On the Consistency of the LIML Estimator of a Spatial Autoregressive Model with Many Instruments*

Xiaodong Liu†

Department of Economics, University of Colorado, Boulder, CO 80309, USA

First version, February 2011

Current version, April 2012

Abstract

This paper derives the LIML estimator for a spatial autoregressive model with endogenous regressors in the presence of many instruments. The LIML estimator is consistent when the number of instruments increases at a slower rate relative to the sample size. Due to spatial correlation, the LIML estimator in general is inconsistent when the number of instruments increases at the same rate as the sample size.

JEL classification: C3, C13

Key words: limited-information maximum likelihood; spatial autoregressive models; many instruments

1 Introduction

Since Bekker's (1994) seminal work, studies on estimation issues in the presence of many instruments have attracted a lot of attention (see, e.g., Donald and Newey, 2001; Chao and Swanson, 2005; Hansen et al., 2008; van Hasselt, 2010). However, much of the current literature has focused on models with independent observations. In a recent paper, Liu and Lee (2010) have considered the estimation of social-interaction effects in a network setting where the population is partitioned into many networks (groups) and observations are correlated within each group. The interaction

*I am grateful to Lung-fei Lee for helpful comments. All remaining errors are my own.

†Phone: 1(303)4927414. Fax: 1(303)4928960. Email: xiaodong.liu@colorado.edu.

among group members can be modelled by a generalized spatial autoregressive (SAR) model, and the various social-interaction effects, namely, endogenous effects, contextual effects and correlated effects (Manski, 1993), can be separately identified by an instrument based on a centrality measure of each network. Thus, the number of potential instruments for this model depends on the number of groups. If the number of groups increases with the sample size in asymptotic analysis, so does the number of instruments. Liu and Lee (2010) have suggested 2SLS and GMM estimators for this network model, and proposed a bias-correction procedure based on the estimated leading-order many-instrument bias.

For the estimation of models with many instruments, the limited-information maximum likelihood (LIML) estimator is of particular importance due to its good properties (see, e.g., Anderson et al., 2010). With independent observations, Bekker (1994) has shown that the LIML estimator is consistent when the number of instruments increases at the same rate as the sample size. This note derives the LIML estimator for an SAR model with endogenous regressors. The estimator can be easily modified to estimate the network model by Liu and Lee (2010). The LIML estimator is shown to be consistent when the number of instruments increases at a slower rate relative to the sample size. However, due to spatial correlation, the LIML estimator, in general, is inconsistent when the number of instruments increases at the same rate as the sample size.

For the rest of the note, Section 2 derives the LIML estimator for an SAR model and studies its consistency. Section 3 briefly concludes. The proofs are given in Appendices.

2 The SAR Model and LIML Estimation

2.1 Derivation of the LIML estimator

Consider an SAR model with endogenous regressors,

$$y_n = \lambda W_n y_n + Y_{1n}\gamma + \epsilon_n, \quad \text{and} \quad Y_{1n} = X_n \Pi + U_n. \tag{1}$$

In this model, n is the total number of spatial units, y_n is an n-dimensional vector of dependent variables, W_n is an $n \times n$ spatial weights matrix of known constants with a zero diagonal, Y_{1n} is an $n \times g$ matrix of explanatory variables that are possibly correlated with ϵ_n, and the innovations $\epsilon_{n1}, \cdots, \epsilon_{nn}$ of the n-dimensional vector ϵ_n are i.i.d. $(0, \sigma_\epsilon^2)$. Let X_n be an $n \times K$ matrix of exogenous variables. For a matrix A_n, let A_{ni} denote its ith row and $A_{n,ij}$denote its (i, j)th entry. We assume

U_{ni}'s are i.i.d. such that $E(U_{ni}) = 0$ and $\mathrm{E}(U'_{ni}U_{ni}) = \Sigma_u$. The correlation between Y_{1n} and ϵ_n is captured by $\mathrm{E}(U'_{ni}\epsilon_{ni}) = \sigma_{u\epsilon}$.

Let $S_n(\lambda) = I_n - \lambda W_n$ and $Y_n(\lambda) = [S_n(\lambda)y_n, Y_{1n}]$. At the true parameter value λ_0, $S_n = S_n(\lambda_0)$ and $Y_n = Y_n(\lambda_0)$. Furthermore, let

$$\Gamma = \begin{bmatrix} 1 & 0 \\ -\gamma & I_g \end{bmatrix}.$$

The model can be written more compactly as $Y_n(\lambda)\Gamma = X_nB + V_n$, where $B = [0, \Pi]$ and $V_n = [\epsilon_n, U_n]$. By assuming the joint normality of the error terms, the density of V_{ni} is given by $(2\pi)^{-(g+1)/2}|\Sigma|^{-1/2}\exp(-\frac{1}{2}V_{ni}\Sigma^{-1}V'_{ni})$, where

$$\Sigma = \mathrm{E}(V'_{ni}V_{ni}) = \begin{bmatrix} \sigma_\epsilon^2 & * \\ \sigma_{u\epsilon} & \Sigma_u \end{bmatrix}.$$

The Jacobian factor for the transformation of V_n into $[y_n, Y_{1n}]$ is

$$\begin{vmatrix} S'_n(\lambda) & \gamma' \otimes I_n \\ 0_{ng\times n} & I_{ng} \end{vmatrix} = |S_n(\lambda)|.$$

Let $\theta = (\lambda, \gamma')'$ and c be a generic constant term that does not involve unknown parameters and may be different for different uses. The log-likelihood function is

$$\begin{aligned} \ln L_n(\theta, \Pi, \Sigma) &= -\frac{n(g+1)}{2}\ln(2\pi) + \ln|S_n(\lambda)| - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\sum_i [Y_{ni}(\lambda)\Gamma - X_{ni}B]\Sigma^{-1}[Y_{ni}(\lambda)\Gamma - X_{ni}B]' \\ &= c + \ln|S_n(\lambda)| - \frac{n}{2}\ln|\Sigma| - \frac{1}{2}\mathrm{tr}(\Sigma^{-1}[Y_n(\lambda)\Gamma - X_nB]'[Y_n(\lambda)\Gamma - X_nB]). \end{aligned}$$

It is convenient to concentrate the log-likelihood function $\ln L_n(\theta, B, \Sigma)$ with respect to Σ^{-1}. As $\partial \ln L_n/\partial\Sigma^{-1} = \frac{n}{2}\Sigma - \frac{1}{2}[Y_n(\lambda)\Gamma - X_nB]'[Y_n(\lambda)\Gamma - X_nB]$, we have $\Sigma = \frac{1}{n}[Y_n(\lambda)\Gamma - X_nB]'[Y_n(\lambda)\Gamma - X_nB]$. The concentrated log-likelihood function is

$$\ln L_n^c(\theta, \Pi) = c + \ln|S_n(\lambda)| - \frac{n}{2}\ln|\frac{1}{n}[Y_n(\lambda)\Gamma - X_nB]'[Y_n(\lambda)\Gamma - X_nB]|. \tag{2}$$

Let $P_X = X_n(X'_nX_n)^{-1}X'_n$ and $M_X = I_n - P_X$. For any given λ, the above log-likelihood function can be further simplified by concentrating B out as

$$|\frac{1}{n}[Y_n(\lambda)\Gamma - X_nB]'[Y_n(\lambda)\Gamma - X_nB]| = [\frac{\epsilon_n'(\theta)M_X\epsilon_n(\theta)}{\epsilon_n'(\theta)\epsilon_n(\theta)}]^{-1}|\frac{1}{n}Y_n'(\lambda)M_XY_n(\lambda)|.$$

where $\epsilon_n(\theta) = S_n(\lambda)y_n - Y_{1n}\gamma$ (see Davidson and MacKinnon, 1993, pp.644-647). Hence, the LIML estimator can be obtained by maximizing the concentrated log-likelihood function

$$\ln L_n^c(\theta) = c + \ln|S_n(\lambda)| - \frac{n}{2}\ln|\frac{1}{n}Y_n'(\lambda)M_XY_n(\lambda)| + \frac{n}{2}\ln\frac{\epsilon_n'(\theta)M_X\epsilon_n(\theta)}{\epsilon_n'(\theta)\epsilon_n(\theta)}. \quad (3)$$

2.2 Consistency

For consistency of the LIML estimator, we inspect the first-order condition of (3). Let $G_n(\lambda) = W_nS_n^{-1}(\lambda)$ and $G_n = G_n(\lambda_0)$. Let e_j be the j column of an identity matrix I_n. At θ_0, the first-order derivative of the log-likelihood function with respective to λ is

$$\frac{1}{n}\frac{\partial \ln L_n^c(\theta_0)}{\partial\lambda} = -\frac{1}{n}\mathrm{tr}(G_n) + e_1'(Y_n'M_XY_n)^{-1}Y_n'M_XW_ny_n + \frac{\epsilon_n'W_ny_n}{\epsilon_n'\epsilon_n} - \frac{\epsilon_n'M_XW_ny_n}{\epsilon_n'M_X\epsilon_n}.$$

As $W_ny_n = G_nX_n\Pi_0\gamma_0 + G_n(U_n\gamma_0+\epsilon_n)$, we have $\mathrm{E}(\epsilon_n'W_ny_n) = (\sigma_{u\epsilon}'\gamma_0+\sigma_\epsilon^2)\mathrm{tr}(G_n)$, $\mathrm{E}(\epsilon_n'M_XW_ny_n) = (\sigma_{u\epsilon}'\gamma_0+\sigma_\epsilon^2)[\mathrm{tr}(G_n) - \mathrm{tr}(G_nP_X)]$ and $\mathrm{E}(\epsilon_n'M_X\epsilon_n) = \sigma_\epsilon^2\mathrm{tr}(M_X) = \sigma_\epsilon^2(n-K)$.[1] Hence,

$$\frac{\mathrm{E}(\epsilon_n'W_ny_n)}{\mathrm{E}(\epsilon_n'\epsilon_n)} - \frac{\mathrm{E}(\epsilon_n'M_XW_ny_n)}{\mathrm{E}(\epsilon_n'M_X\epsilon_n)} = \sigma_\epsilon^{-2}(\sigma_{u\epsilon}'\gamma_0+\sigma_\epsilon^2)\frac{1}{1-K/n}[\frac{1}{n}\mathrm{tr}(G_nP_X) - \frac{K}{n^2}\mathrm{tr}(G_n)]. \quad (4)$$

On the other hand, $Y_n = [S_ny_n, Y_{1n}] = X\Pi_0[\gamma_0, I_g] + [U_n\gamma_0+\epsilon_n, U_n]$ and $W_ny_n = G_nS_ny_n = G_nY_ne_1$. Let $\Upsilon_n = [U_n\gamma_0+\epsilon_n, U_n]$. It follows that $\mathrm{E}(Y_n'M_XY_n) = \mathrm{E}(\Upsilon_{ni}'\Upsilon_{ni})\mathrm{tr}(M_X) = \mathrm{E}(\Upsilon_{ni}'\Upsilon_{ni})(n-K)$ and $\mathrm{E}(Y_n'M_XW_ny_n) = \mathrm{E}(\Upsilon_{ni}'\Upsilon_{ni})e_1[\mathrm{tr}(G_n) - \mathrm{tr}(G_nP_X)]$. It follows that

$$\frac{1}{n}\mathrm{tr}(G_n) - e_1'[\mathrm{E}(Y_n'M_XY_n)]^{-1}\mathrm{E}(Y_n'M_XW_ny_n) = \frac{1}{1-K/n}[\frac{1}{n}\mathrm{tr}(G_nP_X) - \frac{K}{n^2}\mathrm{tr}(G_n)]. \quad (5)$$

Note that $\mathrm{tr}(G_n) = O(n)$ and $\mathrm{tr}(G_nP_X) = O(K)$ by Lemma B.1. Therefore, when $K/n \to 0$, the right hand sides of both (4) and (5) converge to zero as $n \to \infty$. Hence, the LIML estimator can be consistent when the number of instruments K increases at a slower rate than the sample size n.

The consistency of the LIML estimator can be formally established by showing that the LIML

[1] For simplicity, W_n and X_n are assumed to be nonstochastic. Otherwise, the results should be considered as conditional on W_n and X_n.

objective function (3) converges to a function that is uniquely maximized at θ_0, as in the proof of the following result. The regularity assumptions for the asymptotic analysis are summarized in Appendix A.

Proposition 1 *Under Assumptions 1-4, if $K/n \to 0$, the LIML estimator $\hat{\theta}_{liml} = \arg\max \ln L_n^c(\theta)$ is consistent.*

However, due to spatial correlation, the LIML is, in general, inconsistent when the number of instruments increases at the same rate as the sample size. This result can be seen by subtracting (5) from (4), which gives

$$\begin{aligned} & -\frac{1}{n}\mathrm{tr}(G_n) + e_1'[\mathrm{E}(Y_n' M_X Y_n)]^{-1}\mathrm{E}(Y_n' M_X W_n y_n) + \frac{\mathrm{E}(\epsilon_n' W_n y_n)}{\mathrm{E}(\epsilon_n' \epsilon_n)} - \frac{\mathrm{E}(\epsilon_n' M_X W_n y_n)}{\mathrm{E}(\epsilon_n' M_X \epsilon_n)} \\ = \; & \sigma_\epsilon^{-2}\sigma_{u\epsilon}'\gamma_0 \frac{1}{1-K/n}[\frac{1}{n}\mathrm{tr}(G_n P_X) - \frac{K}{n^2}\mathrm{tr}(G_n)]. \end{aligned} \tag{6}$$

In general[2], the right hand side of (6) does not vanish asymptotically when $K/n \to c$ for $0 < c < 1$. Thus, the LIML estimator may be inconsistent in this case.

3 Concluding Remarks

This note considers the LIML estimation of an SAR model with endogenous regressors in the presence of many instruments. The LIML estimator is consistent when the number of instruments increases at a slower rate relative to the sample size, but is inconsistent when the number of instruments increases at the same rate as the sample size due to spatial correlation.

For models with independent observations, it is well known that the LIML estimator can also be derived based on the least variance ratio principle (see, e.g., Christ, 1966; Davidson and MacKinnon, 1993). For the SAR model, the least variance ratio (LVR) estimator can be derived as follows.

From the reduced form equation, the ideal instruments for the SAR model (1) are given by $F_n = \mathrm{E}([W_n y_n, Y_{1n}]) = [G_n X_n \Pi_0 \gamma_0, X_n \Pi_0]$, which are not feasible as they involve unknown parameters. Note F_n can be presented as a linear combination of $[G_n X_n, X_n]$. When W_n is row-normalized and $|\lambda_0| < 1$, $G_n = W_n S_n^{-1} = \sum_{j=0}^{\infty} \lambda_0^j W_n^{j+1} = \sum_{j=0}^{p} \lambda_0^j W_n^{j+1} + \lambda_0^{p+1} W_n^{p+1} G_n$. It follows by Assumption 3 in Appendix A that $||G_n - \sum_{j=0}^{p} \lambda_0^j W_n^{j+1}||_\infty \leq \lambda_0^{p+1} ||W_n^{p+1} G_n||_\infty = o(1)$ as $p \to \infty$, where $||\cdot||_\infty$ is the row-sum matrix norm. Therefore, for $G_X^{(p)} = (W_n X_n, \cdots, W_n^{p+1} X_n)$, the ideal instruments F_n can be approximated by a linear combination of feasible instruments $Z_n = [G_X^{(p)}, X_n]$, with an

[2] An exception would be the case where the regressors Y_{1n} are exogenous so that $\sigma_{u\epsilon} = 0$.

approximation error vanishing in a geometric rate as $p \to \infty$. Suppose Z_n has K_p columns. Let $P_Z = Z_n(Z_n'Z_n)^{-1}Z_n'$ and $M_Z = I_n - P_Z$. The LVR estimator can be obtained by maximizing $\frac{\epsilon_n'(\theta)M_Z\epsilon_n(\theta)}{\epsilon_n'(\theta)\epsilon_n(\theta)}$. By comparing the objective functions, it is easy to see that the LVR estimator is not equivalent to the LIML estimator for the SAR model. This is analogous to the difference between the 2SLS and ML estimators for the SAR model (Lee, 2004, 2007). While the LVR and 2SLS estimators only use linear moment conditions of the disturbances, the LIML and ML estimators also exploit quadratic moment conditions based on the correlation structure of the reduced form disturbances. Furthermore, by a similar argument as in the previous section, it is easy to see that the LVR estimator is consistent if $K_p/n \to 0$ but inconsistent if $K_p/n \to c$ for $0 < c < 1$.

APPENDICES

To simplify notations, let $f_{1n} = \mathrm{E}(W_n y_n) = G_n X_n \Pi_0 \gamma_0$ and $F_n = \mathrm{E}([W_n y_n, Y_{1n}]) = [f_{1n}, X_n\Pi_0]$. Let $\bar{U}_n = U_n\gamma_0 + \epsilon_n$ and $\Upsilon_n = [\bar{U}_n, U_n]$. Let $\sigma_{\bar{u}}^2 = \mathrm{E}(\bar{U}_{ni}^2)$, $\sigma_{\bar{u}\epsilon} = \mathrm{E}(\bar{U}_{ni}\epsilon_{ni})$, $\sigma_{u\bar{u}} = \mathrm{E}(U_{ni}'\bar{U}_{ni})$, and $\Sigma_\Upsilon = \mathrm{E}(\Upsilon_{ni}'\Upsilon_{ni})$. Uniformly bounded in row (column) sums in absolute value of a sequence of square matrices $\{A_n\}$ will be abbreviated as UBR (UBC), and uniformly bounded in both row and column sums in absolute value as UB.

A Regularity Assumptions

Assumption 1 $\epsilon_{ni} \sim (0, \sigma_\epsilon^2)$ and $U_{ni} \sim (0, \Sigma_u)$ are i.i.d. across i, and $\mathrm{E}(U_{ni}'\epsilon_{ni}) = \sigma_{u\epsilon}$. $\mathrm{E}(|\epsilon_{ni}|^4)$, $\mathrm{E}(||U_{ni}||^4)$ and $\mathrm{E}(||U_{ni}\epsilon_{ni}||^2)$ are bounded, uniformly in n.

Assumption 2 The elements of X_n are uniformly bounded constants, X_n has a full column rank, and $\lim_{n\to\infty} \frac{1}{n}X_n'X_n$ exists and is nonsingular for all K.

Assumption 3 The sequences of matrices $\{W_n\}$ and $\{S_n^{-1}\}$ are UB. $\{S_n^{-1}(\lambda)\}$ is either UBR or UBC uniformly in λ in a compact parameter space Λ. The true λ_0 is in the interior of Λ.

Assumption 4 For all K, $\lim_{n\to\infty} \frac{1}{n} f_n' P_X f_n$ is finite, and $\lim_{n\to\infty} \frac{1}{n}F_n'F_n$ is a finite nonsingular matrix.

B Proofs

The following two lemmas are listed for easy reference. Their proofs can be found in Lee (2007) and Liu and Lee (2010).

Lemma B.1 *Suppose $\{A_n\}$ is a UB sequences of nonstochastic matrix. For $B_n = P_X A_n$, $\mathrm{tr}(A_n) = O(n)$, $\mathrm{tr}(B_n) = O(K)$, $\mathrm{tr}(B_n^2) = O(K)$, and $\sum_i (B_{n,ii})^2 = O(K)$.*

Lemma B.2 *Suppose ε_{1n} and ε_{2n} are vectors of i.i.d. innovations such that $\varepsilon_{1n,i} \sim (0, \sigma_1^2)$, $\varepsilon_{2n,i} \sim (0, \sigma_2^2)$ and $\mathrm{E}(\varepsilon_{1n,i}\varepsilon_{2n,i}) = \sigma_{12}$, $\{A_n\}$ and $\{B_n\}$ are UB sequences of nonstochastic matrices, and the elements of matrices C_n and D_n are uniformly bounded constants. Then* (i) $\frac{1}{\sqrt{n}} C_n' A_n \varepsilon_{1n} = O_p(1)$ *and* $\frac{1}{n}[\varepsilon_{1n} A_n \varepsilon_{2n} - \sigma_{12}\mathrm{tr}(A_n)] = o_p(1)$*;* (ii) $\frac{1}{n} C_n' P_X D_n = O(1)$, $\frac{1}{n} C_n' P_X A_n \varepsilon_{1n} = O_p(\sqrt{K/n})$, $\frac{1}{n}\varepsilon_{1n}' B_n' P_X A_n \varepsilon_{2n} = O_p(K/n)$, *and* $\frac{1}{\sqrt{n}}[\varepsilon_{1n}' B_n' P_X A_n \varepsilon_{2n} - \sigma_{12}\mathrm{tr}(B_n' P_X A_n)] = O_p(\sqrt{K/n})$.

Proof of Proposition 1. Let

$$\Psi_n(\theta) = c + \ln|S_n(\lambda)| - \frac{n}{2}\ln|(\lambda_0-\lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} e_1 e_1' + \Sigma_\Upsilon(\lambda)| + \frac{n}{2}\ln \frac{(\lambda_0-\lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} + \sigma_\epsilon^2(\theta)}{(\theta_0-\theta)' \frac{1}{n} F_n' F_n (\theta_0-\theta) + \sigma_\epsilon^2(\theta)},$$

where

$$\sigma_\epsilon^2(\theta) = (\theta_0-\theta)' \begin{bmatrix} \sigma_{\bar{u}}^2 \frac{1}{n}\mathrm{tr}(G_n' G_n) & * \\ \sigma_{u\bar{u}} \frac{1}{n}\mathrm{tr}(G_n) & \Sigma_u \end{bmatrix} (\theta_0-\theta) + 2[\sigma_{\bar{u}\epsilon}\frac{1}{n}\mathrm{tr}(G_n), \sigma_{u\epsilon}'](\theta_0-\theta) + \sigma_\epsilon^2,$$

and

$$\Sigma_\Upsilon(\lambda) = \begin{bmatrix} (\lambda_0-\lambda)^2 \frac{1}{n}\sigma_{\bar{u}}^2 \mathrm{tr}(G_n' G_n) + 2(\lambda_0-\lambda)\frac{1}{n}\sigma_{\bar{u}}^2 \mathrm{tr}(G_n) + \sigma_{\bar{u}}^2 & * \\ (\lambda_0-\lambda)\frac{1}{n}\mathrm{tr}(G_n)\sigma_{u\bar{u}} + \sigma_{u\bar{u}} & \Sigma_u \end{bmatrix}.$$

The consistency of the LIML estimator will follow from the uniform convergence of $\frac{1}{n}[\ln L_n^c(\theta) - \Psi_n(\theta)]$ to zero and the identification condition that, for any $\varepsilon > 0$, $\limsup_{n\to\infty} \max_{\theta \in \bar{N}(\theta_0,\varepsilon)} \frac{1}{n}[\Psi_n(\theta) - \Psi_n(\theta_0)] < 0$, where $\bar{N}(\theta_0, \varepsilon)$ is the complement of an open neighborhood of θ_0 of diameter ε (White 1994, Theorem 3.4). The proof is divided into the following two steps.

Step 1 (Uniform Convergence): Let $d_n(\theta) = F_n(\theta_0-\theta)$ and $e_n(\theta) = [G_n\bar{U}_n, U_n](\theta_0-\theta)$. We have $\epsilon_n(\theta) = d_n(\theta) + e_n(\theta) + \epsilon_n$. Hence, $\epsilon_n'(\theta)\epsilon_n(\theta) = d_n'(\theta)d_n(\theta) + 2l_{1n}(\theta) + q_{1n}(\theta)$, where $l_{1n}(\theta) = d_n'(\theta)e_n(\theta) + d_n'(\theta)\epsilon_n$ and $q_{1n}(\theta) = e_n'(\theta)e_n(\theta) + 2e_n'(\theta)\epsilon_n + \epsilon_n'\epsilon_n$. By Lemma B.2 (i), $\frac{1}{n}l_{1n}(\theta) = o_p(1)$ uniformly in θ and $\frac{1}{n}q_{1n}(\theta) = \sigma_\epsilon^2(\theta) + o_p(1)$ uniformly in θ. Hence, $\frac{1}{n}\epsilon_n'(\theta)\epsilon_n(\theta) = (\theta_0-\theta)'\frac{1}{n}F_n'F_n(\theta_0-\theta) + \sigma_\epsilon^2(\theta) + o_p(1)$ uniformly in θ. Similarly, $\epsilon_n'(\theta)P_X\epsilon_n(\theta) = d_n'(\theta)P_X d_n(\theta) + 2l_{2n}(\theta) + q_{2n}(\theta)$, where $l_{2n}(\theta) = d_n'(\theta)P_X e_n(\theta) + d_n'(\theta)P_X\epsilon_n$ and $q_{2n}(\theta) = e_n'(\theta)P_X e_n(\theta) + 2e_n'(\theta)P_X\epsilon_n + \epsilon_n' P_X \epsilon_n$. When $K/n \to 0$, by Lemma B.2 (ii), $\frac{1}{n}l_{2n}(\theta) = o_p(1)$ and $\frac{1}{n}q_{2n}(\theta) = o_p(1)$. Hence, $\frac{1}{n}\epsilon_n'(\theta)P_X\epsilon_n(\theta) = \frac{1}{n}d_n'(\theta)P_X d_n(\theta) + o_p(1)$. As $F_n = [f_{1n}, X_n\Pi_0]$, we have $\frac{1}{n}d_n'(\theta)M_X d_n(\theta) = \frac{1}{n}(\theta_0-\theta)'F_n' M_X F_n(\theta_0 -$

$\theta) = (\lambda_0 - \lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n}$. It follows that $\frac{1}{n}\epsilon_n'(\theta) M_X \epsilon_n(\theta) = (\lambda_0 - \lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} + \sigma_\epsilon^2(\theta) + o_p(1)$ uniformly in θ. It is straightforward to show that $\sigma_\epsilon^2(\theta)$ is nonzero for all θ by non-singularity of the covariance matrix of $[G_n \bar{U}_n, U_n, \epsilon_n]$. Hence, $\ln \frac{\epsilon_n'(\theta) M_X \epsilon_n(\theta)}{\epsilon_n'(\theta)\epsilon_n(\theta)} - \ln \frac{(\lambda_0-\lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} + \sigma_\epsilon^2(\theta)}{(\theta_0-\theta)' \frac{1}{n} F_n' F_n (\theta_0-\theta) + \sigma_\epsilon^2(\theta)} = o_p(1)$ uniformly in θ.

On the other hand, as $S_n(\lambda) y_n = (\lambda_0 - \lambda) f_{1n} + X_n \Pi_0 \gamma_0 + (\lambda_0 - \lambda) G_n \bar{U}_n + \bar{U}_n$, it follows by Lemma B.2 that, $\frac{1}{n} y_n' S_n'(\lambda) M_X S_n(\lambda) y_n = (\lambda_0 - \lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} + 2(\lambda_0 - \lambda)\frac{1}{n} f_{1n}' M_X [(\lambda_0 - \lambda) G_n \bar{U}_n + \bar{U}_n] + \frac{1}{n}[(\lambda_0 - \lambda) G_n \bar{U}_n + \bar{U}_n]' M_X [(\lambda_0 - \lambda) G_n \bar{U}_n + \bar{U}_n] = (\lambda_0 - \lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} + (\lambda_0 - \lambda)^2 \frac{1}{n} \sigma_{\bar{u}}^2 \mathrm{tr}(G_n' G_n) + 2(\lambda_0 - \lambda)\frac{1}{n}\sigma_{\bar{u}}^2 \mathrm{tr}(G_n) + \sigma_{\bar{u}}^2 + o_p(1)$. Similarly, $\frac{1}{n} y_n' S_n'(\lambda) M_X Y_{1n} = (\lambda_0 - \lambda)\frac{1}{n} f_{1n}' M_X U_n + \frac{1}{n}[(\lambda_0 - \lambda) G_n \bar{U}_n + \bar{U}_n]' M_X U_n = (\lambda_0 - \lambda)\frac{1}{n}\mathrm{tr}(G_n)\sigma_{u\bar{u}}' + \sigma_{u\bar{u}}' + o_p(1)$, and $\frac{1}{n} Y_{1n}' M_X Y_{1n} = \frac{1}{n} U_n' M_X U_n = \Sigma_u + o_p(1)$. Hence,

$$\frac{1}{n} Y_n'(\lambda) M_X Y_n(\lambda) = \frac{1}{n}\begin{bmatrix} y_n' S_n'(\lambda) M_X S_n(\lambda) y_n & * \\ Y_{1n}' M_X S_n(\lambda) y_n & Y_{1n}' M_X Y_{1n} \end{bmatrix} = (\lambda_0 - \lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} e_1 e_1' + \Sigma_\Upsilon(\lambda) + o_p(1),$$

uniformly in λ. It is straightforward to show that $|\Sigma_\Upsilon(\lambda)|$ nonzero for all λ. It follows that $\ln|\frac{1}{n} Y_n'(\lambda) M_X Y_n(\lambda)| - \ln|(\lambda_0 - \lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} e_1 e_1' + \Sigma_\Upsilon(\lambda)| = o_p(1)$ uniformly in λ. Hence, $\sup_\theta \frac{1}{n}|\ln L_n^c(\theta) - \Psi_n(\theta)| = o_p(1)$.

Step 2 (Identification Uniqueness): To prove the identification uniqueness, we consider the following log-likelihood function $\ln L_{pn}(\lambda) = c + \ln|S_n(\lambda)| - \frac{n}{2}\ln|\frac{1}{n}[S_n(\lambda) y_{pn}, U_n]'[S_n(\lambda) y_{pn}, U_n]|$ for the joint distribution of $[y_{pn}, U_n]$, where $y_{pn} = \lambda W_n y_{pn} + \bar{U}_n$. Denote $\Psi_{pn}(\lambda) = \mathrm{E}[\ln L_{pn}(\lambda)]$. It is apparent that $\Psi_{pn}(\lambda) = c + \ln|S_n(\lambda)| - \frac{n}{2}\ln|\Sigma_\Upsilon(\lambda)|$. At θ_0, $\Psi_{pn}(\theta_0) = c + \ln|S_n| - \frac{n}{2}\ln|\Sigma_\Upsilon|$. We have $\Psi_n(\theta) - \Psi_n(\theta_0) = \Delta_1 + \Delta_2 + \Delta_3$, where $\Delta_1 = \Psi_{pn}(\theta) - \Psi_{pn}(\theta_0)$, $\Delta_2 = \frac{n}{2}\ln \frac{(\lambda_0-\lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} + \sigma_\epsilon^2(\theta)}{(\theta_0-\theta)' \frac{1}{n} F_n' F_n(\theta_0-\theta) + \sigma_\epsilon^2(\theta)}$ and $\Delta_3 = -\frac{n}{2}\left[\ln|(\lambda_0 - \lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} e_1 e_1' + \Sigma_\Upsilon(\lambda)| - \ln|\Sigma_\Upsilon(\lambda)|\right]$. By Jensen's inequality, $\Psi_{pn}(\theta) \leq \Psi_{pn}(\theta_0)$, and hence $\Delta_1 \leq 0$. As $(\lambda_0 - \lambda)^2 \frac{1}{n} f_{1n}' M_X f_{1n} = (\theta_0 - \theta)' \frac{1}{n} F_n' M_X F_n (\theta_0 - \theta) \leq (\theta_0 - \theta)' \frac{1}{n} F_n' F_n (\theta_0 - \theta)$, we have $\Delta_2 \leq 0$. Therefore, by Assumption 4, the identification uniqueness condition holds. The desired result thus follows. ■

References

Anderson, T., Kunitomo, N. and Matsushita, Y. (2010). On the asymptotic optimality of the liml estimator with possibly many instruments, *Journal of Econometrics* **157**: 191–204.

Bekker, P. (1994). Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* **62**: 657–681.

Chao, J. and Swanson, N. (2005). Consistent estimation with a large number of weak instruments, *Econometrica* **73**: 1673–1692.

Christ, C. (1966). *Econometric Models and Methods*, Wiley, New York.

Davidson, R. and MacKinnon, J. (1993). *Estimation and Inference in Econometrics*, Oxford University Press, New York, Oxford.

Donald, S. G. and Newey, W. K. (2001). Choosing the number of instruments, *Econometrica* **69**: 1161–1191.

Hansen, C., Hausman, J. and Newey, W. (2008). Estimation with many instrumental variables, *Journal of Business and Economic Statistics* **26**: 398–422.

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.

Liu, X. and Lee, L. F. (2010). Gmm estimation of social interaction models with centrality, *Journal of Econometrics* **159**: 99–115.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem, *The Review of Economic Studies* **60**: 531–542.

van Hasselt, M. (2010). Many instruments asymptotic approximations under nonnormal error distributions, *Econometric Theory* **26**: 633–645.

White, H. (1994). *Estimation, Inference and Specification Analysis*, Cambridge Univsersity Press, New York.