Lecture 4: Bias-Corrected Estimation and Refined Inference for Spatial Linear Regression Model

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

Many econometric models share the following common features:

(i) there are few nonlinear parameters that are the main source of bias in model estimation and main cause of difficulty in bias correction,

(ii) there are many other parameters in the model (linear or scale), but their constrained estimates, given the nonlinear parameters, are either unbiased or can be easily bias-corrected,

(iii) and these constrained estimates possess analytical expressions, rendering the concentrated estimating function (CEF) of nonlinear parameters an analytical form.

These include the spatial linear regression models covered in this lecture,

- the spatial panel models with fixed effects,
- the dynamic regression models,
- the dynamic (spatial) panel model with fixed effects,
- the Box-Cox regression, Weibull duration model, etc.

The above discussions suggest that for bias and variance corrections, one should focus on CEF, for dimension reduction and (more importantly) for capturing the additional variations from the estimation of linear and scale parameters, thus making the bias and variance corrections more effective.

- Yang (2015a) proposed a general method of bias and variance corrections, based on **stochastic expansions** and **bootstrap**, with the former providing tractable approximations to the bias terms and the latter making the 'bias corrections' practically implementable.
- The method gives a satisfactory treatment on a long-researched **bias problem** arising from the estimation of nonlinear parameters (Kiviet, 1995; Hahn and Kuersteiner, 2002; Hahn and Newey, 2004; Bun and Carree, 2005; Bao and Ullah, 2007a,b; Bao, 2013).
- The method also addresses another important issue: the high-order correction on the variance of a bias-corrected estimator.

In this lecture, we

1. introduce the general method for bias and variance corrections on nonlinear estimators, and a bootstrap method for practical implementation of bias and variance corrections;
2. demonstrate the applications of these methods by presenting detailed results for the spatial lag (SL) dependence model, giving,
    - the 2nd- and 3rd-order bias-corrected QMLEs,
    - refined inferences for covariates effect,
    - refined inferences for spatial effects,
    - and Monte Carlo results for their finite sample performance;
3. and outline the results for the closely related spatial models, the spatial error (SE) dependence model, the spatial linear regression (SLR) model with both SL and SE (SLE).

The methods presented apply to all econometric models satisfying the conditions, including the spatial models introduced in this course.

Consider a general class of $\sqrt{n}$-consistent estimators identified by the **moment condition** or **joint estimating equation (JEE)**: $\psi_n(\theta) = 0$, i.e.,

$$\hat{\theta}_n = \arg\{\psi_n(\theta) = 0\},$$

where $\psi_n(\theta) \equiv \psi_n(Z_n, \theta)$ is a $p \times 1$ vector-valued function of the observable data $Z_n = \{Z_i\}_{i=1}^n$ (iid or non-iid) and a parameter vector θ, of the same dimension as θ, and normalized to have order $O_p(n^{-1/2})$.

In studying the finite sample properties of $\hat{\theta}_n$, Bao and Ullah (2007a), extending Rilstone et al. (1996), developed:

- a 3rd-order stochastic expansion, a 2nd-order bias, and a 3rd-order MSE (mean squared error) for $\hat{\theta}_n$, assuming
- $\mathrm{E}\psi_n(\theta_0) = 0$, here θ_0 denotes the true value of the parameter vector θ.
- They adopted an analytical approach for estimating bias and MSE, which involves high-dimension matrices (up to $p \times p^3$) and high-order moments (up to 10th-order) of errors, and thus is limited in the scope of applications.

Yang (2015a, JOE) argued:

- the condition, $E\psi_n(\theta_0) = 0$, is neither necessary nor true in general.
- It is required for achieving asymptotic efficiency but not for achieving consistency (see, e.g., Amemiya, 1985; White, 1994);
- thus, the condition, $E\psi_n(\theta_0) = 0$, needs to be relaxed, in particular, under a constrained estimation framework;
- the vector of parameters θ may contain a set of linear and scale parameters, say α, and a few nonlinear parameters, say δ;
- given δ_0, the constrained estimator $\tilde{\alpha}_n(\delta_0)$ of α_0 possesses an explicit expression, which may be unbiased or easily bias corrected,
- e.g., for $\tilde{\beta}_n(\lambda_0)$ and $\tilde{\sigma}_n^2(\lambda_0)$ in the SL model considered in Lecture 2:

$$\mathrm{E}[\tilde{\beta}_n(\lambda_0)] = \beta_0 \text{ and } \mathrm{E}[\tfrac{n}{n-k}\tilde{\sigma}_n^2(\lambda_0)] = \sigma_0^2.$$

To fix ideas, partition $\psi_n(\theta)$ according to $\theta = (\alpha', \delta')'$, i.e.,

$$\psi_n(\theta) = \{\psi'_{n\alpha}(\alpha, \delta), \psi_{n\delta}(\alpha, \delta)\}'.$$

Then, the estimator $\hat{\delta}_n$ of δ would typically be

$$\hat{\delta}_n = \arg\{\tilde{\psi}_n(\delta) = 0\}, \tag{4.1}$$

where $\tilde{\psi}_n(\delta) \equiv \psi_{n\delta}(\tilde{\alpha}_n(\delta), \delta)$ is the CEF of δ, $\tilde{\alpha}_n(\delta)$ is the estimator of α at a given δ, and $\tilde{\psi}_n(\delta) = 0$ is the concentrated estimating equation (CEE).

Note: $\mathrm{E}[\psi_{n\delta}(\tilde{\alpha}_n(\delta_0), \delta_0)] \neq 0$ **even if** $\mathrm{E}[\psi_{n\delta}(\alpha_0, \delta_0)] = 0$, implying that $\hat{\delta}_n$ is biased **even if** $\tilde{\alpha}_n(\delta_0)$ is unbiased!

Therefore, Yang (2015a) argued that it is more effective to work with the CEF for bias and variance corrections,

- which not only reduces the dimensionality for the bias-correction problem (multi-dimensional to single-dimensional if δ is a scalar),
- but also takes into account the additional variability from the estimation of the 'nuisance' parameters α.
- Once the nonlinear estimator $\hat{\delta}_n$ is bias corrected to give $\hat{\delta}_n^{\mathrm{bc}}$, the bias of the resulting linear estimator $\hat{\alpha}_n = \tilde{\alpha}_n(\hat{\delta}_n^{\mathrm{bc}})$ will be greatly reduced.

Stochastic expansion on $\hat{\delta}_n$ and **bootstrap** provide a feasible tool!

Let $H_{rn}(\delta) = \nabla^r \tilde{\psi}_n(\delta), r = 1, 2, 3$, be the partial derivatives of $\tilde{\psi}_n(\delta)$, carried out sequentially and elementwise with respect to δ'. Let

$$\begin{array}{lllllll} \tilde{\psi}_n & \equiv & \tilde{\psi}_n(\delta_0), & & H_{rn} & \equiv & H_{rn}(\delta_0), r = 1, 2, 3, \\ H^{\circ}_{rn} & = & H_{rn} - \mathrm{E}(H_{rn}), & & \Omega_n & = & -\mathrm{E}(H_{1n})^{-1}. \end{array}$$

Let Δ be the parameter space of δ, and δ_0 be the true value of δ.

Yang (2015a, p.180, p.192) presents a set of sufficient conditions:

Assumption A: Δ is compact with δ_0 being an interior point. $\mathrm{E}(\tilde{\psi}_n) = O(n^{-1})$, and $\hat{\delta}_n$, as the solution of $\tilde{\psi}_n(\delta) = 0$, is a $\sqrt{n}$-consistent estimator of δ_0.

Assumption B: $\tilde{\psi}_n(\delta)$ is differentiable up to rth order for δ in a neighborhood of δ_0, $\mathrm{E}(H_{rn}) = O(1)$, and $H^{\circ}_{rn} = O_p(n^{-\frac{1}{2}}), r = 1, 2, 3$.

Assumption C: $\mathrm{E}(H_{1n})^{-1} = O(1)$, and $H_{1n}^{-1} = O_p(1)$.

Assumption D: $\|H_{rn}(\delta) - H_{rn}(\delta_0)\| \leq \|\delta - \delta_0\| U_n$ for δ in a neighborhood of $\delta_0, r = 1, 2, 3$, and $\mathrm{E}(|U_n|) < C < \infty$ for some constant C.

Theorem 4.1. Under Assumptions A-D, the estimator $\hat{\delta}_n$ defined in (4.1) possesses the following third-order stochastic expansion:

$$\hat{\delta}_n - \delta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(n^{-2}), \tag{4.2}$$

where, $a_{-s/2}$ represents terms of order $O_p(n^{-s/2})$, for $s = 1, 2, 3$, with

$$a_{-1/2} = \Omega_n \tilde{\psi}_n, \tag{4.3}$$

$$a_{-1} = \Omega_n H_{1n}^{\circ} a_{-1/2} + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2} \otimes a_{-1/2}), \tag{4.4}$$

$$\begin{aligned} a_{-3/2} = {} & \Omega_n H_{1n}^{\circ} a_{-1} + \tfrac{1}{2}\Omega_n H_{2n}^{\circ}(a_{-1/2} \otimes a_{-1/2}) \\ & + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) \\ & + \tfrac{1}{6}\Omega_n \mathrm{E}(H_{3n})(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}), \end{aligned} \tag{4.5}$$

and $\otimes$ denotes the Kronecker product. (see Yang, 2015a, Sec. 4).

This CEE-based expansion takes the same form as that given by Bao and Ullah (2007a) based on JEE, but it leads a different expansion for the bias.

Taking expectations on both sides of (4.2) gives an expansion for $\mathrm{Bias}(\hat{\delta}_n)$.

Corollary 4.1. Under Assumptions A-D, assuming further that **"a quantity bounded in probability has a finite expectation"**, we have a third-order expansion for the bias of $\hat{\delta}_n$, $\mathrm{Bias}(\hat{\delta}_n) = \hat{\delta}_n - \delta_0$:

$$\mathrm{Bias}(\hat{\delta}_n) = b_{-1} + b_{-3/2} + O(n^{-2}), \tag{4.6}$$

where $b_{-1} = \mathrm{E}(a_{-1/2} + a_{-1}) = O(n^{-1})$, and $b_{-3/2} = \mathrm{E}(a_{-3/2}) = O(n^{-3/2})$.

The more detailed expressions for the bias terms are:

$$b_{-1} = 2\Omega_n \mathrm{E}(\tilde{\psi}_n) + \Omega_n \mathrm{E}(H_{1n}\Omega_n\tilde{\psi}_n) + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})\mathrm{E}[(\Omega_n\tilde{\psi}_n) \otimes (\Omega_n\tilde{\psi}_n)], \tag{4.7}$$

$$\begin{aligned} b_{-3/2} = {} & \Omega_n \mathrm{E}(H_{1n}^{\circ} a_{-1}) + \tfrac{1}{2}\Omega_n \mathrm{E}[H_{2n}^{\circ}(a_{-1/2} \otimes a_{-1/2})] \\ & + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})\mathrm{E}(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) \\ & + \tfrac{1}{6}\Omega_n \mathrm{E}(H_{3n})\mathrm{E}(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}). \end{aligned} \tag{4.8}$$

- Thus, b_{-1} alone gives a second-order expansion for $\text{Bias}(\hat{\delta}_n)$, and
- $b_{-1} + b_{-3/2}$ gives a third-order expansion for $\text{Bias}(\hat{\delta}_n)$.
- This give options of performing a second-order bias correction on $\hat{\delta}_n$:

$$\hat{\delta}_n^{\text{bc2}} = \hat{\delta}_n - \widehat{b}_{-1}, \tag{4.9}$$

- or a third-order bias correction on $\hat{\delta}_n$:

$$\hat{\delta}_n^{\text{bc3}} = \hat{\delta}_n - \widehat{b}_{-1} - \widehat{b}_{-3/2}, \tag{4.10}$$

  giving a 2nd-order and a 3rd-order bias-corrected estimators.
- $\widehat{b}_{-1}$ and $\widehat{b}_{-3/2}$ are the estimates of b_{-1} and $b_{-3/2}$, respectively.
- Methods for obtaining $\widehat{b}_{-1}$ and $\widehat{b}_{-3/2}$ remain.

Note that $\text{E}(a_{-1/2}) = \Omega_n \text{E}(\tilde{\psi}_n) = O(n^{-1})$, which is a term 'missed' by the stochastic expansions based on joint estimating functions, and is shown to play a pivotal role in bias and variance corrections.

To understand (4.2), suppose δ is a scalar. Assumptions A-D allow Taylor expansions of $\tilde{\psi}_n(\hat{\delta}_n) = 0$ around δ_0 be carried out up to 3rd-order:

$$0 = \tilde{\psi}_n + H_{1n}(\hat{\delta}_n - \delta_0) + O_p(n^{-1}),$$
$$0 = \tilde{\psi}_n + H_{1n}(\hat{\delta}_n - \delta_0) + \tfrac{1}{2}H_{2n}(\hat{\delta}_n - \delta_0)^2 + O_p(n^{-\frac{3}{2}}),$$
$$0 = \tilde{\psi}_n + H_{1n}(\hat{\delta}_n - \delta_0) + \tfrac{1}{2}H_{2n}(\hat{\delta}_n - \delta_0)^2 + \tfrac{1}{6}H_{3n}(\hat{\delta}_n - \delta_0)^3 + O_p(n^{-2}),$$

which give, as $-H_{1n}^{-1} = O_p(1)$ from Assumption C,

$$\hat{\delta}_n - \delta_0 = -H_{1n}^{-1}\tilde{\psi}_n + O_p(n^{-1}),$$
$$\hat{\delta}_n - \delta_0 = -H_{1n}^{-1}\tilde{\psi}_n - \tfrac{1}{2}H_{1n}^{-1}H_{2n}(\hat{\delta}_n - \delta_0)^2 + O_p(n^{-\frac{3}{2}}),$$
$$\hat{\delta}_n - \delta_0 = -H_{1n}^{-1}\tilde{\psi}_n - \tfrac{1}{2}H_{1n}^{-1}H_{2n}(\hat{\delta}_n - \delta_0)^2 - \tfrac{1}{6}H_{1n}^{-1}H_{3n}(\hat{\delta}_n - \delta_0)^3 + O_p(n^{-2}).$$

Under Assumptions B and C, we have $\Omega_n = -\mathrm{E}(H_{1n})^{-1} = O(1)$, $H_{1n}^{-1} = O_p(1)$, and $H_{1n}^{\circ} = H_{1n} - \mathrm{E}(H_{1n}) = O_p(n^{-1/2})$, which lead to:

$$-H_{1n}^{-1} = (\Omega_n^{-1} - H_{1n}^{\circ})^{-1} = (1 - \Omega_n H_{1n}^{\circ})^{-1}\Omega_n = \Omega_n + \Omega_n^2 H_{1n}^{\circ} + \Omega_n^3 H_{1n}^{\circ 2} + O_p(n^{-\frac{3}{2}}),$$

or $-H_{1n}^{-1} = \Omega_n + \Omega_n^2 H_{1n}^{\circ} + O_p(n^{-1})$, or $-H_{1n}^{-1} = \Omega_n + O_p(n^{-1/2})$.

Based on the above results, recursive substitutions of the rth-order form of $-H_{1n}^{-1}$ into the rth-order Taylor expansion of $\hat{\delta}_n - \delta_0$, $r = 1, 2, 3$, lead to a 3rd-order stochastic expansion for $\hat{\delta}_n$, with $a_{-1/2} = \Omega_n \tilde{\psi}_n$,

$$a_{-1} = \Omega_n H_{1n}^{\circ} a_{-1/2} + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2}^2), \tag{4.11}$$

$$\begin{aligned} a_{-3/2} &= \Omega_n H_{1n}^{\circ} a_{-1} + \tfrac{1}{2}\Omega_n H_{2n}^{\circ}(a_{-1/2}^2) \\ &\quad + \Omega_n \mathrm{E}(H_{2n})(a_{-1/2} a_{-1}) + \tfrac{1}{6}\Omega_n \mathrm{E}(H_{3n})(a_{-1/2}^3). \end{aligned} \tag{4.12}$$

(See Yang, 2015a, Theorem 2.1).

Under a scalar δ, the bias terms b_{-1} and $b_{-3/2}$ simplify to:

$$b_{-1} = 2\Omega_n \mathrm{E}(\tilde{\psi}_n) + \Omega_n^2 \mathrm{E}(H_{1n}\tilde{\psi}_n) + \tfrac{1}{2}\Omega_n^3 \mathrm{E}(H_{2n})\mathrm{E}(\tilde{\psi}_n^2), \tag{4.13}$$

$$\begin{aligned} b_{-3/2} &= \Omega_n \mathrm{E}(\tilde{\psi}_n) + 2\Omega_n^2 \mathrm{E}(H_{1n}\tilde{\psi}_n) + \Omega_n^3 \mathrm{E}(H_{2n})\mathrm{E}(\tilde{\psi}_n^2) + \Omega_n^3 \mathrm{E}(H_{1n}^2\tilde{\psi}_n) \\ &\quad + \tfrac{1}{2}\Omega_n^3 \mathrm{E}(H_{2n}\tilde{\psi}_n^2) + \tfrac{3}{2}\Omega_n^4 \mathrm{E}(H_{2n})\mathrm{E}(H_{1n}\tilde{\psi}_n^2) \\ &\quad + \tfrac{1}{2}\Omega_n^5 \mathrm{E}(H_{2n})^2 \mathrm{E}(\tilde{\psi}_n^3) + \tfrac{1}{6}\Omega_n^4 \mathrm{E}(H_{3n})\mathrm{E}(\tilde{\psi}_n^3). \end{aligned} \tag{4.14}$$

How to estimate b_{-1} and $b_{-3/2}$?

- Finding their analytical expressions and plugging-in the parameter estimates as suggested by Bao and Ullah (2007a)?
- It is difficult (if not impossible) to do so, due to the complication of the terms, e.g., $H_{2n}\tilde{\psi}_n^2$, $H_{1n}\tilde{\psi}_n^2$ and $\tilde{\psi}_n^3$, of which expectations are desired,
- in particular when errors are nonnormal and heteroskedastic!
- In case of a vector δ, the analytical approach runs into a greater difficulty as the desired expectations are of more complicated forms, e.g., $\mathrm{E}(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}) = \mathrm{E}(\Omega_n\tilde{\psi}_n \otimes \Omega_n\tilde{\psi}_n \otimes \Omega_n\tilde{\psi}_n)$.
- Thus, the bootstrap is perhaps the ONLY feasible way!

In what follows, we introduce a simple and reliable bootstrap method for bootstrapping the various expected quantities appeared in b_{-1} and $b_{-3/2}$.

Suppose that the model under consideration takes the form:

$$g(Z_n, \theta_0) = e_n,$$

- Z_n represents the 'data' on n cross-sectional units,
- θ the vector model parameters, with θ_0 being its true value,
- the elements $\{e_{n,i}\}$ of e_n are iid with mean 0 and variance 1.

Suppose that the key quantities $\tilde{\psi}_n$ and H_{rn} can be expressed as $\tilde{\psi}_n \equiv \tilde{\psi}_n(e_n, \theta_0)$ and $H_{rn} \equiv H_{rn}(e_n, \theta_0), r = 1, 2, 3$.

Let $\hat{e}_n = g(Z_n, \hat{\theta}_n)$ be the vector of estimated residuals based on the original data, and

$$\hat{e}_n - \text{mean}(\hat{e}_n)1_n,$$

be the 'centered' estimated residuals denoted by $\hat{e}_n$ again to save notation. **Let** $\hat{\mathcal{F}}_n$ be the empirical distribution function (EDF) of $\hat{e}_n$, i.e.,

$$\hat{\mathcal{F}}_n(y) = \frac{1}{n}\sum_{i=1}^{n}\{\hat{e}_{n,i} \le y\}.$$

When δ is a scalar parameter, we see from (4.14) that the quantities needed to be estimated in bias terms take the form:

$$\mathrm{E}(\tilde{\psi}_n^i H_{rn}^j) = \mathrm{E}[\tilde{\psi}_n^i(e_n, \theta_0) H_{rn}^j(e_n, \theta_0)], \quad i, j = 0, 1, 2, \ldots, \quad r = 1, 2, 3.$$

Hence, the corresponding bootstrap estimates of these quantities are:

$$\widehat{\mathrm{E}}(\tilde{\psi}_n^i H_{rn}^j) = \mathrm{E}^*[\tilde{\psi}_n^i(\hat{e}_n^*, \hat{\theta}_n) H_{rn}^j(\hat{e}_n^*, \hat{\theta}_n)], \tag{4.15}$$

for $i, j = 0, 1, 2, \ldots, \quad r = 1, 2, 3,$ where

- E^* denotes the expectation with respect to $\hat{\mathcal{F}}_n$,
- $\hat{e}_n^*$ is a vector of n random draws from $\hat{\mathcal{F}}_n$,
- $\hat{\theta}_n$ is the original estimate of θ_0, now a *constant* with respect to $\hat{\mathcal{F}}_n$.
- Finding the 'exact' bootstrap expectations (4.15) are still practically infeasible as the number of different bootstrap samples can be huge.
- However, with a 'small' number of bootstrap samples, one can obtain accurate approximations to these bootstrap expectations.

Feasible versions of bootstrap estimates can be obtained as follows:

Bootstrap Algorithm 1 (BA-1):

1. Compute the parameter estimate $\hat{\theta}_n$ defined by JEF, the estimated residuals $\hat{e}_n = g(Z_n, \hat{\theta}_n)$, and the EDF $\hat{\mathcal{F}}_n$ of the centered residuals $\hat{e}_n$;
2. Draw a random sample of size n from $\hat{\mathcal{F}}_n$, and denote the resampled vector by $\hat{e}^*_{n,b}$,
3. Compute $\tilde{\psi}_n(\hat{e}^*_{n,b}, \hat{\theta}_n)$ and $H_{rn}(\hat{e}^*_{n,b}, \hat{\theta}_n), r = 1, 2, 3;$
4. Repeat steps 2.-3. for B times, to give approximate bootstrap estimates as,

$$\mathrm{E}^*[\tilde{\psi}^i_n(\hat{e}^*_n, \hat{\theta}_n)H^j_{rn}(\hat{e}^*_n, \hat{\theta}_n)] \doteq \frac{1}{B}\sum_{b=1}^{B} \tilde{\psi}^i_n(\hat{e}^*_{n,b}, \hat{\theta}_n)H^j_{rn}(\hat{e}^*_{n,b}, \hat{\theta}_n),$$

   for $i, j = 0, 1, 2, \ldots, \quad r = 1, 2, 3.$

The approximations in the last step can be made arbitrarily accurate by choosing an arbitrarily large B. In many practical applications, however, $B = 999$ suffices. Other quantities are handled in a similar manner.

Validity of the Bootstrap Method:

Yang (2015a) shows that under certain conditions:

$$
\begin{aligned}
\text{Bias}(\delta_n^{\text{bc2}}) &= \text{Bias}(\hat{\delta}_n) - \text{E}(\hat{b}_{-1}) \\
&= -\text{Bias}(\hat{b}_{-1}) + O(n^{-3/2}) = O(n^{-3/2}), \text{ and} \\
\text{Bias}(\delta_n^{\text{bc3}}) &= \text{Bias}(\hat{\delta}_n) - \text{E}(\hat{b}_{-1}) - \text{E}(\hat{b}_{-3/2}) \\
&= -\text{Bias}(\hat{b}_{-1}) - \text{Bias}(\hat{b}_{-3/2}) + O(n^{-2}) = O(n^{-2}).
\end{aligned}
$$

- Additional variations from the bootstrap approximations do not change the order of the bias.
- The above bootstrap procedures for bias corrections are valid in the sense that it gives the desired order of bias reduction.
- See Yang (2015a) for detailed proofs.

When δ is a vector, the non-stochastic and stochastic quantities are mixed in b_{-1} and $b_{-3/2}$. Yang (2015a) proposed that instead of going through the algebraic procedure to separate the two types of quantities so that the expectations of various quantities can be bootstrapped in one round, the above bootstrap procedure can be revised as follows.

Bootstrap Algorithm 2 (BA-2):

1. Draw B independent random samples, $\{\hat{e}^*_{n,b},\ b = 1, \ldots, B\}$, from $\hat{\mathcal{F}}_n$,
2. Calculate the bootstrap estimates of $\mathrm{E}(H_{1n})$ and $\mathrm{E}(H_{2n})$,

$$\widehat{\mathrm{E}}(H_{1n}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{n} H_{1n}(\hat{e}^*_{n,b}, \hat{\theta}_n) \text{ and } \widehat{\mathrm{E}}(H_{2n}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{n} H_{2n}(\hat{e}^*_{n,b}, \hat{\theta}_n)$$

3. Based on the bootstrap estimates $\hat{\Omega}_n = -\widehat{\mathrm{E}}^{-1}(H_{1n})$ and $\widehat{\mathrm{E}}(H_{2n})$, calculate the bootstrap estimate of, e.g., $\mathrm{E}[H^{\circ}_{2n}(a_{-1/2} \otimes a_{-1/2})]$, as

$$\tfrac{1}{B}\textstyle\sum_{b=1}^{n} \{[H_{2n}(\hat{e}^*_{n,b}, \hat{\theta}_n) - \widehat{\mathrm{E}}(H_{2n})][\hat{\Omega}_n\tilde{\psi}_n(\hat{e}^*_{n,b}, \hat{\theta}_n) \otimes \hat{\Omega}_n\tilde{\psi}_n(\hat{e}^*_{n,b}, \hat{\theta}_n)]\}.$$

The other quantities can be handled in a similar manner.

- This is essentially a two-round bootstrap procedure as it runs the iterations $b = 1, 2, \ldots, B$ two times, based on the same sequence of bootstrap samples.
- Computationally it is slightly more demanding, but algebraically it is much simpler and thus easier to code.
- As noted by Yang (2015a), these procedures are time-efficient as the reestimation of the parameters θ in the bootstrap process is avoided.
- This is in stark contrast to the traditional bootstrap method for bias correction in which every bootstrap iteration has to go through a numerical optimization in order to obtain bootstrap estimates of θ.

In summary, the stochastic expansions coupled with bootstrap provide a **general and feasible methodology** for bias corrections on nonlinear estimators. Once this is done, the linear and scale estimators will 'automatically' become nearly unbiased.

Under a similar set of conditions as those for the bias expansion, we have a third-order expansion for the variance of $\hat{\delta}_n$:

$$\mathrm{Var}(\hat{\delta}_n) = v_{-1} + v_{-3/2} + v_{-2} + O(n^{-\frac{5}{2}}), \tag{4.16}$$

- $v_{-1} = \mathrm{Var}(a_{-1/2})$, the 1st-order (asymptotic) variance,
- $v_{-3/2} = 2\mathrm{Cov}(a_{-1/2},\ a_{-1})$, the 2nd-order variance,
- $v_{-2} = 2\mathrm{Cov}(a_{-1/2},\ a_{-3/2}) + \mathrm{Var}(a_{-1} + a_{-3/2})$, 3rd-order variance,

which are of order $O(n^{-1})$, $O(n^{-3/2})$, $O(n^{-2})$, respectively.

To estimate the second-order variance, the same bootstrap procedure can be followed to give valid bootstrap estimates $\widehat{v}_{-1}$ and $\widehat{v}_{-3/2}$ of $v_{-1} + v_{-3/2}$.

For third-order variance estimate, $\widehat{v}_{-1}$ needs to be further corrected in order to eliminate the third-order bias. See Yang (2015a, p.182) for details.

- An important purpose of bias and variance corrections is to improve the asymptotic inference methods presented in the earlier Lectures.
- There are mainly two types of inferences that could benefit from the bias-corrections on the nonlinear estimators: one is the inference for the nonlinear parameters, and the other for the linear parameters.
- In the framework of linear regressions with spatial dependence, the spatial parameters are the nonlinear parameters, and the regression coefficients are the linear parameters.
- Improved tests for spatial effects have been considered by Baltagi and Yang (2013a,b), Robinson and Rossi (2014a,b), Yang (2010), and Yang (2015b).
- However, the issue of improved inferences for the regression coefficients has not been considered until Liu and Yang (2015a,b).
- Improved Wald-type of tests have not been fully explored.

To fix ideas, we focus on the 2nd-order bias-corrected $\hat{\delta}_n$, the $\hat{\delta}_n^{\text{bc2}}$.

- Let $\hat{\alpha}_n \equiv \tilde{\alpha}_n(\hat{\delta}_n)$, $\hat{\alpha}_n^{\text{bc}} \equiv \tilde{\alpha}_n(\hat{\delta}_n^{\text{bc2}})$, $\hat{\theta}_n = (\hat{\alpha}_n', \hat{\delta}_n')'$ and $\hat{\theta}_n^{\text{bc}} = (\hat{\alpha}_n^{\text{bc}\prime}, \hat{\delta}_n^{\text{bc2}\prime})'$.
- Yang (2015a) argued that estimation of the nonlinear parameters is the main source of bias; once the nonlinear estimators are bias corrected the resulting linear estimators would be nearly unbiased.
- Let $V_n(\theta_0)$ be the asymptotic variance-covariance (VC) matrix of $\hat{\theta}_n$, and $V_{n,\alpha\alpha}(\theta_0)$ be the α-α block and $V_{n,\delta\delta}(\theta_0)$ the δ-δ block of $V_n(\theta_0)$.

Then, an asymptotic *t*-statistic for inference for $c_0'\alpha_0$, a linear contrast of α_0, has the familiar form:

$$\mathrm{t}_n(\alpha_0) = (c_0'\hat{\alpha}_n - c_0'\alpha_0)/\sqrt{c_0' V_{n,\alpha\alpha}(\hat{\theta}_n)c_0}. \quad (4.17)$$

Simply replacing $\hat{\theta}_n$ by $\hat{\theta}_n^{\text{bc}}$, a possibly improved *t*-statistic results:

$$\mathrm{t}_n^{\text{bc}}(\alpha_0) = (c_0'\hat{\alpha}_n^{\text{bc}} - c_0'\alpha_0)/\sqrt{c_0' V_{n,\alpha\alpha}(\hat{\theta}_n^{\text{bc}})c_0}. \quad (4.18)$$

- The statistic $\mathrm{t}_n^{\mathrm{bc}}$ is not fully 2nd-order corrected as it uses the asymptotic variance of $\hat{\alpha}_n$ evaluated at $\hat{\theta}_n^{\mathrm{bc}}$.
- Furthermore, the estimator $\hat{\alpha}_n^{\mathrm{bc}}$ is also not fully 2nd-order bias-corrected, although it can easily be made so.
- Let $\hat{\alpha}_n^{\mathrm{bc2}}$ be the 2nd-order bias-corrected $\hat{\alpha}_n$ or $\hat{\alpha}_n^{\mathrm{bc}}$.
- Let $V_{n,\alpha\alpha}^{\mathrm{bc2}}(\theta_0)$ be the 2nd-order variance of $\hat{\alpha}_n^{\mathrm{bc2}}$, and $\widehat{V}_{n,\alpha\alpha}^{\mathrm{bc2}}$ be its consistent estimate.

A fully 2nd-order corrected t-statistic, using a 2nd-order bias-corrected estimator and its 2nd-order variance estimate, is thus:

$$\mathrm{t}_n^{\mathrm{bc2}}(\alpha_0) = (c_0'\hat{\alpha}_n^{\mathrm{bc2}} - c_0'\alpha_0)\Big/\sqrt{c_0'\,\widehat{V}_{n,\alpha\alpha}^{\mathrm{bc2}}c_0}. \tag{4.19}$$

Typically, $V_{n\,\alpha\alpha}^{\mathrm{bc2}}(\theta_0)$ does not have an explicit expression, but the bootstrap methods described above can be used to give a consistent estimate of it. See the subsequent sections for details.

Inferences for nonlinear parameter δ.

For inferences concerning δ, the bias-corrected estimators (2nd- and 3rd-order) and their variance estimates (2nd- and 3rd-order) are directly obtained in the bootstrap bias and variance corrections process.

Potentially improved Wald statistics for inferences for δ takes the form:

$$\mathrm{t}_n^{\mathrm{bcj}}(\delta_0) = (\hat{\delta}_n^{\mathrm{bcj}} - \delta_0)'(\widehat{V}_{n,\delta\delta}^{\mathrm{bcj}})^{-1}(\hat{\delta}_n^{\mathrm{bcj}} - \delta_0),\ j = 2, 3, \tag{4.20}$$

where $\widehat{V}_{n,\delta\delta}^{\mathrm{bcj}}$ is the estimate of the jth-order variance of $\hat{\delta}_n^{\mathrm{bcj}}$, and $\hat{\delta}_n^{\mathrm{bcj}}$ is the jth-order bias-corrected estimator of δ.

Referring to Section 2.3, Lecture 2, for the QML estimation of SL model based on the concentrated loglikelihood function $\ell_n^c(\delta)$, and letting

$$\tilde{\psi}_n(\lambda) = \frac{1}{n}\frac{\partial}{\partial\lambda}\ell_n^c(\lambda),$$

the $\tilde{\psi}_n(\lambda)$ and its derivatives $H_{sn}(\lambda), s = 1, 2, 3$, required for up to 3rd-order bias correction on the QMLE $\hat{\lambda}_n$ are,

$$\tilde{\psi}_n(\lambda) = -\, T_{0n}(\lambda) + R_{1n}(\lambda), \tag{4.21}$$

$$H_{1n}(\lambda) = -\, T_{1n}(\lambda) - R_{2n}(\lambda) + 2R_{1n}^2(\lambda), \tag{4.22}$$

$$H_{2n}(\lambda) = -\, 2T_{2n}(\lambda) - 6R_{1n}(\lambda)R_{2n}(\lambda) + 8R_{1n}^3(\lambda), \tag{4.23}$$

$$H_{3n}(\lambda) = -\, 6T_{3n}(\lambda) + 6R_{2n}^2(\lambda) - 48R_{1n}^2(\lambda)R_{2n}(\lambda) + 48R_{1n}^4(\lambda), \tag{4.24}$$

where $T_{rn}(\lambda) = n^{-1}\mathrm{tr}(G_n^{r+1}(\lambda)), r = 0, 1, 2, 3,\ G_n(\lambda) = W_n A_n^{-1}(\lambda)$,

$$R_{1n}(\lambda) = \frac{Y_n' A_n'(\lambda) M_n W_n Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n} \text{ and } R_{2n}(\lambda) = \frac{Y_n' W_n' M_n W_n Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n}. \tag{4.25}$$

Two key ratios, R_{1n} and R_{2n}, can be written at θ_0 as:

$$R_{1n}(e_n, \theta_0) = \frac{e_n' M_n G_n e_n + e_n' M_n \eta_n}{e_n' M_n e_n},$$

$$R_{2n}(e_n, \theta_0) = \frac{e_n' G_n' M_n G_n e_n + 2 e_n' G_n' M_n \eta_n + \eta_n' M_n \eta_n}{e_n' M_n e_n},$$

where $e_n = \sigma_0^{-1} \epsilon_n$, and other quantities are defined in Sec. 2.3, Lec. 2.

- Hence, $\tilde{\psi}_n = \tilde{\psi}_n(e_n, \theta_0)$ and $H_{rn} = H_{rn}(e_n, \theta_0),\ r = 1, 2, 3.$
- So, the bootstrap algorithm **BA-1** can be used to obtain bootstrap estimates $\widehat{b}_{-1}$ and $\widehat{b}_{-3/2}$ of b_{-1} and $b_{-3/2}$, to give 2nd- or 3rd-order bias-corrected $\hat{\lambda}_n$:

$$\hat{\lambda}_n^{\text{bc2}} = \hat{\lambda}_n - \widehat{b}_{-1} \ \text{ and } \ \hat{\lambda}_n^{\text{bc3}} = \hat{\lambda}_n - \widehat{b}_{-1} - \widehat{b}_{-3/2}.$$

- In a similar manner, the bootstrap estimates $\widehat{v}_{-1}$ and $\widehat{v}_{-3/2}$ can be obtained to give 2nd- and 3rd-order corrected estimates of the variance of $\hat{\lambda}_n$. See Yang (2015a) for details.

- Although R_{jn} are simply the ratios of quadratic forms in e_n, finding its moments analytically seems extremely difficult, in particular when e_n is allowed to be nonnormal.
- This is made more difficult when more complicated models, such as the SE and SLE models to be considered next, and the FE-SPD (fixed effects spatial panel data) model to be studied in Lecture 7.
- Even for merely a second-order bias correction, it requires $\mathrm{E}(H_{1n}\tilde{\psi}_n)$, and hence $\mathrm{E}[(R_{1n})^3]$, involving up to 6 different moments of $\epsilon_{n,i}$ if a general error distribution is allowed. Besides, estimation of higher-order moments may be numerically unstable.
- In stark contrast, estimation of $\mathrm{E}[(R_{1n})^3]$ using the suggested bootstrap method (**BA-1**) is extremely simple. The method is also very general as it does not require the knowledge of true error distribution.

Table 4.1. Monte Carlo Mean[rmse](sd) of Estimators of λ in SL Model

λ	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 4.2. Monte Carlo Mean[rmse](sd) of Estimators of λ in SL Model

λ	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 4.3. Monte Carlo Mean[rmse](sd) of Estimators of λ in SL Model

λ	n	$\hat{\lambda}_n$	$\hat{\lambda}_n^{\text{bc2}}$	$\hat{\lambda}_n^{\text{bc3}}$
		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)

First, simply replacing $\hat{\lambda}_n$ by $\hat{\lambda}_n^{\mathrm{bc2}}$ in (3.42), we obtain a Wald statistic for β_0:

$$\mathrm{t}_{\mathrm{CE}}^{\mathrm{bc}}(\beta_0) = \frac{c_0' \hat{\beta}_n^{\mathrm{bc}} - c_0' \beta_0}{\sqrt{c_0' \hat{V}_{\mathrm{SL},\beta\beta}^{\mathrm{bc}} c_0}}, \tag{4.26}$$

which is expected to have a better finite sample performance,

- $\hat{V}_{\mathrm{SL},\beta\beta}^{\mathrm{bc}}$ is $V_{\mathrm{SL},\beta\beta}$ in (3.43) but evaluated at $\hat{\lambda}_n^{\mathrm{bc2}}$, $\hat{\beta}_n^{\mathrm{bc}}$, $\hat{\sigma}_n^{2,\mathrm{bc}}$, $\hat{\gamma}_n^{\mathrm{bc}}$, and $\hat{\kappa}_n^{\mathrm{bc}}$,
- where $\hat{\beta}_n^{\mathrm{bc}} = \tilde{\beta}_n(\hat{\lambda}_n^{\mathrm{bc2}})$, $\hat{\sigma}_n^{2,\mathrm{bc}} = \tilde{\sigma}_n^2(\hat{\lambda}_n^{\mathrm{bc2}})$, and
- $\hat{\gamma}_n^{\mathrm{bc}}$ and $\hat{\kappa}_n^{\mathrm{bc}}$ are the estimates of γ and κ, based on residuals at $\hat{\lambda}_n^{\mathrm{bc2}}$.

Obviously, this statistic is not fully second-order bias-corrected.

- However, Monte Carlo results presented in Liu and Yang (2015b) show that it offers a huge improvement over t_{CE} given in.
- This confirms the point made at in Lecture 2.
- However, results also show that when n is not so large, there is still room for further improvement on $\mathrm{t}_{\mathrm{CE}}^{\mathrm{bc}}(\beta_0)$.

To further improve $\mathrm{t}_{\mathrm{CE}}^{\mathrm{bc}}(\beta_0)$, note that

$$\begin{aligned}
&\hat{\beta}_n - \beta_0 \\
&= \tilde{\beta}_n - \beta_0 - (\hat{\lambda}_n - \lambda_0)(X_n'X_n)^{-1}X_n'G_nX_n\beta_0 - (\hat{\lambda}_n - \lambda_0)(X_n'X_n)^{-1}X_n'G_n\epsilon_n \\
&= (X_n'X_n)^{-1}X_n'\big[\epsilon_n - (a_{-1/2} + a_{-1})G_nX_n\beta_0 - a_{-1/2}G_n\epsilon_n\big] + O_p(n^{-3/2}).
\end{aligned} \tag{4.27}$$

This leads immediately to a 2nd-order bias-corrected estimator of β,

$$\hat{\beta}_n^{\mathrm{bc2}} = \hat{\beta}_n - (X_n'X_n)^{-1}X_n'\big(\hat{a}_{-1/2} + \hat{a}_{-1}\big)\hat{G}_n^{\mathrm{bc2}}X_n\hat{\beta}_n,$$

where $\hat{a}_{-1/2}$ and $\hat{a}_{-1}$ are the bootstrap estimates of $a_{-1/2}$ and a_{-1}obtained in the process of obtaining $\hat{\lambda}_n^{\mathrm{bc2}}$, and $\hat{G}_n^{\mathrm{bc2}} = G_n(\hat{\lambda}_n^{\mathrm{bc2}})$.

Similar to (4.27), $\hat{\sigma}_n^2$ can be expressed in $(\hat{\lambda}_n - \lambda_0)$, and hence in $a_{-1/2}$ and a_{-1}, leading to a 2nd-order bias-corrected estimator $\hat{\sigma}_n^{\mathrm{bc2}}$.

Now, a second-order expansion for Var$(\hat{\beta}_n)$ takes the form,

$$\mathrm{Var}(\hat{\beta}_n) = (X_n'X_n)^{-1}\mathrm{Var}(g_n)(X_n'X_n)^{-1} + O(n^{-2}),$$

where $g_n \equiv g_n(e_n, \theta_0) = X_n'\big[\epsilon_n - (a_{-1/2} + a_{-1})G_nX_n\beta_0 - a_{-1/2}G_n\epsilon_n\big]$.

An explicit expression of $\mathrm{Var}(g_n)$ is difficult to obtain, but is not needed as it can be easily estimated by a two-stage bootstrap procedure:

Stage 1: *Compute* $\hat{\theta}_n$ *and the QML residuals* $\hat{e}_n = \hat{\sigma}_n^{-1}(\hat{A}_n Y_n - X_n \hat{\beta}_n)$. *Resample* $\hat{e}_n$ *to give* $\hat{\lambda}_n^{\mathrm{bc2}}$, *and hence* $\hat{\beta}_n^{\mathrm{bc2}}$ *and* $\hat{\sigma}_n^{2,\mathrm{bc2}}$, *using Algorithm* `BA-1` *given in Sec. 4.2.*

Stage 2: *Update the QML residuals as* $\hat{e}_n^{\mathrm{bc2}} = \hat{\sigma}_n^{\mathrm{bc2},-1}(\hat{A}_n^{\mathrm{bc2}} Y_n - X_n \hat{\beta}_n^{\mathrm{bc2}})$ *and compute* $g_{n,b}^* \equiv g(\hat{e}_{n,b}^{\mathrm{bc2}*}, \hat{\theta}_n^{\mathrm{bc2}})$ *for* $b = 1, \ldots, B$, *where* $\hat{e}_{n,b}^{\mathrm{bc2}*}$ *is the bth bootstrap sample drawn from the EDF of* $\hat{e}_n^{\mathrm{bc2}}$, *and* $\hat{\theta}_n^{\mathrm{bc2}} = (\hat{\beta}_n^{\mathrm{bc2}\prime}, \hat{\sigma}_n^{\mathrm{bc2}}, \hat{\rho}_n^{\mathrm{bc2}})'$. *The bootstrap estimate of Var*$(\hat{\beta}_n^{\mathrm{bc2}})$, *unbiased up to* $O(n^{-3/2})$, *is thus,*

$$\widehat{\mathrm{Var}}(g_n) = \frac{1}{B}\sum_{b=1}^{B} g_{n,b}^* g_{n,b}^{*\prime} - \left(\frac{1}{B}\sum_{b=1}^{B} g_{n,b}^*\right)\left(\frac{1}{B}\sum_{b=1}^{B} g_{n,b}^{*\prime}\right).$$

The bootstrap estimate of $\mathrm{Var}(\hat{\beta}_n)$ is thus

$$\widehat{\mathrm{Var}}(\hat{\beta}_n^{\mathrm{bc2}}) = (X_n' X_n)^{-1} \widehat{\mathrm{Var}}(g_n) (X_n' X_n)^{-1},$$

leading to the 2nd-order bias-corrected *t*-statistic:

$$\mathrm{t}_{\mathrm{CE}}^{\mathrm{bc2}}(\beta_0) = \frac{c_0' \hat{\beta}_n^{\mathrm{bc2}} - c_0' \beta_0}{\sqrt{c_0' \widehat{\mathrm{Var}}(\hat{\beta}_n^{\mathrm{bc2}}) c_0}}. \quad (4.28)$$

Some final notes are as follows.

- As $\hat{\beta}_n^{\rm bc}$ and $\hat{\beta}_n^{\rm bc2}$ do not differ much, and $\hat{\sigma}_n^{2,{\rm bc}}$ and $\hat{\sigma}_n^{2,{\rm bc2}}$ also do not differ much, one can simply use $\hat{\beta}_n^{\rm bc}$ and $\hat{\sigma}_n^{2,{\rm bc}}$ in practical applications.
- The same remarks as above apply to the other models to be discussed latter.
- See Yang (2015a), Liu and Yang (2015b), and Yang et al. (2016) for the Monte Carlo results for the SL, SE, SLE, and FE-SPD models.
- Improved *t*-statistic for inference for λ_0, denoted by

  $$\mathrm{t}_{\tt SL}^{\rm bc2}(\lambda_0),$$

  can be obtained directly from $\mathrm{t}_{\tt SL}(\lambda_0)$ given in (3.43), following the ideas laid out around (4.20), as the fully corrected $\hat{\lambda}_n$ and its variance (2nd- and 3rd-order) are available.
- It would be interesting to investigate the finite sample properties of these tests, and compare them with the LM tests.

4.4. Bias Correction and Refined Inference for SE Model

From the concentrated loglikelihood $\ell_n^c(\rho)$ given in Section 2.2, the CEF for δ can be expressed in the following form:

$$\tilde{\psi}_n(\rho) = \frac{1}{n}\frac{\partial}{\partial \rho}\ell_n^c(\rho) = -T_{0n}(\rho) + R_{1n}(\rho), \tag{4.29}$$

The derivatives of $\tilde{\psi}_n(\rho)$ required for up to third-order bias correction are:

$$H_{1n}(\rho) = -T_{1n}(\rho) + R_{2n}(\rho) + 2R_{1n}^2(\rho), \tag{4.30}$$

$$H_{2n}(\rho) = -2T_{2n}(\rho) + R_{3n}(\rho) + 6R_{1n}(\rho)R_{2n}(\rho) + 8R_{1n}^3(\rho), \tag{4.31}$$

$$\begin{aligned} H_{3n}(\rho) = {} & -6T_{3n}(\rho) + R_{4n}(\rho) + 8R_{1n}(\rho)R_{3n}(\rho) + 6R_{2n}^2(\rho) \\ & +48R_{1n}^2(\rho)R_{2n}(\rho) + 48R_{1n}^4(\rho), \end{aligned} \tag{4.32}$$

where $T_{rn}(\rho) = \frac{1}{n}tr(G_n^{r+1}(\rho)), r = 0, 1, 2, 3$, and

$$R_{jn}(\rho) = \frac{Y_n' A_n'(\rho) M_n(\rho) D_{jn}(\rho) M_n(\rho) A_n(\rho) Y_n}{Y_n' A_n'(\rho) M_n(\rho) A_n(\rho) Y_n}, \ j = 1, 2, 3, 4, \tag{4.33}$$

where $D_{1n}(\rho) = G_n(\rho)$, and $D_{jn}(\rho)$, $j = 2, 3, 4$, take more complicated expressions and can be found in Appendix B of Liu and Yang (2015a).

From (36), it is easy to see that at the true parameter ρ_0,

$$R_{jn} \equiv R_{jn}(e_n, \rho_0) = \frac{e_n' \Lambda_{jn}(\rho_0) e_n}{e_n' M_n(\rho_0) e_n}, \; j = 1, 2, 3, 4, \qquad (4.34)$$

the ratios of quadratic forms in e_n, where $\Lambda_{jn}(\rho_0) = M_n(\rho_0) D_{jn}(\rho_0) M_n(\rho_0)$.

- $\tilde{\psi}_n$ and H_{rn} are functions of $R_{jn}, j = 1, \ldots, 4$, as seen from (4.29)-(4.32).
- Thus, to estimate bias, we need to estimate the expectations of R_{jn}, their powers, cross products, and cross products of powers.
- The general bootstrap procedure **BA-1** leads to bootstrap estimates of the bias terms b_{-1} and $b_{-3/2}$ and variance terms v_{-1}, $v_{-3/2}$ and v_{-2}.
- Although R_{jn} is simply the ratio of quadratic forms in e_n, finding its expectation seems difficult, in particular under nonnormality of e_n.
- Even for merely a second-order bias correction, it requires $\mathrm{E}(H_{1n}\tilde{\psi}_n)$, and hence $\mathrm{E}[(R_{1n})^3]$, involving up to 6th moment of $\epsilon_{n,i}$.
- In stark contrast, estimation of $\mathrm{E}[(R_{1n})^3]$ using the suggested bootstrap method (**BA-1**) is extremely simple.

First, simply replacing $\hat{\rho}_n$ by $\hat{\rho}_n^{\rm bc2}$ in (2.13) or (3.23), the 2nd-order bias-corrected $\hat{\rho}$, gives a potentially improved statistic for β_0:

$$\mathrm{t}_{\rm CE}^{\rm bc}(\beta_0) = \frac{c_0'\hat{\beta}_n^{\rm bc} - c_0'\beta_0}{\sqrt{\hat{\sigma}_n^{2,\rm bc} c_0'(X_n'\hat{B}_n^{\rm bc2\prime}\hat{B}_n^{\rm bc2}X_n)^{-1}c_0}}, \tag{4.35}$$

where $\hat{\beta}_n^{\rm bc} = \tilde{\beta}_n(\hat{\rho}_n^{\rm bc2})$, $\hat{\sigma}_n^{2,\rm bc} = \tilde{\sigma}_n^2(\hat{\rho}_n^{\rm bc2})$, and $\hat{B}_n^{\rm bc2} = I_n - \hat{\rho}_n^{\rm bc2}W_n$.

- Obviously, this statistic is not fully second-order bias-corrected. However, Monte Carlo results presented in Liu and Yang (2015b) show that it offers a huge improvement over that in (2.13) or (3.23).
- This confirms the point made in Section 2.2.3, Lecture 2.
- However, results also show that when n is not so large, there is still room for further improvement on $\mathrm{t}_{\rm CE}^{\rm bc}(\beta_0)$, which can be obtained in a similar manner as that for SL model (see also Liu and Yang, 2015b).
- Finally, a t-statistic for ρ_0, $\mathrm{t}_{\rm SE}^{\rm bc2}(\rho_0)$, is given by following (4.20).

From the concentrated loglikelihood function $\ell_n^c(\delta)$ of $\delta = (\lambda, \rho)'$ given in Sec. 2.4.1, Lec. 2, the CEE of δ, $\tilde{\psi}_n(\delta) = \frac{1}{n}\frac{\partial}{\partial\delta}\ell_n^c(\delta)$, has the form:

$$\tilde{\psi}_n(\delta) = \begin{cases} -\frac{1}{n}\mathrm{tr}(F_n(\lambda)) + \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\bar{F}_n(\delta)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \\ -\frac{1}{n}\mathrm{tr}(G_n(\rho)) + \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)G_n(\rho)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\lambda)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \end{cases} \tag{4.36}$$

$F_n(\lambda) = W_{1n}A_n^{-1}(\lambda)$, $G_n(\rho) = W_{2n}B_n^{-1}(\rho)$, and $\bar{F}_n(\delta) = B_n(\rho)F_n(\lambda)B_n^{-1}(\rho)$ as defined in Sec. 2.4.1, along with $\mathbb{Y}_n(\delta)$ and $\mathbb{M}_n(\rho)$.

- This is the key expression for deriving the score-based tests for the spatial effects, and also for performing bias-correction.
- Bias correction can be carried out as an application of the general methods, *Stochastic Expansion and Bootstrap*, laid out in Sec. 4.2 (or Yang 2015a) for a vector of nonlinear estimators.

To derive the higher-order partial derivatives of $\tilde{\psi}_n(\delta)$, $H_{rn}(\delta) = \nabla^r \tilde{\psi}_n(\delta)$, $r = 1, 2, 3$, define,

$$T_{rn}(\lambda) = \text{tr}(F_n^r(\lambda)) \text{ and } K_{rn}(\rho) = \text{tr}(G_n^r(\rho)), r = 0, 1, 2, 3.$$

Also define the following quantities,

$$R_{1n}(\delta) = \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\bar{F}_n(\delta)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)},$$

$$R_{2n}(\delta) = \frac{\mathbb{Y}_n'(\delta)\bar{F}_n'(\delta)\mathbb{M}_n(\rho)\bar{F}_n(\delta)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)},$$

$$S_{rn}(\delta) = \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)D_{rn}(\rho)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \; r = 1, 2, 3, 4,$$

$$Q_{rn}^{\dagger}(\delta) = \frac{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)D_{rn}(\rho)\mathbb{M}_n(\rho)\bar{F}_n(\delta)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \; r = 1, 2, 3,$$

$$Q_{rn}^{\ddagger}(\delta) = \frac{\mathbb{Y}_n'(\delta)\bar{F}_n'(\delta)\mathbb{M}_n(\rho)D_{rn}(\rho)\mathbb{M}_n(\rho)\bar{F}_n(\delta)\mathbb{Y}_n(\delta)}{\mathbb{Y}_n'(\delta)\mathbb{M}_n(\rho)\mathbb{Y}_n(\delta)}, \; r = 1, 2.$$

With these quantities, we have

$$\tilde{\psi}_n(\delta) = (\tilde{\psi}_{1n}(\delta) = -T_{0n}(\lambda) + R_{1n}(\delta),\ \ \tilde{\psi}_{2n}(\delta) = -K_{0n}(\rho) + S_{1n}(\delta))'.$$

$$H_{1n}(\delta) = \begin{pmatrix} -T_{1n}(\lambda) - R_{2n}(\delta) + 2R_{1n}^2(\delta), & -2Q_{1n}^{\dagger}(\delta) + 2R_{1n}(\delta)S_{1n}(\delta) \\ -2Q_{1n}^{\dagger}(\delta) + 2R_{1n}(\delta)S_{1n}(\delta), & -K_{1n}(\rho) + S_{2n}(\delta) + 2S_{1n}^2(\delta) \end{pmatrix}.$$

$$H_{2n}(\delta) = \begin{pmatrix} \tilde{\psi}_{1n}^{\lambda\lambda}(\delta), & \tilde{\psi}_{1n}^{\lambda\rho}(\delta), & \tilde{\psi}_{1n}^{\rho\lambda}(\delta), & \tilde{\psi}_{1n}^{\rho\rho}(\delta) \\ \tilde{\psi}_{2n}^{\lambda\lambda}(\delta), & \tilde{\psi}_{2n}^{\lambda\rho}(\delta), & \tilde{\psi}_{2n}^{\rho\lambda}(\delta), & \tilde{\psi}_{2n}^{\rho\rho}(\delta) \end{pmatrix}, \quad \text{where}$$

- $\tilde{\psi}_{1n}^{\lambda\lambda}(\delta) = -2T_{2n}(\lambda) - 6R_{1n}(\delta)R_{2n}(\delta) + 8R_{1n}^3(\delta)$,
- $\tilde{\psi}_{1n}^{\lambda\rho}(\delta) = 2Q_{1n}^{\ddagger}(\delta) - 8R_{1n}(\delta)Q_{1n}^{\dagger}(\delta) - 2R_{2n}(\delta)S_{1n}(\delta) + 8R_{1n}^2(\delta)S_{1n}(\delta)$,
- $\tilde{\psi}_{1n}^{\rho\rho}(\delta) = -2Q_{2n}^{\dagger}(\delta) - 8S_{1n}(\delta)Q_{1n}^{\dagger}(\delta) + 2R_{1n}(\delta)S_{2n}(\delta) + 8R_{1n}(\delta)S_{1n}^2(\delta)$,
- $\tilde{\psi}_{2n}^{\rho\rho}(\delta) = -2K_{2n}(\rho) + S_{3n}(\delta) + 6S_{1n}(\delta)S_{2n}(\delta) + 8S_{1n}^3(\delta)$
- $\tilde{\psi}_{1n}^{\lambda\rho}(\delta) = \tilde{\psi}_{1n}^{\rho\lambda}(\delta) = \tilde{\psi}_{2n}^{\lambda\lambda}(\delta)$ and $\tilde{\psi}_{1n}^{\rho\rho}(\delta) = \tilde{\psi}_{2n}^{\lambda\rho}(\delta) = \tilde{\psi}_{2n}^{\rho\lambda}(\delta)$.

$H_{3n}(\delta)$ is obtained by taking partial derivatives w.r.t. δ' for every element of $H_{2n}(\delta)$. Its full expression can be found in Liu and Yang (2015a).

The R-, S-, and Q-ratios at δ_0 can all be written as functions of θ_0 and $e_n = \sigma_0^{-1}\epsilon_n$, based on $\mathbb{M}_n B_n X_n = 0$ and $W_{1n} Y_n = F_n(X_n\beta_0 + B_n^{-1}\epsilon_n)$:

$$R_{1n}(e_n, \theta_0) = \frac{e_n'\mathbb{M}_n(\mu_n + \bar{F}_n e_n)}{e_n'\mathbb{M}_n e_n},$$

$$R_{2n}(e_n, \theta_0) = \frac{(\mu_n + \bar{F}_n e_n)'\mathbb{M}_n(\mu_n + \bar{F}_n e_n)}{e_n'\mathbb{M}_n e_n},$$

$$S_{rn}(e_n, \theta_0) = \frac{e_n'\mathbb{M}_n D_{rn}\mathbb{M}_n e_n}{e_n'\mathbb{M}_n e_n}, \; r = 1, 2, 3, 4,$$

$$Q_{rn}^{\dagger}(e_n, \theta_0) = \frac{e_n'\mathbb{M}_n D_{rn}\mathbb{M}_n(\mu_n + \bar{F}_n e_n)}{e_n'\mathbb{M}_n e_n}, \; r = 1, 2, 3,$$

$$Q_{rn}^{\ddagger}(e_n, \theta_0) = \frac{(\mu_n + \bar{F}_n e_n)'\mathbb{M}_n D_{rn}\mathbb{M}_n(\mu_n + \bar{F}_n e_n)}{e_n'\mathbb{M}_n e_n}, \; r = 1, 2,$$

where $\bar{F}_n = \bar{F}_n(\delta_0)$. Thus, $\tilde{\psi}_n = \tilde{\psi}_n(e_n, \theta_0)$ and $H_{rn} = H_{rn}(e_n, \theta_0)$, $r = 1, 2, 3$. The bias terms, b_{-1} and $b_{-3/2}$, can be easily estimated using the general bootstrap procedure (BA-2) for a vector nonlinear parameters.

Replacing $\hat{\delta}_n$ by $\hat{\delta}_n^{\text{bc2}}$ in (3.58), we obtain a statistic which is expected to have a better finite sample performance:

$$\mathrm{t}_{\mathrm{CE}}^{\mathrm{bc}}(\beta_0) = \frac{c_0' \hat{\beta}_n^{\mathrm{bc}} - c_0' \beta_0}{\sqrt{c_0' \hat{V}_{n,\beta\beta}^{\mathrm{bc}} c_0}}, \tag{4.37}$$

- $\hat{V}_{n,\beta\beta}^{\mathrm{bc}}$ is $V_{n,\beta\beta}$ but evaluated at $\hat{\delta}_n^{\mathrm{bc2}}$, $\tilde{\beta}_n^{\mathrm{bc}}$, $\hat{\sigma}_n^{2,\mathrm{bc}}$, $\hat{\gamma}_n^{\mathrm{bc}}$, and $\hat{\kappa}_n^{\mathrm{bc}}$.
- $V_{n,\beta\beta}$ is the asymptotic VC matrix of $\hat{\beta}_n$ given in Lecture 2.
- $\hat{\gamma}_n^{\mathrm{bc}}$, and $\hat{\kappa}_n^{\mathrm{bc}}$ are the estimates of γ and κ using $\hat{\delta}_n^{\mathrm{bc2}}$.

To further improve $t_{\mathrm{CE}}^{\mathrm{bc}}(\beta_0)$ to give a fully 2nd-order corrected *t*-statistics, $t_{\mathrm{CE}}^{\mathrm{bc2}}(\beta_0)$, one follows the procedures leading to the 2nd-order corrected *t*-ratios for the SL and SE models. See Liu and Yang (2015b).

The 2nd-order corrected Wald statistics for δ, λ or ρ can be obtained by working with (3.59), following the ideas laid out around (4.20).

Table 4.4. Size of Test H_0: $\beta_1 = \beta_2$ for SLE Model, $n = 50$, $\lambda = 0.5$, $\sigma = 1$

Group Interact, REG2; Test: $1 = \mathrm{t}_{\mathtt{CE}}$, $2 = \mathrm{t}_{\mathtt{CE}}^{\mathrm{bc}}$, $3 = \mathrm{t}_{\mathtt{CE}}^{\mathrm{bc2}}$

ρ	Test	Normal Error			Normal Mixture			Lognormal		
.50	1	.197	.115	.040	.201	.122	.044	.197	.122	.040
	2	.120	.068	.020	.123	.073	.023	.146	.084	.028
	3	.115	.062	.017	.119	.068	.023	.128	.074	.024
.25	1	.191	.109	.031	.180	.110	.031	.183	.109	.035
	2	.118	.067	.020	.116	.069	.022	.120	.066	.021
	3	.109	.061	.016	.103	.058	.019	.103	.055	.017
.00	1	.191	.110	.031	.177	.099	.028	.191	.114	.037
	2	.111	.054	.015	.100	.054	.016	.117	.065	.021
	3	.098	.047	.012	.095	.046	.013	.100	.055	.018
-.25	1	.173	.100	.025	.170	.096	.027	.184	.108	.033
	2	.094	.048	.011	.098	.049	.016	.111	.059	.020
	3	.108	.048	.009	.108	.051	.013	.090	.047	.016
-.50	1	.182	.104	.030	.162	.085	.023	.177	.100	.034
	2	.097	.049	.013	.085	.043	.010	.102	.059	.019
	3	.100	.048	.011	.091	.052	.009	.092	.046	.014

Table 4.4, Cont'd. $n = 200$

ρ	Test	Normal Error			Normal Mixture			Lognormal		
.50	1	.141	.078	.022	.140	.082	.028	.131	.078	.021
	2	.113	.056	.014	.117	.061	.017	.116	.061	.015
	3	.105	.050	.011	.107	.056	.016	.106	.052	.013
.25	1	.147	.085	.025	.152	.089	.028	.150	.083	.025
	2	.108	.056	.012	.112	.061	.012	.111	.058	.012
	3	.100	.050	.009	.102	.054	.011	.101	.051	.010
.00	1	.150	.089	.026	.137	.075	.016	.138	.084	.020
	2	.104	.055	.012	.116	.061	.014	.124	.066	.017
	3	.097	.050	.010	.102	.051	.010	.105	.052	.012
-.25	1	.158	.093	.030	.131	.074	.018	.120	.062	.014
	2	.108	.054	.013	.123	.068	.019	.118	.066	.017
	3	.099	.049	.012	.102	.055	.010	.095	.054	.010
-.50	1	.127	.072	.020	.120	.061	.013	.119	.063	.013
	2	.115	.066	.017	.122	.063	.015	.135	.074	.019
	3	.105	.060	.013	.095	.046	.009	.091	.052	.009

Table 4.4., Cont'd. $n = 500$

ρ	Test	Normal Error			Normal Mixture			Lognormal		
.50	1	.124	.068	.018	.126	.070	.017	.124	.073	.018
	2	.102	.053	.013	.107	.053	.011	.106	.056	.011
	3	.098	.049	.012	.100	.049	.010	.100	.050	.010
.25	1	.130	.073	.023	.134	.073	.020	.130	.074	.018
	2	.105	.056	.015	.106	.057	.014	.101	.053	.012
	3	.099	.052	.014	.100	.053	.013	.099	.049	.010
.00	1	.138	.075	.021	.135	.072	.019	.133	.076	.020
	2	.106	.055	.013	.099	.052	.009	.106	.054	.012
	3	.103	.053	.011	.099	.049	.009	.101	.053	.010
-.25	1	.131	.074	.020	.135	.077	.022	.132	.075	.022
	2	.101	.055	.013	.102	.053	.012	.098	.053	.011
	3	.096	.051	.012	.099	.051	.011	.100	.051	.010
-.50	1	.128	.071	.018	.144	.076	.022	.129	.072	.019
	2	.094	.046	.011	.107	.054	.014	.093	.050	.011
	3	.092	.045	.011	.103	.051	.013	.099	.050	.010

Table 4.5. Size of Test H_0: $\beta_1 = \beta_2$ in SLE Model, $n = 50$, $\lambda = -.25$, $\sigma = 1$

Group Interact, REG2; Test: $1 = t_{\text{CE}}$, $2 = t_{\text{CE}}^{\text{bc}}$, $3 = t_{\text{CE}}^{\text{bc2}}$

ρ	Test	Normal Error			Normal Mixture			Lognormal		
.50	1	.196	.119	.045	.203	.126	.047	.188	.115	.045
	2	.121	.070	.020	.122	.076	.022	.138	.085	.030
	3	.114	.066	.017	.117	.072	.020	.122	.074	.022
.25	1	.198	.123	.042	.205	.128	.043	.205	.130	.054
	2	.108	.059	.018	.109	.057	.020	.112	.066	.025
	3	.103	.056	.015	.104	.055	.017	.109	.064	.022
.00	1	.192	.115	.037	.180	.109	.038	.199	.127	.051
	2	.115	.065	.017	.118	.065	.017	.106	.062	.020
	3	.103	.058	.014	.101	.056	.015	.104	.059	.020
-.25	1	.196	.114	.032	.186	.108	.038	.194	.115	.042
	2	.107	.052	.016	.109	.060	.019	.114	.069	.022
	3	.099	.050	.013	.098	.051	.015	.098	.057	.017
-.50	1	.188	.113	.040	.188	.111	.037	.186	.116	.040
	2	.111	.061	.018	.089	.049	.014	.098	.055	.015
	3	.095	.051	.015	.093	.055	.013	.099	.051	.015

Table 4.5., Cont'd. $n = 200$

ρ	Test	Normal Error			Normal Mixture			Lognormal		
.50	1	.129	.073	.020	.144	.083	.022	.133	.072	.016
	2	.112	.063	.014	.116	.060	.012	.111	.059	.013
	3	.104	.055	.013	.106	.054	.012	.093	.046	.009
.25	1	.143	.081	.025	.150	.085	.022	.151	.084	.025
	2	.122	.062	.018	.122	.065	.015	.112	.055	.014
	3	.110	.056	.016	.106	.057	.011	.100	.051	.013
.00	1	.144	.086	.023	.129	.075	.016	.156	.091	.028
	2	.123	.066	.017	.114	.065	.015	.113	.056	.013
	3	.110	.059	.014	.100	.053	.011	.103	.050	.012
-.25	1	.136	.075	.023	.125	.065	.018	.153	.090	.026
	2	.123	.068	.018	.120	.062	.017	.106	.056	.011
	3	.112	.060	.015	.101	.048	.012	.097	.052	.011
-.50	1	.120	.064	.016	.114	.055	.011	.150	.091	.025
	2	.117	.063	.015	.126	.063	.015	.105	.051	.012
	3	.106	.055	.012	.099	.045	.009	.097	.049	.011

Table 4.5., Cont'd. $n = 500$

ρ	Test	Normal Error			Normal Mixture			Lognormal		
.50	1	.132	.069	.016	.131	.070	.017	.133	.072	.021
	2	.110	.058	.012	.108	.055	.010	.109	.056	.012
	3	.104	.053	.011	.100	.048	.010	.102	.052	.011
.25	1	.132	.079	.023	.125	.074	.019	.138	.076	.020
	2	.109	.060	.014	.106	.055	.012	.107	.052	.012
	3	.104	.056	.013	.100	.051	.009	.103	.051	.011
.00	1	.135	.077	.025	.129	.077	.020	.128	.071	.019
	2	.105	.056	.015	.099	.049	.012	.100	.050	.011
	3	.101	.053	.013	.100	.049	.011	.099	.050	.011
-.25	1	.139	.082	.026	.139	.079	.022	.130	.077	.020
	2	.106	.059	.014	.104	.053	.011	.099	.050	.012
	3	.101	.056	.014	.099	.050	.010	.100	.050	.012
-.50	1	.143	.085	.023	.140	.084	.024	.126	.074	.023
	2	.107	.054	.012	.111	.059	.014	.098	.053	.013
	3	.105	.053	.011	.108	.055	.012	.099	.052	.012

Refer to Section 2.2.4 for the QML estimation of an SE model fitted to the **Neighborhood Crime** data. We now fit an SL model to the same set of data, based on QML and 2nd-ord bias corrected QML. The results are summarized in Table 4.6, along with OLS estimates.

Table 4.6. Bias-Corrected Estimation of SL Model: Neighborhood Crime

	OLS	QMLE	se	t-Ratio	QMLE-bc2	se-bc2	t-Ratio
constant	38.181	45.078	7.163	6.293	42.316	6.632	6.381
income	-0.866	-1.032	0.304	-3.391	-0.965	0.299	-3.225
hvalue	-0.264	-0.266	0.089	-3.005	-0.265	0.088	-3.011
λ	0.557	0.431	0.118	3.645	0.482	0.105	4.569
σ^2	102.368	95.488	30.571	3.123	94.542	30.982	3.052

- OLS estimation is invalid.
- QMLE of λ is quite biased, and use of 2nd-order bias-corrected QMLE is recommended.
- See Section 2.3.4 for the bias-corrected estimation for Boston House Price data.

References

- Bao, Y., 2013. Finite sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* 29, 68-88.
- Bao, Y., Ullah, A., 2007a. Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* 137, 396-413.
- Bao, Y., Ullah, A., 2007b. The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics* 140, 650-669.
- Bun, M. J. G., Carree, M. A. 2005. Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics* 23, 200-210.
- Hahn J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639-1657.
- Hahn J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295-1319.
- Kiviet, J. F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 53-78.

- 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.
- Rilstone, P., Srivastava, V. K., Ullah, A., 1996. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* 75, 369-395.
- Yang, Z. (2015a). A general method for third-order bias and variance correction on a nonlinear estimator. *Journal of Econometrics* 186, 178-200.
- Yang, Z. (2015b). LM tests of spatial dependence based on bootstrap critical values. *Journal of Econometrics* 185, 33-59.
- Yang, Z. L., Yu, J. H, Liu, S. F. (2016). Bias correction and refined inferences for fixed effects spatial panel data models. *Regional Science and Urban Economics* 61, 52-72.