ANNALS OF ECONOMICS AND FINANCE **11-1**, 155–184 (2010)

A More Efficient Best Spatial Three-stage Least Squares Estimator for Spatial Autoregressive Models *

Zhengyu Zhang

Research Center of Econometrics, Shanghai Academy of Social Sciences
Shanghai, 200020, China
E-mail: zyzhang@sass.org.cn

and

Pingfang Zhu

Research Center of Econometrics, Shanghai Academy of Social Sciences
Shanghai, 200020, China

Lee (2003) proposed for spatial autoregressive (SAR) model the best spatial two-stage least squares estimator (BS2SLSE) as an improvement on Kelejian and Prucha (1998)'s S2SLSE. In this paper, we show that one more step iteration based on BS2SLSE gives a spatial counterpart of the three-stage least squares estimator for a system of J equations. This estimator, named BS3SLSE, is shown to be equivalent to BS2SLSE under normality and more efficient under non-normality. The proposed BS3SLSE can be interpreted as a GMM estimator where the number of moments increases with the sample size n at some slow rate. The asymptotic efficiency of BS3SLSE relative to other previously proposed estimators such as MLE (Lee, 2004) and GMME (Lee, 2007a) is also discussed. As an empirical illustration, we apply these estimation procedures to re-examining the presence of environmental "race-to-the-bottom" effect in competition for FDI across China municipal governments.

Key Words: Spatial autoregressive model; 3SLSE; GMME; Relative asymptotic efficiency.

JEL Classification Numbers: C13; C14; C21.

* The research is supported by NSFC grant "Spatial econometrics: theory and application to the spatial deployment efficiency of modern urban service industry" (Grant No. 70871083). The authors appreciate Professor Shuming Bao from China Data Center, University of Michigan through the financial support of NSFC grant "A study of the applications of remote sensing, GIS and spatial-temporal modeling methods in the analysis of recurrent diseases" (Grant No: 40671146). We also thank Professor Guolin Jiang, Qing Han at Research Center of Econometrics, SASS and an anonymous referee for their valuable comments and suggestions.

1529-7373/2010

1. INTRODUCTION

Recently, topics concerning spatial dependence have received increasing attention in analyzing economic problems using cross sectional data or panel data. Economic models underpinning empirical work in urban, environmental, development, industrial organization and growth convergence often suggest that outcome variables of subjects under investigation are not independent of each other. One form of such dependence arises when the value of the dependent variable corresponding to each cross-sectional unit is assumed, in part, to depend on a weighted average of that dependent variable corresponding to neighboring cross-sectional units. This weighted average is often described in the literature as a spatial lag of the dependent variable, and the resulting model,

$$y_n = \lambda_0 W_n y_n + X_n \beta_0 + u_n, \qquad u_n = \alpha_0 l_n + \sigma_0 \epsilon_n \tag{1}$$

is then referred to as the spatial autoregressive (SAR) model (Anselin 1988). In model (1), $y_n = (y_{n1}, \cdots, y_{nn})'$ is an n-dimensional vector of dependent variables, W_n, a spatial weights matrix of known constants, captures the structure of spatial correlations between cross sectional units scaled by a single parameter λ_0,[1] X_n is an $n \times k$ matrix composed of k columns of spatial-varying regressors, β_0 is a $k \times 1$ regression slopes, $\epsilon_n = (\epsilon_{n1}, \cdots, \epsilon_{nn})'$ are i.i.d. random variables independent of X_n with zero mean and unit variance, l_n is an n-dimensional vector of ones. α_0 and σ_0 are location and scale measure of u_n, respectively.

A number of estimation procedures have been developed for model (1). Among them, Kelejian and Prucha (1998) proposed a spatial two-stage least squares estimator (S2SLSE) and their estimation procedure is then improved by Lee (2003)'s best S2SLSE (BS2SLSE) via choosing asymptotically optimal instruments. The asymptotic properties of QMLE for SAR models were investigated by Lee (2004). Lee (2007a) and Liu et al. (2006) developed for SAR models a computationally simple GMM estimator and show that some distribution-free best GMME using a set of carefully chosen moment conditions can be as efficient as QMLE under normality and will be more efficient relative to QMLE otherwise.

In this paper, an alternative computationally feasible procedure is taken towards efficient estimation of SAR models. The main work of this paper includes that, (a) we show that the asymptotic efficiency of Lee (2003)'s BS2SLSE can be further improved by one more step iteration. The suggested one step iterative estimator, named S3SLSE, is essentially the spatial counterpart of the three-stage least squares estimation procedure of a system of J equations. S3SLSE can also be interpreted in a GMM framework

[1] Throughout this paper, any parameter with subscript zero represents the true one that generates the data.

where the number of moments J is equal to the number of equations in the system and J is assumed to increase with the sample size n very slowly. (b) We give the general feature the spatial weights matrix should have such that the suggested best S3SLSE (BS3SLSE) will be asymptotically equivalent to MLE (Lee 2004) and GMME (Lee 2007a, Liu et al. 2006) and more efficient otherwise.

The paper is organized as follows: S3SLSE is motivated and described in Section 2. Section 3 presents the large sample properties of the suggested estimators and discusses its asymptotic efficiency relative to other estimates such as MLE and GMME. Section 4 reports some Monte Carlo results and Section 5 provides an example where the environmental "race-to-the-bottom" hypothesis in competition for FDI among China municipal governments is re-examined in a newly formulated empirical setting that accounts for both unobservability in the environmental stringency and possible strategic interactions among local policymaking. In contrast to the previous studies, this new framework proposes to test the hypothesis by simultaneously examining whether local government responds actively to the environmental policymaking of its neighboring municipalities and whether lenient environmental regulation is attractive to foreign capitals indeed. Section 6 concludes and all the technical proofs are relegated to the Appendix.

2. A SPATIAL THREE STAGE LEAST SQUARES ESTIMATOR

Lee (2007a) revealed that the S2SLSE, proposed by Kelejian and Prucha (1998), can be interpreted in the GMM framework using the linear moments in the generic form as

$$g_n(\theta, \alpha) = n^{-1} Q_n' \left(u_n(\theta) - \alpha l_n \right) \tag{2}$$

where $\theta = (\lambda, \beta')'$, $u_n(\theta) = y_n - Z_n\theta$, $Z_n = (W_n y_n, X_n)$, Q_n is an $n \times k_x$ IV matrix with a column of ones and is constructed from W_n and X_n. Noting that at the true parameters θ_0 and α_0, $\mathbb{E}\left(g_n(\theta_0, \alpha_0)\right) = 0$, then the resulting S2SLSE is the solution to the following minimization problem,

$$\min_{(\theta,\alpha)\in\Theta\times\mathcal{A}} g_n'(\theta, \alpha) a_n' a_n g_n(\theta, \alpha) \tag{3}$$

where Θ and $\mathcal{A}$ are some compact subsets of $\mathbb{R}^{k+1}$ and $\mathbb{R}$ containing θ_0 and α_0 as the interior respectively. a_n is a matrix with the full row rank greater than or equal to $k+2$. The a_n is assumed to converge to a constant full rank matrix a_0. This corresponds to the Hansen (1982)'s GMM setting, which illustrates the optimal weighting issue. For a given instrumental matrix Q_n, S2SLSE has a closed form as, by setting the optimal weights $a_n' a_n = (Q_n' Q_n)^{-1}$,

$$\begin{aligned}\widehat{\theta}_{s2sls,n} &= \big(Z_n' Q_n (Q_n' Q_n)^{-1} Q_n' \Gamma_n Q_n (Q_n' Q_n)^{-1} Q_n' Z_n\big)^{-1} \\ &\times \; Z_n' Q_n (Q_n' Q_n)^{-1} Q_n' \Gamma_n Q_n (Q_n' Q_n)^{-1} Q_n' y_n \end{aligned} \tag{4}$$

where $\Gamma_n = I_n - l_n l_n'/n$.

Now observe that when the error terms and the regressors are independent, any function of the error term will be uncorrelated with the instruments constructed from any combination of the regressors and the spatial weights. Let $m(\cdot)$ be an arbitrary differentiable function and $\mu = \mathbb{E}(m(u_{n1}))$. Given the Q_n constructed from X_n and W_n, there holds the moment condition,

$$\frac{1}{n}\mathbb{E}\left[Q_n'(m(u_n) - \mu l_n)\right] = 0, \tag{5}$$

where $m(u_n) = (m(u_{n1}), \cdots, m(u_{nn}))'$, representing one of the infinite number of moment restrictions that can be used for estimation. One undesirable consequence on GMM estimates using the moments such as (5) is that the resulting estimators will not be location and scale equivalent. Newey (1988) suggested that this unfortunate property could be fixed by replacing $m(\cdot)$ with the functions that had been location and scale adjusted with preliminary estimates of location and scale parameters, namely, $\widetilde{\overline{m}}(u) = m\big((u - \widetilde{\alpha})/\widetilde{\sigma}\big)$, where $\widetilde{\alpha}$ and $\widetilde{\sigma}$ are estimated location and scale measure, respectively. Let $\widetilde{\theta}$ be any initial estimator of θ_0 such as S2SLSE.[2] Then $\widetilde{\alpha}$ and $\widetilde{\sigma}$ may be the sample mean and sample standard deviation of $u_{ni}(\widetilde{\theta})$'s, $i = 1, \cdots, n$, respectively. It can be shown in the next section that the asymptotic properties of the resulting estimators using $\widetilde{\overline{m}}(\cdot)$ will be identical to that obtained using $\overline{m}(\cdot) = m\big((\cdot - \alpha_0)/\sigma_0\big)$.

Now carefully choose a sequence of J differentiable functions $m_{J1}(\cdot), \cdots, m_{JJ}(\cdot)$. Then θ can be estimated using the following J moments,

$$\widetilde{g}_{J,n}(\delta_J) = \frac{1}{n}\begin{bmatrix} Q_n'\left(\widetilde{\overline{m}}_{J1}(u_n(\theta)) - \mu_{J1} l_n\right) \\ \cdots \\ Q_n'\left(\widetilde{\overline{m}}_{JJ}(u_n(\theta)) - \mu_{JJ} l_n\right)\end{bmatrix} \tag{6}$$

where $\delta_J = (\theta', \mu_J')'$, $\mu_J = (\mu_{J1}, \cdots, \mu_{JJ})'$, $\mu_{Jj} = \mathbb{E}\big(\overline{m}_{Jj}(u_{n1})\big)$. $Q_n = (l_n, q_n)$ is an IV matrix containing a column of ones. The function series $m_{Jj}(\cdot)$'s, $j = 1, 2, \cdots, J$ are composed of differentiable functions that

[2]Throughout this paper, all parameters with symbol $\sim$ indicate some consistent initial estimates.

should meet some regular assumptions specified in the next section. The number of the functions in the series, $J = J(n)$ is some positive integer slowly increasing with sample size n. Further, the GMM setting described above indicates that the suggested estimator can be understood as a spatial generalization of the adaptive GMM estimator developed in Newey (1988).

Likewise, a GMM estimator of δ_J using the moments (6) can be computed by solving the following minimization problem,

$$\min_{\delta_J} \widetilde{g}'_{J,n}(\delta_J)a'_n a_n \widetilde{g}_{J,n}(\delta_J). \tag{7}$$

As $\widetilde{\overline{m}}_{Jj}(u_n(\theta))$ may be highly nonlinear in θ, a first order Taylor expansion around some initial estimator $\widetilde{\theta}$ is useful, which gives

$$\widetilde{\overline{m}}_{Jj}(u_n(\theta)) = \widetilde{\overline{m}}_{Jj}(u_n(\widetilde{\theta})) - \widetilde{\mathcal{M}}_{Jj}(u_n(\bar{\theta}))Z_n(\theta - \widetilde{\theta}), j = 1, \cdots, J, \tag{8}$$

where $\bar{\theta}$ lies between θ_0 and $\widetilde{\theta}$,

$$\widetilde{\mathcal{M}}_{Jj}(u_n(\theta)) = diag\Big\{\dot{\widetilde{\overline{m}}}_{Jj}(u_{n1}(\theta)), \cdots, \dot{\widetilde{\overline{m}}}_{Jj}(u_{nn}(\theta))\Big\},$$

$\dot{\widetilde{\overline{m}}}_{Jj}(u) = \widetilde{\sigma}^{-1}\dot{m}_{Jj}((u - \widetilde{\alpha})/\widetilde{\sigma})$ for $j = 1, \cdots, J$. Further approximate $\widetilde{\mathcal{M}}_{Jj}(u_n(\widetilde{\theta}))$ by the scalar matrix $\widetilde{\mathbb{M}}_{Jj}I_n$, where $\widetilde{\mathbb{M}}_{Jj} = n^{-1}\sum_{i=1}^n \dot{\widetilde{\overline{m}}}_{Jj}(u_{ni}(\widetilde{\theta}))$. Let $\widetilde{\mathbb{M}}_J = (\widetilde{\mathbb{M}}_{J1}, \cdots, \widetilde{\mathbb{M}}_{JJ})'$,

$$\widetilde{\mathcal{Y}}_{J,n} = \begin{bmatrix} \widetilde{\overline{m}}_{J1}(u_n(\widetilde{\theta})) + \widetilde{\mathbb{M}}_{J1}Z_n\widetilde{\theta} \\ \vdots \\ \widetilde{\overline{m}}_{JJ}(u_n(\widetilde{\theta})) + \widetilde{\mathbb{M}}_{JJ}Z_n\widetilde{\theta} \end{bmatrix}, \tag{9}$$

and

$$\widetilde{\mathcal{Z}}_{J,n} = \Big[\widetilde{\mathbb{M}}_J \otimes Z_n, I_J \otimes l_n\Big]. \tag{10}$$

in which $\otimes$ is the Kronecker product. Since now the righthand side of (8) is linear in θ, substituting (8) into (7) and solving δ_J from (7) just like solving θ, α from (3) gives the resulting three stage least squares estimator,

$$\begin{aligned} \widehat{\delta}_{s3sls,J,n} &= \Big[\widetilde{\mathcal{Z}}'_{J,n}\Big(\widetilde{\Sigma}_J^{-1} \otimes Q_n(Q'_nQ_n)^{-1}Q'_n\Big)\widetilde{\mathcal{Z}}_{J,n}\Big]^{-1} \\ &\times\ \widetilde{\mathcal{Z}}'_{J,n}\Big(\widetilde{\Sigma}_J^{-1} \otimes Q_n(Q'_nQ_n)^{-1}Q'_n\Big)\widetilde{\mathcal{Y}}_{J,n} \end{aligned} \tag{11}$$

where $\widetilde{\Sigma}_J$ is the estimated covariance matrix of $(\overline{m}_{J1}(u) - \mu_{J1}l_n, \cdots, \overline{m}_{JJ}(u) - \mu_{JJ}l_n)'$,

$\widetilde{\Sigma}_{J,ij} = n^{-1}\sum_{k=1}^{n}\left(\widetilde{\widetilde{m}}_{Ji}(u_{nk}(\widetilde{\theta})) - \widetilde{\mu}_{Ji}\right)\left(\widetilde{\widetilde{m}}_{Jj}(u_{nk}(\widetilde{\theta})) - \widetilde{\mu}_{Jj}\right)$. Finally, we see from (11) that the representaion of S3SLSE resembles a spatial counterpart of 3SLSE for a system of J equations.

3. ASYMPTOTIC NORMALITY AND BS3SLSE

Let $S_n(\lambda) = I_n - \lambda W_n$, $S_n = S_n(\lambda_0)$, $G_n = W_n S_n^{-1}$. For an $n \times n$ matrix $A_n = [a_{n,ij}]_{i,j=1,\cdots,n}$, let $\|A_n\| = \max_{1\le i\le n, 1\le j\le n} |a_{n,ij}|$. Introduce the following regularity conditions,

Assumption 1. *$\epsilon_{n1}, \cdots, \epsilon_{nn}$ are i.i.d. random variables with density $f(\epsilon)$, zero mean and unit variance. These innovations also possess finite fourth moments. The density $f(\epsilon)$of ϵ is absolutely continuous and has Radon-Nikodym derivative $\dot{f}(\epsilon)$ such that $\int(\dot{f}^2/f)d\epsilon < \infty$.*

Assumption 1 says that the density of ϵ is regular in the sense of Hajek and Sidak (1967). Existence of moment up to the fourth order implies some central limit theorem of simple form can be applied.

Assumption 2. *X_n is some $n \times k$ non-stochastic regressor matrix with full column rank whose elements are uniformly bounded and the limit $\lim_{n\to\infty} \frac{1}{n}X_n'X_n$ exists and is nonsingular.*

Assumption 3. *The spatial weights matrix W_n has zero diagonals. The matrix $S_n = I_n - \lambda_0 W_n$ is nonsingular. The row and column sums of S_n^{-1} and W_n are uniformly bounded in absolute value.* [3]

The non-stochastic X_n and its uniform boundedness of elements are for analytical tractability. Assumption 3 limits the spatial dependence among cross sectional units to a permissable degree and is originated by Kelejian and Prucha (1999). Lee (2004) assumed that the elements of W_n had the uniform order $O(h_n^{-1})$, where $\{h_n\}$ could either be a bounded or a divergent sequence that depended on n. The relevance of Lee's assumption arises when two distinguished spatial scenarios are identified in empirical studies.

[3]The row and column sums of an $n \times n$ matrix P_n are said to be uniformly bounded if we have for all n, there exists a positive constant c independent of n with $\max_i \sum_{j=1}^{n} |P_{n,ij}| < c$ and $\max_j \sum_{i=1}^{n} |P_{n,ij}| < c$. This notion of uniform boundedness can be defined in terms of some matrix norms. The maximum column sum matrix norm $\|\cdot\|_1$ of an $n \times n$ matrix P_n is defined as $\|P_n\|_1 = \max_j \sum_i |P_{n,ij}|$, and the maximum row sum matrix norm $\|\cdot\|_\infty$ is defined as $\|P_n\|_\infty = \max_i \sum_j |P_{n,ij}|$. Thus the uniform boundedness of $\{P_n\}$ in column or row sums is equivalent to the sequence $\{\|P_n\|_1\}$ or $\{\|P_n\|_\infty\}$ being bounded.

When $\{h_n\}$ is a bounded sequence, it implies a cross sectional unit has only a small number of neighbors (Kelejian and Prucha 1999), where the spatial dependence is usually defined based on geographical implications. When $\{h_n\}$ is divergent, it corresponds to the scenario where each unit has a large number of neighbors that often emerges in empirical studies of social interactions and/or cluster sampling data (Case 1991, 1992; Lee 2007b). For example, in Case (1992)'s study on the spillover effect of new technology adoption for farmers, the notion "neighbors" refer to the farmers who live in the same district (in rural Java, Indonesia). This characterization of the neighbors makes the W_n matrix block diagonal. The only non-zero elements appear as a block of households in the same district. Suppose that there are R districts and in each district there are m_r, $r = 1, \cdots, R$ farmers. Case assumed that each neighbor of a member in a district was given equal weight, i.e., $W_r = (m_r - 1)^{-1}(l_{m_r} l'_{m_r} - I_{m_r})$, and $W_n = diag\{W_1, \cdots, W_R\}$. In Case's example, $h_n = O(\overline{m})$, where $\overline{m} = n/R$ is the mean group size, increasing with the sample size n.

Assumption 4. *Q_n is some $n \times k_x$ non-stochastic matrix constructed from X_n and W_n, with the first column being ones and of full column rank. The elements of Q_n are uniformly bounded and the limit $\lim_{n\to\infty} \frac{1}{n} Q'_n Q_n$ exists and is nonsingular. Additionally, the limit $\lim_{n\to\infty} \frac{1}{n} Q'_n Z_n$ exists.*

The regularity conditions satisfied by Q_n are stated in the most general manner to motivate a general GMM estimation framework. As will become evident later, some asymptotically optimally chosen Q_n will meet these regularity conditions automatically.

Assumption 5. *$\sqrt{n}\left(\widetilde{\theta} - \theta_0\right)$, $\sqrt{n}\left(\widetilde{\alpha} - \alpha_0\right)$, $\sqrt{n}\left(\widetilde{\sigma}^2 - \sigma_0^2\right)$ are bounded in probability.*

Assumption 6. *$J(n)$ is chosen such that $J(n) \to \infty$ and $J \cdot \ln J / \ln n \to 0$.*

The growth rate for the number of moment functions specified in Assumption 6 is quite slow, being slower than the natural log of the sample size.

Assumption 7. *$m_{Jj}(\epsilon) = m^j_{J1}(\epsilon)$, for $j = 1, \cdots, J$. The function $m_{J1}(\cdot)$ is continuously differentiable. Also, for any $\sigma_0 > 0$ and α_0 there exists a neighborhood of N of (α_0, σ_0), measurable functions $B_1(u)$ and $B_2(u)$, and $\tau > 0$, such that $\mathbb{E}[\exp(\tau B_1(u))]$ and $\mathbb{E}[B_2^4(u)(1 + u^4)]$ exists* *and for all* u *and* $(\alpha, \sigma) \in N$,

$$\sup_N |m_{J1}((u-\alpha)/\sigma)| \leq B_1(u), \quad \sup_N |\dot{m}_{J1}((u-\alpha)/\sigma)| \leq B_2(u)$$

and

$$|\dot{m}_{J1}((u-\alpha)/\sigma) - \dot{m}_{J1}((u-\alpha_0)/\sigma_0)| \leq B_2(u)\|(\alpha,\sigma)-(\alpha_0,\sigma_0)\| \quad (12)$$

The choice of moment functions and Assumption 7 can impose some restrictions on the distribution of ϵ. For example, if $m_{J1}(\epsilon) = \epsilon$, then the above assumption implies existence of the moment generating function of ϵ. Of course such a choice of moment function would not be appropriate in general. It is possible to choose $m_{J1}(\cdot)$ such that Assumption 7 is satisfied for all distributions, e.g., by choosing $m_{J1}(\cdot)$ to be a bounded function with sufficiently well-behaved first order derivative. For example, $m_{J1}(\epsilon) = \epsilon/(1+\epsilon^2)^{1/2}$ will satisfy Assumption 7 for any distribution of ϵ. The following result gives the limiting distribution of the S3SLSE.

Proposition 1. *Suppose that Assumptions 1-7 hold, then*

$$\sqrt{n}\left(\widehat{\theta}_{s3sls,J,n} - \theta_0\right) \to^d \mathcal{N}\left(0, \sigma_0^2(\mathcal{I}\Omega)^{-1}\right) \quad (13)$$

where $\mathcal{I} = \mathbb{E}(\phi^2(\epsilon))$, $\phi(\epsilon) = -\dot{f}(\epsilon)/f(\epsilon)$,

$$\begin{aligned}\Omega &= \lim_{n\to\infty} \frac{1}{n}(G_n(X_n\beta_0 + \alpha_0 l_n), X_n)' Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_n Q_n(Q_n'Q_n)^{-1}Q_n' \\ &\times (G_n(X_n\beta_0 + \alpha_0 l_n), X_n) \quad (14)\end{aligned}$$

where $G_n = W_n S_n^{-1}$. *Furthermore,* $\left\|\left(\widetilde{\mathcal{I}}\widetilde{\Omega}\right)^{-1} - (\mathcal{I}\Omega)^{-1}\right\| \to^p 0$, *where* $\widetilde{\mathcal{I}} = \widetilde{\sigma}^2 \widetilde{\mathbb{M}}_J' \widetilde{\Sigma}_J^{-1} \widetilde{\mathbb{M}}_J$, $\widetilde{\mathbb{M}}_J$ *and* $\widetilde{\Sigma}_J$ *being defined below (8) and (11) respectively, and*

$$\begin{aligned}\widetilde{\Omega} &= \frac{1}{n}(\widetilde{G}_n(X_n\widetilde{\beta} + \widetilde{\alpha} l_n), X_n)' Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_n Q_n(Q_n'Q_n)^{-1}Q_n' \\ &\times (\widetilde{G}_n(X_n\widetilde{\beta} + \widetilde{\alpha} l_n), X_n)\end{aligned}$$

where $\widetilde{G}_n = W_n(I_n - \widetilde{\lambda} W_n)^{-1}$.

As revealed by Lee (2007), under some regularity conditions, S2SLSE (4) has the limiting distribution

$$\sqrt{n}\left(\widehat{\theta}_{s2sls,n} - \theta_0\right) \to^d \mathcal{N}\left(0, \sigma_0^2 \Omega^{-1}\right) \quad (15)$$

The efficiency gain can be seen via Cauchy-Schwartz inequality, $\mathbb{E}\left(\phi^2\right) \geq \mathbb{E}^2\left(\phi\epsilon\right)/\mathbb{E}\left(\epsilon^2\right) = 1$, and the equality holds if and only if $\phi = c\epsilon$, or simply ϵ is normally distributed.

Lee (2003) pointed out that the generalized two-stage least squares procedure described in Kelejian and Prucha (1998) for model (1) was, however, not asymptotically optimal. Hence Lee proposed a best S2SLSE (BS2SLSE) using the asymptotically optimal IV matrix. For (14), it is straightforward to extend the asymptotically optimal instruments chosen by Lee (2003) to this S3SLSE case by choosing $Q_{o,n} = (l_n, G_n(X_n\beta_0 + \alpha_0 l_n), X_n)$. In practice, with initial consistent estimates $\widetilde{\beta}$, $\widetilde{\lambda}$ and $\widetilde{\alpha}$ (e.g., their BS2SLSE), the best instruments $Q_{o,n}$ can be replaced by their empirical counterparts, $\widetilde{Q}_{o,n} = (l_n, \widetilde{G}_n(X_n\widetilde{\beta}_0 + \widetilde{\alpha}_0 l_n), X_n)$. Finally, the feasible BS3SLSE is defined as

$$\begin{aligned}\widehat{\delta}_{fbs3sls,J,n} &= \left[\widetilde{\mathcal{Z}}'_{J,n}\left(\widetilde{\Sigma}_J^{-1} \otimes \widetilde{Q}_{o,n}(\widetilde{Q}'_{o,n}\widetilde{Q}_{o,n})^{-1}\widetilde{Q}'_{o,n}\right)\widetilde{\mathcal{Z}}_{J,n}\right]^{-1} \\ &\times\ \widetilde{\mathcal{Z}}'_{J,n}\left(\widetilde{\Sigma}_J^{-1} \otimes \widetilde{Q}_{o,n}(\widetilde{Q}'_{o,n}\widetilde{Q}_{o,n})^{-1}\widetilde{Q}'_{o,n}\right)\widetilde{\mathcal{Y}}_{J,n} \qquad (16)\end{aligned}$$

The following proposition shows that the feasible BS3SLSE has the same limiting distribution as the true BS3SLSE.

Proposition 2. *Suppose that Assumptions 1-7 hold, then*

$$\sqrt{n}\left(\widehat{\theta}_{fbs3sls,J,n} - \theta_0\right) \to^d \mathcal{N}\left(0, \sigma_0^2(\mathcal{I}\Omega_b)^{-1}\right) \qquad (17)$$

where

$$\Omega_b = \lim_{n\to\infty} \frac{1}{n}(G_n(X_n\beta_0 + \alpha_0 l_n), X_n)'\Gamma_n(G_n(X_n\beta_0 + \alpha_0 l_n), X_n) \qquad (18)$$

Consider the asymptotic efficiency of BS3SLSE relative to other efficient estimators such as QMLE and GMME. Lee (2007a) revealed that, the QMLE of model (1) was asymptotically equivalent to the GMME via moment restrictions $\mathbb{E}(g_n(\theta, \alpha)) = 0$ with

$$g_n(\theta, \alpha) = \frac{1}{n}\left[(u_n(\theta) - \alpha l_n)'Q_{o,n}, (u_n(\theta) - \alpha l_n)'\left(G_n - \frac{tr(G_n)}{n}I_n\right)(u_n(\theta) - \alpha l_n)\right], \qquad (19)$$

where $Q_{o,n}$ is given above. Denote the GMME of $(\theta'_0, \alpha_0)'$ using the moments (19) as $\left(\widehat{\theta}'_{gmm,n}, \widehat{\alpha}_{gmm,n}\right)'$. Then it suffices to compare the efficiency of BS3SLSE $\widehat{\theta}_{bs3sls,J,n}$ relative to the GMME $\widehat{\theta}_{gmm,n}$. The asymptotic variance matrix of $\left(\widehat{\theta}'_{gmm,n}, \widehat{\alpha}_{gmm,n}\right)'$ computed in Lee (2007a) is

$$\text{AsyVar}\left(\sqrt{n}\widehat{\theta}'_{gmm,n},\sqrt{n}\widehat{\alpha}_{gmm,n}\right)' =$$

$$\begin{pmatrix} \begin{matrix}\frac{1}{n}tr\left(\left(G_n-\frac{tr(G_n)}{n}I_n\right)^s G_n\right) \\ +\frac{1}{\sigma_0^2 n}\left(G_n(X_n\beta_0+\alpha_0 l_n)\right)'\left(G_n(X_n\beta_0+\alpha_0 l_n)\right)\end{matrix} & \frac{1}{\sigma_0^2 n}\left(G_n(X_n\beta_0+\alpha_0 l_n)\right)'(X_n,l_n) \\ \frac{1}{\sigma_0^2 n}(X_n,l_n)'\left(G_n(X_n\beta_0+\alpha_0 l_n)\right) & \frac{1}{\sigma_0^2 n}(X_n,l_n)'(X_n,l_n) \end{pmatrix}^{-1} \tag{20}$$

where $A_n^s = A_n + A_n'$. From (20), we see that generally speaking, efficiency gain is not guaranteed for BS3SLSE relative to GMME as the latter takes into account the quadratic moments. The efficiency gain is guaranteed when the quadratic moment is asymptotically dominated by the linear moments, or simply

$$\frac{1}{n}tr\left(\left(G_n-\frac{tr(G_n)}{n}I_n\right)^s G_n\right)=o(1). \tag{21}$$

This is because in the presence of (21), the asymptotic variance of $\widehat{\theta}_{gmm,n}$ is identical to that of BS2SLSE $\widehat{\theta}_{bs2sls,n}$(see (15)) by the formula of the inverse of a partitioned matrix. One sufficient condition for (21) to hold is that $h_n \to \infty$. According to the remarks below Assumption 3, it implies that each unit can be influenced aggregately by a significant portion of units on the cross section.

4. SOME MONTE CARLO RESULTS

We conduct a small Monte Carlo simulation to evaluate performances of BS3SLSE and its efficiency relative to other estimators such as QMLE (Lee 2004) and distribution-free best GMME (Liu et al. 2006) for small or moderate sample size under various distributions. The data are generated by model (1) with $\theta_0 = (0.3, 1, -1)'$, $\alpha_0 = 0$, $\sigma_0 = 1$. The spatial-varying regressors x_{in}, $i = 1, 2$ are generated by standard normal distribution. The spatial scenario in Case (1991) is used to specify the spatial weights matrix. Assume that there are R districts with m members in each district and the neighbors in the same district are given equal weight, i.e., $W_n = I_R \otimes B_m$, $B_m = (m-1)^{-1}(l_m l_m' - I_m)$. Experiment with $(R, m) = (10, 5), (15, 7)$, $(20, 10)$ and $(30, 15)$. ϵ_{ni}'s, $i = 1, \cdots, n$ are generated by the following three types of distribution: (a) Normal, $\epsilon_{ni} \sim \mathcal{N}(0, 1)$; (b) Bimodal mixture normal, $\epsilon_{ni} = v/\sqrt{10}$, $v \sim 0.5\mathcal{N}(-3, 1) + 0.5\mathcal{N}(3, 1)$; (c) Student t, $\epsilon_{ni} \sim \sqrt{3/5} \cdot t(5)$. All the three distributions are scaled to have unit variance as required by Assumption 1 and they are commonly used in many Monte Carlo studies concerning asymptotic efficiency.

For each design, we compute BS3SLSE using $m_{Jj}(\epsilon) = \left(\epsilon/(1+\epsilon^2)^{1/2}\right)^j$ for four different choices of $J(n)$. As $J(n)$ is supposed to be increasing more slowly than $\ln n$, we choose $J_1 = [\ln n] - 2, J_2 = [\ln n] - 1, J_3 = [\ln n]$ and $J_4 = [\ln n] + 1$, where $[n]$ is the largest integer no greater than n. The commonly used parametric estimators for SAR model (1) include S2SLSE (Kelejian and Prucha 1998), BS2SLSE (Lee 2003), GMME (Lee 2007a), QMLE(Lee 2004) and distribution-free BGMME (Liu et al. 2006). A few remarks concerning their relative asymptotic efficiencies are given below. S2SLSE and BS2SLSE are generally not as efficient as GMME and QMLE. Under normality, QMLE, GMME and the distribution-free BGMME are asymptotically equivalent while the distribution-free BGMME will be more efficient than QMLE and GMME when the disturbances are not normally distributed. Here we compute BS2SLSE, QMLE and distribution free BGMME for comparison of relative efficiency to B3SLSE. All the BS2SLSE, BS3SLSE, and BGMME will use QMLE as the initial estimator. For each case, we do 1000 repetitions and report empirical bias, empirical standard deviation and root mean square error (RMSE). The simulation results are summarized in Table 1-3.

The findings from Table 1-3 include that, (a) for different distributions and different choices of $J(n)$, BS3SLSEs perform quite well for finite samples. They are slightly biased and the biases become smaller when n goes larger. As the sample size increases, both empirical standard deviation and RMSE decline and the magnitude of such decline is generally consistent with the expected $\sqrt{n}$-asymptotics. Furthermore, their biases and standard deviations do not change much with varying $J(n)$'s. (b) Under normality, the BS3SLSEs of the regression slopes in moderate size samples can be as good as their QMLE or BGMME. For the spatial parameter λ, there still exists significant advantage of QMLE in terms of empirical standard error. Intuitively, since a normal distribution can be uniquely determined by its first and second order moment, the quadratic form moment of the error terms may contain a great deal of information useful for estimation. For small or moderate size samples, failure to use the quadratic moments by S3SLSE may significantly undermine the resulting efficiency relative to QMLE. (c) Under symmetric mixture normality, an overwhelming improvement of efficiency is observed in all cases. In contrast to the family of normal distributions, the family of mixture normality is unlikely to be determined by the moments up to the second order.[4] Fundamentally, S3SLSE utilizes much useful information contained in the higher order moments of the underlying distributions that has not yet accounted by either BGMME or QMLE. For the spatial parameter, one will see as much as 40% efficiency gain relative to QMLE or BGMME in terms of RMSE and 60% efficiency gain for the regression slopes. (d) Under Student t distribution,

[4]A general mixture normal distribution can be described as $w_1\mathcal{N}(\mu_1, \sigma_1^2) + (1 - w_1)\mathcal{N}(\mu_2, \sigma^2)$. Hence it will be determined by means of the moments up to the fifth order.

the relative performances of BS3SLSEs are intermediate between normal case and mixture normal case. The regression slopes can be estimated more precisely by 10% while no improvement has been seen for the spatial parameter. (e) In almost each case, BS3SLSE always performs as good as or outperforms B2SLSE, that is quite consistent with our theoretical analysis.

In summary, BS3SLSE will be valuable relative to QMLE and GMME in improving the precision of regression slopes under nonnormality. And it will be particularly useful if a great deal of information of the underlying distribution is characterized by its high order moments.

Finally, a brief discussion of the method of choosing J is needed here. Although addressing the issue of choosing J for small samples is beyond the scope of this paper, it is always of practical concern for any semi-parametric estimates. Choosing J such that the estimated asymptotic variance matrix from that given by Proposition 1 is minimized will not work, because the estimated factor $\widetilde{\mathbb{M}}'_J\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J$ in the asymptotic formula is a positive-definite quadratic form composed of a sum of J^2 terms, whose value generally increases as J grows. One method would be to choose J to minimize the standard errors estimated by the bootstrap. In such a bootstrap re-sampling procedure, the sample distribution of x_{ni}'s is taken as the population distribution, the initial estimates of θ, α and σ are taken as the true parameters, and the sample distribution of the estimated residuals is taken as the population distribution of ϵ. Note that such method would be particular simple for our S3SLSE since the choice involves only a single integer-valued variable J.

5. AN EMPIRICAL EXAMPLE

There has been considerable controversy over the empirical significance of the theoretical prediction that the pressure of economic catching-up in an open economy may encourage developing countries to relax their environmental regulations for attracting more foreign capitals. Strong environmental regulations are viewed to directly drive up production costs by requiring certain equipment, to decrease waste proposal capacity by restricting areas that can be used for landfills and to prohibit certain factor inputs or outputs. Given each of these reasons, multinational enterprises, in order to keep up their international competitiveness, may have the incentives to relocate their polluting production facilities from the country with stronger environmental regulations to the one with weaker regulations. On the other hand, an alternative view, given by Wheeler (2001), argued that the cost effects were so small as to be negligible and the resulting increased environmental cost would be compensated by employees wages. Antweiler et al. (2001) analyzed the impact of international trade upon environment using the data from 44 countries. They observed that laxer environmental regulations induced emission intensive industries to migrate from the richer to the poorer countries. However, on a global basis, it is hard to evaluate whether the net effect is positive or negative. Since the reform and opening up policy, FDI has been viewed as the fundamental driving force behind China's miracle of economic growth. At the same time, deteriorating eco-environment gives rise to much concerns towards foreign capital economy. There are a number of studies that are devoted to empirically examining whether China has become the pollution haven of foreign enterprises. Based on the province level panel data from 1987 to 1998, Ljungwang and Linde-Rahr (2005) concluded that in China the poor regions were more likely to attract FDI at the cost of environment. Dean et al. (2005) collected the data from 2886 foreign funded manufacturing companies and observed that it was only for the capitals from Hong Kong, Taiwan and Macau but not for all the foreign capitals, for example, those from OECD countries that China's lenient regulations would be a sufficient inducement to relocate their production facilities. Furthermore, they argued that this phenomenon could be explained by the production technology differentials from different sources. Similar empirical evidence was recorded in Cole et al. (2007).

Most empirical studies based on China's data seem to support the hypothesis of "pollution haven". However, none of the literature so far has given explanation to the formation of such pollution haven from the perspective of China's special political and fiscal conditions. China is a politically centralized and economically decentralized country. Upper level governments select and appoint officials at lower levels based on their relative performances in developing local economy. The aim to get politically promoted produces sufficient incentives for local officials to court for high growth rate in their own jurisdictions during the term of office (Zhou 2004, 2007). Owing to the inflexible household registration system that largely restricts the population migration and the stringent financial supervision that substantially hampers domestic capital flows, it is not unnatural for local governments to purposefully relax the environmental regulations and even race against each other to the bottom so as to attract more mobile factors such as foreign capitals. On the other hand, China's pollution control is a combination of centralized legislation and decentralized provincial or municipal level implementation and enforcement. Provinces and municipalities in China retain considerable flexibility to align the environmental policies with their own interests and long run development objectives despite the presence of weak vertical administration from the state environmental protection agency. The arguments above suggest that it may be of interest to re-examine the environmental "race-to-the-bottom" hypothesis from the perspective of a decentralized economy accounting for the strategic interactions among local governments.

5.1. A New Empirical Setting

Assume there are n jurisdictions on the cross section, $i = 1, \cdots, n$ and the determination equation of the i-th jurisdiction's FDI inflow is given by

$$d_{ni} = \alpha_0 e_{ni} + x'_{1ni}\gamma_0 + u_{ni}, i = 1, \cdots, n \tag{22}$$

where e_{ni} represents the environmental stringency and x_{ni} includes other important factors than environmental stringency such as market size, company income tax rate and the economic openness, that affect FDI inflows. u_{ni} is the unobservable error term that incorporates other FDI determinants that have not been included in x_{ni}. Traditionally, the literature empirically confirms the presence of environmental "race-to the bottom" effect once the coefficient α_0 is significantly negative. Two characteristics distinguish our new testing procedures that follow from the previous studies. First, since the environmental stringency e_{ni}'s are not directly observed, the equation (22) can not be estimated straightforwardly. Secondly, the issue of interjurisdiction strategic competition is not fully recognized by the equation (22). Spillover effects of local expenditures on education, environment, housing, medical services and infra-structural facilities have been a major focus of theoretical and applied work in public economics. See Brueckner (2003) for an overview of recent literature on strategic interaction among governments and see Murdoch et al. (1997), Fredriksson and Millimet (2002) for empirical evidence of strategic engagement of local environmental policymaking. By accounting for both unobservability in the strictness of environmental regulations and possible strategic interactions among local policymaking, we introduce the following spatial autoregressive specification of e_{ni}'s,

$$e_{ni} = \lambda_0 \bar{e}_{ni} + x'_{2ni}\gamma_{20} + v_{2ni}, i = 1, \cdots, n \tag{23}$$

where $\bar{e}_{ni} = \sum_{j=1}^{n} w_{n,ij} e_{nj}$, x_{2ni} contains a set of variables that can be directly observed and closely related to the environmental stringency at jurisdiction i. Unlike the previous literature, we re-examine the presence of "race-to-the-bottom" hypothesis by simultaneously testing $\alpha_0 < 0$ and $\lambda_0 > 0$ in (22)-(23). The rationale is as follows: a positive λ_0 implies that, as a response to neighboring jurisdictions' tightening or relaxing their environmental regulations, the local government will also actively tighten or relax its own environmental regulations. By simultaneously examining both $\lambda_0 > 0$ and $\alpha_0 < 0$ instead of $\alpha_0 < 0$ only, the implication of "race" in the "race-to-the-bottom" can be better recognized. To derive the estimation equation, substitute (23) into (22) and rewrite it in matrix form as

$$d_n = \lambda_0 W_n d_n + X_{2n}\alpha_0\gamma_{20} + X_{1n}\gamma_{10} + W_n X_{1n}(-\lambda_0\gamma_{10}) + v_n, \tag{24}$$

or more concisely

$$d_n = \lambda_0 W_n d_n + X_{1n}\beta_{10} + W_n X_{1n}\beta_{20} + X_{2n}\beta_{30} + v_n, \tag{25}$$

where $d_n = (d_{n1}, \cdots, d_{nn})'$, $W_n = [w_{n,ij}]_{i,j=1,\cdots,n}$, $X_{ln} = (x_{ln1}, \cdots, x_{lnn})'$, $l = 1, 2$, $v_n = (I_n - \lambda_0 W_n)v_{1n} + \alpha_0 v_{2n}$. Note that in (24)-(25), only λ_0 can be identified from the equation while α_0 can not because $\beta_{30} = \alpha_0\gamma_{20}$. Recall from (23) that γ_{20} represents the slope of X_{2n}. If the variables in X_{2n} can be carefully chosen so that the slope signs of these variables can be determined a priori, then we can deduce the sign of α_0 from the sign of β_{30}. For example, if X_{2n} includes a set of representative pollutant emissions, then γ_{20} is expected to be negative and we can examine whether β_{30} is significantly positive to indirectly test whether α_0 is significantly negative.

We use the data containing 263 cities in 2007 drawn from "China City Statistical Yearbook" to fit the model (25). For our purpose, the log total amount of FDI actually utilized is chosen as the dependent variable. A variety of theoretical studies on FDI have identified many determinants of FDI including differences in the marginal return to capital, market size of host countries, exchange rate risk, trade impediments, market power and these variables may be included in the X_{1n} in (25). See Agarwal (1980) and Caves (1983) for a comprehensive reviews of theories of FDI determination. In summary, the following four variables will be included in X_{1n}, namely, tax rate, market size, internal cash flow and economic openness. Specifically, we use the ratio of company income tax to total profits, GDP per capita, company total profit in 2006 and the ratio of total imports&exports to GDP to proxy tax rate(TR), market size(MS), adequacy of internal cash flow (IC) and economic openness (OP) respectively. By definition, the variables in X_{2n} should be directly observable and closely associated with the strictness of local environmental regulations such as pollutant emissions. The pollutant we consider is city-wide sulfur dioxide emissions (SUL), one of the most significant air pollutants worldwide and one of the variables most commonly used variables to proxy for environmental quality (OECD,1993). As the pollutant emissions are expected to be negatively related to the environmental regulation standards, namely, $\gamma_{20} < 0$ in (23), testing $\lambda_0 > 0$ and $\alpha_0 > 0$ in (22)-(23) then is equivalent to testing $\lambda_0 > 0$ and $\beta_{30} > 0$ in (25). The descriptive statistics of variables are summarized in Table 4. Finally, we have the following equation for estimation,

$$\begin{aligned} FDI_n &= \lambda W_n FDI_n + \beta_{con} l_n + \beta_{MS} MS_n + \beta_{WMS} W_n MS_n + \beta_{TR} TR_n \\ &+ \beta_{WTR} W_n TR_n + \beta_{OP} OP_n + \beta_{WOP} W_n OP_n \\ &+ \beta_{IC} IC_n + \beta_{WIC} W_n IC_n + \beta_{SUL} SUL_n + u_n \end{aligned} \tag{26}$$

5.2. Specify the Spatial Weights Matrix

In applying the spatial autoregressive model, specifying an appropriate spatial weights matrix is quite important. For the current study, an obvious candidate is geographical proximity. Local governments often view other cities in the same province or that share a common boundary as strategic opponents in making environmental policy. Although geography may be relevant, it is not the only factor that determines the neighborhood. Two geographically remote but economically similar cities may be more likely to incorporate each other's environmental policies into their own decision calculus than two geographically nearby cities but with large income gap between them do. Denote the spatial weights matrix based on geographical proximity to be W_n^G, that is generated by the follows steps: (i) $w_{n,ij}^G = 1$ if i and j share a common boundary and $w_{n,ij}^G = 0$ otherwise. (ii) Normalize each row sum of the zero-one matrix to be one. Denote the spatial weights matrix based on similarities in economic characteristics as W_n^I which is generated by: (i) Cities located in different provinces are viewed as spatially uncorrelated. (ii) Within the same province, the closeness of two cities is inversely proportional to their difference in economic levels, namely $w_{n,ij}^I = 1/|\overline{PGDP}_{ni} - \overline{PGDP}_{nj}|/S_{ni}^I$ where $\overline{PGDP}_{ni}$ is per capita GDP in city i, S_{ni}^I is the sum of $1/|\overline{PGDP}_{ni} - \overline{PGDP}_{nj}|$ for j going over all cites other than i within the province containing i. Surely, the resulting W_n^I will be row-normalized automatically. There are two merits for such row normalization. First, row normalization facilitates interpretation of spatial lag term of cross sectional unit as weighted average of its neighbors. Second, it can make different spatial autoregressive parameters comparable. Like any other empirical studies that apply spatial model, the principles to specify a spatial weights matrix seem to be somewhat arbitrary. There is no reason why we should adopt W_n^G instead of W_n^I in the empirical implementation or vice versa. One possible procedure to reduce such arbitrariness in specifying the notion of neighborliness is to nest these two criteria:

$$W_n(\psi) = \psi W_n^G + (1 - \psi) W_n^I \tag{27}$$

By varying ψ between 0 and 1, we can test the hypothesis more robustly and assess the merits of different candidates for neighborliness.

5.3. Empirical Results

The following estimators, (i) MLE, (ii) BS2SLSE, (iii) distribution-free GMME and (iv) BS3SLSE with $J = 5, \cdots, 8$ are computed for equation (26) respectively. As in the Monte Carlo study, all the BS2SLSE, GMME and BS3SLSE's are computed using MLE as the initial estimator and choose $m_{Jj}(\epsilon) = \left(\epsilon/(1+\epsilon^2)^{1/2}\right)^j$ as the basis function series. The empirical setting described in Section 5.1 implies that both λ and β_{SUL} in (26) are of particular interest to us. The estimate results of λ and β_{SUL} and their t-statistics are summarized in Table 5 with ψ ranging from 0 to 1. The findings from the estimation results include that (a) despite a few exceptions, almost all the estimates of λ with different ψ's are significantly positive at 95% level, robustly suggesting the significant empirical presence of strategic interaction of local environmental policymaking. A few irregular signs in BS2SLSEs might be caused by their substantial inefficiency relative to other estimators. (b) As expected, the slope of pollutant emissions are significantly positive for all the estimators with various ψ's, implying that China's lenient environmental regulations remain attractive to foreign capitals. (c) With various estimators and spatial weights matrices that have been tried, both observations above reliably confirm the empirical presence of environmental "race-to-the-bottom" effect for local governments in FDI competition.

6. CONCLUSIONS

In this paper, we develop for spatial autoregressive models a computationally simple three-stage least squares estimator, that can be interpreted as the spatial counterpart of 3SLSE of a system of simultaneous equations, where the number of equations grows with the sample size slowly. With the best chosen instruments, the resulting BS3SLSE is shown to be asymptotically equivalent to BS2SLSE under normality and will be more efficient otherwise. Under certain economic spatial environments where each unit can be influenced aggregately by a significant portion of units in the cross section, our BS3SLSE is shown to be asymptotically equivalent to MLE (Lee, 2004) and GMME (Lee, 2007a) under normality and is more efficient under nonnormal distributions. Monte Carlo results indicate that the suggested BS3SLSE performs quite well in small or medium size samples and will be particularly valuable relative to QMLE and GMME in improving the precision of regression slopes if much information of the underlying distribution is characterized by its high order moments. As an empirical example, we apply BS3SLSE and other previously proposed estimators to re-testing the "race-to-the-bottom" hypothesis for local governments in competition for FDI in a new empirical setting.

For future research, our S3SLSE can be easily extended to other important spatial regression models such as high order spatial autoregressive models and spatial panel data models. It is also of practical interest to investigate optimal choice of the number of moments J used in estimation for small samples.

APPENDIX A

Lemma A.1.-A.3, A.5, A.7-A.9 can be found in Newey (1988). Lemma A.6, Lemma A.10 can be found in Lee (2007a).

Lemma 1. *Consider a sequence* $\{\eta_i(v)\}_{i=1}^{\infty}$ *of measurable functions and a sequence* $\{v_i\}_{j=1}^{\infty}$ *of independent random variables. For each integer* J, *let* $\eta_{J,j} = (\eta_1(v_j), \cdots, \eta_J(v_j))'$. *For some* $1 < \omega \leq 2$ *and* $\{B(J)\}_{J=1}^{\infty}$, *let* $\frac{1}{n}\sum_{i=1}^{J}\sup_n \sum_{j=1}^{n}\mathbb{E}\left(|\eta_i(v_j)|^{1+\omega}\right) = O(B(J))$. *Then* $\left|\frac{1}{n}\sum_{j=1}^{n}\mathbb{E}(\eta_{J,j})\right| = O\left(B^{1/(1+\omega)}(J)\right)$, $\left|\frac{1}{n}\sum_{j=1}^{n}\eta_{J,j}\right| = O_p\left(B^{1/(1+\omega)}(J)\right)$, *and for* $\omega < 1$ *and any sequence* $a_n \to \infty$,

$$\left|\frac{1}{n}\sum_{j=1}^{n}(\eta_{J,j} - E(\eta_{J,j}))\right| = O_p\left(a_n n^{-\omega/(1+\omega)} B^{1/(1+\omega)}(J)\right).$$

Lemma 2. *Let* $\{\eta_i(v,\gamma)\}_{i=1}^{J}$ *be a sequence of functions and* $\{v_i\}_{i=1}^{\infty}$ *a sequence of random variables where* γ *is a Euclidean vector. Suppose that there is a neighborhood of* N *of* γ_0 *and a sequence of measurable functions* $\{\zeta_i(v)\}_{i=1}^{\infty}$, *such that for all* γ *in* N *the Lipschitz condition* $|\eta_i(v_j,\gamma) - \eta_i(v_j,\gamma_0)| \leq \zeta_i(v_j)|\gamma - \gamma_0|$ *holds for all* i *with probability one. Let* $\{\widetilde{\gamma}_{nj}\}_{j=1,n=1}^{n,\infty}$ *be a triangular array of random variables such that* $\sup_{1\leq j\leq n}|\widetilde{\gamma} - \gamma_0| = O_p(n^{-\delta})$ *for some* $\delta > 0$. *Then for any* $\{B(J)\}$ *such that*

$$n^{-1}\sum_{i=1}^{J}\sup_n\sum_{j=1}^{n}\mathbb{E}(\zeta_i(v_j)) = O(B(J)),$$

it follows that for $\eta_{J,j}(\gamma) = (\eta_1(v_j,\gamma), \cdots, \eta_J(v_j,\gamma))'$,

$$\left|\frac{1}{n}\sum_{j=1}^{n}(\eta_{J,j}(\widetilde{\gamma}) - \eta_{J,j}(\gamma_0))\right| = O_p(n^{-\delta}B(J)).$$

Lemma 3. *For any* $\tau > 0$, *any positive integer* $j > 0$, *there exists some positive constant* $C > 0$ *such that* $|s|^j \leq C(j!)^2\exp(\tau|s|)$.

Lemma 4. *Assume that* $\widetilde{\theta}$ *is some* $\sqrt{n}$*-consistent estimator of* $\theta_0 = (\lambda_0, \beta_0')'$ *such that* $\sqrt{n}\left(\widetilde{\theta} - \theta_0\right) = O_p(1)$, *then* $z_{ni}'(\widetilde{\theta} - \theta_0) = O_p(n^{-1/2})$ *uniformly for* $i = 1, \cdots, n$, *where* $z_{ni} = (\sum_{j=1}^{n} w_{n,ij}y_{nj}, x_{ni}')'$.

Proof. Write $Z_n(\widetilde{\theta}-\theta) = (\widetilde{\lambda}-\lambda_0)G_n u_n + X_n(\widetilde{\beta}-\beta_0) + (\widetilde{\lambda}-\lambda_0)G_n X_n \beta_0$. For the moment, let $[c_n]_i$ be the i-th element of n-dimensional vector c_n. Then $\sqrt{n}z'_{ni}(\widetilde{\theta}-\theta_0) = \sqrt{n}(\widetilde{\lambda}-\lambda_0)\,[G_n u_n]_i + x'_{ni}\sqrt{n}(\widetilde{\beta}-\beta_0) + \sqrt{n}(\widetilde{\lambda}-\lambda_0)\,[G_n X_n \beta_0]_i$. As the elements of X_n are uniformly bounded and the row and column sums of G_n are uniformly bounded (by c, for example), it remains to show that $[G_n u_n]_i = O_p(1)$ uniformly in $i = 1, \cdots, n$. Both the uniform boundedness of the row and column sums of G_n and Cauchy-Schwartz inequality give that $\mathbb{E}|[G_n u_n]_i| = \mathbb{E}\left(|\sum_{j=1}^n g_{n,ij}u_{nj}|\right) \leq \sum_{j=1}^n |g_{n,ij}|\cdot\mathbb{E}|u_{ni}| \leq c\mathbb{E}^{1/2}\left(u_{ni}^2\right)$ uniformly in $i = 1, \cdots, n$. The desired result follows by Markov inequality. ▮

LEMMA 5. *For any positive constant C and δ that is sufficiently small, under Assumption 6, it follows that $J^{CJ}n^{-\delta} \to 0$.*

LEMMA 6. *Suppose that A_n is an $n \times n$ matrix with its column sums being uniformly bounded in absolute value, elements of the $n \times k$ matrix C_n are uniformly bounded, and $u_{n1}, \cdots, u_{nn}$ are i.i.d. with zero mean and finite variance σ_0^2. Then, $1/\sqrt{n}C'_n A_n u_n = O_p(1)$, $1/nC'_n A_n u_n = o_p(1)$ and $1/\sqrt{n}C'_n A_n u_n \to^d \mathcal{N}(0, \sigma_0^2 \lim_{n\to\infty} \frac{1}{n}C'_n A_n A'_n C_n)$ if the limit of $\frac{1}{n}C'_n A_n A'_n C_n$ exists and is positive definite.*

LEMMA 7. *For positive definite matrix A and a conformable matrix S, $\underline{\rho}(S'AS) \geq \underline{\rho}(S'S)\underline{\rho}(A)$, where $\underline{\rho}(\cdot)$ is the smallest eigenvalue of some matrix.*

LEMMA 8. *Assume the density function $f(\epsilon)$ of ϵ is continuous so that there is an interval on which it is bounded away from zero. $m(\cdot)$ is some continuously differentiable function. Let $q(\epsilon) = (1, m(\epsilon), \cdots, m^v(\epsilon))'$, then $\underline{\rho}\left(\mathbb{E}\left[q'(\epsilon)q(\epsilon)\right]\right) \geq Cv^{-cv}$.*

LEMMA 9. *Let A and B denote $r \times r$ symmetric and $r \times k$ matrices, respectively, and consider corresponding random matrices $\widetilde{A}_n$ and $\widetilde{B}_n$, and non-random matrices A_n and B_n, where $r = r_n$, is allowed to depend on n. Suppose that and that $|A_n| = O(\Delta_n)$, $\left|[\widetilde{A}_n, \widetilde{B}_n] - [A_n, B_n]\right| = O_p(\delta_n)$, and that $\underline{\rho}(A_n) = \rho_n > 0$. If $r_n\delta_n/\rho_n = o(1)$, then $\left|\widetilde{A}_n^{-1}\widetilde{B}_n - A_n^{-1}B_n\right| = O_p(r_n^3\rho_n^{-2}\Delta_n\delta_n)$, $\left|A_n^{-1}B_n\right| = O(r_n\Delta_n/\rho_n)$.*

Lemma 10. *Consider $S_n(\lambda) = I_n - \lambda W_n$. Suppose that $\{\|S_n^{-1}\|\}$ and $\{\|W_n\|\}$, where $\|\cdot\|$ is a matrix norm, are bounded. Then, $\{\|S_n(\lambda)\|\}$ are uniformly bounded in a neighborhood of λ_0.*

Proof of Proposition 1. To isolate the finite dimensional parameter θ from the infinite-dimensional nuisance parameters μ_J of δ in (11), the formula of the inverse of a partitioned matrix gives that

$$\left[\widetilde{\mathcal{Z}}'_{J,n}\left(\widetilde{\Sigma}_J^{-1}\otimes Q_n(Q_n'Q_n)^{-1}Q_n'\right)\widetilde{\mathcal{Z}}_{J,n}\right]^{-1}$$

$$= \begin{bmatrix} \widetilde{\mathbb{M}}'_J\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J\otimes\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'Z_n\right) & \widetilde{\mathbb{M}}'_J\widetilde{\Sigma}_J^{-1}\otimes\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'l_n\right) \\ \widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J\otimes\left(l_n'Q_n(Q_n'Q_n)^{-1}Q_n'Z_n\right) & \widetilde{\Sigma}_J^{-1}\otimes\left(l_n'Q_n(Q_n'Q_n)^{-1}Q_n'l_n\right) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} A_1 & A_2 \\ * & * \end{bmatrix},$$

where $A_1 = \left(\widetilde{\mathbb{M}}'_J\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J\right)^{-1}\otimes\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'Z_n\right)^{-1}$ and $A_2 = -\left(\widetilde{\mathbb{M}}'_J\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J\right)^{-1}\widetilde{\mathbb{M}}'_J\otimes n^{-1}\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'Z_n\right)^{-1} \times\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'l_n\right)$. Together with

$$\widetilde{\mathcal{Z}}'_{J,n}\left(\widetilde{\Sigma}_J^{-1}\otimes Q_n(Q_n'Q_n)^{-1}Q_n'\right)\widetilde{\mathcal{Y}}_{J,n}$$

$$= \begin{bmatrix} \widetilde{\mathbb{M}}'_J\widetilde{\Sigma}_J^{-1}\otimes\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\right) \\ \widetilde{\Sigma}_J^{-1}\otimes\left(l_n'Q_n(Q_n'Q_n)^{-1}Q_n'\right) \end{bmatrix}\widetilde{\mathcal{Y}}_{J,n},$$

the S3SLSE given by (11) can be written equivalently as

$$\begin{aligned} \widehat{\theta}_{s3sls,J,n} &= \widetilde{\theta} \qquad\qquad \text{(A.1)} \\ &+ \left(\widetilde{\mathbb{M}}'_J\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J\right)^{-1}\widetilde{\mathbb{M}}'_J\widetilde{\Sigma}_J^{-1}\otimes\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'Z_n\right)^{-1} \\ &\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'\right)\begin{bmatrix} \widetilde{\widetilde{m}}_{J1}(u_n(\widetilde{\theta})) - \widetilde{\mu}_{J1}l_n \\ \cdots \\ \widetilde{\widetilde{m}}_{JJ}(u_n(\widetilde{\theta})) - \widetilde{\mu}_{JJ}l_n \end{bmatrix}. \end{aligned}$$

A Taylor expansion of the righthand side of (A.1) around θ_0 yields

$$
\begin{aligned}
&\sqrt{n}(\widehat{\theta}_{s3sls,J,n}-\theta_0)= \\
&\Bigg\{ I_{k+1}-\left[\left(\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J\right)^{-1}\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'Z_n\right)^{-1}\right] \\
&\left[\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1}\otimes\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'\right)\right]\cdot\begin{bmatrix}\widetilde{\mathcal{M}}_{J1}(u_n(\overline{\theta}))\\ \vdots \\ \widetilde{\mathcal{M}}_{JJ}(u_n(\overline{\theta}))\end{bmatrix} Z_n\Bigg\} \\
&\times\sqrt{n}(\widetilde{\theta}-\theta_0) \\
+\ &\sqrt{n}\left[\left(\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J\right)^{-1}\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'Z_n\right)^{-1}\right] \\
&\left[\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1}\otimes\left(Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'\right)\begin{bmatrix}\widetilde{\widetilde{m}}_{J1}(u_n(\theta_0))-\widetilde{\mu}_{J1}l_n\\ \cdots \\ \widetilde{\widetilde{m}}_{JJ}(u_n(\theta_0))-\widetilde{\mu}_{JJ}l_n\end{bmatrix}\right].
\end{aligned}
$$

The proof will be completed by the steps (i)-(v) established below.

(i) Denote $\mathbb{M}_J=\sigma_0^{-1}\mathbb{E}\left(\dot{m}_{J1}(\epsilon),\cdots,\dot{m}_{JJ}(\epsilon)\right)'$, $\rho_J(\epsilon)=\left(\dot{m}_{J1}(\epsilon)-\mu_{J1},\cdots,\dot{m}_{JJ}(\epsilon)-\mu_{JJ}\right)'$ and $\Sigma_J=\mathbb{E}\left(\rho_J(\epsilon)\rho_J'(\epsilon)\right)$, where $\epsilon=(u-\alpha_0)/\sigma_0$. Then $\mathbb{M}_J=-\sigma_0^{-1}\mathbb{E}\left(\rho_J(\epsilon)\phi(\epsilon)\right)$. First we show that $\sigma_0^2\mathbb{M}_J'\Sigma_J^{-1}\mathbb{M}_J\to\mathbb{E}\left(\phi^2(\epsilon)\right)$ as $J\to\infty$. Let $d_J=\Sigma_J^{-1}\mathbb{E}\left(\rho_J(\epsilon)\phi(\epsilon)\right)=-\sigma_0\Sigma_J^{-1}\mathbb{M}_J$, then $d_J'\rho_J(\epsilon)$ is the least squares projection of $\phi(\epsilon)$ on $\rho_J(\epsilon)$. By Freud (1971), the raw moments of the random variable $v=m_{J1}(\epsilon)$ characterize its distribution if and only if the non-negative powers of v to form a basis for the Hilbert space of measure functions of v that have finite squared expectation. Since $\mathbb{E}\left(\phi^2(m_{J1}^{-1}(v))\right)=\mathbb{E}\left(\phi^2(\epsilon)\right)$ is finite, where $m_{J1}^{-1}(\cdot)$ is the inverse function of $m_{J1}(\cdot)$, there exists a triangular array $\{\widetilde{c}_{Jj}\}_{j\le J}$ such that

$$
\mathbb{E}\left(\left\{\phi(m_{J1}^{-1}(v))-\sum_{j=0}^{J}\widetilde{c}_{Jj}v^j\right\}^2\right)=\mathbb{E}\left(\left\{\phi(\epsilon)-\sum_{j=0}^{J}\widetilde{c}_{Jj}m_{J1}^j(\epsilon)\right\}^2\right)\to 0,\quad \text{as}\quad J\to\infty.
$$

As $\mathbb{E}\left(\phi(\epsilon)\right)=0$, it follows that the least squares projection of $\phi(\epsilon)$ on $\rho_J(\epsilon)$ will be equal the least squares projection of $\phi(\epsilon)$ on $\left(1,m_{J1}(\epsilon),\cdots,m_{J1}^J(\epsilon)\right)$.

Therefore $\lim_{J\to\infty}\mathbb{E}\left(\{\phi(\epsilon)-d_J'\rho_J(\epsilon)\}^2\right)=0$ follows from

$$\begin{aligned}\mathbb{E}\left(\{\phi(\epsilon)-d_J'\rho_J(\epsilon)\}^2\right) &= \min_{c_0,\cdots,c_J}\mathbb{E}\left(\left\{\phi(\epsilon)-\sum_{j=0}^{J}c_j m_{J1}^j(\epsilon)\right\}^2\right)\\ &\le \mathbb{E}\left(\left\{\phi(\epsilon)-\sum_{j=0}^{J}\widetilde{c}_{Jj} m_{J1}^j(\epsilon)\right\}^2\right). \qquad (A.2)\end{aligned}$$

Finally, $\sigma_0^2\mathbb{M}_J'\Sigma_J^{-1}\mathbb{M}_J=d_J'\Sigma_J d_J=\mathbb{E}\left(d_J'\rho_J(\epsilon)\rho_J'(\epsilon)d_J\right)=\mathbb{E}\left([d_J'\rho_J(\epsilon)]^2\right)\to\mathbb{E}\left(\phi^2(\epsilon)\right)$.

(ii) Next show that $\left\|\widetilde{\sigma}^2\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J-\sigma_0^2\mathbb{M}_J'\Sigma_J^{-1}\mathbb{M}_J\right\|\to^p 0$. Let $\eta_{J,ni}^1(\theta,\alpha,\sigma)=\left(\dot{m}_{J1}\left((u_{ni}(\theta)-\alpha)/\sigma\right),\cdots,\dot{m}_{JJ}\left((u_{ni}(\theta)-\alpha)/\sigma\right)\right)'$. For any $j=1,\cdots,J$ and θ,α,σ lying in the some neighborhood of $\theta_0,\alpha_0,\sigma_0$, we have by Assumption 7 that

$$\begin{aligned}&\left|\dot{m}_{Jj}\left((u_{ni}(\theta)-\alpha)/\sigma\right)-\dot{m}_{Jj}\left((u_{ni}-\alpha_0)/\sigma_0\right)\right|\\ =\ & j\Big|m_{J1}^{j-1}\left((u_{ni}(\theta)-\alpha)/\sigma\right)\dot{m}_{J1}\left((u_{ni}(\theta)-\alpha)/\sigma\right)\\ &-m_{J1}^{j-1}\left((u_{ni}-\alpha_0)/\sigma_0\right)\dot{m}_{J1}\left((u_{ni}-\alpha_0)/\sigma_0\right)\Big|\\ \le\ & j\left|m_{J1}^{j-1}\left((u_{ni}(\theta)-\alpha)/\sigma\right)\right|\left|\dot{m}_{J1}\left((u_{ni}(\theta)-\alpha)/\sigma\right)-\dot{m}_{J1}\left((u_{ni}-\alpha_0)/\sigma_0\right)\right|\\ +\ & j\left|\dot{m}_{J1}\left((u_{ni}-\alpha_0)/\sigma_0\right)\right|\left|m_{J1}^{j-1}\left((u_{ni}(\theta)-\alpha)/\sigma\right)-m_{J1}^{j-1}\left((u_{ni}-\alpha_0)/\sigma_0\right)\right|\\ \le\ & J\left(B_1^{J-1}(u_{ni})B_2(u_{ni})+JB_1^{J-2}(u_{ni})B_2^2(u_{ni})\right)\left|\left(z_{ni}'\left(\theta-\theta_0\right)+\alpha-\alpha_0,\sigma-\sigma_0\right)\right|\end{aligned}$$

where $z_{ni}=\left(\sum_{j=1}^n w_{n,ij}y_{nj},x_{ni}'\right)'$. By Lemma A.3 and Assumption 7, both $\mathbb{E}\left(B_1^{J-1}(u_{ni})\right)$ and $\mathbb{E}\left(B_1^{J-2}(u_{ni})\right)$ are smaller than $(J!)^2\mathbb{E}\left(\exp(\tau B_1(u))\right)\le J^{2J}$. Letting $B(J)=J^{CJ}$ for some sufficiently large positive constant C, $v_j=u_{nj}$ and $\delta=1/2$, both Lemma A.4 and Lemma A.2 apply and give that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\eta_{J,ni}^1(\widetilde{\theta},\widetilde{\alpha},\widetilde{\sigma})-\eta_{J,ni}^1(\theta_0,\alpha_0,\sigma_0)\right)\right|=O_p\left(n^{-1/2}J^{CJ}\right). \qquad (A.3)$$

Further as for $j=1,\cdots,J$, $\left|\dot{m}_{Jj}((u_{ni}-\alpha_0)/\sigma_0)\right|^2=\left|j\cdot m_{J1}^{j-1}((u_{ni}-\alpha_0)/\sigma_0)\dot{m}_{J1}((u_{ni}-\alpha_0)/\sigma_0)\right|^2\le J^2B_1^{2J}(u_{ni})B_2^2(u_{ni})$, still we have $\mathbb{E}\left(\eta^{1\prime}_{J,ni}(\theta_0,\alpha_0,\sigma_0)\eta^1_{J,ni}(\theta_0,\alpha_0,\sigma_0)\right) = O\left(J^{CJ}\right)$. Letting $\omega = 1$, $B(J) = J^{CJ}$, Lemma A.1 applies and gives

$$\left|\frac{1}{n}\sum_{i=1}^{n}\left(\eta^1_{J,ni}(\theta_0,\alpha_0,\sigma_0) - \mathbb{E}\left(\eta^1_{J,ni}(\theta_0,\alpha_0,\sigma_0)\right)\right)\right| = O_p\left(n^{-1/2}J^{CJ}\right). \tag{A.4}$$

and

$$\left|\mathbb{E}\left(\eta^1_{J,ni}(\theta_0,\alpha_0,\sigma_0)\right)\right| = O\left(J^{CJ}\right). \tag{A.5}$$

Then $\|\widetilde{\sigma}\widetilde{\mathbb{M}}_J - \sigma_0\mathbb{M}_J\| \to^p 0$ follows from Lemma A.5, (A.3) and (A.4).

As $\widetilde{\mu}_{Ji} = n^{-1}\sum_{k=1}^{n}\widetilde{\widetilde{m}}(u_{nk}(\widetilde{\theta}))$, $\Sigma_{J,ij} = n^{-1}\sum_{k=1}^{n}\widetilde{\widetilde{m}}_{Ji}(u_{nk}(\widetilde{\theta}))\widetilde{\widetilde{m}}_{Jj}(u_{nk}(\widetilde{\theta})) - n^{-1}\sum_{k=1}^{n}\widetilde{\widetilde{m}}_{Ji}(u_{nk}(\widetilde{\theta}))n^{-1}\sum_{l=1}^{n}\widetilde{\widetilde{m}}_{Ji}(u_{nl}(\widetilde{\theta}))$. Define a $J\times J$ matrix function $\eta^2_{J,n,i}$ on θ,α and σ with its (j,k)-th element being $\eta^2_{J,n,i,jk}(\theta,\alpha,\sigma) = m_{Jj}\left((u_{ni}(\theta)-\alpha)/\sigma\right)m_{Jk}\left((u_{ni}(\theta)-\alpha)/\sigma\right)$ and a J-dimensional vector function $\eta^3_{J,n,i}$ on θ,α and σ with its j-th element being $\eta^3_{J,n,i,j}(\theta,\alpha,\sigma) = m_{Jj}\left((u_{ni}(\theta)-\alpha)/\sigma\right)$. Then for any θ,α,σ lying in the some neighborhood of $\theta_0,\alpha_0,\sigma_0$, we have

$$\begin{aligned}
&\left|\eta^2_{J,n,i,jk}\left(\widetilde{\theta},\widetilde{\alpha},\widetilde{\sigma}\right) - \eta^2_{J,n,i,jk}(\theta_0,\alpha_0,\sigma_0)\right| \\
\le\ &\left|m_{Jj}\left((u_{ni}(\widetilde{\theta})-\widetilde{\alpha})/\widetilde{\sigma}\right)\right|\left|m_{Jk}\left((u_{ni}(\widetilde{\theta})-\widetilde{\alpha})/\widetilde{\sigma}\right) - m_{Jk}\left((u_{ni}-\alpha_0)/\sigma_0\right)\right| \\
+\ &\left|m_{Jk}\left((u_{ni}-\alpha)/\sigma\right)\right|\left|m_{Jj}\left((u_{ni}(\widetilde{\theta})-\widetilde{\alpha})/\widetilde{\sigma}\right) - m_{Jj}\left((u_{ni}-\alpha_0)/\sigma_0\right)\right| \\
\le\ &2B_1^J(u_{ni})JB_1^{J-1}(u_{ni})B_2(u_{ni})\left|\left(z'_{ni}\left(\theta-\theta_0\right)+\alpha-\alpha_0,\sigma-\sigma_0\right)\right|,
\end{aligned}$$

$$\begin{aligned}
&\left|\eta^3_{J,n,i,j}\left(\widetilde{\theta},\widetilde{\alpha},\widetilde{\sigma}\right) - \eta^3_{J,n,i,j}(\theta_0,\alpha_0,\sigma_0)\right| \\
\le\ &JB_1^{J-1}(u_{ni})B_2(u_{ni})\left|\left(z'_{ni}\left(\theta-\theta_0\right)+\alpha-\alpha_0,\sigma-\sigma_0\right)\right|,
\end{aligned}$$

$$\mathbb{E}\left(tr\left(\eta^{2\prime}_{J,ni}(\theta_0,\alpha_0,\sigma_0)\eta^2_{J,ni}(\theta_0,\alpha_0,\sigma_0)\right)\right) = O\left(J^{CJ}\right)$$

and

$$\mathbb{E}\left(\eta^{3\prime}_{J,ni}(\theta_0,\alpha_0,\sigma_0)\eta^3_{J,ni}(\theta_0,\alpha_0,\sigma_0)\right) = O\left(J^{CJ}\right).$$

Applying Lemma A.1-A.2 and by similar argument in proving $\left\|\widetilde{\sigma}\widetilde{\mathbb{M}}_J - \sigma_0\mathbb{M}_J\right\| \to^p 0$, we have $\left\|\widetilde{\Sigma}_J - \Sigma_J\right\| \to^p 0$. By Assumption 1, the density $f(\epsilon)$ is continuous so that there is an interval on which it is bounded away from zero. Let $q(\epsilon_{ni}) = (1, m_{J1}(\epsilon_{ni}), \cdots, m_{JJ}(\epsilon_{ni}))'$, it follows by $\Sigma_J = (-\mu_J, I_J)\,\mathbb{E}\left(q(\epsilon_{ni})q'(\epsilon_{ni})\right)(-\mu_J, I_J)'$, Lemma A.7-A.8 that $\underline{\rho}(\Sigma_J) \geq CJ^{-CJ}$. Further the assumptions of Lemma A.9 are satisfied with $A_n = \Sigma_J$ and $B_n = \mathbb{M}_J$, $\Delta_n = J^{-CJ}$, $\delta_n = n^{-1/2}J^{-CJ}$ and $\rho_n = J^{-CJ}$. Then $\left\|\widetilde{\sigma}^2\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J - \sigma_0^2\mathbb{M}_J'\Sigma_J^{-1}\mathbb{M}_J\right\| = \left\|\widetilde{d}_J'\widetilde{\mathbb{M}}_J - d_J'\mathbb{M}_J\right\| \leq \left\|\widetilde{d}_J' - d_J'\right\|\left\|\widetilde{\mathbb{M}}_J\right\| + \|d_J'\|\left\|\widetilde{\mathbb{M}}_J - \mathbb{M}_J\right\| = O_p\left(n^{-1/2}J^{CJ}\right) \to 0$.

(iii) We claim that

$$\left[\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1} \otimes \left(\frac{1}{n}Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n(Q_n'Q_n)^{-1}Q_n'\right)\right] \cdot \begin{bmatrix} \widetilde{\mathcal{M}}_{J1}(u_n(\overline{\theta})) \\ \vdots \\ \widetilde{\mathcal{M}}_{JJ}(u_n(\overline{\theta})) \end{bmatrix} Z_n$$

$$\begin{aligned}\to^p \lim_{n\to\infty} \sigma_0^{-2}\mathcal{I} \cdot \frac{1}{n}\left(G_n(X_n\beta_0 + \alpha_0 l_n)\right)' Q_n(Q_n'Q_n)^{-1}Q_n'\Gamma_nQ_n \\ \times(Q_n'Q_n)^{-1}Q_n'\left(G_n(X_n\beta_0 + \alpha_0 l_n)\right)\end{aligned}$$

By Lemma A.6, $\frac{1}{n}Q_n'G_n\epsilon_n = O_p\left(n^{-1/2}\right)$, $\frac{1}{n}Z_n'Q_n \to^p \frac{1}{n}\left(G_n(X_n\beta_0 + \alpha_0 l_n)\right)' Q_n$. It suffices to verify that

$$\left\|\frac{1}{n}\left(I_J \otimes Q_n'\right)\begin{bmatrix} \widetilde{\mathcal{M}}_{J1}(u_n(\overline{\theta})) \\ \vdots \\ \widetilde{\mathcal{M}}_{JJ}(u_n(\overline{\theta})) \end{bmatrix} Z_n - \frac{1}{n}\left(\mathbb{M}_J \otimes Q_n'Z_n\right)\right\| \to^p 0$$

Write $Z_n = \overline{Z}_n + (\sigma_0 G_n\epsilon_n, 0)$, where $\overline{Z}_n = (G_n(X_n\beta_0 + \alpha_0 l_n), X_n)$. Denote $Q_n = (q_{n1}, \cdots, q_{nn})'$, where q_{ni} is the i-th row of Q_n and analogously, $\overline{Z}_n = (\overline{z}_{n1}, \cdots, \overline{z}_{nn})'$. Let $\eta_{J,ni}^4(\theta, \alpha, \sigma) = (\dot{m}_{J1}\left((u_{ni}(\theta) - \alpha)/\sigma\right), \cdots, \dot{m}_{JJ}\left((u_{ni}(\theta) - \alpha)/\sigma\right))' \otimes q_{ni}\overline{z}_{ni}'$. Noting that the elements of Q_n and $\overline{Z}_n$ are regarded as uniformly bounded constants, by analogous argument to proving $\left\|\widetilde{\sigma}\widetilde{\mathbb{M}}_J - \sigma_0\mathbb{M}_J\right\| \to^p 0$, we have

$$\left\|\frac{1}{n}\left(I_J \otimes Q_n'\right)\begin{bmatrix} \widetilde{\mathcal{M}}_{J1}(u_n(\overline{\theta})) \\ \vdots \\ \widetilde{\mathcal{M}}_{JJ}(u_n(\overline{\theta})) \end{bmatrix} \overline{Z}_n - \frac{1}{n}\left(\mathbb{M}_J \otimes Q_n'\overline{Z}_n\right)\right\| \to^p 0$$

For the rest, let $\widetilde{\mathfrak{M}}(\bar{\theta}) = diag\Big\{\widetilde{\sigma}^{-1}\dot{m}((u_{n1}(\bar{\theta})-\widetilde{\alpha})/\widetilde{\sigma}) - \sigma_0^{-1}E\left(\dot{m}(\epsilon)\right), \cdots,$ $\widetilde{\sigma}^{-1}\dot{m}((u_{nn}(\bar{\theta})-\widetilde{\alpha})/\widetilde{\sigma}) - \sigma_0^{-1}E\left(\dot{m}(\epsilon)\right)\Big\}$ and $\bar{g}_{ni} = \sum_{j=1}^{n} g_{n,ij}\epsilon_{nj}$. As

$$\begin{aligned}
\sup_{1\le i\le n}\mathbb{E}|\bar{g}_{ni}| &\le \sup_{1\le i\le n}\left(\sum_{j=1}^{n}|g_{n,ij}|\right)\mathbb{E}|\epsilon_{ni}| \\
&\le \sup_{1\le i\le n}\left(\sum_{j=1}^{n}|g_{n,ij}|\right)\mathbb{E}^{1/2}\left(\epsilon_{ni}^2\right) \le C,
\end{aligned}$$

$\bar{g}_{ni} = O_p(1)$ uniformly in $i = 1, \cdots, n$ by Markov inequality. Then by the similar argument to (A.3), we have

$$\begin{aligned}
&\left\|\frac{1}{n}\left(I_J\otimes Q_n'\right)\begin{bmatrix}\widetilde{\mathcal{M}}_{J1}(u_n(\bar{\theta}))\\ \vdots \\ \widetilde{\mathcal{M}}_{JJ}(u_n(\bar{\theta}))\end{bmatrix}G_n\epsilon_n - \frac{1}{n}\left(\mathbb{M}_J\otimes Q_n'G_n\epsilon_n\right)\right| \\
\le&\left|\frac{1}{n}\begin{bmatrix}Q_n'\widetilde{\mathfrak{M}}_{J1}(u_n(\bar{\theta}))G_n\epsilon_n\\ \cdots \\ Q_n'\widetilde{\mathfrak{M}}_{JJ}(u_n(\bar{\theta}))G_n\epsilon_n\end{bmatrix}\right\| \\
\le&\left\|\frac{1}{n}\begin{bmatrix}\sum_{i=1}^{n}|q_{ni}|\,|\bar{g}_{ni}|\left|\widetilde{\sigma}^{-1}\dot{m}_{J1}((u_{ni}(\bar{\theta})-\widetilde{\alpha})/\widetilde{\sigma})-\sigma_0^{-1}E\left(\dot{m}_{J1}(\epsilon)\right)\right|\\ \cdots \\ \sum_{i=1}^{n}|q_{ni}|\,|\bar{g}_{ni}|\left|\widetilde{\sigma}^{-1}\dot{m}_{JJ}((u_{ni}(\bar{\theta})-\widetilde{\alpha})/\widetilde{\sigma})-\sigma_0^{-1}E\left(\dot{m}_{J1}(\epsilon)\right)\right|\end{bmatrix}\right\| \\
\le&\ C\left\|\frac{1}{n}\begin{bmatrix}\sum_{i=1}^{n}\left|\widetilde{\sigma}^{-1}\dot{m}_{J1}((u_{ni}(\bar{\theta})-\widetilde{\alpha})/\widetilde{\sigma})-\sigma_0^{-1}E\left(\dot{m}_{J1}(\epsilon)\right)\right|\\ \cdots \\ \sum_{i=1}^{n}\left|\widetilde{\sigma}^{-1}\dot{m}_{JJ}((u_{ni}(\bar{\theta})-\widetilde{\alpha})/\widetilde{\sigma})-\sigma_0^{-1}E\left(\dot{m}_{J1}(\epsilon)\right)\right|\end{bmatrix}\right\| \\
=&\ O_p\left(n^{-1/2}J^{CJ}\right).
\end{aligned}$$

(iv) We show that

$$\frac{1}{\sqrt{n}}Q_n'\Gamma_n\left(\widetilde{\phi}_J\left(u_n\right)-\phi(\epsilon_n)\right) = o_p(1),$$

where $\widetilde{\phi}_J(u) = -\widetilde{\mathbb{M}}_J\widetilde{\Sigma}_J^{-1}\begin{bmatrix}\widetilde{\widetilde{m}}_{J1}(u)-\widetilde{\mu}_{J1}\\ \cdots \\ \widetilde{\widetilde{m}}_{JJ}(u)-\widetilde{\mu}_{JJ}\end{bmatrix}$, $\widetilde{\phi}_J(u_n) = \left(\widetilde{\phi}_J(u_{n1}), \cdots, \widetilde{\phi}_J(u_{nn})\right)'$ and $\phi(u_n) = (\phi(u_{n1}), \cdots, \phi(u_{nn}))'$. Denote $\phi_J(\epsilon) = d_J'\rho_J(\epsilon)$ and $\phi_J(\epsilon_n) = (\phi_J(\epsilon_{n1}), \cdots, \phi_J(\epsilon_{nn}))'$. Note that $\mathbb{E}\left(\frac{1}{\sqrt{n}}Q_n'\Gamma_n\left(\phi_J(\epsilon_n)-\phi(\epsilon_n)\right)\right) = 0$ and use the independence of $\epsilon_{n1},\cdots,\epsilon_{nn}$, we have $\mathbb{E}\left(\frac{1}{n}Q_n'\Gamma_n\left(\phi_J(\epsilon_n)-\phi(\epsilon_n)\right)\left(\phi_J(\epsilon_n)-\phi(\epsilon_n)\right)'\Gamma_nQ_n\right) = \mathbb{E}\left(\phi_J(\epsilon)-\phi(\epsilon)\right)^2\cdot\frac{1}{n}Q_n'\Gamma_nQ_n = o(1)$ by (A.2). By Markov inequality,

$$\frac{1}{\sqrt{n}}Q_n'\Gamma_n\left(\phi_J(\epsilon_n)-\phi(\epsilon_n)\right)=o_p(1). \tag{A.6}$$

Letting $\eta^5_{J,ni}(\alpha,\sigma) = \left(m_{J1}\left((u_{ni}-\alpha)/\sigma\right),\cdots,m_{JJ}\left((u_{ni}-\alpha)/\sigma\right)\right)'$, $\gamma = (\alpha,\sigma)'$, we have

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(q_{ni}-\overline{q}\right)\widetilde{d}_J'\left(\widetilde{\rho}_J(u_{ni})-\rho_J(\epsilon_{ni})\right)\right\|\\
=&\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(q_{ni}-\overline{q}\right)\widetilde{d}_J'\left(m_J((u_{ni}-\widetilde{\alpha})/\widetilde{\sigma})-m_J((u_{ni}-\alpha_0)/\sigma_0)+\widetilde{\mu}_J-\mu_J\right)\right\|\\
=&\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(q_{ni}-\overline{q}\right)\widetilde{d}_J'\left(\frac{\partial\eta^5_{J,ni}(\overline{\gamma})}{\partial\gamma'}\left(\widetilde{\gamma}-\gamma_0\right)+\widetilde{\mu}_J-\mu_J\right)\right\| \qquad\qquad (A.7)\\
=&\left|\frac{1}{n}\sum_{i=1}^n\left(q_{ni}-\overline{q}\right)\widetilde{d}_J'\left(\frac{\partial\eta^5_{J,ni}(\overline{\gamma})}{\partial\gamma'}-\mathbb{E}\left(\frac{\partial\eta^5_{J,ni}(\overline{\gamma})}{\partial\gamma'}\right)\right)\sqrt{n}\left(\widetilde{\gamma}-\gamma_0\right)\right\|\\
\leq&\left\|\widetilde{d}_J'\right\|\left\|\sqrt{n}\left(\widetilde{\gamma}-\gamma_0\right)\right\|\left\|\frac{1}{n}\sum_{i=1}^n\left(q_{ni}-\overline{q}\right)\left(\frac{\partial\eta^5_{J,ni}(\overline{\gamma})}{\partial\gamma'}-\mathbb{E}\left(\frac{\partial\eta^5_{J,ni}(\overline{\gamma})}{\partial\gamma'}\right)\right)\right\|
\end{aligned}$$

$\partial m\left((u-\alpha)/\sigma\right)/\partial\alpha = -\sigma^{-1}\dot{m}\left((u-\alpha)/\sigma\right)$ and $\partial m\left((u-\alpha)/\sigma\right)/\partial\sigma = -\sigma^{-2}(u-\alpha)\dot{m}\left((u-\alpha)/\sigma\right)$. By analogous argument in proving $\left\|\widetilde{\sigma}\widetilde{\mathbb{M}}_J-\sigma_0\mathbb{M}_J\right\|\to^p 0$, the right hand of (A.7) will be no greater than $O_p\left(n^{-1/2}J^{CJ}\right)=o_p(1)$. The desired result then follows by

$$\begin{aligned}
&\frac{1}{\sqrt{n}}Q_n'\Gamma_n\left(\widetilde{\phi}_J(u_n)-\phi_J(\epsilon_n)\right)=\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(q_{ni}-\overline{q}\right)\left(\widetilde{d}_J'-d_J'\right)\rho_J(\epsilon_{ni})\right\|\\
\leq&\left\|\widetilde{d}_J-d_J\right\|\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(q_{ni}-\overline{q}\right)\rho_J(\epsilon_{ni})\right\|=O_p\left(n^{-1/2}J^{CJ}\right)\cdot J\cdot O_p(1)=o_p(1).
\end{aligned}$$

(v) Finally,

$$\frac{1}{\sqrt{n}}Q_n'\Gamma_n\phi(\epsilon_n)\to^d\mathcal{N}\left(0,\mathcal{I}\cdot\frac{1}{n}Q_n'\Gamma_nQ_n\right)$$

by Lemma A.6. ▮

Proof of Proposition 2. The proof essentially follows that of Proposition 1 and will be completed by the steps (i)-(vi) established below.

(i) By direct computation,

$$\begin{aligned} & (G_n(X_n\beta_0+\alpha_0 l_n),X_n)'Q_{o,n}(Q_{o,n}'Q_{o,n})^{-1} \\ \times\ & Q_{o,n}'\Gamma_n Q_{o,n}(Q_{o,n}'Q_{o,n})^{-1}Q_{o,n}'(G_n(X_n\beta_0+\alpha_0 l_n),X_n) \\ =\ & (G_n(X_n\beta_0+\alpha_0 l_n),X_n)'\Gamma_n(G_n(X_n\beta_0+\alpha_0 l_n),X_n), \end{aligned}$$

suggesting that the specific asymptotic variance matrix for BS3SLSE can be obtained by just replacing $Q_n = Q_{o,n}$ in (13).

(ii) By exactly the same proof to that of the part (i)-(ii) in Proposition 1, we can show that $\widetilde{\sigma}^2\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1}\widetilde{\mathbb{M}}_J \to^p \mathbb{E}(\phi^2(\epsilon))$.

(iii) Write

$$\begin{aligned} \widetilde{G}_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right) &= G_n\left(X_n\beta_0+\alpha_0 l_n\right)+\left(\widetilde{\lambda}-\lambda_0\right)\widetilde{G}_nG_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right) \\ &+\ G_n\left(X_n\left(\widetilde{\beta}-\beta_0\right)+\left(\widetilde{\alpha}-\alpha_0\right)l_n\right). \end{aligned} \tag{A.8}$$

It is straightforward to verify that $\frac{1}{n}\widetilde{Q}_{o,n}'\widetilde{Q}_{o,n} \to^p \frac{1}{n}Q_{o,n}{\prime}Q_{o,n}$ and $\frac{1}{n}Z_n{\prime}\widetilde{Q}_n \to^p \frac{1}{n}\left(G_n(X_n\beta_0+\alpha_0 l_n),X_n\right)' Q_{o,n}$ by referring to Lee (2003).

(iv)

$$\begin{aligned} & \left\| \frac{1}{n}\left(I_J\otimes\left(\widetilde{G}_nG_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right)\right)'\right)\begin{bmatrix}\widetilde{\mathcal{M}}_{J1}(u_n(\bar{\theta}))\\ \vdots \\ \widetilde{\mathcal{M}}_{JJ}(u_n(\bar{\theta}))\end{bmatrix}G_n\epsilon_n-\frac{1}{n}\left(\mathbb{M}_J\otimes\left(\widetilde{G}_nG_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right)\right)'G_n\epsilon_n\right)\right\| \\ \le\ & \left\|\frac{1}{n}\begin{bmatrix}\left(\widetilde{G}_nG_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right)\right)'\widetilde{\mathfrak{M}}_{J1}(u_n(\bar{\theta}))G_n\epsilon_n \\ \cdots \\ \left(\widetilde{G}_nG_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right)\right)'\widetilde{\mathfrak{M}}_{JJ}(u_n(\bar{\theta}))G_n\epsilon_n\end{bmatrix}\right\| \\ \le\ & \left\|\frac{1}{n}\begin{bmatrix}\sum_{i=1}^n|\kappa_{ni}|\,|\bar{g}_{ni}|\left|\widetilde{\sigma}^{-1}\dot{m}_{J1}((u_{ni}(\bar{\theta})-\widetilde{\alpha})/\widetilde{\sigma})-\sigma_0^{-1}\mathbb{E}\left(\dot{m}_{J1}(\epsilon)\right)\right\| \\ \cdots \\ \sum_{i=1}^n\|\kappa_{ni}|\,|\bar{g}_{ni}|\left|\widetilde{\sigma}^{-1}\dot{m}_{JJ}((u_{ni}(\bar{\theta})-\widetilde{\alpha})/\widetilde{\sigma})-\sigma_0^{-1}\mathbb{E}\left(\dot{m}_{J1}(\epsilon)\right)\right\|\end{bmatrix}\right\| \\ =\ & O_p\left(n^{-1/2}J^{CJ}\right) \end{aligned}$$

where $\kappa_{ni} = \left[\widetilde{G}_nG_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right)\right]_i$ being the i-th element of $\widetilde{G}_nG_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right)$, is uniformly bounded by a constant C independent of n by Lemma A.10. By analogous proof to that of part (iii) in Proposition 1, it is straightforward to show that

$$\begin{aligned}
&\left[\widetilde{\mathbb{M}}_J'\widetilde{\Sigma}_J^{-1}\otimes\left(\frac{1}{n}Z_n'\widetilde{Q}_{o,n}(\widetilde{Q}_{o,n}'\widetilde{Q}_{o,n})^{-1}\widetilde{Q}_{o,n}'\Gamma_n\widetilde{Q}_{o,n}(\widetilde{Q}_{o,n}'\widetilde{Q}_{o,n})^{-1}\widetilde{Q}_{o,n}'\right)\right]\\
\times\quad&\begin{bmatrix}\widetilde{\mathcal{M}}_{J1}(u_n(\overline{\theta}))\\ \vdots\\ \widetilde{\mathcal{M}}_{JJ}(u_n(\overline{\theta}))\end{bmatrix}Z_n\\
\rightarrow^p\quad&\sigma_0^{-2}\mathcal{I}\cdot\frac{1}{n}\left(G_n(X_n\beta_0+\alpha_0l_n)\right)'\Gamma_n\left(G_n(X_n\beta_0+\alpha_0l_n)\right).
\end{aligned}$$

(v) We claim that

$$\frac{1}{\sqrt{n}}\widetilde{Q}_{o,n}'\Gamma_n\left(\widetilde{\phi}_J\left(u_n\right)-\phi(\epsilon_n)\right)=o_p(1).$$

By referring to part (iv) of Proposition 1 and (A.8), it suffices to show that

$$\frac{1}{\sqrt{n}}(\widetilde{\lambda}-\lambda_0)\left(\widetilde{G}_nG_n\left(X_n\widetilde{\beta}+\widetilde{\alpha}l_n\right)\right)'\Gamma_n\left(\widetilde{\phi}_J\left(u_n\right)-\phi(\epsilon_n)\right)=o_p(1).\quad\text{(A.9)}$$

Write $\widetilde{G}_nG_n=G_n^2+(\widetilde{\lambda}-\lambda_0)\widetilde{G}_nG_n^2$. It can be concluded by the analogous argument to that of part (iv) in Proposition 1 that

$$\frac{1}{\sqrt{n}}\left(G_n^2X_n\right)'\Gamma_n\left(\widetilde{\phi}_J\left(u_n\right)-\phi(\epsilon_n)\right)=o_p(1)$$

and

$$\frac{1}{\sqrt{n}}\left(G_n^2l_n\right)'\Gamma_n\left(\widetilde{\phi}_J\left(u_n\right)-\phi(\epsilon_n)\right)=o_p(1).$$

Then (A.9) is given by

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{n}}(\widetilde{\lambda}-\lambda_0)^2\left(\widetilde{G}_nG_n^2X_n\widetilde{\beta}\right)'\Gamma_n\left(\widetilde{\phi}_J\left(u_n\right)-\phi(\epsilon_n)\right)\right\|\\
\le\;& n^{-1/2}\left\|n\left(\widetilde{\lambda}-\lambda_0\right)^2\right\|\left|\sup_{1\le i\le n}\left(\Gamma_n\widetilde{G}_nG_n^2X_n\widetilde{\beta}\right)_i\right|\\
&\times\frac{1}{n}\sum_{i=1}^{n}\left\|\widetilde{\phi}_J\left(u_{ni}\right)-\phi(\epsilon_{ni})\right\|\\
\le\;& O_p\left(n^{-1/2}\right).
\end{aligned}$$

(vi) Finally, it can be established that

$$\frac{1}{\sqrt{n}}\widetilde{Q}'_{o,n}\Gamma_n\phi(\epsilon_n) \to^d \mathcal{N}\left(0, \mathcal{I}\cdot\frac{1}{n}Q'_{o,n}\Gamma_n Q_{o,n}\right)$$

by Lemma A.6 and similar reasoning to proving (A.9). ∎

REFERENCES

Agarwal. J. P, 1980. Determinants of Foreign Direct Investment: A Survey. *Weltwirtschaftliches Archive* **116**, 739-773.

Anselin. L, 1988. *Spatial Econometrics: Methods and Models*, Kluwer Academic, The Netherlands.

Antweiler. W., B. Copeland and S. Taylor, 2001. Is Free Trade Good For the Rnvironment? *American Economic Review* **91**, 877-908.

Brueckner. J. k, 2003. Strategic Interactions Among Governments: An Overview of the Empirical Literature. *International Regional Science Review* **26**, 175 188.

Case. A. C, 1991. Spatial Pattern in Household Demand. *Econometrica* **59**, 953-965.

Case. A. C, 1992. Neighborhood Influence and Technological Change. *Regional Science and Urban Economics* **22** 491-508.

Caves. R. E, 1982. *Multinational Enterprise and Economics Analysis.* Cambridge, UK, Cambridge University Press.

Cole. M. A, R. Elliott. and J. Zhang, 2007. Environmental Regulation, Anticorruption, Government Efficiency and FDI Location in China: A Province Level Analysis. Department of Economics, University of Birmingham, Working Paper.

Dean. J. M., M. E. Lovely and H. Wang, 2005. Are Foreign Investors Attracted to Weak Environmental Regulations? World Bank Policy Research Working Paper.

Freud. G, 1971. *Orthogonal polynomials* Pergamon Press, Oxford.

Fredriksson. P. G. and D. L. Millimet, 2002. Strategic Interaction And the Determinants of Environmental Policy Across US States. *Journal of Urban Economics* **51**, 101-22.

Hajek. J. and Z. Sidak, 1967. *Theory of Rank Tests* Academic Press, New York.

Hansen. L. P, 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* **50**, 1029-1054.

Kelejian. H. H. and Prucha. I. R, 1998. A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbance. *Journal of Real Estate Finance and Economics* **17**, 99-121.

Kelejian, H. H. and Prucha. I. R, 1999. A generalized Moments Estimator For the Autoregressive Parameter in a Spatial Model. *International Economic Review* **40**, 509-533.

Lee. L. F, 2003. Best Spatial Two Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econometric Reviews* **22**, 307-335.

Lee. L. F, 2004. Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Econometric Models. *Econometrica* **72**, 1899-1926.

Lee. L. F, 2007. GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models. *Journal of Econometrics* **137**, 489-514

Lee. L. F, 2007. Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects. *Journal of Econometrics* **140**, 333-374.

Ljungwall. C. and M. Linde-Rahr, 2005. Environmental Policy and the Location of FDI in China. CCER Working Paper.

Liu. X., Lee. L. F. and Bollinger. C, 2006. Improved Efficient Quasi- maximum Likelihood Estimator of Spatial Autoregressive Models, Working Paper.

Murdoch. J. C., T. Sandler and K. Sargent. 1997. A Tale of Two Collectives: Sulphur Versus Nitrogen Oxide Emission Reduction in Europe. *Economica* **64**, 381-401.

Newey. W. K, 1988. Adaptive Estimation of Regression Models via Moment Restrictions. *Journal of Econometrics* **38**, 301-339.

OECD, 1993. *OECD Environmental Date*, Paris, OECD.

Wheeler. D. 2001. Race to the Bottom? Foreign Investment and Air Pollution in Developing Countries. *Journal of Environment and Development* **10**, 225-245.

Zhou, L, 2004. The Incentive and Cooperation of Government Officials in the Political Tournaments: An Interpretation of the Prolonged Local Protectionism and Duplicative Investments in China. *Economic Research Journal* **6**, 33-40.

Zhou, L, 2007. Governing China's Local Officials: An Analysis of Promotion Tournament Model. *Economic Research Journal* **7**, 36-50.