# Lecture 2: Spatial Linear Regression Models

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

Lecture 2 introduces popular spatial linear regression models and their estimation through quasi maximum likelihood (QML) method, and GM or GMM method, which include the linear regression models with

1. spatial lag (SL) dependence,
2. spatial error (SE) dependence,
3. both SL and SE (SLE),

Where spatial Durbin (SD) effect can be added to any of the three models.

- Asymptotic properties of the QML estimators are presented.
- Method of estimating standard errors of the QMLE is introduced.
- Finite sample performance of the QML-based methods is discussed based on Monte Carlo results.
- Two examples are presented to illustrate the applications of QML-based methods.

**Model.** Spatial linear regression (SLR) model with spatial error (SE) dependence, or the SE model in short, takes the following form:

$$Y_n = X_n\beta + u_n, \quad u_n = \rho W_n u_n + \epsilon_n, \tag{2.1}$$

where the SE structure is a spatial autoregressive (SAR) process. It can also be a spatial moving average (SMA) process: $u_n = \rho W_n \epsilon_n + \epsilon_n$,

- $Y_n$: $n \times 1$ vector of observations on $n$ spatial units,
- $X_n$: an $n \times k$ matrix containing the values of $k$ regressors,
- $W_n$: $n \times n$ matrix summarizing interactions among $n$ spatial units, called the spatial weight matrix or the spatial interaction matrix,
- $\epsilon_n$: $n \times 1$ vector of independent and identically distributed (iid) idiosyncratic errors with mean zero and variance $\sigma^2$,
- $\rho$: the spatial error parameter,
- $\beta$: $k \times 1$ vector of regression coefficients.

**Durbin-SE Model.** The model (2.1) can be extended by adding a spatial Durbin term $W_n X_n^*$, where $X_n^*$ contains **some** regressors, referred to in this course as Durbin-SE model:

$$Y_n = X_n\beta + W_n X_n^* \beta^* + u_n, \quad u_n = \rho W_n u_n + \epsilon_n. \tag{2.2}$$

The term '**spatial Durbin model**' was first appeared in Anselin (1988) for its analogy with Durbin (1960) for time series. See also Elhorst (2014, p.7).

- Elhorst (2014) interprets the spatial Durbin effect as the *exogenous interaction effects*, where the dependent variable of a particular unit depends on independent variables of other units.
- By defining a new regressor matrix $\mathbb{X}_n = [X_n,\ W_n X_n^*]$, and a new vector of regression coefficients $\boldsymbol{\beta} = (\beta', \beta^{*\prime})'$, Durbin-SE model has the same form as the regular SE model, and hence inference methods remain the same.
- However, a problem of particular interest is to test the existence of spatial Durbin effect, i.e., testing $H_0$: $\beta^* = 0$, which can be carried out based on the usual tests for covariate effects, to be presented in Lecture 3.

For QML estimation of spatial models, the following results are useful:

**(i)** An $n$-dimensional random vector $\mathbf{Y}$ is said to have a multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance-covariance (VC) matrix $\boldsymbol{\Sigma}$, denoted as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its joint probability density function (pdf) takes the form:

$$f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right). \tag{2.3}$$

where $|\cdot|$ denotes the determinant of a square matrix.

**(ii)** For two $n \times n$ matrices $A$ and $B$, and a scalar $c$:

$$|A'| = |A|, \quad |A^{-1}| = |A|^{-1}, \quad |cA| = c^n|A|, \quad |AB| = |A||B|;$$

**(iii)** For a matrix function $A(\rho)$ of scalar $\rho$, positive definite (p.d.),

$\frac{\partial}{\partial\rho}A(\rho)^{-1} = -A(\rho)^{-1}[\frac{\partial}{\partial\rho}A(\rho)]A(\rho)^{-1}$, Horn and Johnson (1985).

$\frac{\partial}{\partial\rho}\log|A(\rho)| = \mathrm{tr}[A(\rho)^{-1}\frac{\partial}{\partial\rho}A(\rho)]$, $\mathrm{tr}(\cdot) =$ trace of a matrix.

Now, define $B_n(\rho) = I_n - \rho W_n$. Then, $u_n = B_n^{-1}(\rho)\epsilon_n$, and

$$\Sigma_n = \mathrm{Var}(u_n) = \sigma^2 B_n^{-1}(\rho) B_n'^{-1}(\rho).$$

By (2.3), the **quasi** Gaussian loglikelihood function of $\theta = (\beta', \sigma^2, \rho)'$ for the SE or Durbin-SE model, **as if** $\epsilon_{n,i}$ are iid normal, is given by,

$$\ell_n(\theta) = -\tfrac{n}{2}\log(2\pi\sigma^2) + \log|B_n(\rho)| - \tfrac{1}{2\sigma^2} u_n'(\beta) B_n'(\rho) B_n(\rho) u_n(\beta). \quad (2.4)$$

where $u_n(\beta) = Y_n - X_n\beta$. It is assumed: $|B_n(\rho)| > 0$, and $B_n^{-1}(\rho)$ exists.

- Maximizing $\ell_n(\theta)$ gives the MLE $\hat{\theta}_n$ of $\theta$ if the errors are indeed Gaussian, otherwise the QMLE.
- Letting $G_n(\rho) = W_n B_n^{-1}(\rho)$, maximizing $\ell_n(\theta)$ is equivalent to solving $S_n(\theta) = 0$, where the score function has the form:

$$S_n(\theta) = \frac{\partial}{\partial\theta}\ell_n(\theta) = \begin{cases} \frac{1}{\sigma^2} X_n' B_n'(\rho) B_n(\rho) u_n(\beta), \\ \frac{1}{2\sigma^4} u_n'(\beta) B_n'(\rho) B_n(\rho) u_n(\beta) - \frac{n}{2\sigma^2}, \\ \frac{1}{\sigma^2} u_n'(\beta) B_n'(\rho) W_n u_n(\beta) - \mathrm{tr}[G_n(\rho)]. \end{cases} \quad (2.5)$$

The QML estimation process proceeds as follows:

- Solving the first two sets of equations of (2.5) for a given $\rho$ gives the constrained QMLEs of $\beta$ and $\sigma^2$,

$$\tilde{\beta}_n(\rho) = [X_n' B_n'(\rho) B_n(\rho) X_n]^{-1} X_n' B_n'(\rho) B_n(\rho) Y_n, \text{ and} \quad (2.6)$$

$$\tilde{\sigma}_n^2(\rho) = \tfrac{1}{n} Y_n' B_n'(\rho) M_n(\rho) B_n(\rho) Y_n, \quad (2.7)$$

  where $M_n(\rho) = I_n - B_n(\rho) X_n [X_n' B_n'(\rho) B_n(\rho) X_n]^{-1} X_n' B_n'(\rho)$.

- The concentrated log-likelihood function for $\rho$ upon substituting the constrained QMLEs $\tilde{\beta}_n(\rho)$ and $\tilde{\sigma}_n^2(\rho)$ into $\ell(\theta)$:

$$\ell_n^c(\rho) = -\tfrac{n}{2}[\log(2\pi) + 1] + \log|B_n(\rho)| - \tfrac{n}{2}\log(\tilde{\sigma}_n^2(\rho)). \quad (2.8)$$

- Maximising $\ell_n^c(\rho)$ numerically gives the unconstrained QMLE $\hat{\rho}_n$ of $\rho$,
- which upon substitutions gives the unconstrained QMLEs of $\beta$ and $\sigma^2$ as, $\hat{\beta}_n \equiv \tilde{\beta}_n(\hat{\rho}_n)$ and $\tilde{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\hat{\rho}_n)$.
- Thus, the QMLE of the full parameter vector $\theta$ is $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\rho}_n)'$.

Let $\theta_0$ be the true value of the parameter vector $\theta$ that generates the data; $\mathrm{E}(\cdot)$ and $\mathrm{Var}(\cdot)$ correspond to $\theta_0$; and $B_n \equiv B_n(\lambda_0)$, $G_n \equiv G_n(\lambda_0)$, etc.

**Theorem 2.1**. Under regularity conditions, we have $\hat{\theta}_n \xrightarrow{p} \theta_0$, and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} n\mathcal{J}_n^{-1}\mathcal{I}_n\mathcal{J}_n^{-1}\big), \tag{2.9}$$

(Liu & Yang 2015a), where $\mathcal{J}_n = -\mathrm{E}[\frac{\partial}{\partial\theta} S_n(\theta_0)]$ and $\mathcal{I}_n = \mathrm{Var}[S_n(\theta_0)]$, with

$$\mathcal{J}_n = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' B_n' B_n X_n & 0 & 0 \\ \sim & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) \\ \sim & \sim & \mathrm{tr}(G_n^s G_n) \end{pmatrix}, \text{ and} \tag{2.10}$$

$$\mathcal{I}_n = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' B_n' B_n X_n & \frac{1}{2\sigma_0^3}\gamma_0 X_n' B_n' \iota_n & \frac{1}{\sigma_0}\gamma_0 X_n' B_n' g_n \\ \sim & \frac{n}{4\sigma_0^4}(\kappa_0 + 2) & \frac{1}{2\sigma_0^2}(\kappa_0+2)\mathrm{tr}(G_n) \\ \sim & \sim & \kappa_0 g_n' g_n + \mathrm{tr}(G_n^s G_n) \end{pmatrix}, \tag{2.11}$$

where $\iota_n$ is a vector of ones, $\gamma_0$ and $\kappa_0$ are the measures of skewness and excess kurtosis of $\epsilon_{n,i}$, $g_n = \mathrm{diagv}(G_n)$, $G_n = G_n(\rho_0)$, and $G_n^s = G_n + G_n'$.

**Remarks:**

- $\mathrm{diagv}(G_n)$ forms a column vector by the diagonal elements of $G_n$.
- Note that in deriving the last component (the $\rho$-$\rho$ component) of $\mathcal{J}_n$, we have used the following matrix identity (given above):

$$\frac{\partial}{\partial\rho}B_n^{-1}(\rho) = -B_n^{-1}(\rho)[\frac{\partial}{\partial\rho}B_n(\rho)]B_n^{-1}(\rho). \quad (2.12)$$

- Clearly, when $\epsilon_{n,i}$ are iid normal, $\gamma_0 = \kappa_0 = 0$, and the asymptotic result reduces to $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} n\mathcal{J}_n^{-1}\big)$.

For deriving the expression for $\mathcal{I}_n = \mathrm{Var}[S_n(\theta_0)]$, we have by (2.5),

$$S_n(\theta_0) = \begin{cases} \frac{1}{\sigma_0^2}X_n'B_n'\epsilon_n, \\ \frac{1}{2\sigma_0^4}\epsilon_n'\epsilon_n - \frac{n}{2\sigma_0^2}, \\ \frac{1}{\sigma_0^2}\epsilon_n'G_n\epsilon_n - \mathrm{tr}(G_n), \end{cases}$$

of which the elements are linear or quadratic forms in $\epsilon_n$.

Based on the expression of $S_n(\theta_0)$ and using the following lemma, one can easily derive the analytical expression for $\mathcal{I}_{ns} = \mathrm{Var}[S_n(\theta_0)]$.

**Lemma 2.1.** Let $A_n$ and $B_n$ be $n \times n$ non-stochastic matrices and $c_n$ be $n \times 1$ non-stochastic vector. For $n \times 1$ random vector $\epsilon_n$ of iid elements with mean 0, variance $\sigma_0^2$, skewness $\gamma_0$, and finite excess kurtosis $\kappa_0$, we have

$$(i)\ \mathrm{Cov}(c_n'\epsilon_n, \epsilon_n' A_n \epsilon_n) = \sigma_0^3 \gamma_0 c_n' a_n,$$

$$(ii)\ \mathrm{Cov}(\epsilon_n' A_n \epsilon_n, \epsilon_n' B_n \epsilon_n) = \sigma_0^4 \kappa_0 a_n' b_n + \sigma_0^4 \mathrm{tr}(A_n B_n^s),$$

where $a_n = \mathrm{diagv}(A_n)$, $b_n = \mathrm{diagv}(B_n)$, and $B_n^s = B_n' + B_n$.

Note from Lemma 2.1 (*ii*), we can obtain $\mathrm{Var}(\epsilon_n' A_n \epsilon_n)$ by letting $A_n = B_n$.

# Standard error estimate of the QMLE of SE model

With the results of Theorem 2.1, the asymptotic variance-covariance (VC) matrix $\mathcal{J}_n^{-1}\mathcal{I}_n\mathcal{J}_n^{-1}$ of $\hat{\theta}_n$ is estimated as follows:

1. estimate $\mathcal{J}_n$ by $\hat{\mathcal{J}}_n = \mathcal{J}_n(\hat{\theta}_n)$,
2. estimate $\mathcal{I}_n$ by $\hat{\mathcal{I}}_n = \mathcal{I}(\hat{\theta}_n)$, the plug-in estimators,
3. and estimate $\gamma_0$ and $\kappa_0$ in $\mathcal{I}_n$ by the sample skewness and excess kurtosis of the QML residuals: $\hat{\epsilon}_n = \epsilon_n(\hat{\rho}_n, \hat{\beta}_n) = B_n(\hat{\rho}_n)(Y_n - X_n\hat{\beta}_n)$.

The square roots of the diagonal elements of $\mathcal{J}_n^{-1}(\hat{\theta}_n)\mathcal{I}_n(\hat{\theta}_n)\mathcal{J}_n^{-1}(\hat{\theta}_n)$ give the estimated standard errors of $\hat{\theta}_n$, robust against nonnormality.

- *t*-ratios for the elements of $\theta_0$ can then be constructed, from which one can
- judge whether the elements of $\theta_0$ are significantly different from 0, and thus
- make a conclusion on whether a covariate or a spatial effect is significant.

A hybrid method, combination of generalized least squares (GLS) and generalized moments (GM), for estimating the SE model is introduced by Kelejian and Prucha (1999). The ideas are

- first to develop GM estimators $\tilde{\rho}_n$ and $\tilde{\sigma}_n^2$ for $\rho$ and $\sigma^2$, based on a consistent 'predictor' $\tilde{u}_n$ for $u_n$;
- then to estimate $\Sigma_n = \mathrm{Var}(u_n)$ by $\widetilde{\Sigma}_n = \tilde{\sigma}_n^2 B_n^{-1}(\tilde{\rho}_n) B_n'^{-1}(\tilde{\rho}_n)$, leading to a feasible GLS estimate for $\beta$ as
$$\tilde{\beta}_n = (X_n' \widetilde{\Sigma}_n^{-1} X_n)^{-1} X_n' \widetilde{\Sigma}_n^{-1} Y_n;$$
- Under typical conditions, the GLS estimator of $\beta$ based on $\Sigma_n$ is consistent and asymptotically normal;
- Under additional conditions, e.g., $\tilde{\rho}_n$ and $\tilde{\sigma}_n^2$ are consistent, the feasible GLS estimator is asymptotically equivalent to the GLS estimator, and thus is also consistent and asymptotically normal.

**GM estimation.** The generalized moments (GM) estimation of $\rho$ and $\sigma^2$ is based on the following three moment conditions:

$$\begin{aligned}
&\mathrm{E}(\tfrac{1}{n}\epsilon_n'\epsilon_n) = \sigma^2, \\
&\mathrm{E}(\tfrac{1}{n}\epsilon_n'W_n'W_n\epsilon_n) = \sigma^2 n^{-1}\mathrm{tr}(W_n'W_n), \\
&\mathrm{E}(\tfrac{1}{n}\epsilon_n'W_n'\epsilon_n) = 0.
\end{aligned}$$

By $\epsilon_n = B_n(\rho)u_n$, we obtain the following sample moment conditions:

$$\mathbf{g}_n(\rho,\sigma^2) = \begin{cases} \frac{1}{n}\tilde{u}_n'B_n'(\rho)B_n(\rho)\tilde{u}_n - \sigma^2, \\ \frac{1}{n}\tilde{u}_n'B_n'(\rho)W_n'W_nB_n(\rho)\tilde{u}_n - \frac{1}{n}\sigma^2\mathrm{tr}(W_n'W_n), \\ \frac{1}{n}\tilde{u}_n'B_n'(\rho)W_n'B_n(\rho)\tilde{u}_n. \end{cases}$$

The GM estimators of $\rho$ and $\sigma^2$ are thus:

$$(\tilde{\rho}_n, \tilde{\sigma}_n^2) = \mathrm{argmin}\ [\mathbf{g}_n'(\rho,\sigma^2)\mathbf{g}_n(\rho,\sigma^2)],$$

**Remark:** $\tilde{u}_n$ can be the ordinary least squares (OLS) residuals, i.e., the residuals obtained from regressing $Y_n$ on $X_n$ (Kelejian and Prucha, 1999).

Intuitively, spatial error dependence causes the disturbances $u_n$ to lose 'a lot' of degrees of freedom (df). As a result, the QML estimation of $\rho$ and $\sigma^2$ may suffer from finite sample bias. This issue needs attention.

Liu and Yang (2015a) demonstrate based on Monte Carlo experiments:

- $\hat{\rho}_n$ can be severely downward biased, but the bias of $\hat{\rho}_n$ does not spill over much to $\hat{\beta}_n$;
- However, the bias of $\hat{\rho}_n$ does spill over to the estimate of $\text{Var}(\hat{\beta}_n)$;
- This makes the usual $t$-ratios for (the elements of) $\beta_0$ more variable than $N(0,1)$ and inferences for $\beta_0$ based on it unreliable.

From the asymptotic results given in (2.9)-(2.11), we see that $\hat{\beta}_n$ follows asymptotically $N(\beta_0, \sigma_0^2(X_n' B_n' B_n X_n)^{-1})$. Thus, inference for the linear contrast in $\beta_0$: $c_0'\beta_0$, is carried out based on the following $t$-ratio:

$$t_{\text{SE}}(\beta_0) = \frac{c_0'\hat{\beta}_n - c_0'\beta_0}{\sqrt{\hat{\sigma}_n^2 c_0' \left[X_n' B_n'(\hat{\rho}_n) B_n(\hat{\rho}_n) X_n\right]^{-1} c_0}}. \quad (2.13)$$

From the $t$-ratio given above, we see that

- downward bias of $\hat{\rho}_n$ causes $\hat{\sigma}_n^2$ to be downward biased, and
- severe bias of $\hat{\rho}_n$ may cause $X_n'\hat{B}_n'\hat{B}_nX_n$ to be severely biased for the estimation of $X_n'B_n'B_nX_n$ as seen from the expression:

$$X_n'B_n'(\hat{\rho}_n)B_n(\hat{\rho}_n)X_n = X_n'B_n'B_nX_n - (\hat{\rho}_n - \rho_0)X_n'(W_n'B_n + B_n'W_n)X_n + (\hat{\rho}_n - \rho_0)^2X_n'W_n'W_nX_n.$$

- If $X_n'(W_n'B_n + B_n'W_n)X_n \geq 0$, then
  $X_n'B_n'(\hat{\rho}_n)B_n(\hat{\rho}_n)X_n$ overestimates $X_n'B_n'B_nX_n$,
  $\hat{\sigma}_n^2c_0'[X_n'B_n'(\hat{\rho}_n)B_n(\hat{\rho}_n)X_n]^{-1}c_0$ underestimates $\text{Var}(c_0'\hat{\beta}_n)$.
- $\Rightarrow t_{\text{SE}}(\beta_0)$ tends to be 'larger' than $N(0,1)$ (or more variable),
  $\Rightarrow$ confidence interval for $c_0'\beta_0$ has low coverage,
  $\Rightarrow$ test of $c_0'\beta_0 = 0$ over rejects.

See Liu and Yang (2015b) or Lecture 4 for details on this issue. colorblueA Monte Carlo comparison of QML and GLS-GM estimators is of interest.

**Neighborhood Crime.** In illustrating the applications of spatial cross-sectional models, Anselin (1988, p.187) used the neighborhood crime data corresponding to 49 contiguous neighborhood in Columbus, Ohio, in 1980. These neighborhood correspond to census tracts, where

- `Crime`: the response variable pertaining to the combined total of residential burglaries and vehicle thefts per thousand household in the neighborhood.
- `Income` and housing values (`House`), are the explanatory variables both in thousand dollars.
- A dummy variable `East` indicates whether the 'neighborhood' in the east or west of a main north-south transportation axis.

The estimation results for the SLR-SE model are summarized in Table 2.1.

- Both `Income` and `House` have significant and negative effects on `Crime`.
- Data show a strong positive spatial error correlation in `Crime` among the 'neighbors' in Columbus, Ohio, in 1980.

**Table 2.1.** Estimation of SLR-SE Model: Neighborhood Crime

| | QMLE | se | t-Ratio | rse | rt-Ratio |
|---|---|---|---|---|---|
| constant | 59.8924 | 5.3662 | 11.1611 | 5.3662 | 11.1611 |
| income | -0.9413 | 0.3306 | -2.8477 | 0.3306 | -2.8477 |
| hvalue | -0.3023 | 0.0905 | -3.3407 | 0.0905 | -3.3407 |
| $\sigma^2$ | 95.5737 | 19.8735 | 4.8091 | 27.1596 | 3.5190 |
| $\rho$ | 0.5618 | 0.1339 | 4.1963 | 0.1343 | 4.1835 |

- The dummy variable `East` can be added to the model to see whether there is a significant difference between east and west in neighborhood crime.
- Spatial Durbin terms (of `Income` and/or `House`) can be added to the model to 'see' if there are contextual effects on `Crime`.
- There may be model specifications that can better reflect 'spillover effects' of crimes in Columbus, Ohio.
- See Lab1 for details on Matlab implementation of the estimation and inference procedures introduced.

**Boston House Price.** The data, form Harrison and Rubinfeld (1978), corrected and augmented with longitude and latitude by Gilley and Pace (1996), contains 506 observations (1 observation per census tract) from Boston Metropolitan Statistical Area (Click for Data). The variables are:

- `MEDV`: the median value (corrected) of owner-occupied homes in 1000's;
- `crime`: per capita crime rate by town;
- `zoning`: proportion of residential land zoned for lots over 25,000 square feet;
- `industry`: proportion of non-retail business acres per town;
- `charlesr`: Charles River dummy variable (= 1 if tract bounds river; 0 o.w.);
- `nox`: nitric oxides concentration (parts per 10 million);
- `room`: average number of rooms per dwelling;
- `houseage`: proportion of owner-occupied units built prior to 1940;
- `distance`: weighted distances to five Boston employment centres;
- `access`: index of accessibility to radial highways;
- `taxrate`: full-value property-tax rate per 10,000;
- `ptratio`: pupil-teacher ratio by town;
- `blackpop`: $1000(Bk - 0.63)^2$ where $Bk$ is the proportion of blacks by town;
- `lowclass`: lower status of the population proportion.

The spatial weight matrix is constructed using the Euclidean distance in terms of longitude and latitude. A threshold distance e.g., 0.05, is chosen, which gives a $W_n$ matrix with 19.08% non-zero elements.

The results from fitting a SLR-SE model is summarized in Table 2.2.

- The variables `crime`, `noxsq`, `distance`, `taxrate`, `ptratio`, and `lowclass` all have strong (highly significant) negative effects on house price.
- The variables `zoning`, `rooms2`, `access`, and `blackpop` all have strong positive effects on house price.
- Data show a strong positive SE correlation among neighboring regions.
- SD effects can be added, and their significance can be inferred.
- Alternative model specifications can be used.
- See Lab2 for details on Matlab implementation of the estimation and inference procedures introduced.

**Table 2.2.** Estimation of SED Model: Boston House Price (MEDV)

| | QMLE | se | t-Ratio | rse | rt-Ratio |
|---|---|---|---|---|---|
| constant | 29.6250 | 5.4956 | 5.3907 | 5.4956 | 5.3907 |
| crime | -0.1318 | 0.0276 | -4.7693 | 0.0276 | -4.7693 |
| zoning | 0.0379 | 0.0141 | 2.6887 | 0.0141 | 2.6887 |
| industry | -0.0139 | 0.0729 | -0.1909 | 0.0729 | -0.1909 |
| charlesr | -0.4975 | 0.8794 | -0.5658 | 0.8794 | -0.5658 |
| noxsq | -19.2666 | 5.2686 | -3.6568 | 5.2686 | -3.6568 |
| rooms2 | 4.2812 | 0.3643 | 11.7516 | 0.3643 | 11.7516 |
| houseage | -0.0259 | 0.0139 | -1.8604 | 0.0139 | -1.8604 |
| distance | -1.6095 | 0.3021 | -5.3276 | 0.3021 | -5.3276 |
| access | 0.3174 | 0.0767 | 4.1407 | 0.0767 | 4.1407 |
| taxrate | -0.0130 | 0.0036 | -3.6137 | 0.0036 | -3.6137 |
| ptratio | -0.6143 | 0.1523 | -4.0344 | 0.1523 | -4.0344 |
| blackpop | 0.0106 | 0.0031 | 3.4261 | 0.0031 | 3.4261 |
| lowclass | -0.4270 | 0.0514 | -8.2999 | 0.0514 | -8.2999 |
| $\sigma^2$ | 14.9219 | 0.9697 | 15.3874 | 1.7700 | 8.4306 |
| $\rho$ | 0.6947 | 0.0420 | 16.5461 | 0.0420 | 16.5332 |

**The model.** The spatial lag (SL) dependence model takes the form:

$$Y_n = \lambda W_n Y_n + X_n\beta + \epsilon_n, \tag{2.14}$$

where $Y_n$, $X_n$, and $W_n$ are as in (2.1). The errors $\epsilon_{n,i}$ are iid$(0, \sigma^2)$.

- The term $\lambda W_n Y_n$ says that the dependent variable of a specific spatial unit may depend on the dependent variables of other spatial units.
- This model captures the possible endogenous interaction effects.
- A major difference between the SL and SE models is that the spatial interactions in SE model changes only the variance of $Y_n$, whereas in SL model, it changes both the mean and the variance of $Y_n$:

  SE model: $\mathrm{E}(Y_n) = X_n\beta, \;\; \mathrm{Var}(Y_n) = \sigma^2 B_n^{-1}(\rho)B_n'^{-1}(\rho)$,

  SL model: $\mathrm{E}(Y_n) = A_n^{-1}(\lambda)X_n\beta, \;\; \mathrm{Var}(Y_n) = \sigma^2 A_n^{-1}(\lambda)A_n'^{-1}(\lambda)$,

  where $A_n(\lambda) = I_n - \lambda W_n$, with it inverse being assumed to exist.

**Durbin-SL model.** Similar to the SE model, the SL model (2.14) can also be extended by adding a spatial Durbin term $W_n X_n^*$, where $X_n^*$ contains a subset of regressors (excluding, e.g., the constant term), giving the so-called Durbin-SL model of the form (Elhorst 2014, p.7):

$$Y_n = \lambda W_n Y_n + X_n \beta + W_n X_n^* \beta^* + \epsilon_n. \tag{2.15}$$

By defining a new regressor matrix $\mathbb{X} = [X_n,\ W_n X_n^*]$, and a new vector of regression coefficients $\boldsymbol{\beta} = (\beta', \beta^{*\prime})'$,

- Durbin-SL model (2.15) has an identical form as the regular SL model, and all the estimation and inference methods remain the same.
- However, a problem of particular interest is to infer the significance of spatial Durbin effect as in the Durbin-SE model, i.e., whether the data provide sufficient evidence to infer $\beta^* \neq 0$,
- which can be carried out based on confidence intervals or tests for covariate effects. The latter is to be introduced in Lecture 3.

For ease of exposition, we proceed with Model (2.14). Using the matrix $A_n(\lambda)$, Model (2.14) has a reduced form:

$$A_n(\lambda)Y_n = X_n\beta + \epsilon_n.$$

**Assuming** $\epsilon_{n,i}$ are iid $N(0, \sigma^2)$, the joint pdf of $\epsilon_n$ is

$$(2\pi\sigma^2)^{-\frac{n}{2}} \exp(-\tfrac{1}{2\sigma^2}\epsilon_n'\epsilon_n),$$

which gives the joint pdf of $Y_n$ (or likelihood function) as the **Jacobian** of transformation ($\epsilon_n \to Y_n$) equals $|\frac{\partial \epsilon_n}{\partial Y_n}| = |A_n(\lambda)|$, assumed to be positive.

Thus, the log-likelihood function of $\theta = (\beta', \sigma^2, \lambda)'$ is

$$\begin{aligned}\ell_n(\theta) = & -\tfrac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\lambda)| \\ & - \tfrac{1}{2\sigma^2}\left[A_n(\lambda)Y_n - X_n\beta\right]'\left[A_n(\lambda)Y_n - X_n\beta\right]. \end{aligned} \tag{2.16}$$

Maximizing $\ell_n(\theta)$ gives the MLE $\hat{\theta}_n$ of $\theta$ if $\epsilon_{n,i}$ are truly iid $N(0, \sigma^2)$, otherwise the QMLE if $\epsilon_{n,i}$ are merely iid$(0, \sigma^2)$.

Similar to the case of SE model, the MLE or QMLE $\hat{\theta}_n$ can be obtained by solving the score-based estimating equations $S_n(\theta) = 0$, where

$$S_n(\theta) = \frac{\partial}{\partial\theta}\ell_n(\theta) = \begin{cases} \frac{1}{\sigma^2} X_n' \epsilon_n(\beta, \lambda), \\ \frac{1}{2\sigma^4} \epsilon_n'(\beta, \lambda)\epsilon_n(\beta, \lambda) - \frac{n}{2\sigma^2}, \\ \frac{1}{\sigma^2} Y_n' W_n' \epsilon_n(\beta, \lambda) - \mathrm{tr}[F_n(\lambda)], \end{cases} \tag{2.17}$$

$\epsilon_n(\beta, \lambda) = A_n(\lambda) Y_n - X_n\beta$, and $F_n(\lambda) = W_n A_n^{-1}(\lambda)$. Thus, the process of obtaining $\hat{\theta}_n$ can be simplified:

- Solving the first two sets of equations in (2.17) for a given $\lambda$, we obtain the constrained QMLEs for $\beta$ and $\sigma^2$, letting $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$:

$$\tilde{\beta}_n(\lambda) = (X_n'X_n)^{-1}X_n'A_n(\lambda)Y_n, \;\; \tilde{\sigma}_n^2(\lambda) = \frac{1}{n} Y_n'A_n'(\lambda)M_nA_n(\lambda)Y_n. \tag{2.18}$$

- Substituting $\tilde{\beta}_n(\lambda)$ and $\tilde{\sigma}_n^2(\lambda)$ back into (2.16) for $\beta$ and $\sigma^2$, we obtain the partially maximized or the concentrated loglikelihood of $\lambda$:

$$\ell_n^c(\lambda) = -\frac{n}{2}[\log(2\pi) + 1] - \frac{n}{2}\log\hat{\sigma}_n^2(\lambda) + \log|A_n(\lambda)|. \tag{2.19}$$

- Maximizing $\ell_n^c(\lambda)$ gives $\hat{\lambda}_n$, $\hat{\beta}_n \equiv \tilde{\beta}_n(\hat{\lambda}_n)$, $\hat{\sigma}_n^2 \equiv \tilde{\sigma}_n^2(\hat{\lambda}_n)$, and hence $\hat{\theta}_n$.

A similar set of notation is followed, e.g., $A_n \equiv A_n(\lambda_0)$, and $F_n \equiv F_n(\lambda_0)$.

**Theorem 2.2.** Under regularity conditions, we have $\hat{\theta}_n \xrightarrow{p} \theta_0$, and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(\theta_0,\ \lim_{n\to\infty} n\mathcal{J}_n^{-1}\mathcal{I}_n\mathcal{J}_n^{-1}), \tag{2.20}$$

(Lee 2004), where $\mathcal{J}_n = -\mathrm{E}[\frac{\partial}{\partial\theta'} S_n(\theta_0)]$ and $\mathcal{I}_n = \mathrm{Var}[S_n(\theta_0)]$, with

$$\mathcal{J}_n = \begin{pmatrix} \frac{1}{\sigma_0^2}X_n'X_n & 0 & \frac{1}{\sigma_0}X_n'\eta_n \\ 0 & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\mathrm{tr}(F_n) \\ \frac{1}{\sigma_0}\eta_n'X_n & \frac{1}{\sigma_0^2}\mathrm{tr}(F_n) & \eta_n'\eta_n + \mathrm{tr}(F_n^sF_n) \end{pmatrix}, \text{ and} \tag{2.21}$$

$$\mathcal{I}_n = \begin{pmatrix} \frac{1}{\sigma_0^2}X_n'X_n & \frac{1}{2\sigma_0^3}\gamma_0X_n'\iota_n & \frac{1}{\sigma_0}X_n'\eta_n + \frac{1}{\sigma_0}\gamma_0X_n'f_n \\ \sim & \frac{n}{2\sigma_0^4} + \frac{n}{4\sigma_0^4}\kappa_0 & \frac{1}{\sigma_0^2}\mathrm{tr}(F_n) + \frac{1}{2\sigma_0^2}\gamma_0\iota_n'\eta_n + \frac{1}{2\sigma_0^2}\kappa_0\mathrm{tr}(F_n) \\ \sim & \sim & \eta_n'\eta_n + \mathrm{tr}(F_n^sF_n) + \kappa_0f_n'f_n + 2\gamma_0f_n'\eta_n \end{pmatrix}, \tag{2.22}$$

where $f_n = \mathrm{diagv}(F_n)$, $F_n^s = F_n + F_n'$, $\eta_n = \sigma_0^{-1}G_nX_n\beta_0$, and $\gamma_0$ and $\kappa_0$ are the skewness and excess kurtosis of $\epsilon_{n,i}$.

The results (2.20)-(2.22) provide QML-based statistical inferences concerning the parameters of the SLR model with SL dependence.

- Again, $\mathcal{J}_n$ can be estimated by the plug-in estimator $\mathcal{J}_n(\hat{\theta}_n)$, or simply by its sample counterpart $-\frac{\partial}{\partial\theta} S_n(\theta)|_{\theta=\hat{\theta}_n}$.
- The variance of the score $\mathcal{I}_n$ can be consistently estimated by the plug-in method, i.e., plugging
  - $\hat{\theta}_n$ in $\mathcal{I}_n$ for $\theta_0$, and
  - $\hat{\gamma}_n$ and $\hat{\kappa}_n$ for $\gamma_0$ and $\kappa_0$,
  - where $\hat{\gamma}_n$ and $\hat{\kappa}_n$ are the sample skewness and excess kurtosis of the estimated errors $\hat{\epsilon}_n = \epsilon_n(\hat{\beta}_n, \hat{\lambda}_n) = A_n(\hat{\lambda}_n) Y_n - X_n \hat{\beta}_n$.
- These results give the estimates of robust standard errors of the QMLE $\hat{\theta}_n$ – robust against nonnormality of the error distribution.
- When it is known that $\epsilon_n$ is normally distributed, then $\mathcal{I}_n = \mathcal{J}_n$, and $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\big(\theta_0,\ \lim_{n\to\infty} n\mathcal{J}_n^{-1}\big)$.

The added regressor $W_n Y_n$ in the SLD model (2.14) is endogenous, as

$$\mathrm{E}[(W_n Y_n)' \epsilon_n] = \sigma_0^2 \mathrm{tr}(F_n(\lambda_0)) \neq 0,$$

i.e., the elements of $W_n Y_n$ are correlated with the elements of $\epsilon_n$. This can easily seen from the expression $W_n Y_n = F_n(\lambda_0) X_n \beta_0 + F_n(\lambda_0)\epsilon_n$, which follows from the reduced form of (2.14): $Y_n = A_n^{-1}(\lambda_0) X_n \beta + A_n^{-1}(\lambda_0)\epsilon_n$.

- Therefore, OLS estimate of $\vartheta = (\beta', \lambda)'$ is not consistent.
- If $W_n$ is row normalized and $|\lambda| < 1$, then $(I_n - \lambda W_n)^{-1} = \sum_{i=0}^{\infty} \lambda^i W_n^i$,
  $\Rightarrow$ instrumental variables (IVs) for $W_n Y_n$: $W_n X_n, W_n^2 X_n, \cdots,$
  $\Rightarrow$ linear moments: $[X_n,\ Q_n]' \epsilon_n(\vartheta)$,
    - $\epsilon_n(\vartheta) = A_n(\lambda) Y_n - X_n \beta$,
    - $Q_n$ contains columns of $[W_n X_n, W_n^2 X_n, \cdots]$, linearly independent of $X_n$.
- For $n \times n$ matrices $P_{jn}$ with $\mathrm{tr}(P_{jn}) = 0$, $\mathrm{E}(\epsilon_n' P_{jn} \epsilon_n) = \mathrm{tr}(P_{jn} \mathrm{E}(\epsilon_n \epsilon_n')) = 0$,
  $\Rightarrow$ quadratic moments: $\epsilon_n'(\vartheta) P_{jn} \epsilon_n(\vartheta),\ j = 1, \cdots, m$.

**Intuitions:** Note that $W_n Y_n = F_n(\lambda_0)X_n\beta_0 + F_n(\lambda_0)\epsilon_n$.

- The chosen $Q_n$ is correlated with $F_n X_n \beta_0$ but uncorrelated with $\epsilon_n$, $\Rightarrow Q$ instruments the mean of $W_n Y_n$;
- $P_{jn}\epsilon_n$ is uncorrelated with $\epsilon_n$, and thus it instruments the error of $W_n Y_n$, if $P_{jn}$ is chosen such that $P_{jn}\epsilon_n$ is correlated with $F_n\epsilon_n$, $\Rightarrow P_{jn}$ can be $W_n$, $W_n^2 - \frac{1}{n}\mathrm{tr}(W_n^2)I_n$, etc.

Letting $\mathbb{Q}_n = (X_n, Q_n)$, the **GMM estimator** of $\vartheta_0$ is

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

where $\mathbf{g}_n(\vartheta) = \left\{\epsilon_n'(\vartheta)\mathbb{Q}_n,\ \epsilon_n'(\vartheta)P_{1n}\epsilon_n(\vartheta),\ \cdots,\ \epsilon_n'(\vartheta)P_{mn}\epsilon_n(\vartheta)\right\}'$, and $\Omega_n$ is the GMM weight matrix. The asymptotic VC matrix of $\tilde{\vartheta}_n$ is given as:

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

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

$$\Sigma_n = \begin{pmatrix} \mathbb{Q}_n' X_n & \mathbb{Q}_n' F_n X_n \beta_0 \\ 0 & \sigma_0^2 \mathrm{tr}(P_{1n}^s F_n) \\ \vdots & \vdots \\ 0 & \sigma_0^2 \mathrm{tr}(P_{mn}^s F_n) \end{pmatrix}, \tag{2.23}$$

$$\Gamma_n = \begin{pmatrix} \sigma_0^2 \mathbb{Q}_n' \mathbb{Q}_n & \sigma_0^3 \gamma_0 \mathbb{Q}_n' \omega_{nm} \\ \sigma_0^3 \gamma_0 \omega_{nm}' \mathbb{Q}_n & \sigma_0^4 (\Lambda_{mn} + \kappa_0 \omega_{nm}' \omega_{nm}) \end{pmatrix}, \tag{2.24}$$

where $P_{kn}^s = P_{kn}' + P_{kn}, k = 1, \cdots, m,\ \omega_{nm} = \{\mathrm{diagv}(P_{1n}), \cdots, \mathrm{diagv}(P_{mn})\}$, and $\Lambda_{mn} = \{\mathrm{tr}(P_{jn} P_{kn}^s),\ j, k = 1, \cdots, m\}$.

- Simplest GMM: $\Omega_n = I_n$;
- Optimal GMM: $\Omega_n = \Gamma_n^{-1}$, not feasible as $\Gamma_n$ contains $\sigma_0^2$, $\gamma_0$ and $\kappa_0$;
- Feasible optimal GMM: $\Omega_n = \tilde{\Gamma}_n^{-1}$, where $\tilde{\Gamma}_n$ is a 'consistent estimate' of $\Gamma_n$, based on initial consistent estimate of $\vartheta_0$;
- Best OPGMM: choose 'best' moment functions, . . . (see Lee, 2007).

A simple choice for $\tilde{\Gamma}_n$ is $\tilde{\Gamma}_{\rm 2SLS}$ based on the 2SLS estimation:

- (i) let $Z_n = (X_n, W_n Y_n)$, and compute the 2SLSE of $\vartheta_0$:

$$\tilde{\vartheta}_{\rm 2SLS} = [Z_n' \mathbb{Q}_n (\mathbb{Q}_n' \mathbb{Q}_n)^{-1} \mathbb{Q}_n' Z_n]^{-1} Z_n' \mathbb{Q}_n (\mathbb{Q}_n' \mathbb{Q}_n)^{-1} \mathbb{Q}_n' Y_n,$$

- (ii) compute the 2SLE residuals $\tilde{\epsilon}_{\rm 2SLS} = A_n(\tilde{\lambda}_{\rm 2SLS}) - X_n \tilde{\beta}_{\rm 2SLS}$, and the sample variance, skewness and excess kurtosis of $\tilde{\epsilon}_{\rm 2SLS}$ to give consistent estimates of $\sigma^2$, $\gamma_0$ and $\kappa_0$.

The feasible OGMM estimator of $\vartheta_0$ is

$$\tilde{\vartheta}_n^{\circ} = \operatorname{argmin} \mathbf{g}_n'(\vartheta) \tilde{\Gamma}_{\rm 2SLS}^{-1} \mathbf{g}_n(\vartheta),$$

and a consistent estimate of the asymptotic VC matrix of $\tilde{\vartheta}_n^{\circ}$ is

$$(\tilde{\Sigma}_{\rm 2SLS}' \tilde{\Gamma}_{\rm 2SLS}^{-1} \tilde{\Sigma}_{\rm 2SLS})^{-1}.$$

- For more details on GMM estimation of SL model, see Lee (2007).
- A thorough Monte Carlo comparison of QMLE and GMME are desirable, based on ‘strong’ or ‘weak’ instruments.

Like the case of SE dependence, the existence of SL dependence also causes the QMLE of the spatia lag parameter to be biased.

- Yang (2015) presents a rigorous study on the finite sample properties of the QML estimator of $\lambda$.
- The QMLE $\hat{\lambda}_n$ is downward biased – the denser is the spatial weight matrix the more its is downward biased.
- Yang (2015) proposes a general method of bias correction, which is shown to be quite effective in removing the bias.
- Liu and Yang (2015b) show that the usual $t$-statistics for covariate effects tend to reject the null hypothesis of 'no effect' too often. They proposed finite sample improved test statistics based on the bias-correction method of Yang (2015).
- These methods will be introduced in the subsequent lectures.

**Boston House Price.** See Sec. 2.2.4 for detailed description of the data and the construction of spatial weight matrix.

The results from fitting a SLR-SL model is summarized in Table 2.3.

- Similar to the SLR-SE model, `crime`, `noxsq`, `distance`, `taxrate`, `ptratio`, and `lowclass` all have strong negative effects on house price.
- The variables `zoning`, `rooms2`, `access`, and `blackpop` all have strong positive effects on house price.
- Data show a strong positive SL dependence among neighboring regions.
- SD effects can be added, and their significance can be inferred.
- Alternative model specifications can be used.
- QMLE-bc (bias-corrested QMLE of Yang 2015) results are slightly different.
- See Lab2 for details on Matlab implementation of the estimation and inference procedures introduced.

**Table 2.3.** QML Estimation of SL Model: Boston House Price

| | QMLE | rse | t-Ratio | QMLE-bc | rse-bc | t-Ratio-bc |
|---|---|---|---|---|---|---|
| constant | 14.5396 | 5.0687 | 2.8685 | 13.4682 | 5.0337 | 2.6756 |
| crime | -0.0831 | 0.0300 | -2.7678 | -0.0819 | 0.0300 | -2.7326 |
| zoning | 0.0448 | 0.0124 | 3.6167 | 0.0447 | 0.0124 | 3.6151 |
| industry | 0.0353 | 0.0555 | 0.6368 | 0.0361 | 0.0554 | 0.6506 |
| charlesr | 1.1181 | 0.7850 | 1.4243 | 1.0414 | 0.7835 | 1.3291 |
| noxsq | -11.4788 | 3.5190 | -3.2620 | -11.1715 | 3.5104 | -3.1824 |
| rooms2 | 3.7066 | 0.3800 | 9.7543 | 3.7016 | 0.3794 | 9.7564 |
| houseage | 0.0021 | 0.0119 | 0.1733 | 0.0021 | 0.0119 | 0.1792 |
| distance | -1.2240 | 0.1842 | -6.6431 | -1.2117 | 0.1838 | -6.5921 |
| access | 0.2553 | 0.0604 | 4.2258 | 0.2529 | 0.0603 | 4.1921 |
| taxrate | -0.0110 | 0.0034 | -3.2225 | -0.0109 | 0.0034 | -3.2075 |
| ptratio | -0.5349 | 0.1247 | -4.2880 | -0.5145 | 0.1242 | -4.1427 |
| blackpop | 0.0078 | 0.0024 | 3.2216 | 0.0078 | 0.0024 | 3.1968 |
| lowclass | -0.3679 | 0.0483 | -7.6215 | -0.3602 | 0.0481 | -7.4952 |
| $\sigma^2$ | 18.3358 | 2.2937 | 7.9938 | 18.2882 | 2.2845 | 8.0053 |
| $\lambda$ | 0.3833 | 0.0399 | 9.6044 | 0.4020 | 0.0386 | 10.4204 |

**The model.** Adding both spatial lag and error (SLE) into the linear regression models, we have a more general SLR model:

$$Y_n = \lambda W_{1n} Y_n + X_n \beta + u_n, \quad u_n = \rho W_{2n} u_n + \epsilon_n. \tag{2.25}$$

It is also called the SARAR model in the literature, which emphasizes that the model contains a spatial autoregressive (SAR) term in response and a SAR term in error in line with the terms used in time series model.

- As for the SE and SL models, a spatial Durbin term, $W_{3n} X_n^* \beta^*$, can also be added into the model to capture the so-called contextual effects.
- Again, it is of interest to infer if $\beta^* = 0$.
- The $W$ matrices are in general different to capture different types of spatial interactions corresponding to SL, SE and SD effects, but they are allowed to be the same as far as the methods are concerned.
- For ease of exposition, we work with Model (2.25), thinking that the SD term, if any, has already been merged into the covariates effect $X_n \beta$.

The model (2.25) has the reduced form:

$$B_n(\rho)A_n(\lambda)Y_n = B_n(\rho)X_n\beta + \epsilon_n.$$

where $A_n(\lambda) = I_n - \lambda W_{1n}$ and $B_n(\rho) = I_n - \rho W_{2n}$ as defined earlier.

- The Jacobian of the transformation ($\epsilon_n \to Y_n$) is $|B_n(\rho)A_n(\lambda)|$.
- The quasi Gaussian loglikelihood function of $\theta = (\beta', \sigma^2, \lambda, \rho)'$ is

$$\begin{aligned}\ell_n(\theta) = &- \tfrac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\lambda)| + \log|B_n(\rho)| \\ &- \tfrac{1}{2\sigma^2}\left[\mathbb{Y}_n(\delta) - \mathbb{X}_n(\rho)\beta\right]'\left[\mathbb{Y}_n(\delta) - \mathbb{X}_n(\rho)\beta\right], \end{aligned} \quad (2.26)$$

  where $\delta = (\lambda, \rho)'$, $\mathbb{X}_n(\rho) = B_n(\rho)X_n$, and $\mathbb{Y}_n(\delta) = B_n(\rho)A_n(\lambda)Y_n$.
- Maximizing $\ell_n(\theta)$ gives the MLE $\hat{\theta}_n$ of $\theta$ if $\{\epsilon_i\}$ are iid normal. Otherwise, $\hat{\theta}_n$ is the QMLE of $\theta$.
- Assumptions: $|A_n(\lambda)| > 0$, $|B_n(\rho)| > 0$, and $A_n^{-1}(\lambda)$ and $B_n^{-1}(\rho)$ exist.

The (quasi) score function, $S_n(\theta) = \frac{\partial}{\partial\theta}\ell_n(\theta)$, has the form:

$$S_n(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbb{X}_n'(\rho)\epsilon_n(\beta,\delta), \\ \frac{1}{2\sigma^4}\epsilon_n'(\beta,\lambda)\epsilon_n(\beta,\delta) - \frac{n}{2\sigma^2}, \\ \frac{1}{\sigma^2}Y_n'W_{1n}'B_n'(\rho)\epsilon_n(\beta,\delta) - \mathrm{tr}[F_n(\lambda)], \\ \frac{1}{\sigma^2}\epsilon_n'(\beta,\delta)G_n(\rho)\epsilon_n(\beta,\delta) - \mathrm{tr}[G_n(\rho)], \end{cases} \tag{2.27}$$

- $\epsilon_n(\beta,\delta) = \mathbb{Y}_n(\delta) - \mathbb{X}_n(\rho)\beta$, $F_n(\lambda) = W_{1n}A_n^{-1}(\lambda)$, and $G_n(\rho) = W_{2n}B_n^{-1}(\rho)$.

Given $\delta$, solving the first two components of the quasi score equations, $S_n(\theta) = 0$, we obtain the constrained (Q)MLEs of $\beta$ and $\sigma^2$:

$$\tilde{\beta}_n(\delta) = [\mathbb{X}_n'(\rho)\mathbb{X}_n(\rho)]^{-1}\mathbb{X}_n'(\rho)\mathbb{Y}_n(\delta), \tag{2.28}$$

$$\tilde{\sigma}_n^2(\delta) = \frac{1}{n}\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)_n(\rho)\mathbb{Y}_n(\delta), \tag{2.29}$$

where $\mathbb{M}_n(\rho) = I_n - \mathbb{X}_n(\rho)[\mathbb{X}_n'(\rho)\mathbb{X}_n(\rho)]^{-1}\mathbb{X}_n'(\rho)$. Substituting $\tilde{\beta}_n(\delta)$ and $\tilde{\sigma}_n^2(\delta)$ back into (2.26) gives the concentrated quasi loglikelihood for $\delta$:

$$\ell_n^c(\delta) = -\frac{n}{2}[\ln(2\pi) + 1] - \frac{n}{2}\ln(\hat{\sigma}_n^2(\delta)) + \ln|A_n(\lambda)| + \ln|B_n(\rho)|. \tag{2.30}$$

Maximizing (2.30) gives the QMLE $\hat{\delta}_n$ of $\delta$, and thus the QMLEs of $\beta$ and $\sigma^2$ as $\hat{\beta}_n \equiv \tilde{\beta}_n(\hat{\delta}_n)$ and $\hat{\sigma}_n^2 \equiv \tilde{\sigma}_n^2(\hat{\delta}_n)$. Write $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\delta}_n')'$.

Plugging $\tilde{\beta}_n(\delta)$ and $\tilde{\sigma}_n^2(\delta)$ into the $\delta$-component of $S_n(\theta)$ and simplifying, we have the concentrated quasi score (CQS) function of $\delta$:

$$S_n^c(\delta) = \begin{cases} -\mathrm{tr}(G_{1n}(\lambda)) + \dfrac{n\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\bar{F}_n(\delta)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)M_n(\rho)\mathbb{Y}_n(\delta)}, \\ -\mathrm{tr}(G_n(\rho)) + \dfrac{n\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)G_n(\rho)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\lambda)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \end{cases} \tag{2.31}$$

where $\bar{F}_n(\delta) = B_n(\rho)F_n(\lambda)B_n^{-1}(\rho)$. Maximizing $\ell_n^c(\delta) \Leftrightarrow$ solving $S_n^c(\delta) = 0$.

The CQS function $S_n^c(\delta)$ is the key expression for

- deriving the score-based tests for the spatial effects,
- for performing bias-correction on the QMLE $\hat{\delta}_n$,

to be introduced in the subsequent lectures.

**Theorem 2.3.** Under some regularity conditions, we have $\hat{\theta}_n \xrightarrow{p} \theta_0$, and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\big(\theta_0,\ \lim_{n\to\infty} n\mathcal{J}_n^{-1}\mathcal{I}_n\mathcal{J}_n^{-1}\big), \tag{2.32}$$

(Jin and Lee 2013), where $\mathcal{J}_n = -\mathrm{E}[\frac{\partial}{\partial\theta'}S_n(\theta_0)]$ and $\mathcal{I}_n = \mathrm{Var}[S_n(\theta_0)]$, with

$$\mathcal{J}_n = \begin{pmatrix} \frac{1}{\sigma_0^2}\mathbb{X}_n'\mathbb{X}_n & 0 & \frac{1}{\sigma_0}\mathbb{X}_n'\mu_n & 0 \\ \sim & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\mathrm{tr}(F_n) & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) \\ \sim & \sim & \mu_n'\mu_n + \mathrm{tr}(\bar{F}_n^s\bar{F}_n) & \mathrm{tr}(G_n^s\bar{F}_n) \\ \sim & \sim & \sim & \mathrm{tr}(G_n^sG_n) \end{pmatrix}, \tag{2.33}$$

$$\mathcal{I}_n = \mathcal{J}_n + \begin{pmatrix} 0 & \frac{\gamma_0}{2\sigma_0^3}\mathbb{X}_n'\iota_n & \frac{\gamma_0}{\sigma_0}\mathbb{X}_n'\bar{f}_n & \frac{\gamma_0}{\sigma_0}\mathbb{X}_n'g_n \\ \sim & \frac{n\kappa_0}{4\sigma_0^4} & \frac{\kappa_0}{2\sigma_0^2}\mathrm{tr}(F_n) + \frac{\gamma_0}{2\sigma_0^2}\iota_n'\mu_n & \frac{\kappa_0}{2\sigma_0^2}\mathrm{tr}(G_n) \\ \sim & \sim & \kappa_0\bar{f}_n'\bar{f}_n + 2\gamma_0\bar{f}_n'\mu_n & \kappa_0 g_n'\bar{f}_n + \gamma_0 g_n'\mu_n \\ \sim & \sim & \sim & \kappa_0 g_n'g_n \end{pmatrix}, \tag{2.34}$$

where $\mu_n = \sigma_0^{-1}B_nF_nX_n\beta_0$, $\bar{f}_n = \mathrm{diag}(\bar{F}_n)$, $\bar{F}_n^s = \bar{F}_n + \bar{F}_n'$, $\bar{F}_n = \bar{F}_n(\delta_0) = B_nF_nB_n^{-1}$, and all others quantities are defined earlier.

The results (2.32)-(2.34) form the base for statistical inferences, e.g., confidence intervals, LM tests, Wald tests, for the elements of $\theta$.

- $\mathcal{J}_n$ is estimated by the plug-in estimator $\mathcal{J}_n(\hat{\theta}_n)$, or by $-\frac{\partial}{\partial\theta}S_n(\theta)|_{\theta=\hat{\theta}_n}$.
- $\mathcal{I}_n$ is estimated by the 'plug-in' estimator as well, i.e., $\mathcal{I}_n(\hat{\theta}_n, \hat{\gamma}_n, \hat{\kappa}_n)$, obtained by plugging in $\hat{\theta}_n$, $\hat{\gamma}_n$ and $\hat{\kappa}_n$,
- $\hat{\gamma}_n$ and $\hat{\kappa}_n$ are the sample skewness and excess kurtosis of the QML residuals $\hat{\epsilon}_n = \mathbb{Y}_n(\hat{\delta}_n) - \mathbb{X}_n(\hat{\rho}_n)\hat{\beta}_n$.
- Square roots of diagonal elements of $\mathcal{J}_n^{-1}(\hat{\theta}_n)\mathcal{I}_n(\hat{\theta}_n, \hat{\gamma}_n, \hat{\kappa}_n)\mathcal{J}_n^{-1}(\hat{\theta}_n)$ give the estimated standard errors of the elements of $\hat{\theta}_n$.

Lee and Liu (2010) extend the GMM framework for the SL model to a high-order spatial linear regression model, SARAR$(p, q)$, including the SLE model, or SARAR$(1, 1)$, discussed above. The set of moment functions for SARAR(1,1) remains in a similar form as for SL model:

$$\mathbf{g}_n(\vartheta) = \left\{\epsilon_n'(\vartheta)\mathbb{Q}_n,\ \epsilon_n'(\vartheta)P_{1n}\,\epsilon_n(\vartheta),\ \cdots,\ \epsilon_n'(\vartheta)P_{mn}\,\epsilon_n(\vartheta)\right\}',$$

except that now $\vartheta = (\beta', \lambda, \rho)'$ and $\epsilon_n(\vartheta) = B_n(\rho)[A_n(\lambda)Y_n - X_n\beta]$.

Now, letting again $\Sigma_n = -\mathrm{E}[\frac{\partial}{\partial\vartheta'}\mathbf{g}_n(\vartheta_0)]$ and $\Gamma_n = \mathrm{Var}[\mathbf{g}_n(\vartheta_0)]$, we have

$$\Sigma_n = \begin{pmatrix} \mathbb{Q}_n'\mathbb{X}_n, & \mathbb{Q}_n'\bar{F}_n\mathbb{X}_n\beta_0, & 0 \\ 0, & \sigma_0^2\mathrm{tr}(P_{1n}^s\bar{F}_n), & \sigma_0^2\mathrm{tr}(P_{1n}^s G_n) \\ \vdots & \vdots & \vdots \\ 0, & \sigma_0^2\mathrm{tr}(P_{mn}^s\bar{F}_n), & \sigma_0^2\mathrm{tr}(P_{mn}^s G_n) \end{pmatrix}, \tag{2.35}$$

and $\Gamma_n$ identical in expression as that given in (2.24).

The feasible OGMM estimate of $\vartheta_0$ is thus

$$\tilde{\vartheta}_n = \text{argmin } \mathbf{g}_n'(\vartheta)\tilde{\Gamma}_n^{-1}\mathbf{g}_n(\vartheta)$$

where $\tilde{\Gamma}_n$ is a consistent estimate of $\Gamma_n$, which can be obtained based on the generalized 2SLS estimate of Kelejian and Prucha (1998):

- (i) let $Z_n = (X_n, W_n Y_n)$, and compute the G2SLSE of $\vartheta_0$:

$$\begin{aligned}\tilde{\vartheta}_{\text{G2SLS}} &= [Z_n' B_n'(\tilde{\rho}_n)\mathbb{Q}_n(\mathbb{Q}_n'\mathbb{Q}_n)^{-1}\mathbb{Q}_n' B_n(\tilde{\rho}_n) Z_n]^{-1} \\ &\quad Z_n' B_n'(\tilde{\rho}_n)\mathbb{Q}_n(\mathbb{Q}_n'\mathbb{Q}_n)^{-1}\mathbb{Q}_n' B_n(\tilde{\rho}_n) Y_n,\end{aligned}$$

  where $\tilde{\rho}_n$ is a consistent initial estimator of $\rho$.
- (ii) compute the G2SLE residuals $\epsilon(\tilde{\vartheta}_{\text{G2SLS}})$, and the sample variance, skewness and excess kurtosis of $\epsilon(\tilde{\vartheta}_{\text{G2SLS}})$ to give consistent estimates of $\sigma_0^2$, $\gamma_0$ and $\kappa_0$.

Finally, the IV matrix $\mathbb{Q}_n$ can be taken as a subset of linearly independent columns of $\{X_n, W_{1n}X_n, W_{1n}^2 X_n, W_{2n}X_n, W_{2n}^2 X_n\}$ (containing $X_n$), and the $P_{jn}$ matrices can be constructed similarly to these for the SL model.

With both SL and SE effects in the model, the finite sample performance of the QMLEs may be of more concern.

- Liu and Yang (2015b) extended the bias-correction procedure of Yang (2015) from SL model to SLE model. They focus on improving the finite sample performance of $t$-ratios for covariates effects.
- Their bootstrap methods can be followed for a Monte Carlo comparison of QMLE $\hat{\delta}_n$ and bias-corrected $\hat{\hat{\delta}}_n$ on their finite sample performance.
- Details on bias-corrected estimation of spatial autocorrelation will be given in Lecture 4.
- A Monte Carlo comparison of QML-type and GMM-type estimators is desired.

**Boston House Price.** See Sec. 2.2.4 for detailed description of the data and the construction of spatial weight matrix.

The results from fitting a SLR-SLE model is summarized in Table 2.4.

- The conclusions on covariates effects are largely unchanged.
- Interestingly, spatial error effect is strong and positive as in the SLR-SE model, but with the existence of SE effect the SL effects becomes insignificant.
- Alternative model specifications with different types of spatial weight matrices for SL and SE effects may help addressing this issue.
- SD effects can be added, and their significance can be inferred.
- QMLE-bc (bias-corrested QMLE of Yang 2015) results are slightly different.
- See Lab2 for details on Matlab implementation of the estimation and inference procedures introduced.

**Table 2.4.** Estimation of SLE Model: Boston House Price

| | QMLE | rse | t-Ratio | QMLE-bc | rse-bc | t-Ratio-bc |
|---|---|---|---|---|---|---|
| constant | 27.8251 | 5.9648 | 4.6649 | 27.8251 | 5.9648 | 4.6649 |
| crime | -0.1301 | 0.0279 | -4.6660 | -0.1301 | 0.0279 | -4.6660 |
| zoning | 0.0390 | 0.0141 | 2.7763 | 0.0390 | 0.0141 | 2.7763 |
| industry | -0.0071 | 0.0723 | -0.0976 | -0.0071 | 0.0723 | -0.0976 |
| charlesr | -0.3970 | 0.8775 | -0.4524 | -0.3970 | 0.8775 | -0.4524 |
| noxsq | -18.5059 | 5.1765 | -3.5750 | -18.5059 | 5.1765 | -3.5750 |
| rooms2 | 4.2958 | 0.3680 | 11.6746 | 4.2958 | 0.3680 | 11.6746 |
| houseage | -0.0240 | 0.0139 | -1.7300 | -0.0240 | 0.0139 | -1.7300 |
| distance | -1.6254 | 0.2897 | -5.6097 | -1.6254 | 0.2897 | -5.6097 |
| access | 0.3177 | 0.0756 | 4.2047 | 0.3177 | 0.0756 | 4.2047 |
| taxrate | -0.0131 | 0.0036 | -3.6294 | -0.0131 | 0.0036 | -3.6294 |
| ptratio | -0.6201 | 0.1513 | -4.0984 | -0.6201 | 0.1513 | -4.0984 |
| blackpop | 0.0104 | 0.0031 | 3.3818 | 0.0104 | 0.0031 | 3.3818 |
| lowclass | -0.4234 | 0.0521 | -8.1350 | -0.4234 | 0.0521 | -8.1350 |
| $\sigma^2$ | 15.0920 | 1.8042 | 8.3648 | 15.0920 | 1.8042 | 8.3648 |
| $\lambda$ | 0.0569 | 0.0679 | 0.8381 | 0.0506 | 0.0679 | 0.7456 |
| $\rho$ | 0.6595 | 0.0593 | 11.1309 | 0.6893 | 0.0593 | 11.6339 |

# Appendix: Basics on Consistency and Asymptotic Normality

This appendix presents some basics for establishing consistency and asymptotic normality of a QML estimator or an M-estimator. Some Monte Carlo results are also presented in supporting the asymptotic arguments.

We use SLR-SE model to demonstrate typical assumptions required for the asymptotic analysis.

**Assumption 1:** *The true $\rho_0$ is in the interior of the compact parameter set $\mathcal{P}$.*

**Assumption 2:** *$\{\epsilon_{n,i}\}$ are iid with mean 0, variance $\sigma^2$, and $\mathrm{E}|\epsilon_{n,i}|^{4+\delta} < \infty, \forall\delta > 0$.*

**Assumption 3:** *$X_n$ has full column rank $k$, its elements are uniformly bounded constants, and $\lim_{n\to\infty} \frac{1}{n} X_n' B_n'(\rho) B_n(\rho) X_n$ exists and is non-singular for $\rho$ near $\rho_0$.*

**Assumption 4:** *The elements $\{w_{ij}\}$ of $W_n$ are at most of order $h_n^{-1}$ uniformly for all $i$ and $j$, where $h_n$ can be bounded or divergent but subject to $\lim_{n\to\infty} \frac{h_n}{n} = 0$; $W_n$ is bounded in both row and column sum norms and its diagonal elements are zero.*

**Assumption 5:** *$B_n$ is non-singular and $B_n^{-1}$ is bounded in both row and column sum norms. Further, $B_n^{-1}(\rho)$ is bounded in either row or column sum norm, uniformly in $\rho \in \mathcal{P}$.*

**Assumption 6:** $\lim_{n\to\infty} \frac{h_n}{n}[\log|\sigma_0^2 B_n^{-1} B_n'^{-1}| - \log|\sigma_n^2(\rho) B_n^{-1}(\rho) B_n'^{-1}(\rho)|] \neq 0, \forall \rho \neq \rho_0$.

**Consistency.** Let $\delta$ be the parameter vector in the concentrated quasi loglikelihood $\ell_n^c(\delta)$, obtained from the full loglikelihood $\ell_n(\theta)$ by replacing $\beta$ and $\sigma^2$ by their constrained QMLEs.

- Let $\delta_0$ be the true value of $\theta$ which takes values in the parameter space $\Theta$. The expectation operator 'E' corresponds the true $\theta_0$.
- It is only necessary to prove the consistency of $\hat{\delta}_n$ as the consistency of $\hat{\beta}_n$ and $\hat{\sigma}_n$ immediately follow if $\text{plim}\frac{1}{n}X_n'X_n$ exists and is invertible.
- Define $\bar{\ell}_n(\theta) = \text{E}[\ell_n(\theta)]$, the expected loglikelihood at a general $\theta$.
- Define $\bar{\ell}_n^c(\delta) = \max_{\beta,\sigma^2} \bar{\ell}_n(\theta)$.

The consistency of $\hat{\delta}_n$ follows, i.e., $\hat{\delta}_n \xrightarrow{p} \delta_0$, if

**(a)** $\sup_{\delta: d(\delta,\delta_0)\geq\varepsilon} \bar{\ell}_n^c(\delta) < \bar{\ell}_n^c(\delta_0)$, for every $\varepsilon > 0$,

**(b)** $\sup_{\theta\in\Theta} \frac{1}{n}|\ell_n^c(\delta) - \bar{\ell}_n^c(\delta)| \xrightarrow{p} 0$.

The condition (a) is the so-called identification uniqueness condition for $\delta_0$ which ensures that $\bar{\ell}_n^c(\delta)$ has a identifiably unique maximizer that converges to $\delta_0$.

This together with condition (b) ensures that the difference between the maximizer $\hat{\theta}_n$ of $\ell_n^c(\delta)$ and the maximizer of $\bar{\ell}_n^c(\delta)$ gets smaller and smaller as $n$ goes large.

See White (1994, Theorem 3.4) or van der Vaar (1998, Theorem 5.7).

For detailed applications of these theorems, see Liu and Yang (2015a) for the case of SE model, Lee (2004) for the case of SL model, and Jin and Lee (2013) for the case of SLE model.

**Asymptotic normality.** Assume the QMLE $\hat{\theta}_n$ which maximizes $\ell_n(\theta)$, can be equivalently obtained by solving the estimating equation $S_n(\theta) = 0$, where $S_n(\theta) = \frac{\partial}{\partial\theta}\ell_n(\theta)$ is the score function. Assume $\hat{\theta}_n \xrightarrow{p} \theta_0$. Then, by **Taylor series expansion**, we have

$$0 = \frac{1}{\sqrt{n}} S_n(\hat{\theta}_n) = \frac{1}{\sqrt{n}} S_n(\theta_0) + \left[\frac{1}{n}\frac{\partial}{\partial\theta'} S_n(\theta^*)\right] \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1),$$

where $\theta^*$ lies elementwise between $\hat{\theta}_n$ and $\theta_0$. The asymptotic normlity of $\hat{\theta}_n$ follows if

**(a)** $\frac{1}{\sqrt{n}} S_n(\theta_0) \xrightarrow{D} N(0,\ \lim_{n\to\infty} \frac{1}{n}\mathcal{I}_n)$,

**(b)** $\frac{1}{n}\left[\frac{\partial}{\partial\theta'} S_n(\theta^*) - \frac{\partial}{\partial\theta'} S_n(\theta_0)\right] \xrightarrow{p} 0$, and

**(c)** $\frac{1}{n}\left[\frac{\partial}{\partial\theta'} S_n(\theta_0) - \mathrm{E}\left(\frac{\partial}{\partial\theta'} S_n(\theta_0)\right)\right] \xrightarrow{p} 0$.

As the elements of $S_n(\theta_0)$ are typically linear or quadratic forms in $\epsilon_n$ due to the use of Gaussian quasi likelihood, the result (a) can be proved using the central limit theorem (CLT) for linear-quadratic forms of Kelejian and Prucha (2001), and Cramér-Wold devise.

The conditions (b) and (c) can be proved by applying some law of large numbers (LLN).

See for details Liu and Yang (2015a) for the case of SE model, Lee (2004) for SL model, and Jin and Lee (2013) for SLE model.

Clearly, if the above conditions hold, then

$$\begin{aligned} 0 &= \tfrac{1}{\sqrt{n}} S_n(\theta_0) + \left[\tfrac{1}{n}\mathrm{E}\left(\tfrac{\partial}{\partial\theta'} S_n(\theta_0)\right)\right] \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1) \\ &= \tfrac{1}{\sqrt{n}} S_n(\theta_0) - \left[\tfrac{1}{n}\mathcal{J}_n(\theta_0)\right] \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1) \end{aligned}$$

It follows that $\sqrt{n}(\hat{\theta}_n - \theta_0) = \left[\tfrac{1}{n}\mathcal{J}_n(\theta_0)\right]^{-1} \tfrac{1}{\sqrt{n}} S_n(\theta_0) + o_p(1)$. It follows that

$$\mathrm{Var}[\sqrt{n}(\hat{\theta}_n - \theta_0)] = \left[\tfrac{1}{n}\mathcal{J}_n(\theta_0)\right]^{-1} \mathrm{Var}[\tfrac{1}{\sqrt{n}} S_n(\theta_0)] \left[\tfrac{1}{n}\mathcal{J}_n(\theta_0)\right]^{-1}.$$

If further $\mathrm{E}[S_n(\theta_0)] = 0$ or at least $\frac{1}{n}\mathrm{E}[S_n(\theta_0)] \to 0$, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \lim_{n\to\infty} n\mathcal{J}_n^{-1}\mathcal{I}_n\mathcal{J}_n^{-1}).$$

Tables A1-A3 presents some Monte Carlo results for the finite sample performance of $\hat{\lambda}_n$, $\hat{\lambda}_n^{\rm bc2}$ (2nd-order bias-corrected $\hat{\lambda}_n$), and $\hat{\lambda}_n^{\rm bc3}$ (3rd-order bias-corrected $\hat{\lambda}_n$), of the SLR-SL model. The results reveals:

- $\hat{\lambda}_n$ can be quite biased,
- $\hat{\lambda}_n^{\rm bc2}$ almost removes the bias,
- $\hat{\lambda}_n^{\rm bc3}$ does not improve much over $\hat{\lambda}_n^{\rm bc2}$,
- More results from Yang (2015) show that with a denser spatial weight matrix, the bias of $\hat{\lambda}_n$ becomes larger.
- When sample size increases, all three estimators 'converge' to the true value of $\lambda$.
- When sample size is not large, it is necessary to carry out the bias-correction, in particular on the QMLE of the spatial parameter(s).

**Table A1.** Monte Carlo Mean[rmse](sd) of Estimators of $\lambda$ in SL Model

| $\lambda$ | $n$ | $\hat{\lambda}_n$ | $\hat{\lambda}_n^{\text{bc2}}$ | $\hat{\lambda}_n^{\text{bc3}}$ |
|---|---|---|---|---|
| | | Queen Contiguity, Normal Errors | | |
| .50 | 50 | .411 [.195](.174) | .492 [.175](.175) | .497 [.175](.175) |
| | 100 | .459 [.123](.116) | .498 [.117](.117) | .500 [.117](.117) |
| | 200 | .480 [.078](.076) | .499 [.075](.075) | .499 [.075](.075) |
| | 500 | .493 [.049](.048) | .501 [.048](.048) | .501 [.048](.048) |
| .25 | 50 | .163 [.222](.204) | .242 [.209](.209) | .246 [.210](.210) |
| | 100 | .212 [.146](.140) | .248 [.142](.142) | .250 [.143](.143) |
| | 200 | .231 [.094](.092) | .250 [.093](.093) | .250 [.093](.093) |
| | 500 | .242 [.060](.060) | .250 [.060](.060) | .250 [.060](.060) |
| .00 | 50 | -.078 [.229](.216) | -.006 [.224](.224) | -.003 [.226](.226) |
| | 100 | -.034 [.157](.153) | -.002 [.156](.156) | -.001 [.157](.157) |
| | 200 | -.018 [.106](.104) | -.000 [.105](.105) | .000 [.105](.105) |
| | 500 | -.008 [.068](.067) | -.000 [.068](.068) | -.000 [.068](.068) |
| -.25 | 50 | -.317 [.233](.223) | -.255 [.236](.236) | -.254 [.237](.237) |
| | 100 | -.279 [.164](.161) | -.253 [.166](.166) | -.253 [.166](.166) |
| | 200 | -.266 [.112](.111) | -.252 [.112](.112) | -.251 [.112](.112) |
| | 500 | -.256 [.073](.072) | -.250 [.073](.073) | -.250 [.073](.073) |

**Table A2.** Monte Carlo Mean[rmse](sd) of Estimators of $\lambda$ in SL Model

| $\lambda$ | $n$ | $\hat{\lambda}_n$ | $\hat{\lambda}_n^{\text{bc2}}$ | $\hat{\lambda}_n^{\text{bc3}}$ |
|---|---|---|---|---|
| | | Queen Contiguity, Normal Mixture Errors | | |
| .50 | 50 | .420 [.182](.164) | .494 [.165](.165) | .498 [.165](.165) |
| | 100 | .462 [.120](.114) | .499 [.114](.114) | .500 [.114](.114) |
| | 200 | .482 [.076](.074) | .500 [.074](.074) | .500 [.074](.074) |
| | 500 | .492 [.049](.048) | .500 [.048](.048) | .500 [.048](.048) |
| .25 | 50 | .169 [.207](.190) | .241 [.195](.195) | .244 [.195](.195) |
| | 100 | .213 [.140](.135) | .248 [.136](.136) | .249 [.137](.137) |
| | 200 | .230 [.092](.090) | .249 [.090](.090) | .249 [.090](.090) |
| | 500 | .242 [.060](.060) | .250 [.060](.060) | .250 [.060](.060) |
| .00 | 50 | -.070 [.217](.206) | -.004 [.213](.213) | -.002 [.214](.214) |
| | 100 | -.032 [.150](.147) | -.002 [.150](.150) | -.001 [.150](.150) |
| | 200 | -.018 [.104](.103) | -.001 [.103](.103) | -.001 [.103](.103) |
| | 500 | -.008 [.068](.067) | -.001 [.067](.067) | -.001 [.067](.067) |
| -.25 | 50 | -.314 [.223](.213) | -.258 [.224](.224) | -.257 [.225](.225) |
| | 100 | -.275 [.155](.153) | -.251 [.157](.157) | -.250 [.157](.157) |
| | 200 | -.263 [.111](.110) | -.249 [.112](.112) | -.249 [.112](.112) |
| | 500 | -.257 [.072](.072) | -.251 [.072](.072) | -.251 [.072](.072) |

**Table A3.** Monte Carlo Mean[rmse](sd) of Estimators of $\lambda$ in SL Model

| $\lambda$ | $n$ | $\hat{\lambda}_n$ | $\hat{\lambda}_n^{\mathrm{bc2}}$ | $\hat{\lambda}_n^{\mathrm{bc3}}$ |
|---|---|---|---|---|
| | | Queen Contiguity, Lognormal Errors | | |
| .50 | 50 | .426 [.163](.146) | .491 [.146](.146) | .493 [.146](.146) |
| | 100 | .465 [.110](.105) | .498 [.105](.105) | .498 [.105](.105) |
| | 200 | .482 [.072](.069) | .499 [.069](.069) | .499 [.069](.069) |
| | 500 | .491 [.047](.046) | .499 [.046](.046) | .499 [.046](.046) |
| .25 | 50 | .179 [.185](.171) | .241 [.174](.174) | .244 [.174](.174) |
| | 100 | .216 [.128](.124) | .247 [.126](.125) | .248 [.126](.125) |
| | 200 | .232 [.087](.085) | .249 [.085](.085) | .249 [.085](.085) |
| | 500 | .242 [.058](.057) | .249 [.057](.057) | .249 [.057](.057) |
| .00 | 50 | -.067 [.198](.186) | -.011 [.192](.191) | -.008 [.192](.192) |
| | 100 | -.029 [.139](.136) | -.003 [.138](.138) | -.002 [.138](.138) |
| | 200 | -.017 [.099](.097) | -.002 [.098](.098) | -.001 [.098](.098) |
| | 500 | -.007 [.065](.064) | -.000 [.065](.065) | .000 [.065](.065) |
| -.25 | 50 | -.307 [.199](.191) | -.258 [.198](.198) | -.256 [.199](.199) |
| | 100 | -.272 [.142](.140) | -.252 [.144](.144) | -.251 [.144](.144) |
| | 200 | -.264 [.105](.104) | -.251 [.105](.105) | -.250 [.105](.105) |
| | 500 | -.256 [.070](.070) | -.250 [.070](.070) | -.250 [.070](.070) |

# References


- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. The Netherlands: Kluwer Academic Publishers.
- Durbin, J. 1960. Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Soceity B* 22, 139-153
- Elhorst, J. P., 2014. *Spatial Econometrics From Cross-Sectional Data to Spatial Panels*. Springer.
- Horn, R. A., Johnson, C. R., 1985. *Matrix Analysis*. Cambridge: Cambridge University Press.
- Gilley, O. W., Pace, R. K., 1996. On the Harrison and Rubinfeld Data. *Journal of Environmental Economics and Management* 31, 403-405.
- Harrison, D., Rubinfeld, D. L., 1978. Hedonic housing prices and the demand for clear air. *Journal of Environmental Economics and Management* 5, 81-102.

- Jin, F., Lee, L. F., 2013. Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. *Regional Science and Urban Economics* 43, 590-616.
- 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., 2001. On the asymptotic distribution of the Moran *I* test statistic with applications. *Journal of Econometrics* 104, 219-257.
- Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.
- Lee, L. F., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489-514.
- Lee, L. F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26, 187-230.

- Liu, S. F., Yang, Z. L., 2015a. Asymptotic distribution and finite-sample bias correction of QML estimators for spatial error dependence model. *Econometrics* 3, 376-411.
- Liu, S. F., Yang, Z. L., 2015b. Improved inferences for spatial regression models. *Regional Science and Urban Economics* 55, 55-67.
- van der Vaart, A. W., 1998. *Asymptotic Statistics.* Cambridge: Cambridge University Press.
- White, H., 1994. *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.
- Yang, Z. (2015). A general method for third-order bias and variance correction on a nonlinear estimator. *Journal of Econometrics* 186, 178-200.