\mathrm{SE}}^{\mathrm{MD}} = \frac{\tilde{\epsilon}_n' W_{2n} \tilde{\epsilon}_n}{(\tilde{\epsilon}_n^{2\,\prime}\, \tilde{\xi}_{2n}^2)^{\frac{1}{2}}}, \text{ and} \tag{6.37}$$

$$\mathrm{LM}_{\mathrm{SLE}}^{\mathrm{MD}} = \begin{pmatrix} \tilde{\epsilon}_n' W_{1n} Y_n, \\ \tilde{\epsilon}_n' W_{2n} \tilde{\epsilon}_n \end{pmatrix}' \begin{pmatrix} \tilde{\epsilon}_n^{2\,\prime} \tilde{\xi}_{1n}^2 & \tilde{\epsilon}_n^{2\,\prime}(\tilde{\xi}_{1n} \odot \tilde{\xi}_{2n}) \\ \sim & \tilde{\epsilon}_n^{2\,\prime} \tilde{\xi}_{2n}^2 \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\epsilon}_n' W_{1n} Y_n \\ \tilde{\epsilon}_n' W_{2n} \tilde{\epsilon}_n \end{pmatrix}, \tag{6.38}$$

where $\odot$ denotes Hadamard product, square of a vector, e.g., $\tilde{\epsilon}_n^2 = \tilde{\epsilon}_n \odot \tilde{\epsilon}_n$,

$$\tilde{\xi}_{1n} = (W_{1n}^{u\prime} + W_{1n}^{l})\tilde{\epsilon}_n + M_n W_n X_n \tilde{\beta}_n, \tag{6.39}$$

$$\tilde{\xi}_{2n} = (W_{2n}^{u\prime} + W_{2n}^{l})\tilde{\epsilon}_n, \tag{6.40}$$

with W_{rn}^u and W_{rn}^l: the upper and lower triangular matrices of W_{rn}.

- The three $\mathrm{LM}^{\mathrm{MD}}$ statistics are robust against heteroskedasticity and non-normality, due to fact that the diagonal elements of W_{rn} are zero and the use of OPMD variance estimates.
- The three $\mathrm{LM}^{\mathrm{FI}}$ statistics are robust only against nonnormality.

The OPMD variants of the LM tests considered by Born and Breitung (2011) (as well as the original LM tests) do not take into account the estimation of β and σ^2, and hence may suffer from the problems of size distortion due mainly to the lack of centering and rescaling.

Note that the numerators of the two sets of tests above are of the forms:

$$\tilde{\epsilon}_n' W_{1n} Y_n = \epsilon_n' M_n W_{1n} \epsilon_n + \epsilon_n' M_n \eta_n \equiv \epsilon_n' A_{1n} \epsilon_n + \epsilon_n' M_n \eta_n,$$
$$\tilde{\epsilon}_n' W_{2n} \tilde{\epsilon}_n = \epsilon_n' M_n W_{2n} M_n \epsilon_n \equiv \epsilon_n' A_{2n} \epsilon_n.$$

When the errors are heteroskedastic, i.e., $\mathrm{Var}(\epsilon_{n,i}) = \sigma_i^2, i = 1, \ldots, n$,

$$\mathrm{E}(\tilde{\epsilon}_n' W_{1n} Y_n) = \sum_{i=1}^n \sigma_i^2 a_{1n,i} \neq 0 \quad \text{and} \quad \mathrm{E}(\tilde{\epsilon}_n' W_{2n} \tilde{\epsilon}_n) = \sum_{i=1}^n \sigma_i^2 a_{2n,i} \neq 0,$$

where $\{a_{rn,i}\}$ are the diagonal elements of $A_{rn}, r = 1, 2$.

While these non-zero means are asymptotically negligible, they may have significant effect on the finite sample performance of these test statistics — finite sample null distributions differ from limiting distributions, i.e., $N(0, 1)$ or χ_2^2.

Standardization and OPMD estimate of variance may help (i) improving the finite sample performance of the LM tests, and (ii) making the tests robust against unknown heteroskedasticity and nonnormality.

To center the key quantities so that they have zero mean under CH, write

$$\tilde{\epsilon}_n' W_{1n} Y_n = Y_n' M_n W_{1n} Y_n \equiv Y_n' M_n A_{1n} Y_n, \tag{6.41}$$

$$\tilde{\epsilon}_n' W_{2n} \tilde{\epsilon}_n = Y_n' M_n W_{2n} M_n Y_n \equiv Y_n' M_n A_{2n} Y_n, \tag{6.42}$$

where $A_{1n} = W_{1n}$ and $A_{2n} = W_{2n} M_n$.

Referring to the transitions from (6.23) to (6.24), define

$$A_{rn}^{\circ} = A_{rn} - \mathrm{diag}(M_n)^{-1} \mathrm{diag}(M_n A_{rn}), \; r = 1, 2.$$

Replacing A_{rn} in $Y_n' M_n A_{rn} Y_n$ by A_{rn}°, we obtain the centered quantities:

$$Y_n' M_n A_{rn}^{\circ} Y_n, \quad r = 1, 2,$$

which form the base for constructing standardized LM tests robust to CH.

To estimate the variances of $Y_n' M_n A_{rn}^{\circ} Y_n$ for constructing the standardized LM tests, either the plug-in method or the OPMD method can be followed.

The OPMD method is given below.

Note that $\mathrm{diag}(M_n A_{rn}^{\circ}) = 0$ and $\mathrm{E}(Y_n' M_n A_{rn}^{\circ} Y_n) = 0$ under the null. Decompose $M_n A_{rn}^{\circ} \equiv \mathbb{A}_{rn}^{\circ} = \mathbb{A}_{rn}^{\circ u} + \mathbb{A}_{rn}^{\circ l}$. Let

$$\xi_{1n}^{\circ} = (\mathbb{A}_{rn}^{\circ u\prime} + \mathbb{A}_{rn}^{\circ l})\epsilon_n + c_{rn}, \quad r = 1, 2,$$

where $c_{rn} = M_n A_{rn}^{\circ} X_n \beta_0, \quad r = 1, 2.$ Therefore,

$$Y_n' M_n A_{rn}^{\circ} Y_n = \sum_{i=1}^{n} \epsilon_{ni} \xi_{rni}, \quad r = 1, 2.$$

It is easy to see that $\{\epsilon_{ni}\xi_{rni}\}$ form an M.D. sequence, and hence

$$\mathrm{Var}(Y_n' M_n A_{rn}^{\circ} Y_n) = \sum_{i=1}^{n} \mathrm{E}(\epsilon_{ni}^2 \xi_{rni}^2), \ r = 1, 2;$$

and that $\{\epsilon_{ni}\xi_{1ni}, \ \epsilon_{ni}\xi_{2ni}\}$ form a vector M.D. sequence, and

$$\mathrm{Var}\begin{pmatrix} Y_n' M_n A_{1n}^{\circ} Y_n \\ Y_n' M_n A_{2n}^{\circ} Y_n \end{pmatrix} = \sum_{i=1}^{n} \mathrm{E}\begin{pmatrix} \epsilon_{ni}^2 \xi_{1ni}^2 & \epsilon_{ni}^2 \xi_{1ni}\xi_{2ni} \\ \epsilon_{ni}^2 \xi_{1ni}\xi_{2ni} & \epsilon_{ni}^2 \xi_{2ni}^2 \end{pmatrix}.$$

These lead to the standardized LM tests for testing $H_0^{\texttt{SL}}$, $H_0^{\texttt{SE}}$ and $H_0^{\texttt{SLE}}$, robust to both nonnormality and unknown heteroskedasticity:

$$\mathrm{SLM}_{\texttt{SL}}^{\texttt{MDh}} = \frac{Y_n' M_n A_{1n}^{\circ} Y_n}{(\tilde{\epsilon}_n^{2\,\prime}\, \tilde{\xi}_{1n}^{*2})^{\frac{1}{2}}}, \tag{6.43}$$

$$\mathrm{SLM}_{\texttt{SE}}^{\texttt{MDh}} = \frac{Y_n' M_n A_{2n}^{\circ} Y_n}{(\tilde{\epsilon}_n^{2\,\prime}\, \tilde{\xi}_{2n}^{*2})^{\frac{1}{2}}}, \text{ and} \tag{6.44}$$

$$\mathrm{SLM}_{\texttt{SLE}}^{\texttt{MDh}} = S_n' \begin{pmatrix} \tilde{\epsilon}_n^{2\,\prime} \tilde{\xi}_{1n}^{\circ 2} & \tilde{\epsilon}_n^{2\,\prime}(\tilde{\xi}_{1n}^{\circ} \odot \tilde{\xi}_{2n}^{\circ}) \\ \tilde{\epsilon}_n^{2\,\prime}(\tilde{\xi}_{1n}^{\circ} \odot \tilde{\xi}_{2n}^{\circ}) & \tilde{\epsilon}_n^{2\,\prime} \tilde{\xi}_{2n}^{\circ 2} \end{pmatrix}^{-1} S_n, \tag{6.45}$$

where $S_n = \{Y_n' M_n A_{1n}^{\circ} Y_n; \;\; Y_n' M_n A_{2n}^{\circ} Y_n\}$, and $\tilde{\epsilon}_n$ and $\tilde{\xi}_{rn}^{\circ}$ are the OLS estimates of ϵ_n and ξ_{rn}°, $r = 1, 2$.

- These tests are similar to the three SLM tests given in Baltagi & Yang (2013b); but they further simplify and potentially improve the latter.
- Under the assumptions given in Baltagi and Yang (2013b) and H_0, $\mathrm{SLM}_{\texttt{SL}}^{\texttt{MDh}} \xrightarrow{D} N(0,1)$, $\mathrm{SLM}_{\texttt{SE}}^{\texttt{MDh}} \xrightarrow{D} N(0,1)$, and $\mathrm{SLM}_{\texttt{SLE}}^{\texttt{MDh}} \xrightarrow{D} \chi_2^2$.

Here '`MDh`' denotes 'heteroskedasticity robust M.D.' form of an SLM test.

Baltagi and Yang (2013b) perform extensive Monte Carlo experiments to investigate finite sample performance of various LM-type tests. We replicate here (Table 6.2) some of their results for the case of testing H_0^{SLE} using two variants of LM tests and the SLM test they proposed.

The results reveal the following:

- The $\mathrm{SLM}_{\mathrm{SLE}}^{\mathrm{MD}}$ test dominates $\mathrm{LM}_{\mathrm{SLE}}^{\mathrm{FI}}$ and $\mathrm{LM}_{\mathrm{SLE}}^{\mathrm{MD}}$, with Monte Carlo means, sds, and rejection rates being very close to their nominal values: (2, 2, 10%, 5%, 1%).
- The $\mathrm{LM}_{\mathrm{SLE}}^{\mathrm{FI}}$ test is not robust against unknown CH, which is clearly shown by the results in Table 6.3.
- The $\mathrm{LM}_{\mathrm{SLE}}^{\mathrm{MD}}$ does not perform well enough even when n is quite large, although it does seem to be robust against nonnormality and unknown heteroskedasticity, consistent with what the theory suggests.

Table 6.3. Mean, sd, and Rejection Frequency: Joint LM Tests for SLE Dependence
W_{1n}=Queen, $r = 5$; W_{2n}=Group, $g = n^{0.5}$; XVal-B; **Normal Errors**

n	Heteroskedasticity = $\|X_1\|$					Heteroskedasticity = $2\|X_1\|$				
	mean	sd	10%	5%	1%	mean	sd	10%	5%	1%
50	2.2886	1.4830	.0613	.0180	.0018	2.4416	2.1771	.1081	.0472	.0138
	2.6891	1.8687	.1494	.0592	.0051	2.4771	1.8823	.1364	.0517	.0052
	2.2328	1.9698	.1201	.0539	.0080	2.2385	1.8381	.1149	.0455	.0037
100	2.3192	2.1102	.1014	.0503	.0116	2.2836	1.8988	.0903	.0378	.0091
	2.5021	2.0250	.1450	.0686	.0081	2.5870	2.0651	.1557	.0724	.0091
	2.1200	1.8979	.1038	.0478	.0063	2.1832	1.9700	.1117	.0498	.0081
200	2.5567	2.0947	.1286	.0505	.0103	2.8150	2.9793	.1686	.0766	.0185
	2.4096	2.0472	.1366	.0683	.0099	2.5063	2.1511	.1528	.0763	.0123
	2.2554	2.1731	.1246	.0653	.0140	2.1532	2.0063	.1121	.0540	.0087
500	2.7570	2.8166	.1743	.0934	.0244	2.6424	2.7786	.1593	.0820	.0211
	2.3415	2.1658	.1389	.0697	.0130	2.2700	2.1095	.1305	.0636	.0113
	2.1228	2.0155	.1130	.0557	.0088	2.1090	2.0027	.1098	.0539	.0089
1000	2.3977	2.3921	.1318	.0676	.0181	2.5871	2.5542	.1605	.0831	.0209
	2.2587	2.1948	.1284	.0687	.0137	2.2352	2.1468	.1264	.0657	.0126
	2.0670	2.0385	.1059	.0556	.0107	2.0765	2.0116	.1094	.0520	.0102

Three rows under each n: $\text{LM}^{\text{FI}}_{\text{SLE}}$, $\text{LM}^{\text{MD}}_{\text{SLE}}$ and $\text{SLM}^{\text{MD}}_{\text{SLE}}$ (Baltagi & Yang 2013b).

Table 6.3, Cont'd, Normal Mixture Errors

	Heteroskedasticity = $\|X_1\|$					Heteroskedasticity = $2\|X_1\|$				
n	mean	sd	10%	5%	1%	mean	sd	10%	5%	1%
50	2.1706	1.4789	.0546	.0164	.0018	2.1743	1.8616	.0895	.0394	.0074
	2.5894	1.7446	.1265	.0465	.0032	2.3768	1.7502	.1132	.0437	.0023
	2.1809	1.8714	.1055	.0479	.0073	2.1529	1.7201	.0939	.0347	.0020
100	2.1496	1.9206	.0894	.0430	.0089	2.1344	1.7725	.0838	.0347	.0056
	2.3976	1.8953	.1244	.0555	.0058	2.4772	1.9166	.1339	.0583	.0059
	2.1028	1.7966	.0969	.0392	.0049	2.1594	1.8613	.1032	.0430	.0061
200	2.3802	1.9236	.1119	.0439	.0080	2.5839	2.3989	.1525	.0737	.0153
	2.3394	1.9379	.1228	.0562	.0071	2.4137	1.9868	.1349	.0622	.0082
	2.2335	2.1201	.1207	.0601	.0117	2.1232	1.8890	.1025	.0454	.0064
500	2.6161	2.6828	.1591	.0845	.0209	2.5565	2.4927	.1556	.0791	.0189
	2.2481	1.9890	.1179	.0550	.0080	2.2760	2.0051	.1237	.0566	.0097
	2.0777	1.9016	.0989	.0467	.0073	2.1270	1.9423	.1064	.0498	.0075
1000	2.3695	2.3712	.1342	.0676	.0158	2.5007	2.5717	.1535	.0831	.0200
	2.2267	2.0895	.1216	.0611	.0101	2.1974	2.0589	.1197	.0590	.0098
	2.0651	1.9807	.1061	.0500	.0085	2.0582	1.9657	.1044	.0483	.0083

Three rows under each n: $\text{LM}_{\texttt{SLE}}^{\texttt{FI}}$, $\text{LM}_{\texttt{SLE}}^{\texttt{MD}}$ and $\text{SLM}_{\texttt{SLE}}^{\texttt{MD}}$ (Baltagi & Yang 2013b).

Table 6.3, Cont'd, Lognormal Errors

n	Heteroskedasticity = $\lvert X_1 \rvert$					Heteroskedasticity = $2\lvert X_1 \rvert$				
	mean	sd	10%	5%	1%	mean	sd	10%	5%	1%
50	2.1689	1.5147	.0527	.0169	.0017	2.0630	1.8949	.0755	.0330	.0085
	2.5160	1.7353	.1234	.0444	.0027	2.3862	1.8016	.1196	.0478	.0028
	2.0377	1.7615	.0890	.0343	.0049	2.1161	1.7444	.0926	.0392	.0037
100	2.1353	2.2535	.0821	.0434	.0143	2.1039	1.9902	.0768	.0338	.0086
	2.4211	1.9868	.1300	.0642	.0088	2.5107	2.0459	.1426	.0703	.0103
	2.1930	1.9544	.1164	.0524	.0083	2.2294	2.0110	.1099	.0545	.0110
200	2.5451	2.5838	.1261	.0561	.0168	2.4693	2.3420	.1321	.0578	.0129
	2.5126	2.1653	.1462	.0792	.0143	2.4749	2.1352	.1432	.0683	.0142
	2.3739	2.3203	.1383	.0754	.0200	2.2189	2.0168	.1157	.0518	.0102
500	2.5566	2.7368	.1442	.0771	.0241	2.3771	2.5619	.1310	.0631	.0172
	2.3736	2.1520	.1340	.0685	.0134	2.3298	2.1134	.1322	.0649	.0122
	2.1850	2.0532	.1157	.0558	.0105	2.1413	1.9842	.1070	.0533	.0087
1000	2.2785	2.6453	.1146	.0591	.0167	2.4533	2.8658	.1379	.0706	.0211
	2.2545	2.1161	.1254	.0647	.0133	2.3116	2.1562	.1295	.0688	.0133
	2.0782	1.9999	.1052	.0534	.0098	2.1387	2.0244	.1103	.0541	.0091

Three rows under each n: $\text{LM}^{\text{FI}}_{\text{SLE}}$, $\text{LM}^{\text{MD}}_{\text{SLE}}$ and $\text{SLM}^{\text{MD}}_{\text{SLE}}$ (Baltagi & Yang 2013b).

6.7. Empirical Applications

References

- Amemiya, T., 1985. *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.
- Anselin, L. 1988a. Lagrange multiplier test diagnostics for spatial dependence and heterogeneity. *Geographical Analysis* **20**, 1-17.
- Anselin, L., 1988b. *Spatial Econometrics: Methods and Models*. The Netherlands: Kluwer Academic Publishers.
- Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26, 153-166.
- Baltagi, B. H., Pirotte, A. and Yang, Z. L. (2021). Diagnostic tests for homoscedasticity in spatial cross-sectional or panel models. *Journal of Econometrics* 224, 245-270.
- Baltagi, B., Yang, Z. L., 2013a. Standardized LM tests for spatial error dependence in linear or panel regressions. *The Econometrics Journal* 16 103-134.

- Baltagi, B., Yang, Z. L., 2013b. Heteroskedasticity and non-normality robust LM tests of spatial dependence. *Regional Science and Urban Economics* 43, 725-739.
- Born, B., Breitung, J., 2011. Simple regression based tests for spatial dependence. *Econometrics Journal* 14, 330-342.
- Breusch, T., Pagan, A., 1979. A simple test for heteroskedasticity and random coefficient variation. *Econometrica* 47, 1287-1294.
- Burridge, P. 1980. On the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society* B 42, 107-8.
- Glaeser, E. L., Sacerdote, B., and Scheinkman, J. A. (1996). Crime and social interactions. *Quarterly Journal of Economics*, **111**, 507-548.
- Kelejian H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.
- Kelejian H. H., Prucha, I. R., 2010 Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53-67.

- Lee, L. F., 2004a. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.
- Lee, L. F., 2007a. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489-514.
- Lee, L. F., 2007b. The method of elimination and substitution in the GMM estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 140, 155-189.
- LeSage, J. P., Pace, R.K., 2009. *Introduction to Spatial Econometrics*. CRC Press, Taylor & Francis Group, London.
- Lin, X., Lee, L. F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157, 34-52.
- Liu, S. F., Yang, Z. L., 2015. Modified QML Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity and Nonnormality. *Regional Science and Urban Economics*. 52, 50-70.
- Ord, J., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.