UNIVERSITY *of York*

Discussion Papers in Economics

No. 15/24

QML Estimation of the Spatial Weight Matrix in the MR-SAR Model

Saruta Benjanuvatra and Peter Burridge

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

QML Estimation of the Spatial Weight Matrix in the MR-SAR Model†

Saruta Benjanuvatra and Peter Burridge*

Ministry of Commerce, Thailand, and

Department of Economics and Related Studies,

University of York, UK

September 28, 2015

Abstract

We investigate QML estimation of a parametric form for the spatial weight matrix, W, appearing in the mixed regressive, spatial autoregressive (MR-SAR) model and extend the identifiability, consistency, and asymptotic Normality results given by Lee (2004, 2007) to the case when W depends on an unknown parameter, γ, that is to be estimated from a single cross-section. Numerical experiments illustrate that the QML estimator works quite well in moderate sized samples, yielding well-behaved parameter estimates and t-statistics with approximately correct size in most cases. These findings should open the door to a much more flexible approach to the construction of spatial regression models. Finally, the QML estimator using two types of sub-models for the spatial weights is applied to the cross-sectional dataset used in Ertur and Koch (2007), to illustrate the utility of the approach.

Keywords: Spatial autoregressive model, estimated spatial weight matrix, quasi-maximum likelihood estimator, growth spillovers.

JEL Classification: C13, C15, C21, R15

† Saruta Benjanuvatra presented an earlier version of this paper at the VII World Congress of the Spatial Econometrics Association, Washington DC, USA, July 2013.

* Corresponding Author: peter.burridge@york.ac.uk

1 Introduction

An important issue in spatial econometrics is the specification of the spatial weight matrix, and it has received much attention, particularly in the past few years. Case et al. (1993) consider notions of geographic, economic or demographic distance. Numerous forms of spatial weight matrix have been introduced, as reviewed by Anselin (1988a) and Anselin and Bera (1998), and discussed more recently by Partridge et al (2012) and Pinkse and Slade(2010) among others. Cliff and Ord (1973, 1981) introduce weights that are a combination of distances and relative border length between units; other popular forms are Rook contiguity where the weights are zero unless the two regions share a common border, and Queen contiguity where the weight is positive if two regions share a common side or vertex, inverse distance, or n-nearest neighbours. Others are geostatistical in form, such as the Spherical Variogram, Gaussian Variogram, and Exponential Variogram, described by Getis and Aldstadt (2004).

It is widely asserted that an inappropriate choice of weights can lead to estimation problems of varying degrees of seriousness, with Paez et al. (2008) showing that errors in the weight matrix can lead to biased estimates. More generally, inefficiency of the estimator is to be expected (Cliff and Ord, 1973). However, proper specification of the weight matrix has been regarded as difficult and controversial (Bavaud, 1998). Practitioners sometimes choose a weight matrix based on empirical convenience that may not capture the dependence structure properly. In this paper we introduce a sub-model for the spatial weights and estimate a spatial weight matrix for the mixed regressive, spatial autoregressive (MR-SAR) model by maximum Gaussian likelihood (QML). The maximum likelihood estimator in spatial regression models is studied by Ord (1975), Anselin (1988a) and Anselin and Bera (1998). Ord (1975) also presents a computational scheme extended to the MR-SAR models. Asymptotic properties of the MLE and QMLE are developed by Lee (2004a) for the spatial autoregressive models with fixed sequences of weights. Our approach relies heavily on the development in Lee (2004a) and Lee (2002), and we establish the identifiability of the parameter defining the weights and the consistency as well as the asymptotic distribution of the QMLE under appropriate conditions that extend those given by Lee (2004a). Small sample properties of the estimator are studied in a Monte Carlo experiment.

The utility on a real spatial data set of our QML estimator using two forms of sub-models for the spatial weights that satisfy the identifiability, consistency and asymptotic normality conditions, is illustrated by application to the cross-sectional data of 91 countries used in Ertur and Koch (2007).

In the remainder of the paper, Section 2 describes the MR-SAR model, Section 3 contains our large-sample theoretical results for the QMLE in the form of three theorems, Section 4 describes the Monte Carlo, followed by the empirical example in Section 5, and some brief conclusions in Section 6. Proofs of the theorems are in the Appendices, together with some supporting Lemmas and miscellaneous notation to assist referees.

2 Mixed Regressive, Spatial Autoregressive Model

The first-order MR-SAR model (Ord, 1975 and Anselin, 1988a) is described as follows

$$Y_n = X_n\beta + \lambda W_n(\gamma)Y_n + \varepsilon_n \tag{1}$$

where Y_n is an $n \times 1$ vector of observations of the dependent variable, X_n is an $n \times k$ matrix of values of k exogenous explanatory variables with $[1, 1, ..., 1]'$ in the first column, β is a $k \times 1$ vector of parameters, ε_n is an $n \times 1$ vector of disturbances that are independently distributed with mean 0 and variance σ^2 and independent of X_n, λ is the spatial autoregressive parameter, and n is the total number of spatial units. $W_n(\gamma)$ is an $n \times n$ matrix of spatial weights that represent the impact of the output variable in location j on location i (Ord, 1975). The weight elements are standardised to have row-sums unity, and are thus specified as

$$w_{n,ij}(\gamma) = \frac{w^*_{n,ij}(\gamma)}{\sum_j w^*_{n,ij}(\gamma)}, (i \neq j), \qquad = 0, (i = j)$$

where $w^*_{n,ij}(\gamma) = f(\gamma, d_{ij})$ is a function of fixed non-negative distances, d_{ij}, between spatial units i and j, and γ is a positive scalar parameter. The subscript n indicates that each component of the model depends on n, which is necessary to allow for the total number of spatial units to increase to deliver the asymptotics.

The objective is to estimate $\theta = (\beta', \lambda, \gamma, \sigma^2)'$. Our approach follows Lee (2004a) and Lee (2002), and we extend his notation as follows. Let $S_n(\lambda, \gamma) = I_n - \lambda W_n(\gamma)$, so that after rearrangement equation (1) becomes

$$S_n(\lambda, \gamma)Y_n = X_n\beta + \varepsilon_n \tag{2}$$
$$Y_n = S_n^{-1}(\lambda, \gamma)(X_n\beta + \varepsilon_n)$$

where $S_n(\lambda, \gamma)Y_n$ is a spatially filtered dependent variable. Denote by $\theta_0 = (\beta_0', \lambda_0, \gamma_0, \sigma_0^2)'$ the vector of true parameter values. At the true values, we shall write $S_n = S_n(\lambda_0, \gamma_0)$ and $W_n = W_n(\gamma_0)$ for notational convenience. Under the further assumption that the disturbance is Normal, the log-likelihood function of equation (1) is given by

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \ln|det(S_n(\lambda, \gamma))| - \frac{1}{2\sigma^2}\varepsilon_n'(\delta)\varepsilon_n(\delta) \tag{3}$$

where $\varepsilon_n(\delta) = Y_n - X_n\beta - \lambda W_n(\gamma)Y_n$, with $\delta = (\beta', \lambda, \gamma)'$.

The quasi-maximum likelihood estimator is obtained by maximising (3) with respect to the parameters. It is useful to first concentrate out β and σ^2. Thus, for given λ and γ, the QMLE of β is

$$\hat{\beta}_n(\lambda, \gamma) = (X_n'X_n)^{-1}(X_n'Y_n - \lambda X_n'W_n(\gamma)Y_n) = (X_n'X_n)^{-1}X_n'S_n(\lambda, \gamma)Y_n. \tag{4}$$

Inserting this $\hat{\beta}_n(\lambda, \gamma)$ into the foc for σ_n^2, gives the QMLE of σ^2 as

$$\begin{aligned}\hat{\sigma}_n^2(\lambda,\gamma) &= \frac{1}{n}[(S_n(\lambda,\gamma)Y_n - X_n\hat{\beta}(\lambda,\gamma))'(S_n(\lambda,\gamma)Y_n - X_n\hat{\beta}(\lambda,\gamma)] \\ &= \frac{1}{n}[Y_n'S_n'(\lambda,\gamma)M_nS_n(\lambda,\gamma)Y_n]\end{aligned} \tag{5}$$

where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. Inserting (4) and (5) back into the log-likelihood delivers the concentrated log-likelihood function of λ and γ:

$$\ln L_n(\lambda,\gamma) = -\frac{n}{2}(\ln(2\pi)+1) + \ln|det(S_n(\lambda,\gamma))| - \frac{n}{2}\ln\hat{\sigma}_n^2(\lambda,\gamma). \tag{6}$$

To obtain the QMLEs $\hat{\lambda}_n$ and $\hat{\gamma}_n$, maximise (6) with respect to λ and γ, which in turn yields the QMLEs of β and σ^2 as $\hat{\beta}_n(\hat{\lambda}_n, \hat{\gamma}_n)$ and $\hat{\sigma}_n^2(\hat{\lambda}_n, \hat{\gamma}_n)$.

3 Properties of the QMLE

Before stating our three theorems, we list the assumptions on which these results rely. The assumptions mirror very closely those introduced by Lee (op. cit.) with extensions to accommodate the dependence of the $w_{n,ij}$ sequences on the parameter, γ.

Assumption 1 *the elements of ε_n, that is $\varepsilon_{n,1}, \ldots, \varepsilon_{n,n}$ are independently and identically distributed with mean 0 and finite variance σ^2 for every n. Some moment of ε_n higher than the fourth exists; the third and fourth moments of ε_n are denoted by μ_3 and μ_4.*

Assumption 2 *Let $\Theta = \Lambda \otimes \Gamma$ be the compact and continuous parameter space in which the concentrated log-likelihood function is log-concave. The true values of λ and γ denoted by λ_0 and γ_0 respectively, are in the interior of Θ.*

Assumption 3 *The elements $x_{n,ij}$ of X_n for $i, j = 1, \ldots, n$, are uniformly bounded constants for all n. The $\lim_{n\to\infty} \frac{X_n'X_n}{n}$ is finite and nonsingular.*

Assumption 4 *The sequence, h_n, is such that $\frac{h_n}{n} \to 0$ as $n \to \infty$, where n is the total number of spatial units.*

Assumption 5 *The distance d_{ij} between spatial units i and j is a bounded nonnegative constant for all n, and γ is bounded away from zero.*

Assumption 6 *The elements $w_{n,ij}(\gamma)$ of $W_n(\gamma)$ are $O(\frac{1}{h_n})$ uniformly in all i and j, where $\{h_n\}$ can be bounded or divergent. There exists an open neighbourhood $\eta_n(\gamma_0)$ of γ_0 such that $w_{n,ij}(\gamma) = \frac{f(\gamma,d_{ij})}{\sum_j f(\gamma,d_{ij})}$ for $i \neq j$ is continuous in $\gamma \in \eta_n(\gamma_0)$ uniformly in n. The first-, second-, and third-order derivatives of $W_n(\gamma)$ with respect to γ are uniformly bounded and continuous on $\eta_n(\gamma_0)$.*

Assumption 7 *The matrix $S_n = I_n - \lambda_0 W_n$ is nonsingular on $\Lambda \otimes \Gamma$, where $0 < |\lambda_0| < 1$.*

Assumption 8 *The sequences $\{W_n\}$ and $\{S_n^{-1}\}$ are uniformly bounded in both row and column sums.*

Assumption 9 *$\{S_n^{-1}(\lambda, \gamma)\}$ and $\{W_n(\gamma)\}$ are uniformly bounded in either row or column sums, uniformly in λ and γ in $\Lambda \otimes \Gamma$. The true λ_0 and γ_0 are in the interior of $\Lambda \otimes \Gamma$.*

Assumption 1 is completely standard and is required for the application of the CLT of Kelejian and Prucha (2001). Assumption 2 imposes a restriction on the parameter space. The compactness of the parameter space is needed because we work with the concentrated log-likelihood function, which is nonlinear in λ and γ. It is also one of the two sufficient conditions to assure that the maximum of the limit of the log-likelihood is the limit of the maximum likelihood estimator, of which the second condition is that the convergence is uniform (Amemiya, 1985). Note that we do not need to impose any restriction on the parameter space for β and σ^2 as QML estimates for β and σ^2 can be obtained from (4) and (5), and their identifiable uniqueness follows from that of λ_0 and γ_0.

Assumption 3 ensures that there is no multicollinearity among the regressors and Lee (2004a) shows that this implies that $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$ and $(I_n - M_n)$ are uniformly bounded in both row and column sums. Assumptions 5 and 6 provide the characteristics of the spatial weight matrix and the functional form of its elements.

Assumption 7 is sufficient to ensure that S_n is nonsingular such that (1) has an equilibruim with the equilibrium vector $Y_n = S_n^{-1}(X_n\beta_0 + \varepsilon_n)$, the mean $S_n^{-1}X_n\beta_0$ and the variance $\sigma_0^2 S_n^{-1}S_n^{-1\prime}$, where σ_0^2 is the true variance of ε_n. Assumption 8 assures that the degree of spatial correlation (Kelejian and Prucha, 1999), which is captured in S_n^{-1}, is limited. The uniform boundedness of S_n^{-1} at (λ_0, γ_0), and of W_n at γ_0 implies that $S_n^{-1}(\lambda, \gamma)$ and $W_n(\gamma)$ are uniformly bounded in both row and column sums, uniformly in the neighbourhood of λ_0 and γ_0. Finally, as our weight matrix is nonnegative and row-normalised, Assumption 9 implies that $S_n^{-1}(\lambda, \gamma)$ is uniformly bounded in row sums uniformly in λ and γ in $\Lambda \otimes \Gamma$ where Λ is a closed subset in $(-1, 1)$ (Lee 2003, Lemma 1).

Together with a rank condition to be introduced in a moment, these assumptions will enable us to establish the identifiability of the parameters and the consistency of the QML estimator. At the true values, $S_n^{-1} = (I_n - \lambda_0 W_n)^{-1} = I_n + \lambda_0 G_n$, where $G_n = W_n S_n^{-1}$ (Lee, 2004a). Then, equation (2) can be rewritten as

$$Y_n = (I_n + \lambda_0 G_n)(X_n\beta_0 + \varepsilon_n) = X_n\beta_0 + \lambda_0 G_n X_n \beta_0 + S_n^{-1}\varepsilon_n. \tag{7}$$

Let $Q_n(\lambda, \gamma) = max_{\beta,\sigma^2} E[\ln L_n(\theta)]$. To prove that the QML estimator $\hat{\theta}_n$ is consistent, we need to show that the identifiable uniqueness condition holds and that $\frac{1}{n}\ln L_n(\lambda, \gamma) - \frac{1}{n}Q_n(\lambda, \gamma)$ converges to zero in probability uniformly on the parameter space (White 1996, Theorem 3.4). Formally, $\frac{1}{n}\ln L_n(\lambda, \gamma)$ converges in probability uniformly to $\frac{1}{n}Q_n(\lambda, \gamma)$ if $sup_{(\lambda,\gamma)\in\Lambda\otimes\Gamma}|\frac{1}{n}\ln L_n(\lambda, \gamma) - \frac{1}{n}Q_n(\lambda, \gamma)| = o_p(1)$. An intuition behind this is that the log-likelihood function will be close to the expected log-likelihood function, so we may expect the QML estimator to be close to the maximum of the expected log-likelihood as well.

As already mentioned, the second sufficient condition for the maximum of the limit to be the limit of the maximum is that the convergence is uniform. It ensures that the maximum is close to the true value for all λ and γ, that is, $\frac{1}{n}\ln L_n(\lambda,\gamma)$ will be uniformly close to $\frac{1}{n}Q_n(\lambda,\gamma)$. Uniform convergence also implies that if $\ln L_n(\lambda,\gamma)$ is continuous on the parameter space, then so is the limit function $Q_n(\lambda,\gamma)$.

Now let $Z_n = \frac{\partial W_n(\gamma_0)}{\partial \gamma}$ denote the first-order derivative of the W-matrix evaluated at γ_0, and write $T_n = Z_n S_n^{-1}$; we make the following additional assumption:

Assumption 10 *The following limits exist and are nonsingular.*

$$\lim_{n\to\infty}\frac{1}{n}(X_n, G_nX_n\beta_0)'(X_n, G_nX_n\beta_0), \qquad \lim_{n\to\infty}\frac{1}{n}(X_n, T_nX_n\beta_0)'(X_n, T_nX_n\beta_0) \text{ and}$$
$$\lim_{n\to\infty}\frac{1}{n}(X_n, G_nX_n\beta_0)'(X_n, T_nX_n\beta_0).$$

This assumption ensures that $G_nX_n\beta_0$ in (7) and $T_nX_n\beta_0$ are not asymptotically multicollinear with X_n. It implies that $\lim_{n\to\infty}\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ and $\lim_{n\to\infty}\frac{1}{n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0)$ are positive, and $\lim_{n\to\infty}\frac{1}{n}(G_nX_n\beta_0)'M_n(T_nX_n\beta_0)$ is not zero. Note that the condition that $\lim_{n\to\infty}\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ exists and is positive is a sufficient condition for identification of θ_0

Maximise $E[\ln L_n(\theta)]$ with respect to β and σ^2 and, as in Lee (2004a), we get the following solutions

$$\beta_n^*(\lambda,\gamma) = (X_n'X_n)^{-1}X_n'S_n(\lambda,\gamma)S_n^{-1}X_n\beta_0 \tag{8}$$

and

$$\sigma_n^{2*}(\lambda,\gamma) = \frac{1}{n}[(\lambda_0-\lambda)^2(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \sigma_0^2 tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})] \tag{9}$$

Substitute (8) and (9) into the log-likelihood, then we get

$$Q_n(\lambda,\gamma) = -\frac{n}{2}(\ln(2\pi)+1) + \ln|det(S_n(\lambda,\gamma))| - \frac{n}{2}\ln\sigma_n^{2*}(\lambda,\gamma) \tag{10}$$

and it is concave and continuous in $\theta \in \Theta$. The following two theorems are proved in the Appendix:

Theorem 1 *Under Assumptions 1 - 10, θ_0 is identifiably unique.*

This theorem guarantees that no other value or sequence of values of θ yields $Q_n(\lambda,\gamma)$ arbitrarily close to Q_n when $n\to\infty$ (White 1996, Definition 3.3). Therefore, in the limit, $Q_n(\lambda,\gamma)$ is uniquely maximised at θ_0.

Theorem 2 *Under Assumptions 1 - 10, $\hat{\theta}_n$ is a consistent estimator of θ_0.*

Now we turn to the question of asymptotic normality of the QML estimator $\hat{\theta}_n$. Formally, we show that a consistent root of $\frac{\partial \ln L_n(\hat{\theta}_n)}{\partial\theta} = 0$ at θ_0 is asymptotically normal.

The first-order derivatives of the log-likelihood function at θ_0 are:

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \beta} = \frac{1}{\sigma_0^2\sqrt{n}}X_n'\varepsilon_n \tag{11}$$

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \lambda} = \frac{1}{\sigma_0^2\sqrt{n}}[(G_nX_n\beta_0)'\varepsilon_n + \varepsilon_n'G_n\varepsilon_n - \sigma_0^2 tr(G_n)] \tag{12}$$

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \gamma} = \frac{\lambda_0}{\sigma_0^2\sqrt{n}}[(T_nX_n\beta_0)'\varepsilon_n + \varepsilon_n'T_n\varepsilon_n - \sigma_0^2 tr(T_n)] \tag{13}$$

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \sigma^2} = \frac{1}{2\sigma_0^4\sqrt{n}}(\varepsilon_n'\varepsilon_n - n\sigma_0^2). \tag{14}$$

These derivatives contain linear and quadratic forms in ε_n. Now, as the elements of X_n are bounded and the matrices G_n and T_n are uniformly bounded in row sums, the elements of $G_nX_n\beta_0$ and $T_nX_n\beta_0$ for all n are uniformly bounded by Lemma A.6 in Lee (2004b). Thus, if $\{h_n\}$ that appears in Assumptions 4 and 6 is a bounded sequence, we can use the central limit theorem introduced in Kelejian and Prucha (2001) to derive the asymptotic distribution of the estimator. If $\{h_n\}$ is divergent, then we can apply the Kolmogorov central limit theorem to $\frac{\sqrt{n}}{n}\frac{\partial \ln L_n(\theta_0)}{\partial \theta}$ (Lee, 2004a).

With $\theta_0 = (\beta_0', \lambda_0, \gamma_0, \sigma_0^2)'$, we obtain

$$Var(\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta}) = \begin{cases} -E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) & \text{if } \varepsilon_{i,n} \text{ is normally distributed} \\ -E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) + \Omega_{\theta,n} & \text{if } \varepsilon_{i,n} \text{ is only i.i.d.} \end{cases}$$

where

$$\Omega_{\theta,n} = E(\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta}\cdot\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta'}) + E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}). \tag{15}$$

Writing $P_n = G_nX_n\beta_0$, $R_n = T_nX_n\beta_0$, $G_n^s = G_n + G_n'$, and $T_n^s = T_n + T_n'$, we have

$$-E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) = \tag{16}$$

$$\begin{pmatrix} \frac{1}{\sigma_0^2 n}X_n'X_n & \frac{1}{\sigma_0^2 n}X_n'P_n & \frac{\lambda_0}{\sigma_0^2 n}X_n'R_n & 0 \\ \frac{1}{\sigma_0^2 n}P_n'X_n & \frac{1}{\sigma_0^2 n}P_n'P_n + \frac{1}{n}tr(G_n^sG_n) & \frac{\lambda_0}{\sigma_0^2 n}[P_n'R_n + \sigma_0^2 tr(G_n^sT_n)] & \frac{1}{\sigma_0^2 n}tr(G_n) \\ \frac{\lambda_0}{\sigma_0^2 n}R_n'X_n & \frac{\lambda_0}{\sigma_0^2 n}[R_n'P_n + \sigma_0^2 tr(T_n^sG_n)] & \frac{\lambda_0^2}{\sigma_0^2 n}[R_n'R_n + \sigma_0^2 tr(T_n^sT_n)] & \frac{\lambda_0}{\sigma_0^2 n}tr(T_n) \\ 0 & \frac{1}{\sigma_0^2 n}tr(G_n) & \frac{\lambda_0}{\sigma_0^2 n}tr(T_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

Next, writing $l_n = [1, 1, ...1]'$ for the $n-$ vector of 1's, $G_{n,i.}$ for the i^{th} row of G_n and similarly $T_{n,i.}$ and $x_{n,i.}$ while $G_{n,ij}$ and $T_{n,ij}$ are the (i, j) entries of G_n and T_n, respectively and μ_3 and μ_4 are the third and fourth moments of ε_n, we have

$$\Omega_{\theta,n} = \tag{17}$$

$$\begin{pmatrix} 0 & * & * & * \\ \frac{\mu_3}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}x_{n,i.} & \begin{array}{c}\frac{2\mu_3}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}G_{n,i.}X_n\beta_0 \\ +\frac{(\mu_4-3\sigma_0^4)}{\sigma_0^4 n}\sum_{i=1}^n G_{n,ii}^2\end{array} & * & * \\ \frac{\lambda_0\mu_3}{\sigma_0^4 n}\sum_{i=1}^n T_{n,ii}x_{n,i.} & \begin{array}{c}\frac{\lambda_0}{\sigma_0^4 n}[(\mu_4-3\sigma_0^4)\sum_{i=1}^n T_{n,ii}G_{n,ii} \\ +\mu_3\sum_{i=1}^n G_{n,ii}T_{n,i.}X_n\beta_0 \\ +\mu_3\sum_{i=1}^n T_{n,ii}G_{n,i.}X_n\beta_0]\end{array} & \begin{array}{c}\frac{\lambda_0^2}{\sigma_0^4 n}[(\mu_4-3\sigma_0^4)\sum_{i=1}^n T_{n,ii}^2 \\ +2\mu_3\sum_{i=1}^n T_{n,ii}T_{n,i.}X_n\beta_0]\end{array} & * \\ \frac{\mu_3}{2\sigma_0^6 n}l_n'X_n & \begin{array}{c}\frac{1}{2\sigma_0^6 n}[\mu_3 l_n' G_n X_n\beta_0 \\ +(\mu_4-3\sigma_0^4)tr(G_n)]\end{array} & \begin{array}{c}\frac{\lambda_0}{2\sigma_0^6 n}[\mu_3 l_n' T_n X_n\beta_0 \\ +(\mu_4-3\sigma_0^4)tr(T_n)]\end{array} & \frac{(\mu_4-3\sigma_0^4)}{4\sigma_0^8} \end{pmatrix}$$

The matrix $\Omega_{\theta,n}$ above is symmetric about the leading diagonal, such terms being indicated by the asterisks (*).

If the $\varepsilon_{i,n}$ are i.i.d.,

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} \xrightarrow{D} N[0, -E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) + \Omega_{\theta,n}] \tag{18}$$

and, consequently, writing $\Sigma_\theta = -\lim_{n\to\infty} E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})$, which is nonsingular by Assumption 10, we obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_{\theta,n}\Sigma_\theta^{-1}]. \tag{19}$$

For $\varepsilon_{i,n}$ normally distributed, $\Omega_{\theta,n} \equiv 0$ and we get

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} \xrightarrow{D} N[0, \Sigma_\theta].$$

and, hence,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1}]. \tag{20}$$

Given the above development, we state the following theorem:

Theorem 3 *Under Assumptions 1 - 10, the QML estimator $\hat{\theta}_n$ satisfies*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1}] \tag{21}$$

where $\Omega_\theta = \lim_{n\to\infty}\Omega_{\theta,n}$ and $\Sigma_\theta = -\lim_{n\to\infty} E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})$ exist. If the $\varepsilon_{i,n}$ are normally distributed, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1}]. \tag{22}$$

Results obtained from Theorems 1 - 3 are valid for both bounded and divergent $\{h_n\}$. Note that when $\{h_n\}$ is divergent, the matrices in (16) and (17) can be simplified to

$$\Sigma_\theta = -\lim_{n\to\infty} E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) = \begin{pmatrix} \frac{1}{\sigma_0^2 n}X_n'X_n & \frac{1}{\sigma_0^2 n}X_n'P_n & \frac{\lambda_0}{\sigma_0^2 n}X_n'R_n & 0 \\ \frac{1}{\sigma_0^2 n}P_n'X_n & \frac{1}{\sigma_0^2 n}P_n'P_n & \frac{\lambda_0}{\sigma_0^2 n}P_n'R_n & 0 \\ \frac{\lambda_0}{\sigma_0^2 n}R_n'X_n & \frac{\lambda_0}{\sigma_0^2 n}R_n'P_n & \frac{\lambda_0^2}{\sigma_0^2 n}R_n'R_n & 0 \\ 0 & 0 & 0 & \frac{1}{2\sigma_0^4} \end{pmatrix}$$

and

$$\Omega_\theta = \lim_{n\to\infty}\Omega_{\theta,n} = \begin{pmatrix} 0 & 0 & 0 & \frac{\mu_3}{2\sigma_0^6 n}X_n'l_n \\ 0 & 0 & 0 & \frac{\mu_3}{2\sigma_0^6 n}P_n'l_n \\ 0 & 0 & 0 & \frac{\lambda_0\mu_3}{2\sigma_0^6 n}R_n'l_n \\ \frac{\mu_3}{2\sigma_0^6 n}l_n'X_n & \frac{\mu_3}{2\sigma_0^6 n}l_n'P_n & \frac{\lambda_0\mu_3}{2\sigma_0^6 n}l_n'R_n & \frac{(\mu_4-3\sigma_0^4)}{4\sigma_0^8} \end{pmatrix}.$$

This is because when $\{h_n\}$ is divergent, $G_{n,ij}$ and $T_{n,ij}$ are $O(\frac{1}{h_n})$ and, consequently, $\lim_{n\to\infty}\frac{1}{n}tr(G_n)$ and $\lim_{n\to\infty}\frac{1}{n}tr(T_n)$ become zero. Then the QMLE $\hat{\lambda}_n$ and $\hat{\gamma}_n$ become asymptotically independent of $\hat{\sigma}_n^2$, whereas they are asymptotically dependent on $\hat{\sigma}_n^2$ when $\{h_n\}$ is bounded because $\lim_{n\to\infty}\frac{1}{n}tr(G_n)$ and $\lim_{n\to\infty}\frac{1}{n}tr(T_n)$ may not be zero.

4 Monte Carlo Results

4.1 Experiment Design

We investigate small sample properties of our estimator using two functional forms of the sub-model for the spatial weights. The elements of the row-standardised weight matrix $W_n(\gamma)$ are defined as

$$w_{n,ij}(\gamma) = \frac{w_{n,ij}^*(\gamma)}{\sum_j w_{n,ij}^*(\gamma)}$$

with

$$w_{n,ij}^*(\gamma) = \begin{cases} 0 & \text{if } i=j \\ e^{-\gamma d_{ij}} & \text{if } i\neq j \end{cases} \text{(exponential) or } w_{n,ij}^*(\gamma) = \begin{cases} 0 & \text{if } i=j \\ d_{ij}^{-\gamma} & \text{if } i\neq j \end{cases} \text{(inverse power)} \quad (23)$$

where γ is a positive scalar parameter to be estimated and d_{ij} is a fixed nonnegative distance between spatial units i and j.

We simulate the distributions of the MLE and associated t-statistics in samples of size $n = 100$, or $n = 400$, in each case laid out as a square grid, using 4096 replications. Parameter values trialled are $(\lambda, \gamma) \in (0.2, 0.4, 0.6, 0.8) \times (1.5, 2.0, 2.5, 3.0)$ with the regressors, $\mathbf{X}_n$, consisting of 3 columns with associated coefficients, $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$. For each replication the first column of $\mathbf{X}_n$ is a vector of ones and the other 2 columns are independent draws from the standard $n-$ variate Normal distribution, while the independent Normal vector of disturbances, has variance, $\sigma_n^2 = 0.25$. The qualitative results are not sensitive to this value.

4.2 Estimates of λ and γ

Tables 1 and 2 contain results for the exponential weight specification, and Tables 3 and 4 those for the inverse power specification (see 23). Table 1 for $n = 100$ shows the empirical significance levels of t-tests at nominal 5% and 10% of the (true) null hypotheses, $\lambda = \lambda_0$ and $\gamma = \gamma_0$ constructed using standard errors estimated from the numerical Hessian evaluated at the MLE. Also reported are the medians of the empirical distributions of the MLEs of λ and γ. The rows labelled, "λ", or "γ" in the "unknown" column illustrate behaviour of the MLE when the other parameter is treated (unrealistically) as known and equal to the true value. The means and mean square errors are reported for the estimated λ in all cases but omitted for the estimated γ in those cases where they are inflated by a few very extreme outcomes, and thus are very inaccurately estimated. In an effort to reduce the incidence of wild γ estimates, a number of different starting values were tried for every numerical search; this revealed the presence of multiple local maxima of the likelihood, and selecting the best solution for which $\hat{\lambda}$ was less than 1 in magnitude led to the results in the tables.

λ_0, γ_0	unknown	λ5%	λ10%	γ5%	γ10%	$\bar{\lambda}$	$\bar{\gamma}$	λ_{mse}	γ_{mse}	λ_{med}	γ_{med}
0.2, 1.5	λ, γ	**.08**	**.11**	**.12**	**.14**	**.16**	*	**.04**	*	**.17**	**2.04**
	λ	.05	.10	-	-	.17	-	.02	-	.17	-
	γ	-	-	.18	.21	-	3.56	-	*	-	1.55
0.4, 1.5	λ, γ	**.07**	**.11**	**.05**	**.07**	**.35**	*	**.02**	*	**.35**	**1.87**
	λ	.06	.11	-	-	.36	-	.02	-	.37	-
	γ	-	-	.10	.12	-	1.77	-	*	-	1.49
0.6, 1.5	λ, γ	**.10**	**.15**	**.03**	**.05**	**.54**	*	**.02**	*	**.55**	**1.74**
	λ	.06	.11	-	-	.56	-	.01	-	.57	-
	γ	-	-	.07	.10	-	1.54	-	.47	-	1.50
0.8, 1.5	λ, γ	**.10**	**.16**	**.02**	**.04**	**.74**	**2.61**	**.01**	*	**.75**	**1.67**
	λ	.06	.12	-	-	.76	-	.01	-	.77	-
	γ	-	-	.06	.10	-	1.52	-	.09	-	1.49
0.2, 2.0	λ, γ	**.06**	**.08**	**.15**	**.17**	**.19**	*	**.03**	*	**.19**	**2.47**
	λ	.05	.10	-	-	.18	-	.01	-	.18	-
	γ	-	-	.17	.19	-	5.12	-	*	-	2.07
0.4, 2.0	λ, γ	**.05**	**.09**	**.08**	**.10**	**.37**	*	**.02**	*	**.37**	**2.40**
	λ	.06	.11	-	-	.37	-	.01	-	.38	-
	γ	-	-	.09	.12	-	2.75	-	*	-	2.01
0.6, 2.0	λ, γ	**.07**	**.12**	**.04**	**.07**	**.56**	*	**.01**	*	**.57**	**2.26**
	λ	.06	.11	-	-	.57	-	.01	-	.58	-
	γ	-	-	.07	.10	-	2.19	-	2.08	-	2.01
0.8, 2.0	λ, γ	**.09**	**.14**	**.03**	**.05**	**.76**	**2.99**	**.01**	*	**.77**	**2.19**
	λ	.06	.11	-	-	.77	-	.005	-	.78	-
	γ	-	-	.06	.10	-	2.05	-	.18	-	2.00

Table 1A

Negative Exponential Weights n=100

λ, γ	unknown	$\lambda 5\%$	$\lambda 10\%$	$\gamma 5\%$	$\gamma 10\%$	$\bar{\lambda}$	$\bar{\gamma}$	λ_{mse}	γ_{mse}	λ_{med}	γ_{med}
0.2, 2.5	λ, γ	**.05**	**.07**	**.17**	**.19**	**.20**	*	**.03**	*	**.20**	**2.88**
	λ	.05	.10	-	-	.18	-	.01	-	.18	-
	γ	-	-	.16	.18	-	*	-	*	-	2.63
0.4, 2.5	λ, γ	**.04**	**.07**	**.09**	**.11**	**.39**	*	**.01**	*	**.38**	**2.91**
	λ	.06	.11	-	-	.38	-	.01	-	.38	-
	γ	-	-	.09	.12	-	3.91	-	*	-	2.53
0.6, 2.5	λ, γ	**.06**	**.10**	**.06**	**.08**	**.57**	*	**.01**	*	**.59**	**2.79**
	λ	.06	.11	-	-	.58	-	.01	-	.58	-
	γ	-	-	.07	.10	-	2.92	-	*	-	2.51
0.8, 2.5	λ, γ	**.08**	**.13**	**.04**	**.06**	**.77**	**3.89**	**.005**	*	**.78**	**2.72**
	λ	.06	.11	-	-	.79	-	.003	-	.79	-
	γ	-	-	.06	.09	-	2.63	-	.64	-	2.51
0.2, 3.0	λ, γ	**.04**	**.06**	**.17**	**.20**	**.20**	*	**.02**	*	**.20**	**3.40**
	λ	.05	.10	-	-	.18	-	.01	-	.19	-
	γ	-	-	.16	.18	-	*	-	*	-	3.15
0.4, 3.0	λ, γ	**.03**	**.06**	**.10**	**.13**	**.39**	*	**.01**	*	**.39**	**3.46**
	λ	.06	.11	-	-	.38	-	.01	-	.38	-
	γ	-	-	.10	.13	-	*	-	*	-	3.06
0.6, 3.0	λ, γ	**.04**	**.09**	**.08**	**.10**	**.58**	*	**.01**	*	**.58**	**3.34**
	λ	.06	.11	-	-	.58	-	.005	-	.59	-
	γ	-	-	.08	.11	-	3.84	-	*	-	3.03
0.8, 3.0	λ, γ	**.07**	**.12**	**.05**	**.07**	**.78**	*	**.004**	*	**.78**	**3.26**
	λ	.06	.11	-	-	.78	-	.003	-	.79	-
	γ	-	-	.07	.10	-	3.31	-	3.33	-	3.01

Table 1B

Negative Exponential Weights n=100

The salient features of Tables 1 A and B are:

(i) When λ is small, say 0.2 or 0.4, and unknown, the shape parameter, γ, is poorly estimated, even though the median is not far from the true value. Conversely, if λ is known then $\hat{\gamma}$ is almost median-unbiased, even for small λ values.

(ii) The sampling distribution of $\hat{\gamma}$ is skewed to the right, especially for smaller λ values; the skew is much reduced when λ is known.

(iii) The empirical sizes of $t-$ tests reflect the skewed distribution of $\hat{\gamma}$ in that the over-sizing at nominal 5%, where present, is much worse than that at nominal 10%.

(iv) When γ is known (so $\mathbf{W}$ is known) the empirical sizes of the $t-$ tests for λ are correct, while such tests may be either liberal or conservative otherwise.

(v) Broadly speaking, accommodating uncertainty about γ by estimating it from the sample doubles the mean square error of the estimate of λ.

(vi) Similarly, removing uncertainty about λ dramatically reduces the MSE of $\hat{\gamma}$; however, in only a few cases, notably when $\lambda = 0.8$, is the MSE likely to be accurately estimated here.

Comparing with Tables 2A and B for $n = 400$ we immediately see the improvement due to the increased sample size, though the essential pattern remains. Bearing in mind that in applications of this approach, it is the results in the bold rows that are most relevant, there is certainly some encouragement to be found from Tables 2A,B; here we find that with the exception of cases in which $\lambda = 0.2$ (quite weak spatial dependence) the $t-$ tests have approximately correct empirical levels, and the estimators are nearly median-unbiased.

λ,γ	unknown	λ5%	λ10%	γ5%	γ10%	$\bar{\lambda}$	$\bar{\gamma}$	λ_{mse}	γ_{mse}	λ_{med}	γ_{med}
0.2, 1.5	λ,γ	**.06**	**.10**	**.10**	**.12**	**.20**	*	**.01**	*	**.19**	**1.68**
	λ	.05	.11	-	-	.19	-	.005	-	.19	-
	γ	-	-	.10	.12	-	1.74	-	*	-	1.50
0.4, 1.5	λ,γ	**.08**	**.13**	**.05**	**.07**	**.39**	*	**.01**	*	**.39**	**1.59**
	λ	.06	.11	-	-	.39	-	.004	-	.39	-
	γ	-	-	.05	.09	-	1.53	-	.13	-	1.50
0.6, 1.5	λ,γ	**.07**	**.12**	**.04**	**.07**	**.58**	**1.65**	**.005**	*	**.59**	**1.56**
	λ	.06	.11	-	-	.59	-	.002	-	.59	-
	γ	-	-	.05	.10	-	1.51	-	.04	-	1.50
0.8, 1.5	λ,γ	**.06**	**.12**	**.04**	**.09**	**.78**	**1.56**	**.002**	**.04**	**.79**	**1.54**
	λ	.06	.11	-	-	.79	-	.001	-	.79	-
	γ	-	-	.05	.10	-	1.50	-	.02	-	1.50
0.2, 2.0	λ,γ	**.04**	**.08**	**.10**	**.13**	**.21**	*	**.008**	*	**.20**	**2.17**
	λ	.05	.11	-	-	.19	-	.003	-	.20	-
	γ	-	-	.08	.11	-	2.63	-	*	-	2.01
0.4, 2.0	λ,γ	**.07**	**.12**	**.05**	**.08**	**.39**	**4.0**	**.005**	*	**.39**	**2.10**
	λ	.05	.11	-	-	.39	-	.003	-	.40	-
	γ	-	-	.05	.08	-	2.07	-	.24	-	2.00
0.6, 2.0	λ,γ	**.06**	**.12**	**.04**	**.07**	**.59**	**2.23**	**.003**	*	**.59**	**2.06**
	λ	.06	.10	-	-	.59	-	.002	-	.60	-
	γ	-	-	.05	.09	-	2.03	-	.08	-	2.00
0.8, 2.0	λ,γ	**.06**	**.12**	**.04**	**.08**	**.79**	**2.07**	**.001**	**.07**	**.79**	**2.04**
	λ	.06	.11	-	-	.79	-	.001	-	.80	-
	γ	-	-	.05	.09	-	2.01	-	.03	-	2.00

Table 2A

Negative Exponential Weights n=400

λ, γ	unknown	$\lambda 5\%$	$\lambda 10\%$	$\gamma 5\%$	$\gamma 10\%$	$\bar{\lambda}$	$\bar{\gamma}$	λ_{mse}	γ_{mse}	λ_{med}	γ_{med}
0.2, 2.5	λ, γ	**.03**	**.06**	**.11**	**.14**	**.21**	*	**.006**	*	**.20**	**2.71**
	λ	.05	.10	-	-	.20	-	.003	-	.20	-
	γ	-	-	.08	.11	-	3.66	-	*	-	2.53
0.4, 2.5	λ, γ	**.06**	**.11**	**.06**	**.08**	**.40**	*	**.004**	*	**.39**	**2.61**
	λ	.05	.11	-	-	.39	-	.002	-	.40	-
	γ	-	-	.06	.08	-	2.63	-	.59	-	2.51
0.6, 2.5	λ, γ	**.06**	**.11**	**.04**	**.07**	**.59**	**2.72**	**.002**	*	**.59**	**2.57**
	λ	.05	.10	-	-	.59	-	.001	-	.60	-
	γ	-	-	.05	.09	-	2.55	-	.16	-	2.50
0.8, 2.5	λ, γ	**.06**	**.11**	**.04**	**.08**	**.79**	**2.59**	**.001**	**.13**	**.79**	**2.56**
	λ	.06	.11	-	-	.79	-	.001	-	.80	-
	γ	-	-	.05	.09	-	2.52	-	.07	-	2.50
0.2, 3.0	λ, γ	**.03**	**.06**	**.12**	**.15**	**.21**	*	**.005**	*	**.20**	**3.20**
	λ	.05	.10	-	-	.20	-	.002	-	.20	-
	γ	-	-	.09	.12	-	4.85	-	*	-	3.03
0.4, 3.0	λ, γ	**.05**	**.10**	**.06**	**.09**	**.40**	*	**.003**	*	**.40**	**3.12**
	λ	.05	.11	-	-	.40	-	.002	-	.40	-
	γ	-	-	.06	.09	-	3.28	-	2.57	-	3.01
0.6, 3.0	λ, γ	**.06**	**.11**	**.05**	**.07**	**.59**	**3.48**	**.002**	*	**.60**	**3.08**
	λ	.05	.11	-	-	.60	-	.001	-	.60	-
	γ	-	-	.05	.08	-	3.09	-	.32	-	3.01
0.8, 3.0	λ, γ	**.06**	**.11**	**.04**	**.07**	**.79**	**3.13**	**.001**	**.25**	**.79**	**3.06**
	λ	.06	.10	-	-	.80	-	.001	-	.80	-
	γ	-	-	.05	.09	-	3.05	-	.14	-	3.00

Table 2B

Negative Exponential Weights n=400

λ_0, γ_0	unknown	$\lambda 5\%$	$\lambda 10\%$	$\gamma 5\%$	$\gamma 10\%$	$\bar{\lambda}$	$\bar{\gamma}$	λ_{mse}	γ_{mse}	λ_{med}	γ_{med}
0.2, 1.5	λ, γ	**.16**	**.19**	**.11**	**.12**	**.03**	*	**.11**	*	**.09**	**2.79**
	λ	.02	.06	-	-	.12	-	.07	-	.13	-
	γ	-	-	.02	.03	-	1.72	-	*	-	1.73
0.4, 1.5	λ, γ	**.21**	**.25**	**.06**	**.07**	**.15**	*	**.12**	*	**.17**	**2.79**
	λ	.03	.08	-	-	.27	-	.08	-	.29	-
	γ	-	-	.03	.06	-	1.03	-	5.73	-	1.52
0.6, 1.5	λ, γ	**.27**	**.32**	**.03**	**.03**	**.27**	*	**.15**	*	**.27**	**2.62**
	λ	.07	.14	-	-	.43	-	.10	-	.46	-
	γ	-	-	.05	.09	-	1.20	-	2.11	-	1.49
0.8, 1.5	λ, γ	**.34**	**.42**	**.02**	**.02**	**.39**	*	**.21**	*	**.40**	**2.49**
	λ	.20	.29	-	-	.64	-	.12	-	.62	-
	γ	-	-	.05	.09	-	1.31	-	1.09	-	1.48
0.2, 2.0	λ, γ	**.12**	**.14**	**.09**	**.10**	**.08**	*	**.08**	*	**.12**	**3.07**
	λ	.03	.07	-	-	.16	-	.03	-	.17	-
	γ	-	-	.03	.02	-	3.43	-	*	-	2.16
0.4, 2.0	λ, γ	**.14**	**.18**	**.02**	**.03**	**.24**	*	**.06**	*	**.24**	**3.10**
	λ	.04	.10	-	-	.33	-	.04	-	.34	-
	γ	-	-	.01	.03	-	1.91	-	5.83	-	2.00
0.6, 2.0	λ, γ	**.20**	**.26**	**.01**	**.01**	**.39**	*	**.08**	*	**.39**	**2.87**
	λ	.06	.12	-	-	.51	-	.04	-	.53	-
	γ	-	-	.02	.06	-	1.91	-	0.72	-	1.98
0.8, 2.0	λ, γ	**.25**	**.33**	**.004**	**.006**	**.55**	*	**.09**	*	**.56**	**2.71**
	λ	.10	.16	-	-	.70	-	.03	-	.72	-
	γ	-	-	.04	.08	-	1.97	-	0.17	-	1.99

Table 3A

Inverse Power Weights n=100

λ, γ	unknown	λ5%	λ10%	γ5%	γ10%	$\bar{\lambda}$	$\bar{\gamma}$	λ_{mse}	γ_{mse}	λ_{med}	γ_{med}
0.2, 2.5	λ, γ	**.08**	**.09**	**.07**	**.08**	**.13**	*	**.06**	*	**.15**	**3.45**
	λ	.04	.07	-	-	.18	-	.02	-	.18	-
	γ	-	-	.01	.02	-	*	-	*	-	2.64
0.4, 2.5	λ, γ	**.09**	**.13**	**.02**	**.02**	**.30**	*	**.04**	*	**.29**	**3.45**
	λ	.05	.10	-	-	.36	-	.02	-	.37	-
	γ	-	-	.02	.04	-	2.72	-	*	-	2.50
0.6, 2.5	λ, γ	**.15**	**.21**	**.01**	**.01**	**.47**	*	**.04**	*	**.47**	**3.19**
	λ	.06	.11	-	-	.55	-	.02	-	.56	-
	γ	-	-	.03	.07	-	2.51	-	.33	-	2.49
0.8, 2.5	λ, γ	**.19**	**.25**	**.003**	**.006**	**.65**	*	**.04**	*	**.66**	**3.05**
	λ	.08	.14	-	-	.74	-	.01	-	.76	-
	γ	-	-	.04	.09	-	2.50	-	.11	-	2.49
0.2, 3.0	λ, γ	**.06**	**.08**	**.07**	**.09**	**.16**	*	**.04**	*	**.17**	**3.89**
	λ	.04	.08	-	-	.18	-	.01	-	.18	-
	γ	-	-	.01	.02	-	*	-	*	-	3.14
0.4, 3.0	λ, γ	**.06**	**.09**	**.03**	**.04**	**.34**	*	**.02**	*	**.33**	**3.88**
	λ	.06	.11	-	-	.37	-	.01	-	.38	-
	γ	-	-	.03	.06	-	3.71	-	*	-	3.00
0.6, 3.0	λ, γ	**.11**	**.16**	**.02**	**.03**	**.52**	*	**.02**	*	**.52**	**3.62**
	λ	.06	.11	-	-	.57	-	.01	-	.57	-
	γ	-	-	.04	.08	-	3.14	-	2.14	-	3.00
0.8, 3.0	λ, γ	**.15**	**.21**	**.005**	**.01**	**.70**	*	**.02**	*	**.71**	**3.48**
	λ	.06	.11	-	-	.76	-	.01	-	.77	-
	γ	-	-	.05	.09	-	3.00	-	.15	-	2.98

Table 3B

Inverse Power Weights n=100

λ_0, γ_0	unknown	λ5%	λ10%	γ5%	γ10%	$\bar{\lambda}$	$\bar{\gamma}$	λ_{mse}	γ_{mse}	λ_{med}	γ_{med}
0.2, 1.5	λ, γ	**.23**	**.27**	**.11**	**.11**	**.06**	*	**.08**	*	**.08**	**2.26**
	λ	.04	.08	-	-	.14	-	.04	-	.15	-
	γ	-	-	.03	.07	-	.53	-	*	-	1.53
0.4, 1.5	λ, γ	**.25**	**.29**	**.03**	**.04**	**.21**	*	**.08**	*	**.19**	**2.16**
	λ	.05	.11	-	-	.32	-	.04	-	.33	-
	γ	-	-	.04	.08	-	1.21	-	2.13	-	1.50
0.6, 1.5	λ, γ	**.26**	**.31**	**.01**	**.01**	**.35**	*	**.10**	**-**	**.35**	**2.02**
	λ	.06	.12	-	-	.51	-	.04	-	.52	-
	γ	-	-	.04	.07	-	1.42	-	.35	-	1.49
0.8, 1.5	λ, γ	**.28**	**.36**	**.01**	**.03**	**.50**	*	**.12**	*	**.51**	**1.93**
	λ	.07	.14	-	-	.68	-	.03	-	.70	-
	γ	-	-	.04	.08	-	1.47	-	.07	-	1.49
0.2, 2.0	λ, γ	**.14**	**.18**	**.06**	**.07**	**.13**	*	**.04**	*	**.12**	**2.55**
	λ	.05	.09	-	-	.18	-	.01	-	.18	-
	γ	-	-	.01	.02	-	1.58	-	*	-	2.02
0.4, 2.0	λ, γ	**.17**	**.21**	**.01**	**.02**	**.31**	*	**.03**	*	**.29**	**2.42**
	λ	.05	.10	-	-	.37	-	.01	-	.38	-
	γ	-	-	.02	.06	-	1.97	-	.23	-	2.00
0.6, 2.0	λ, γ	**.17**	**.23**	**.01**	**.01**	**.48**	*	**.04**	*	**.48**	**2.30**
	λ	.06	.11	-	-	.57	-	.01	-	.57	-
	γ	-	-	.04	.09	-	1.99	-	.05	-	2.00
0.8, 2.0	λ, γ	**.17**	**.25**	**.01**	**.04**	**.66**	**2.59**	**.04**	*	**.67**	**2.25**
	λ	.06	.12	-	-	.76	-	.01	-	.77	-
	γ	-	-	.04	.10	-	1.99	-	.02	-	2.00

Table 4A

Inverse Power Weights n=400

λ, γ	unknown	λ5%	λ10%	γ5%	γ10%	$\bar{\lambda}$	$\bar{\gamma}$	λ_{mse}	γ_{mse}	λ_{med}	γ_{med}
0.2, 2.5	λ, γ	**.08**	**.11**	**.05**	**.06**	**.18**	*	**.02**	*	**.16**	**2.98**
	λ	.05	.11	-	-	.19	-	.01	-	.19	-
	γ	-	-	.004	.01	-	2.58	-	8.65	-	2.51
0.4, 2.5	λ, γ	**.13**	**.17**	**.02**	**.04**	**.36**	*	**.02**	*	**.35**	**2.80**
	λ	.05	.11	-	-	.39	-	.01	-	.39	-
	γ	-	-	.03	.08	-	2.51	-	.13	-	2.50
0.6, 2.5	λ, γ	**.13**	**.18**	**.02**	**.03**	**.54**	**3.63**	**.02**	*	**.54**	**2.70**
	λ	.06	.11	-	-	.58	-	.005	-	.59	-
	γ	-	-	.04	.10	-	2.50	-	.05	-	2.50
0.8, 2.5	λ, γ	**.11**	**.17**	**.02**	**.06**	**.74**	**2.77**	**.01**	*	**.74**	**2.66**
	λ	.06	.11	-	-	.78	-	.003	-	.78	-
	γ	-	-	.05	.10	-	2.50	-	.02	-	2.50
0.2, 3.0	λ, γ	**.05**	**.08**	**.06**	**.08**	**.20**	*	**.01**	*	**.18**	**3.43**
	λ	.05	.11	-	-	.19	-	.005	-	.19	-
	γ	-	-	.01	.04	-	3.58	-	*	-	3.00
0.4, 3.0	λ, γ	**.10**	**.14**	**.04**	**.05**	**.38**	*	**.01**	*	**.37**	**3.25**
	λ	.05	.11	-	-	.39	-	.004	-	.39	-
	γ	-	-	.04	.08	-	3.04	-	.20	-	3.00
0.6, 3.0	λ, γ	**.10**	**.15**	**.03**	**.04**	**.57**	**3.90**	**.01**	*	**.57**	**3.16**
	λ	.06	.10	-	-	.59	-	.003	-	.59	-
	γ	-	-	.04	.10	-	3.01	-	.07	-	3.00
0.8, 3.0	λ, γ	**.09**	**.14**	**.02**	**.06**	**.77**	**3.17**	**.005**	**.18**	**.77**	**3.12**
	λ	.06	.11	-	-	.79	-	.002	-	.79	-
	γ	-	-	.05	.10	-	3.00	-	.03	-	3.00

Table 4B

Inverse Power Weights n=400

As Tables 3 and 4 reveal, there is a significant difference between the estimators' behaviour for exponential weights and inverse power weights, results for the latter being substantially less favourable. A partial explanation for this could lie in the fact that the decline over distance shown by the exponential weights is steeper than that of the power weights for the parameter values tested. Whatever the cause,the most obvious differences are the severely over-sized t-statistics for λ and undersized statistics for γ inTables 3 and 4. The problem is less severe in Tables 3B and 4B than in Tables 3A and 4A, suggesting a connection to the steepness of the weight function. The results for the two benchmark cases, of known λ or known γ, show that the steepness of the weights is at best a partial explanation, however.

5 An Empirical Exercise

We now apply our QML estimator to data from Ertur and Koch (2007)[1], originally from the Penn World Tables version 6.1 (Heston et al., 2002). The data comprise a single cross-section of 7 variables for 91 countries for the period 1960-1995. The countries are from the non-oil sample in Mankiw et al. (1992), see Table 1 in the Appendix for a list and their ISO codes. The variables are listed below in Table 5.

Variable description	**Code name**
level of income per worker in 1960	*lny60*
level of income per worker in 1995	*lny95*
average growth rate 1960-1995	*gy = (lny95-lny60)/35*
average share of real investment in GDP 1960-1995	*lns*
average growth rate of working-age population (n_p) plus[a] $(g+\delta)$	*lnngd*
longitude of capital city	*xlong*
latitude of capital city	*ylat*

Table 5: List of variables and their code names

a : $(g+\delta)$ is the rate of technical progress plus the depreciation rate

The first five variables are used to evaluate the impact of saving, population growth and neighbourhood spillovers on economic growth. The longitude and latitude are used to construct the distance matrix of which the elements d_{ij} are great-circle, geographical distances between country capitals. The growth rates differ widely across countries; of the 91 countries in the sample, 17 have negative average growth rates, while that of Hong Kong, at 6.24%, is the highest, and the Democratic Republic of the Congo is the lowest, with -3.43%. Real investment shares are also highly variable, ranging from 1.9% in Uganda to 41% in Singapore. Finally, the average rates of growth of the working-age population (15 - 64 years) (n_p) plus $(g+\delta)$ (which is assumed to equal 0.05 as in Mankiw et al. (1992, p. 413)) differ widely, ranging from the highest, Jordan (9.3%) to the lowest, Austria (5.3%).

[1] See http://qed.econ.queensu.ca/jae/2007-v22.6/ertur-koch/ for detail.

For country $i \in (1, ..., 91)$, the model is

$$gy_i = \beta_1 + \beta_2 lny60_i + \beta_3 lns_i + \beta_4 lnngd_i + \lambda \sum_{j \neq i}^{n} w_{ij}(\gamma_l) gy_j + \varepsilon_i. \quad (24)$$

Defining weight matrices $W1_n(\gamma_1) = \{w_{ij}(\gamma_1)\}$ and $W2_n(\gamma_2) = \{w_{ij}(\gamma_2)\}$, $W1_n(\gamma_1)$ has exponential weights and $W2_n(\gamma_2)$ power weights of the form,

$$w1_{n,ij}(\gamma_1) = \begin{cases} 0 & \text{if } i = j \\ \frac{e^{-\gamma_1 d_{ij}}}{\sum_j e^{-\gamma_1 d_{ij}}} \geq 0 & \text{if } i \neq j \end{cases} ; w2_{n,ij}(\gamma_2) = \begin{cases} 0 & \text{if } i = j \\ \frac{d_{ij}^{-\gamma_2}}{\sum_j d_{ij}^{-\gamma_2}} \geq 0 & \text{if } i \neq j \end{cases}$$

as in (23).

Ertur and Koch (2007) set both γ_1 and γ_2 equal to 2.0 to obtain the weight matrices employed in their study. When the γ parameter determining $W1$ or $W2$ in the SLM is freely estimated by QML, we find the results in the table below:

Weight structure	γ_{qml}	γ *se.*	$t(\gamma = 2)$	λ_{qml}	λ *se.*
Inverse power	2.48	.92	0.52	0.25	0.10
Negative exponential	0.8	.35	-3.45	0.47	0.14

Table 6: QML estimates of γ and λ in the SLM applied to the Ertur/Koch data.

Clearly, there is no evidence against the choice of $\gamma = 2$ when inverse power weights are used, but quite strong evidence against this value if the exponential weights are used. Looking back at the results for the exponential weight experiments with $n = 100$ and $(\lambda, \gamma) = (0.4, 2.0)$ or $(0.6, 2.0)$ we see that the $t-$ statistic probabilities were fairly reliable. On the other hand, the results for the power weights with $(\lambda, \gamma) = (0.2, 2.0)$ suggest the $t-$ statistics are somewhat liberal, reinforcing the insignificant result in the table.

6 Conclusion

In this paper we introduce a sub-model for the spatial weights matrix and estimate a variable spatial weight matrix in the MR-SAR model by the maximum Gaussian likelihood (QML). We give identifiability conditions for the parameter defining the weights as well as establishing the consistency and the asymptotic distribution of the QML estimator under appropriate conditions that extend those given in Lee (2004a). Finite sample properties of the QMLE are studied in a Monte Carlo experiment. The performance of the estimator is then illustrated on a small sample of real data.

The Monte Carlo results show that the QML estimator of the parameter defining the spatial weights, γ, behaves reasonably well in samples of quite modest size, particularly when the negative exponential functional form is adopted. Extension of these results to more general model forms is an area deserving attention. The MATLAB code used for the empirical exercise is available from the authors' web pages (i.e. it will be, when the paper is published).

7 Appendix

This appendix contains three sections. In the first, to assist the reader, we collect frequently used notation, and a number of lemmata, definitions and theorems that are used repeatedly in the proofs. The second section establishes the useful properties of some intermediate objects, then gives proofs of Theorems 1 - 3. The third section contains a table of countries and their ISO codes. The first two sections follow material in Lee (2004a,b) very closely in many places; this is inevitable given that, roughly speaking, our task is to show that properties established by Lee for a sequence, $\{W_n\}$ hold for the family of sequences, $\{W_n(\gamma)\} \quad \gamma \in \mathbf{\Gamma}$.

7.1 Frequently used notation and background results

7.1.1 Notation

The following notation is used in the main text and in this appendix. In most cases it is an obvious extension of that adopted by Lee (2004a,b).

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|det(S_n(\lambda,\gamma))|$$
$$- \frac{1}{2\sigma^2}(S_n(\lambda,\gamma)Y_n - X_n\beta)'(S_n(\lambda,\gamma)Y_n - X_n\beta)$$

$$\ln L_n(\lambda,\gamma) = -\frac{n}{2}(\ln(2\pi)+1) + \ln|det(S_n(\lambda,\gamma))| - \frac{n}{2}\ln\hat{\sigma}_n^2(\lambda,\gamma)$$

$$S_n(\lambda,\gamma) = I_n - \lambda W_n(\gamma)$$

$$S_n = I_n - \lambda_0 W_n$$

$$G_n = W_n S_n^{-1}$$

$$T_n = Z_n S_n^{-1}$$

$$C_n = A_n S_n^{-1}$$

$$V_n = B_n S_n^{-1}$$

$$Z_n = \frac{\partial W_n}{\partial\gamma}$$

$$A_n = \frac{\partial Z_n}{\partial\gamma}$$

$$B_n = \frac{\partial A_n}{\partial\gamma}$$

$$Q_n(\lambda,\gamma) = max_{\beta,\sigma^2}E[\ln L_n(\theta)] = -\frac{n}{2}(\ln(2\pi)+1) + \ln|det(S_n(\lambda,\gamma))| - \frac{n}{2}\ln\sigma_n^{2*}(\lambda,\gamma)$$

$$\hat{\sigma}_n^2(\lambda,\gamma) = \frac{1}{n}[Y_n'S_n'(\lambda,\gamma)M_nS_n(\lambda,\gamma)Y_n]$$

$$\sigma_n^{2*}(\lambda,\gamma) = \frac{1}{n}[(\lambda_0-\lambda)^2(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \sigma_0^2 tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})]$$

$$\sigma_n^2(\lambda,\gamma) = \frac{\sigma_0^2}{n}tr[S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1}]$$

$$M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$$

7.1.2 Useful lemmata, a definition and theorem (not for publication)

Note that notation in the lemmata, definition and theorem below follows the sources cited in the references.

Lee (2002) - Lemma A.2: Let $V_n = [v_1, v_2, ..., v_n]'$ be a vector of iid random variables with zero mean and finite variance, σ^2, and suppose its first three absolute moments exist. Let A_n be a square matrix with uniformly bounded column sums and suppose that elements of the $n \times k$ matrix C_n are uniformly bounded. Then, $(1/\sqrt{n})C_n'A_nV_n = O(1)$. Furthermore, if the limit of $(1/n)C_n'A_nA_n'C_n$ exists and it is positive definite, then $(1/\sqrt{n})C_n'A_nV_n \xrightarrow{D} N(0, \sigma^2 \lim_{n\to\infty}(1/n)C_n'A_nA_n'C_n)$.

Proof: (Lee (2002)) Let $a_{n,j}$ denote the jth column of A_n. It follows that $(1/\sqrt{n})C_n'A_nV_n = (1/\sqrt{n})\sum_{j=1}^n q_{nj}\nu_j$ where $q_{nj} = C_n'a_{n,j}$. The first result follows from Chebyshev's inequality because q_{nj} are uniformly bounded and $var((1/\sqrt{n})C_n'A_nV_n = (\sigma^2/n)\sum_{j=1}^n q_{nj}q_{nj}'$. The second result follows from the Liapounov double array CLT and the Cramér-Wold device (Billingsley, 1995, Theorem 27.3 and Theorem 29.4). To check the Liapounov condition, let α be a nonzero row vector of constants and $B_n = var(\alpha C_n'A_nV_n) = \sigma^2\alpha C_n'A_nA_n'C_n\alpha'$. The assumptions imply that $\lim_{n\to\infty}(1/n)B_n^2 > 0$ and there exists a constant c such that $|\alpha q_{nj}| < c$, for all n and j. Hence, the Liapounov condition $\sum_{j=1}^n(1/B_n^3)E(|\alpha q_{nj}\nu_j|^3) \leq c^3E|\nu^3|/((1/n)B_n^2)^{3/2}n^{1/2} \to 0$ holds.

Lee (2003) - Lemma 1: Suppose that all elements of the spatial weights matrices W_n are nonnegative. If W_n are row-normalized, then $(I_n - \eta W_n)^{-1}$ are uniformly bounded in row sums uniformly in η in Λ, where Λ is any closed set in $(-1, 1)$.

Lee (2004b) - Lemma A.6: Suppose that the elements of the sequences of vectors $P_n = (p_{n1}, \cdots, p_{nn})'$ and $Q_n = (q_{n1}, \cdots, q_{nn})'$ are uniformly bounded for all n.

1. If the sequence of $n \times n$ matrices $\{A_n\}$ are uniformly bounded in either row or column sums, then $|Q_n'A_nP_n| = O(n)$.

2. If the row sums of $\{A_n\}$ and $\{Z_n\}$ are uniformly bounded, $|z_{i,n}A_nP_n| = O(1)$ uniformly in i, where $z_{i,n}$ is the ith row of Z_n.

Proof (Lee (2004b)): Introduce constants c_1 and c_2 such that $|p_{ni}| \leq c_1$ and $|q_{ni}| \leq c_2$. For 1), there exists a constant such that $\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n|a_{n,ij}| \leq c_3$. Hence, $|Q_n'A_nP_n| = |\sum_{i=1}^n\sum_{j=1}^n a_{n,ij}q_{ni}p_{nj}| \leq c_1c_2\sum_{i=1}^n\sum_{j=1}^n|a_{n,ij}| \leq nc_1c_2c_3$. For 2), let c_4 be a constant such that $\sum_{j=1}^n|a_{n,ij}| \leq c_4$ for all n and i. It follows that $|e_{ni}'A_nP_n| = |\sum_{j=1}^n a_{n,ij}p_{nj}| \leq c_1\sum_{j=1}^n|a_{n,ij}| \leq c_1c_4$ where e_{ni} is the ith unit column vector. Because $\{Z_n\}$ is uniformly bounded in row sums, $\sum_{j=1}^n|z_{n,ij}| \leq c_z$ for some constant c_z. It follows that $|z_{i,n}A_nP_n| \leq \sum_{j=1}^n|z_{n,ij}|\cdot|e_{nj}'A_nP_n| \leq (\sum_{j=1}^n|z_{n,ij}|)c_1c_4 \leq c_zc_1c_4$. Q.E.D.

Lee (2004b) - Lemma A.8: Suppose that the elements $a_{n,ij}$ of the sequence of $n \times n$ matrices $\{A_n\}$, where $A_n = [a_{n,ij}]$, are $O(\frac{1}{h_n})$ uniformly in all i and j; and $\{B_n\}$ is a sequence of conformable $n \times n$ matrices.

1. If $\{B_n\}$ are uniformly bounded in column sums, the elements of A_nB_n have the uniform order $O(\frac{1}{h_n})$.

2. If $\{B_n\}$ are uniformly bounded in row sums, the elements of B_nA_n have the uniform order $O(\frac{1}{h_n})$.

For both cases (1) and (2), $tr(A_nB_n) = tr(B_nA_n) = O(\frac{n}{h_n})$.

Proof (Lee (2004b)): Consider (1). Let $a_{n,ij} = \frac{c_{n,ij}}{h_n}$. Because $a_{n,ij} = O(\frac{1}{h_n})$ uniformly in i and j, there exists a constant c_1 so that $|c_{n,ij}| \leq c_1$ for all i, j and n. Because $\{B_n\}$ is uniformly bounded in column sums, there exists a constant c_2 such that $\sum_{k=1}^n |b_{n,kj}| \leq c_2$ for all n and j. Let $a_{i,n}$ be the ith row of A_n and $b_{n,l}$ be the lth column of B_n. It follows that $|a_{i,n} b_{n,l}| \leq \frac{1}{h_n} \sum_{j=1}^n |c_{n,ij} b_{n,jl}| \leq \frac{c_1}{h_n} \sum_{j=1}^n |b_{n,jl}| \leq \frac{c_1 c_2}{h_n}$, for all i and l. Furthermore, $|tr(A_n B_n)| = |\sum_{i=1}^n a_{i,n} b_{n,i}| \leq \sum_{i=1}^n |a_{i,n} b_{n,i}| \leq c_1 c_2 \frac{n}{h_n}$. These prove the results in (1). The results in (2) follow from (1) because $(B_n A_n)' = A_n' B_n'$ and the uniform boundedness in row sums of $\{B_n\}$ is equivalent to the uniform boundedness in column sums of $\{B_n'\}$. Q.E.D.

Lee (2004b) - Lemma A.11: Let $A_n = [a_{ij}]$ be an n-dimensional square matrix of constants. Let $V_n = [v_1, v_2, ..., v_n]'$ be a vector of iid random variables with zero mean, variance, σ^2, and finite fourth moment, μ_4.Then

1. $E(V_n' A_n V_n) = \sigma^2 tr(A_n)$,
2. $E(V_n' A_n V_n)^2 = (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4 [tr^2(A_n) + tr(A_n A_n') + tr(A_n^2)]$,
3. $var(V_n' A_n V_n) = (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4 [tr(A_n A_n') + tr(A_n^2)]$.

In particular, if ν's are normally distributed, then $E(V_n' A_n V_n)^2 = \sigma^4 [tr^2(A_n) + tr(A_n A_n') + tr(A_n^2)]$ and $var(V_n' A_n V_n) = \sigma^4 [tr(A_n A_n') + tr(A_n^2)]$.

Proof (Lee (2004b)): The result in 1) is trivial. For the second moment,

$E(V_n' A_n V_n)^2 = E(\sum_{i=1}^n \sum_{j=1}^n a_{ij} \nu_i \nu_j)^2 = E(\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n a_{ij} a_{kl} \nu_i \nu_j \nu_k \nu_l)$.

Because ν's are i.i.d. with zero mean, $E(\nu_i \nu_j \nu_k \nu_l)$ will not vanish only when $i = j = k = l, (i = j) \neq (k = l), (i = k) \neq (j = l)$, and $(i = l) \neq (j = k)$. Therefore,

$$
\begin{aligned}
E(V_n' A_n V_n)^2 &= \sum_{i=1}^n a_{ii}^2 E(\nu_i^4) + \sum_{i=1}^n \sum_{j \neq i}^n a_{ii} a_{jj} E(\nu_i^2 \nu_j^2) + \sum_{i=1}^n \sum_{j \neq i}^n a_{ij}^2 E(\nu_i^2 \nu_j^2) \\
&+ \sum_{i=1}^n \sum_{j \neq i}^n a_{ij} a_{ji} E(\nu_i^2 \nu_j^2) \\
&= (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4 [\sum_{i=1}^n \sum_{j=1}^n a_{ii} a_{jj} + \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 + \sum_{i=1}^n \sum_{j=1}^n a_{ij} a_{ji}] \\
&= (\mu_4 - 3\sigma^4) \sum_{i=1}^n a_{ii}^2 + \sigma^4 [tr^2(A_n) + tr(A_n A_n') + tr(A_n^2)]
\end{aligned}
$$

The result 3) follows from $var(V_n' A_n V_n) = E(V_n' A_n V_n)^2 - E^2(V_n' A_n V_n)$ and those of 1) and 2). When ν's are normally distributed, $\mu_4 = 3\sigma^2$. Q.E.D.

Lee (2004b) - Lemma A.12: Suppose that the sequence of $n \times n$ matrices $\{A_n\}$ are uniformly bounded in either row and column sums, and the elements $a_{n,ij}$ of A_n are $O(\frac{1}{h_n})$ uniformly in all i and j. Then,

$E(V_n'A_nV_n) = O(\frac{n}{h_n})$, $var(V_n'A_nV_n) = O(\frac{n}{h_n})$ and $V_n'A_nV_n = O_p(\frac{n}{h_n})$. Furthermore, if $\lim_{n\to\infty} \frac{h_n}{n} = 0$, $\frac{h_n}{n}V_n'A_nV_n - \frac{h_n}{n}E(V_n'A_nV_n) = o_p(1)$.

Proof (Lee (2004b)): $E(V_n'A_nV_n) = \sigma^2 tr(A_n) = O(\frac{n}{h_n})$. From Lemma A.11, the variance of $V_n'A_nV_n$ is $var(V_n'A_nV_n) = (\mu_4 - 3\sigma^4)\sum_{i=1}^n a_{n,ii}^2 + \sigma^4[tr(A_nA_n') + tr(A_n^2)]$. Lemma A.8 implies that $tr(A_n^2)$ and $tr(A_nA_n')$ are $O(\frac{n}{h_n})$. As $\sum_{i=1}^n a_{n,ii}^2 \leq tr(A_nA_n')$, it follows that $\sum_{i=1}^n a_{n,ii}^2 = O(\frac{n}{h_n})$. Hence, $var(V_n'A_nV_n) = O(\frac{n}{h_n})$. As $E((V_n'A_nV_n)^2) = var(V_n'A_nV_n) + E^2(V_n'A_nV_n) = O((\frac{n}{h_n})^2)$, the generalized Chebyshev inequality implies that $P(\frac{h_n}{n}|V_n'A_nV_n| \geq M) \leq \frac{1}{M^2}(\frac{h_n}{n})^2 E((V_n'A_nV_n)^2) = \frac{1}{M^2}O(1)$ and, hence, $\frac{h_n}{n}V_n'A_nV_n = O_p(1)$.

Finally, because $var(\frac{h_n}{n}V_n'A_nV_n) = O(\frac{h_n}{n}) = o(1)$ when $\lim_{n\to\infty}\frac{h_n}{n} = 0$, the Chebyshev inequality implies that $\frac{h_n}{n}V_n'A_nV_n - \frac{h_n}{n}E(V_n'A_nV_n) = o_p(1)$. Q.E.D.

White (1996) Definition 3.3 (Identifiable Uniqueness): Let $\bar{Q}_n : \Theta \to \bar{\Re}$ be continuous on Θ, a compact subset of $\Re^p$, $p \in \aleph$, and let Θ_n be a non-empty compact subset of Θ, $n = 1, 2, \ldots$. Suppose that $\bar{Q}_n(\theta)$ has a maximum on Θ_n at θ_n^*, $n = 1, 2, \ldots$. Let $s_n(\varepsilon)$ be an open sphere in $\Re^p$ centered at θ_n^* with fixed radius $\varepsilon > 0$. For each $n = 1, 2, \ldots$ define the neighbourhood $\eta_n(\varepsilon) = s_n(\varepsilon) \cap \Theta_n$ with compact complement $\eta_n^c(\varepsilon)$ in Θ_n. The sequence of maximizers $\theta^* \equiv \{\theta_n^*\}$ is said to be *identifiably unique on* $\{\Theta_n\}$ if either for all $\varepsilon > 0$ and all n, $\eta_n^c(\varepsilon)$ is empty, or for all $\varepsilon > 0$

$$\lim_{n\to\infty} \sup[\max_{\theta\in\eta_n^c(\varepsilon)} \bar{Q}_n(\theta) - \bar{Q}_n(\theta_n^*)] < 0.$$

■

White (1996) Theorem 3.4: Let (Ω, F, P) be a complete probability space, let Θ be a compact subset of $\Re^p, p \in \aleph$ and let $\{\Theta_n\}$ be a sequence of compact subsets of Θ. Let $\{Q_n\}$ be a sequence of random functions continuous on Θ a.s. - P and let $\hat{\theta}_n = argmax_{\Theta_n} Q_n(\cdot, \theta)$ a.s. - P. If $Q_n(\cdot,\theta) - \bar{Q}_n(\theta) \to 0$ as $n \to \infty$ a.s. - P (prob-P) uniformly on Θ and if $\{\bar{Q}_n : \Theta \to \bar{\Re}\}$ has identifiably unique maximizers θ^* on $\{\Theta_n\}$ then $\hat{\theta}_n - \theta_n^* \to 0$ as $n \to \infty$ a.s. - P (prob - P). ■

7.2 Proofs of theorems

In this section, we first state some properties that we frequently use in our proofs. We show the properties of $\ln|det(S_n(\lambda,\gamma))|$, $\sigma_n^2(\lambda,\gamma)$, $Q_n(\lambda,\gamma)$, and an auxiliary model $Q_{p,n}(\lambda,\gamma)$. Detailed proofs of the identifiable uniqueness, consistency and normality of the QML estimator $\hat{\theta}_n$ are shown in the subsequent sections. The proofs are carried out following the approach in Lee (2004a). Note that, for notational convenience, we omit the parameters in the parentheses when the parameters are at their true values. For example, we write W_n for $W_n(\gamma_0)$.

7.2.1 Properties of $\ln|det(S_n(\lambda,\gamma))|$

Let λ_1 and λ_2 be in Λ and γ_1 and γ_2 in Γ, and all of them belong to $\Lambda\otimes\Gamma$. By the mean value theorem,

$$\begin{aligned}&\frac{1}{n}(\ln|det(S_n(\lambda_2,\gamma_2))|-\ln|det(S_n(\lambda_1,\gamma_1))|)\\&=-\frac{1}{n}tr(W_n(\bar{\gamma}_n)S_n^{-1}(\bar{\lambda}_n,\bar{\gamma}_n))[\lambda_2-\lambda_1]-\frac{\bar{\lambda}_n}{n}tr(Z_n(\bar{\gamma}_n)S_n^{-1}(\bar{\lambda}_n,\bar{\gamma}_n))[\gamma_2-\gamma_1]\\&=-\frac{1}{n}tr(G_n(\bar{\lambda}_n,\bar{\gamma}_n))[\lambda_2-\lambda_1]-\frac{\bar{\lambda}_n}{n}tr(T_n(\bar{\lambda}_n,\bar{\gamma}_n))[\gamma_2-\gamma_1]\end{aligned}\tag{25}$$

where $\bar{\lambda}_n$ lies between λ_1 and λ_2, and $\bar{\gamma}_n$ lies between γ_1 and γ_2. Note that $G_n=W_nS_n^{-1}$ and $T_n=Z_nS_n^{-1}$. As $\{S_n^{-1}(\lambda,\gamma)\}$ is uniformly bounded in either row or column sums uniformly in λ and γ by Assumption 9, and elements of $W_n(\gamma)$ are assumed to be $O(\frac{1}{h_n})$ by Assumption 5, then Lemma A.8 in Lee (2004b) implies that $\frac{1}{n}tr(G_n(\bar{\lambda},\bar{\gamma}))=O(\frac{1}{h_n})$. The term $Z_n(\bar{\gamma}_n)$ on the right hand side of (25), which is the first-order derivative of $W_n(\gamma)$ with respect to γ at $\bar{\gamma}_n$, is continuous and uniformly bounded by Assumption 5, then $\frac{\bar{\lambda}_n}{n}tr(T_n(\bar{\lambda},\bar{\gamma}))=O(\frac{1}{h_n})$ as well. Hence, $\frac{1}{n}\ln|det(S_n(\lambda,\gamma))|$ is uniformly equicontinuous in λ and γ in $\Lambda\otimes\Gamma$. Because $\Lambda\otimes\Gamma$ is a compact set, $\frac{1}{n}(\ln|det(S_n(\lambda_2,\gamma_2))|-\ln|det(S_n(\lambda_1,\gamma_1))|)=O(1)$ uniformly in λ_1 and λ_2, and γ_1 and γ_2 in $\Lambda\otimes\Gamma$.

7.2.2 An Auxiliary Model $Q_{p,n}(\lambda,\gamma)$

As in Lee (op. cit.) we introduce the following pure SAR model, and corresponding log likelihood:

$$\begin{aligned}Y_n=\lambda W_n(\gamma)Y_n+\varepsilon_n\qquad&\varepsilon_n\sim N(0,\sigma_0^2I_n)\\\ln L_{p,n}(\lambda,\gamma,\sigma^2)=-\frac{n}{2}\ln(2\pi)-&\frac{n}{2}\ln\sigma^2+\ln|det(S_n(\lambda,\gamma))|\\-&\frac{1}{2\sigma^2}Y_n'S_n'(\lambda,\gamma)S_n(\lambda,\gamma)Y_n.\end{aligned}\tag{26}$$

Then, defining

$$Q_{p,n}(\lambda,\gamma)=\max_{\sigma^2}E[\ln L_{p,n}(\lambda,\gamma,\sigma^2)]$$

and

$$\begin{aligned}\sigma_n^2(\lambda,\gamma)&=\frac{\sigma_0^2}{n}tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})\\&=\sigma_0^2[1+2(\lambda_0-\lambda)\frac{1}{n}tr(G_n)+(\lambda_0-\lambda)^2\frac{1}{n}tr(G_nG_n')],\end{aligned}\tag{27}$$

we obtain

$$Q_{p,n}(\lambda,\gamma)=-\frac{n}{2}(\ln 2\pi+1)-\frac{n}{2}\ln\sigma_n^2(\lambda,\gamma)+\ln|det(S_n(\lambda,\gamma))|.\tag{28}$$

Note that, by the Jensen inequality, we have $Q_{p,n}(\lambda,\gamma)\le E[\ln L_{p,n}(\lambda_0,\gamma_0,\sigma_0^2)]=Q_{p,n}$ for all λ and γ, which implies that $\frac{1}{n}[Q_{p,n}(\lambda,\gamma)-Q_{p,n}]\le 0$ for all λ and γ.

7.2.3 Properties of $\sigma_n^2(\lambda, \gamma)$

We show that $\sigma_n^2(\lambda, \gamma)$ defined in (27) is uniformly bounded away from zero on $\Lambda \otimes \Gamma$. We prove this by a counter argument. If $\sigma_n^2(\lambda, \gamma)$ were not uniformly bounded away from zero on $\Lambda \otimes \Gamma$, then there would exist sequences $\{\lambda_n\}$ and $\{\gamma_n\}$ in $\Lambda \otimes \Gamma$ such that $\lim_{n\to\infty} \sigma_n^2(\lambda_n, \gamma_n) = 0$. As established earlier, $\frac{1}{n}[Q_{p,n}(\lambda, \gamma) - Q_{p,n}] \leq 0$ for all λ and γ. This means that

$$-\frac{1}{2}\ln \sigma_n^2(\lambda, \gamma) \leq -\frac{1}{2}\ln \sigma_0^2 + \frac{1}{n}(\ln |det(S_n)| - \ln |det(S_n(\lambda, \gamma))|). \tag{29}$$

We have shown that $\frac{1}{n}(\ln |det(S_n)| - \ln |det(S_n(\lambda, \gamma))|) = O(1)$ and it implies that $-\frac{1}{2}\ln \sigma_n^2(\lambda_n, \gamma_n)$ is bounded from above. This contradicts
$\lim_{n\to\infty} \sigma_n^2(\lambda_n, \gamma_n) = 0$, and so $\sigma_n^2(\lambda, \gamma)$ is bounded away from zero uniformly on $\Lambda \otimes \Gamma$.

7.2.4 Properties of $Q_n(\lambda, \gamma)$

Finally, we show that $\frac{1}{n}Q_n(\lambda, \gamma)$ is uniformly equicontinuous on $\Lambda\otimes\Gamma$. Note that $\frac{1}{n}Q_n(\lambda, \gamma) = -\frac{1}{2}(\ln(2\pi)+1) - \frac{1}{2}\ln \sigma_n^{*2}(\lambda, \gamma) + \frac{1}{n}\ln |det(S_n(\lambda, \gamma))|$. Substituting (27) into σ_n^{*2}, we have

$$\begin{aligned}\sigma_n^{*2}(\lambda, \gamma) =& \frac{1}{n}[(\lambda_0 - \lambda)^2 (G_n X_n \beta_0)' M_n (G_n X_n \beta_0) \\ &+ \sigma_0^2[1 + 2(\lambda_0 - \lambda)\frac{1}{n}tr(G_n) + (\lambda_0 - \lambda)^2 \frac{1}{n}tr(G_n G_n')] \\ =& \frac{1}{n}[(\lambda_0 - \lambda)^2 (G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + \sigma_n^2(\lambda, \gamma)].\end{aligned}$$

It is quadratic in λ and its coefficients $\frac{1}{n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0)$, $\frac{1}{n}tr(G_n)$, and $\frac{1}{n}tr(G_n G_n')$ are bounded by Lemma A.6 and Lemma A.8 in Lee (2004b), so $\sigma_n^{*2}(\lambda, \gamma)$ is uniformly continuous on $\Lambda \otimes \Gamma$. The uniform continuity of $\ln \sigma_n^{*2}(\lambda, \gamma)$ on $\Lambda \otimes \Gamma$ follows because $\frac{1}{\sigma_n^{*2}(\lambda,\gamma)}$ is uniformly bounded on $\Lambda \otimes \Gamma$. It will also be shown later that $\sigma_n^{*2}(\lambda, \gamma)$ is uniformly bounded away from zero. Therefore, $\frac{1}{n}Q_n(\lambda, \gamma)$ is uniformly equicontinuous on $\Lambda \otimes \Gamma$.

In the following sections, we give detailed proofs of the identifiable uniqueness, consistency and asymptotic normality of $\hat{\theta}_n$.

7.2.5 Proof of Theorem 1: Identifiable Uniqueness

We show that

$$\lim_{n\to\infty} \sup[\max_{\theta\in\eta_n^c(\nu)} \{\frac{1}{n}Q_n(\lambda, \gamma) - \frac{1}{n}Q_n\}] < 0 \tag{30}$$

where $\eta_n^c(\nu)$ is the compact complement of the neighbourhood $\eta_n(\nu) = s_n(\nu) \cap \Theta_n$, with $s_n(\nu)$ an open sphere centred at θ_0 with fixed radius $\nu > 0$. Note that, for notational convenience, we omit the parameters in the parentheses when the functions are at the true values. For example, we write Q_n for $Q_n(\lambda_0, \gamma_0)$. We have

$$\frac{1}{n}Q_n(\lambda, \gamma) - \frac{1}{n}Q_n = \frac{1}{n}(\ln |det(S_n(\lambda, \gamma))| - \ln |det(S_n)|) - \frac{1}{2}(\ln \sigma_n^{*2}(\lambda, \gamma) - \ln \sigma_n^{*2}). \tag{31}$$

Note that

$$\sigma_n^{*2}(\lambda_0, \gamma_0) = \frac{\sigma_0^2}{n} tr[S_n^{-1\prime} S_n(\lambda_0, \gamma_0)' S_n(\lambda_0, \gamma_0) S_n^{-1}]$$
$$= \frac{\sigma_0^2}{n} tr[S_n^{-1\prime} S_n' S_n S_n^{-1}] = \sigma_0^2.$$

Add and subtract $\frac{1}{2}\ln\sigma_n^2(\lambda, \gamma)$ on the rhs of (31) and rearrange the terms, to obtain

$$\frac{1}{n}Q_n(\lambda, \gamma) - \frac{1}{n}Q_n = \frac{1}{n}(\ln|det(S_n(\lambda, \gamma))| - \ln|det(S_n)|) - \frac{1}{2}(\ln\sigma_n^2(\lambda, \gamma) - \ln\sigma_n^2)$$
$$- \frac{1}{2}(\ln\sigma_n^{*2}(\lambda, \gamma) - \ln\sigma_n^2(\lambda, \gamma))$$
$$= \frac{1}{n}(Q_{p,n}(\lambda, \gamma) - Q_{p,n}) - \frac{1}{2}(\ln\sigma_n^{*2}(\lambda, \gamma) - \ln\sigma_n^2(\lambda, \gamma)).$$

We prove this theorem by a counter example. Suppose that the condition of identifiable uniqueness did not hold, then there would exist $\nu > 0$ and sequences $\{\lambda_n\}$ and $\{\gamma_n\}$ in $\eta_n^c(\nu)$ such that $\lim_{n\to\infty}(\frac{1}{n}Q_n(\lambda_n, \gamma_n) - \frac{1}{n}Q_n) = 0$.

As $\eta_n^c(\nu)$ is the compact complement set of $\eta_n(\nu)$, there exist convergent subsequences $\{\lambda_{n_m}\}$ of $\{\lambda_n\}$, and $\{\gamma_{n_m}\}$ of $\{\gamma_n\}$. Let λ_+ and γ_+ denote the limit points of $\{\lambda_{n_m}\}$ and $\{\gamma_{n_m}\}$ in $\Lambda \otimes \Gamma$, respectively. Because $\frac{1}{n}Q_n(\lambda, \gamma)$ is uniformly equicontinuous in λ and γ, then $\lim_{n_m\to\infty}(\frac{1}{n_m}Q_{n_m}(\lambda_+, \gamma_+) - \frac{1}{n_m}Q_{n_m}) = 0$. However, because $-(\ln\sigma_n^{*2}(\lambda, \gamma) - \ln\sigma_n^2(\lambda, \gamma)) \leq 0$ and $\frac{1}{n}(Q_{p,n}(\lambda, \gamma) - Q_{p,n}) \leq 0$, which lead to $\lim_{n\to\infty}(\frac{1}{n}Q_n(\lambda, \gamma) - \frac{1}{n}Q_n) \leq 0$, this limit can be equal to zero only when $\lim_{n_m\to\infty}(\frac{1}{n_m}Q_{p,n_m}(\lambda_+, \gamma_+) - \frac{1}{n_m}Q_{p,n_m}) = 0$ and $\lim_{n_m\to\infty}(\sigma_{n_m}^{*2}(\lambda_+, \gamma_+) - \sigma_{n_m}^2(\lambda_+, \gamma_+)) = 0$. However, $\lim_{n_m\to\infty}(\sigma_{n_m}^{*2}(\lambda_+, \gamma_+) - \sigma_{n_m}^2(\lambda_+, \gamma_+)) = 0$, contradicts Assumption 10 that guarantees that $\lim_{n\to\infty}\frac{1}{n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0)$ exists and is positive. Hence, identifiable uniqueness must hold. Q.E.D.

7.2.6 Proof of Theorem 2: Consistency

The consistency of $\hat{\theta}_n$ follows from the identifiable uniqueness and uniform convergence (White 1996, Theorem 3.4). We have proved that θ_0 is uniquely identifiable, so we now need to prove that $\frac{1}{n}\ln L_n(\lambda, \gamma) - \frac{1}{n}Q_n(\lambda, \gamma)$ converges to zero in probability uniformly on $\Lambda\otimes\Gamma$. In other words, we show that $sup_{(\lambda,\gamma)\in\Lambda\otimes\Gamma}|\frac{1}{n}\ln L_n(\lambda, \gamma) - \frac{1}{n}Q_n(\lambda, \gamma)| = o_p(1)$. The first step is to show that $\hat{\sigma}_n^2(\lambda, \gamma) - \sigma_n^{*2}(\lambda, \gamma) = o_p(1)$ uniformly on $\Lambda \otimes \Gamma$, then we show that $|\ln\hat{\sigma}_n^2(\lambda, \gamma) - \ln\sigma_n^{*2}(\lambda, \gamma)| = o_p(1)$.

Clearly, $\frac{1}{n}\ln L_n(\lambda, \gamma) - \frac{1}{n}Q_n(\lambda, \gamma) = -\frac{1}{2}(\ln\hat{\sigma}_n^2(\lambda, \gamma) - \ln\sigma_n^{*2}(\lambda, \gamma))$, and we show that $\hat{\sigma}_n^2(\lambda, \gamma) - \sigma_n^{*2}(\lambda, \gamma) = o_p(1)$ uniformly on $\Lambda \otimes \Gamma$. Recall that

$$\sigma_n^{*2}(\lambda, \gamma) = (\lambda_0 - \lambda)^2 \frac{1}{n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + \frac{\sigma_0^2}{n} tr(S_n^{-1} S_n'(\lambda, \gamma) S_n(\lambda, \gamma) S_n^{-1})$$

and

$$\hat{\sigma}_n^2(\lambda, \gamma) = \frac{1}{n} Y_n' S_n'(\lambda, \gamma) M_n S_n(\lambda, \gamma) Y_n = \frac{1}{n}(M_n S_n(\lambda, \gamma) Y_n)'(M_n S_n(\lambda, \gamma) Y_n)$$

Because $M_nS_n(\lambda,\gamma)Y_n = (\lambda_0-\lambda)M_nG_nX_n\beta_0 + M_nS_n(\lambda,\gamma)S_n^{-1}\varepsilon_n$, then

$$\hat{\sigma}_n^2(\lambda,\gamma) = (\lambda_0-\lambda)^2\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + 2(\lambda_0-\lambda)H_{1n}(\lambda,\gamma) + H_{2n}(\lambda,\gamma)$$

where

$$H_{1n}(\lambda,\gamma) = \frac{1}{n}(G_nX_n\beta_0)'M_nS_n(\lambda,\gamma)S_n^{-1}\varepsilon_n \tag{32}$$

and

$$H_{2n}(\lambda,\gamma) = \frac{1}{n}\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)M_nS_n(\lambda,\gamma)S_n^{-1}\varepsilon_n. \tag{33}$$

Thus,

$$\hat{\sigma}_n^2(\lambda,\gamma) - \sigma_n^{*2}(\lambda,\gamma) = 2(\lambda_0-\lambda)H_{1n}(\lambda,\gamma) + H_{2n}(\lambda,\gamma) - \frac{\sigma_0^2}{n}tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})$$

and we show that the terms on the right hand side are all $o_p(1)$. We split (32) as follows

$$H_{1n}(\lambda,\gamma) = \frac{1}{n}(G_nX_n\beta_0)'M_n\varepsilon_n + (\lambda_0-\lambda)\frac{1}{n}(G_nX_n\beta_0)'M_nG_n\varepsilon_n \tag{34}$$

and by Lemma A.2 in Lee (2002) and linearity of $H_{1n}(\lambda,\gamma)$ in λ, we have $H_{1n}(\lambda,\gamma) = o_p(1)$ uniformly in $(\lambda,\gamma)\in\Lambda\otimes\Gamma$. Next,

$$\begin{aligned} H_{2n}(\lambda,\gamma) - \sigma_n^2(\lambda,\gamma) &= \frac{1}{n}\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)M_nS_n(\lambda,\gamma)S_n^{-1}\varepsilon_n - \sigma_n^2(\lambda,\gamma) \\ &= \frac{1}{n}\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1}\varepsilon_n \\ &- \frac{\sigma_0^2}{n}tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1}) - H_{3n}(\lambda,\gamma) \end{aligned} \tag{35}$$

where $H_{3n}(\lambda,\gamma) = \frac{1}{n}\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)X_n(X_n'X_n)^{-1}X_n'S_n(\lambda,\gamma)S_n^{-1}\varepsilon_n$. Note that, by Lemma A.2 in Lee (2002), we have

$$\frac{1}{\sqrt{n}}X_n'S_n(\lambda,\gamma)S_n^{-1}\varepsilon_n = \frac{1}{\sqrt{n}}X_n'S_n^{-1}\varepsilon_n - \frac{\lambda}{\sqrt{n}}X_n'G_n\varepsilon_n = O_p(1) \tag{36}$$

Therefore,

$H_{3n}(\lambda,\gamma) = \frac{1}{n}[(\frac{1}{\sqrt{n}}X_n'S_n(\lambda,\gamma)S_n^{-1}\varepsilon_n)'(\frac{X_n'X_n}{n})^{-1}(\frac{1}{\sqrt{n}}X_n'S_n(\lambda,\gamma)S_n^{-1}\varepsilon_n)] = o_p(1)$. Finally, by Lemma A.12 in Lee (2004b),

$$\frac{1}{n}[\varepsilon_n'S_n'^{-1}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1}\varepsilon_n - \sigma_0^2tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})] = o_p(1) \tag{37}$$

uniformly in $(\lambda,\gamma)\in\Lambda\otimes\Gamma$. Consequently, we have $H_{2n}(\lambda,\gamma) - \sigma_n^2(\lambda,\gamma) = o_p(1)$. We have shown earlier that $H_{1n}(\lambda,\gamma) = o_p(1)$, therefore, $\hat{\sigma}_n^2(\lambda,\gamma) - \sigma_n^{*2}(\lambda,\gamma) = o_p(1)$ uniformly on $\Lambda\otimes\Gamma$.

Next, we show that $|\ln\hat{\sigma}_n^2(\lambda,\gamma) - \ln\sigma_n^{*2}(\lambda,\gamma)| = o_p(1)$. Expand the Taylor series, $|\ln\hat{\sigma}_n^2(\lambda,\gamma) - \ln\sigma_n^{*2}(\lambda,\gamma)| = \frac{|\hat{\sigma}_n^2(\lambda,\gamma)-\sigma_n^{*2}(\lambda,\gamma)|}{\tilde{\sigma}_n^2(\lambda,\gamma)}$, where $\tilde{\sigma}_n^2(\lambda,\gamma)$ lies between $\hat{\sigma}_n^2(\lambda,\gamma)$ and $\sigma_n^{*2}(\lambda,\gamma)$. We have shown above that $\sigma_n^2(\lambda,\gamma)$ is uniformly bounded away from zero on $\Lambda\otimes\Gamma$, then $\sigma_n^{*2}(\lambda,\gamma)$ is also uniformly bounded away from zero on $\Lambda\otimes\Gamma$. This is because $\sigma_n^{*2}(\lambda,\gamma)\geq\sigma_n^2(\lambda,\gamma)$ as

$\sigma_n^{*2}(\lambda,\gamma) = (\lambda_0-\lambda)^2\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \frac{\sigma_0^2}{n}tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})$

$= (\lambda_0 - \lambda)^2 \frac{1}{n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + \sigma_n^2(\lambda, \gamma)$.

Besides, as we have shown that $\hat{\sigma}_n^2(\lambda,\gamma) - \sigma_n^{*2}(\lambda,\gamma) = o_p(1)$ uniformly on $\Lambda \otimes \Gamma$, and $\sigma_n^{*2}(\lambda,\gamma)$ is uniformly bounded away from zero on $\Lambda \otimes \Gamma$, then so is $\hat{\sigma}_n^2(\lambda,\gamma)$. Finally, these yield $|\ln \hat{\sigma}_n^2(\lambda,\gamma) - \ln \sigma_n^{*2}(\lambda,\gamma)| = o_p(1)$ uniformly on $\Lambda \otimes \Gamma$ and, hence, $\sup_{(\lambda,\gamma)\in\Lambda\otimes\Gamma} |\frac{1}{n}\ln L_n(\lambda,\gamma) - \frac{1}{n}Q_n(\lambda,\gamma)| = o_p(1)$.

We have proved that identifiable uniqueness holds and that $\frac{1}{n}\ln L_n(\lambda,\gamma) - \frac{1}{n}Q_n(\lambda,\gamma)$
converges in probability to zero uniformly on $\Lambda \otimes \Gamma$. Consequently, the consistency of $\hat{\lambda}_n$ and $\hat{\gamma}_n$, and thus, $\hat{\theta}_n$ follow. Q.E.D.

7.2.7 Proof of Theorem 3: Asymptotic Normality

To prove the asymptotic normality of the QML estimator $\hat{\theta}_n$, we need to show that
$\Sigma_\theta = -\lim_{n\to\infty} E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})$ is nonsingular,
$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} 0$,
and $\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) \xrightarrow{p} 0$.

Nonsingularity of Σ_θ First we show that Σ_θ is nonsingular. Let $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)'$ be a column vector of constants such that $\Sigma_\theta \alpha = 0$. Here we need to show that $\alpha = 0$. From the first row of the linear equation system $\Sigma_\theta\alpha = 0$ using the definition of Σ_θ, we have

$$0 = \lim_{n\to\infty} \frac{1}{\sigma_0^2 n} X_n' X_n \alpha_1 + \frac{1}{\sigma_0^2 n} \lim_{n\to\infty} X_n' G_n X_n \beta_0 \alpha_2 + \lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n} X_n'(T_n X_n \beta_0)\alpha_3.$$

Consequently,

$$\alpha_1 = -\lim_{n\to\infty} (X_n'X_n)^{-1} X_n'(G_n X_n \beta_0)\alpha_2 - \lim_{n\to\infty} \lambda_0 (X_n'X_n)^{-1} X_n'(T_n X_n \beta_0)\alpha_3. \quad (38)$$

From the fourth equation of the linear system, we have

$$0 = \lim_{n\to\infty} \frac{1}{\sigma_0^2 n} tr(G_n)\alpha_2 + \lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n} tr(T_n)\alpha_3 + \lim_{n\to\infty} \frac{1}{2\sigma_0^4}\alpha_4.$$

Rearrange the terms and solve for α_4, we get

$$\alpha_4 = -\lim_{n\to\infty} \frac{2\sigma_0^2}{n} tr(G_n)\alpha_2 - \lim_{n\to\infty} \frac{2\lambda_0\sigma_0^2}{n} tr(T_n)\alpha_3. \quad (39)$$

From the second equation of the linear system, we have

$$\begin{aligned} 0 = &\lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_n X_n \beta_0)' X_n \alpha_1 + \lim_{n\to\infty} [\frac{1}{\sigma_0^2 n}(G_n X_n\beta_0)'(G_n X_n \beta_0) + \frac{1}{n} tr(G_n^S G_n)]\alpha_2 \\ &+ \lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}[(G_n X_n \beta_0)'(T_n X_n \beta_0) + \sigma_0^2 tr(G_n^S T_n)]\alpha_3 + \lim_{n\to\infty} \frac{1}{\sigma_0^2 n} tr(G_n)\alpha_4 \end{aligned}$$

for $\lambda_0 \neq 0$ and $G_n^S = G_n + G_n'$. Substituting α_1 in (38) and α_4 in (39) into the above equation, we get

$$\begin{aligned} 0 = &\left[\lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + \lim_{n\to\infty} \frac{1}{n}[tr(G_n^S G_n) - \frac{2}{n} tr^2(G_n)]\right]\alpha_2 \\ &+ \left[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(G_n X_n \beta_0)' M_n (T_n X_n \beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(G_n^S T_n) - \frac{2}{n} tr(G_n) tr(T_n)]\right]\alpha_3. \quad (40) \end{aligned}$$

Rearrange the terms and solve for α_2

$$\begin{aligned}\alpha_2 = &- \left[\lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty} \frac{1}{n}[tr(G_n^SG_n) - \frac{2}{n}tr^2(G_n)]\right]^{-1} \\ &\times \left[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(G_n^ST_n) - \frac{2}{n}tr(G_n)tr(T_n)]\right]\alpha_3 \end{aligned} \tag{41}$$

Note that the inverse in equation (41) above exists as Assumption 10 implies that $\lim_{n\to\infty} \frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ is positive and that $[tr(G_n^SG_n) - \frac{2}{n}tr^2(G_n)] = \frac{1}{2}tr[(\Psi_n' + \Psi_n)(\Psi_n' + \Psi_n)'] \geq 0$, where $\Psi_n = G_n - (tr(G_n)/n)I_n$ (Lee, 2004a).

From the third equation of the linear system, we have

$$\begin{aligned}0 = &\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(T_nX_n\beta_0)'X_n\alpha_1 + \lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}[(T_nX_n\beta_0)'(G_nX_n\beta_0) + \sigma_0^2 tr(T_n^SG_n)]\alpha_2 \\ &+ \lim_{n\to\infty} \frac{\lambda_0^2}{\sigma_0^2 n}[(T_nX_n\beta_0)'(T_nX_n\beta_0) + \sigma_0^2 tr(T_n^ST_n)]\alpha_3 + \lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}tr(T_n)\alpha_4\end{aligned}$$

where $T_n^S = T_n + T_n'$. Substituting α_1 and α_4 into the above equation, we get

$$\begin{aligned}0 = &\left[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(T_n^SG_n) - \frac{2}{n}tr(T_n)tr(G_n)]\right]\alpha_2 \\ &+ \left[\lim_{n\to\infty} \frac{\lambda_0^2}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0^2}{n}[tr(T_n^ST_n) - \frac{2}{n}tr^2(T_n)]\right]\alpha_3\end{aligned}$$

Rearrange the terms and solve for α_2

$$\begin{aligned}\alpha_2 = &- \left[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(T_n^SG_n) - \frac{2}{n}tr(T_n)tr(G_n)]\right]^{-1} \\ &\times \left[\lim_{n\to\infty} \frac{\lambda_0^2}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0^2}{n}[tr(T_n^ST_n) - \frac{2}{n}tr^2(T_n)]\right]\alpha_3. \end{aligned} \tag{42}$$

The inverse in equation (42) above exists for $\lambda_0 \neq 0$ as Assumption 10 implies that $\lim_{n\to\infty} \frac{1}{n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ exists. If this limit is positive, then the sum of the terms in the inverse will exist and be positive whereas if this limit is negative, then the sum of the terms in the inverse will exist and be either negative or positive. Finally, combine equations (41) with (42), to get

$$\begin{aligned}0 = \Bigg\{ &\left[\lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty} \frac{1}{n}[tr(G_n^SG_n) - \frac{2}{n}tr^2(G_n)]\right]^{-1} \\ &\times \left[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(G_n^ST_n) - \frac{2}{n}tr(G_n)tr(T_n)]\right] && (43) \\ &- \left[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(T_n^SG_n) - \frac{2}{n}tr(T_n)tr(G_n)]\right]^{-1} && (44) \\ &\times \left[\lim_{n\to\infty} \frac{\lambda_0^2}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0^2}{n}[tr(T_n^ST_n) - \frac{2}{n}tr^2(T_n)]\right]\Bigg\}\alpha_3 && (45)\end{aligned}$$

We show that the products of the above equation are nonzero. First of all, Assumption 10 implies that $\lim_{n\to\infty} \frac{1}{n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ exists, and $\lim_{n\to\infty} \frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ and $\lim_{n\to\infty} \frac{1}{n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0)$ are positive. As stated earlier, because $[tr(G_n^SG_n) - \frac{2}{n}tr^2(G_n)] = \frac{1}{2}tr[(\Psi_n' + \Psi_n)(\Psi_n' + \Psi_n)'] \geq 0$, then the first and fourth lines of equation (46) above are positive while the second and third lines exist and can be either positive or negative.

Next, as the limits above are scalars, rearrange the terms to eliminate the inverses as follows.

$$\begin{aligned}0 =& \Big\{ \Big[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(G_n X_n \beta_0)' M_n (T_n X_n \beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(G_n^S T_n) - \frac{2}{n} tr(G_n) tr(T_n)]\Big] \\ &\times \Big[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(T_n X_n \beta_0)' M_n (G_n X_n \beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(T_n^S G_n) - \frac{2}{n} tr(T_n) tr(G_n)]\Big] \\ &- \Big[\lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) + \lim_{n\to\infty} \frac{1}{n}[tr(G_n^S G_n) - \frac{2}{n} tr^2(G_n)]\Big] \\ &\times \Big[\lim_{n\to\infty} \frac{\lambda_0^2}{\sigma_0^2 n}(T_n X_n \beta_0)' M_n (T_n X_n \beta_0) + \lim_{n\to\infty} \frac{\lambda_0^2}{n}[tr(T_n^S T_n) - \frac{2}{n} tr^2(T_n)]\Big] \Big\} \alpha_3. \end{aligned} \tag{46}$$

Recall that $G_n = W_n S_n^{-1}$ and $T_n = Z_n S_n^{-1}$, where Z_n is the first order derivative of W_n with respect to γ and $Z_n \neq W_n$, the product of the first two lines is not equal to the product of the third and fourth lines. Thus, α_3 must be zero. This leads to $\alpha_2 = 0$ and, consequently, $\alpha = 0$.

$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} 0$ In this subsection we show that $\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}$ converges in probability to zero. In other words, we show that differences between the second-order derivatives of the log-likelihood function at $\hat{\theta}_n$ and θ_0 with respect to each parameter converge in probability to zero. The second-order derivatives, which are assumed to exist and be continuous in the neighbourhood of θ_0, for each parameter are as follows.

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2} X_n' X_n, \tag{47}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\lambda} = -\frac{1}{\sigma^2} X_n' W_n(\gamma) Y_n, \tag{48}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\gamma} = -\frac{\lambda}{\sigma^2} X_n' Z_n(\gamma) Y_n, \tag{49}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4} X_n' \varepsilon_n(\delta), \tag{50}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda^2} = -tr(G_n^2(\lambda,\gamma)) - \frac{1}{\sigma^2} Y_n' W_n'(\gamma) W_n(\gamma) Y_n, \tag{51}$$

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\gamma} =& - tr(T_n(\lambda,\gamma)) - \lambda tr(G_n(\lambda,\gamma) T_n(\lambda,\gamma)) \\ &- \frac{\lambda}{\sigma^2} Y_n' Z_n'(\gamma) W_n(\gamma) Y_n, \end{aligned} \tag{52}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\sigma^2} = -\frac{1}{\sigma^4} Y_n' W_n'(\gamma) \varepsilon_n(\delta), \tag{53}$$

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial\gamma^2} =& - \lambda tr(C_n(\lambda,\gamma)) - \lambda^2 tr(T_n^2(\lambda,\gamma)) \\ &- \frac{\lambda^2}{\sigma^2} Y_n' Z_n'(\gamma) Z_n(\gamma) Y_n, \end{aligned} \tag{54}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\gamma\partial\sigma^2} = -\frac{\lambda}{\sigma^4} Y_n' Z_n'(\gamma)\varepsilon_n(\delta), \tag{55}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\varepsilon_n'(\delta)\varepsilon_n(\delta) \tag{56}$$

We now show that the differences between each of the above derivatives at $\hat{\theta}_n$ and their counterparts at θ_0 converge in probability to zero. First, as $\frac{1}{n}X_n'X_n = O(1)$ and $\hat{\theta}_n \xrightarrow{p} \theta_0$, the difference between (47) at $\hat{\theta}_n$ and its counterpart at θ_0 becomes

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\beta\partial\beta'} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\beta\partial\beta'} = (\frac{1}{\sigma_0^2} - \frac{1}{\hat{\sigma}_n^2})\frac{X_n'X_n}{n} = o_p(1).$$

Next, the difference between (48) at $\hat{\theta}_n$ and at θ_0 is

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\beta\partial\lambda} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\beta\partial\lambda} = \frac{1}{\sigma_0^2}\frac{X_n'W_nY_n}{n} - \frac{1}{\hat{\sigma}_n^2}\frac{X_n'W_n(\hat{\gamma}_n)Y_n}{n} \tag{57}$$

To show that $\frac{1}{n}X_n'W_n(\hat{\gamma}_n)Y_n \xrightarrow{p} \frac{1}{n}X_n'W_nY_n$, we use the mean value theorem for a vector-valued function. Then, for $\bar{\gamma}_n$ that lies between $\hat{\gamma}_n$ and γ_0, we have

$$\left|\left|\frac{X_n'W_n(\hat{\gamma}_n)Y_n}{n} - \frac{X_n'W_nY_n}{n}\right|\right| \leq \sup_{\gamma\in\Gamma}\left|\left|\frac{X_n'Z_n(\bar{\gamma}_n)Y_n}{n}\right|\right||\hat{\gamma}_n - \gamma_0| \tag{58}$$

where $Z_n(\gamma)$ is the first-order derivative of $W_n(\gamma)$ and $||\cdot||$ is a matrix norm. As $\frac{1}{n}X_n'Z_n(\bar{\gamma}_n)Y_n = O_p(1)$ and $\hat{\gamma}_n \xrightarrow{p} \gamma_0$, $\left|\left|\frac{X_n'W_n(\hat{\gamma}_n)Y_n}{n} - \frac{X_n'W_nY_n}{n}\right|\right| \xrightarrow{p} 0$. This implies that $\frac{1}{n}X_n'W_n(\hat{\gamma}_n)Y_n \xrightarrow{p} \frac{1}{n}X_n'W_nY_n$. Then, (57) above becomes

$$\begin{aligned}\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta})}{\partial\beta\partial\lambda} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\beta\partial\lambda} &= \frac{1}{\sigma_0^2}\frac{X_n'W_nY_n}{n} - \frac{1}{\hat{\sigma}_n^2}\frac{X_n'W_nY_n}{n} + o_p(1)\\ &= (\frac{1}{\sigma_0^2} - \frac{1}{\hat{\sigma}_n^2})\frac{X_n'W_nY_n}{n} + o_p(1) = o_p(1).\end{aligned}$$

Next, for (49), we first show that

$$\left|\left|\frac{X_n'Z_n(\hat{\gamma}_n)Y_n}{n} - \frac{X_n'Z_nY_n}{n}\right|\right| \leq \sup_{\gamma\in\Gamma}\left|\left|\frac{X_n'A_n(\bar{\gamma}_n)Y_n}{n}\right|\right||\hat{\gamma}_n - \gamma_0| = o_p(1) \tag{59}$$

where $A_n(\gamma) = \frac{\partial Z_n(\gamma)}{\partial\gamma}$ and $\frac{1}{n}X_n'A_n(\bar{\gamma}_n)Y_n = O_p(1)$. Therefore,

$$\begin{aligned}\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\beta\partial\gamma} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\beta\partial\gamma} &= \frac{\lambda_0}{\sigma_0^2}\frac{X_n'Z_nY_n}{n} - \frac{\hat{\lambda}_n}{\hat{\sigma}_n^2}\frac{X_n'Z_n(\hat{\gamma}_n)Y_n}{n}\\ &= (\frac{\lambda_0}{\sigma_0^2} - \frac{\hat{\lambda}_n}{\hat{\sigma}_n^2})\frac{X_n'Z_nY_n}{n} + o_p(1) = o_p(1).\end{aligned}$$

For the above equation, note that as $\hat{\lambda}_n \xrightarrow{p} \lambda_0$ and $\hat{\sigma}_n^2 \xrightarrow{p} \sigma_0^2$, the continuous mapping theorem implies that $\frac{\hat{\lambda}_n}{\hat{\sigma}_n^2} \xrightarrow{p} \frac{\lambda_0}{\sigma_0^2}$, provided that σ_0^2 and $\hat{\sigma}_n^2$ are nonzero. Further, for (50), we first look at the following equation.

$$\varepsilon_n(\delta_n) = Y_n - X_n\beta_n - \lambda_nW_n(\gamma_n)Y_n = X_n(\beta_0 - \beta_n) + [\lambda_0W_n - \lambda_nW_n(\gamma_n)]Y_n + \varepsilon_n,$$

where $\delta_n = (\beta_n', \lambda_n, \gamma_n)'$. Substitute this equation into (50) and as we have shown in (58) that $\frac{1}{n}X_n'W_n(\hat{\gamma}_n)Y_n \xrightarrow{p} \frac{1}{n}X_n'W_nY_n$, the difference of (50) evaluated at $\hat{\theta}_n$ and θ_0 becomes

$$\begin{aligned}
&\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\beta\partial\sigma^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\beta\partial\sigma^2} \\
&= (\frac{1}{\sigma_0^4} - \frac{1}{\hat{\sigma}_n^4})\frac{X_n'\varepsilon_n}{n} + \frac{X_n'X_n}{\hat{\sigma}_n^4 n}(\hat{\beta}_n - \beta_0) + \frac{1}{\hat{\sigma}_n^4 n}[\hat{\lambda}_n X_n'W_n(\hat{\gamma}_n)Y_n - \lambda_0 X_n'W_nY_n] \\
&= (\frac{1}{\sigma_0^4} - \frac{1}{\hat{\sigma}_n^4})\frac{X_n'\varepsilon_n}{n} + \frac{X_n'X_n}{\hat{\sigma}_n^4 n}(\hat{\beta}_n - \beta_0) + (\hat{\lambda}_n - \lambda_0)\frac{X_n'W_nY_n}{\hat{\sigma}_n^4 n} + o_p(1) = o_p(1)
\end{aligned}$$

for $\hat{\theta}_n \xrightarrow{p} \theta_0$. Next, for (53), we have

$$\begin{aligned}
&\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\lambda\partial\sigma^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda\partial\sigma^2} = \frac{1}{\sigma_0^4 n}Y_n'W_n'\varepsilon_n - \frac{1}{\hat{\sigma}_n^4 n}Y_n'W_n'(\hat{\gamma}_n)\varepsilon_n(\hat{\delta}_n) \\
&= \frac{1}{n}[\frac{Y_n'W_n'\varepsilon_n}{\sigma_0^4} - \frac{Y_n'W_n'(\hat{\gamma}_n)\varepsilon_n}{\hat{\sigma}_n^4}] + \frac{1}{\hat{\sigma}_n^4 n}Y_n'W_n'(\hat{\gamma}_n)X_n(\hat{\beta}_n - \beta_0) \\
&+ \frac{1}{\hat{\sigma}_n^4 n}[\hat{\lambda}_n Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n - \lambda_0 Y_n'W_n'(\hat{\gamma}_n)W_nY_n].
\end{aligned}$$

To show that the difference above converges in probability to zero, we first apply the mean value theorem and show that $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)\varepsilon_n \xrightarrow{p} \frac{1}{n}Y_n'W_n'\varepsilon_n$.

$$\Big|\Big|\frac{Y_n'W_n'(\hat{\gamma}_n)\varepsilon_n}{n} - \frac{Y_n'W_n'\varepsilon_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'Z_n'(\bar{\gamma}_n)\varepsilon_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1). \tag{60}$$

For $\hat{\gamma}_n \xrightarrow{p} \gamma_0$ and $\bar{\gamma}_n$ lies between $\hat{\gamma}_n$ and γ_0, (60) above implies that $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)\varepsilon_n \xrightarrow{p} \frac{1}{n}Y_n'W_n'\varepsilon_n$. Next, we show that $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n$ and $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)W_nY_n$ converge in probability to $\frac{1}{n}Y_n'W_n'W_nY_n$. Apply the mean value theorem for vector-valued function, we have

$$\begin{aligned}
&\Big|\Big|\frac{Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n} - \frac{Y_n'W_n'W_nY_n}{n}\Big|\Big| \\
&\leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'Z_n'(\bar{\gamma}_n)W_n(\bar{\gamma}_n)Y_n}{n} + \frac{Y_n'W_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1)
\end{aligned} \tag{61}$$

and

$$\Big|\Big|\frac{Y_n'W_n'(\hat{\gamma}_n)W_nY_n}{n} - \frac{Y_n'W_n'W_nY_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'Z_n'(\bar{\gamma}_n)W_nY_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1), \tag{62}$$

where $\frac{1}{n}Y_n'Z_n'(\bar{\gamma}_n)W_n(\bar{\gamma}_n)Y_n + \frac{1}{n}Y_n'W_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n$ and $\frac{1}{n}Y_n'Z_n'(\bar{\gamma}_n)W_nY_n$ are $O_p(\frac{1}{h_n})$. Hence, by (58) and $\hat{\theta}_n \xrightarrow{p} \theta_0$, the difference of (53) evaluated at $\hat{\theta}_n$ and θ_0 becomes

$$\begin{aligned}
&\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\lambda\partial\sigma^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda\partial\sigma^2} = (\frac{1}{\sigma_0^4} - \frac{1}{\hat{\sigma}_n^4})\frac{Y_n'W_n'\varepsilon_n}{n} \\
&+ \frac{1}{\hat{\sigma}_n^4 n}Y_n'W_n'X_n(\hat{\beta}_n - \beta_0) + (\hat{\lambda}_n - \lambda_0)\frac{Y_n'W_n'W_nY_n}{\hat{\sigma}_n^4 n} + o_p(1) = o_p(1).
\end{aligned}$$

For (55), the convergence is as follows

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\gamma\partial\sigma^2}-\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\gamma\partial\sigma^2}=\frac{\lambda_0}{\sigma_0^4 n}Y_n'Z_n'\varepsilon_n-\frac{\hat{\lambda}_n}{\hat{\sigma}_n^4 n}Y_n'Z_n'(\hat{\gamma}_n)\varepsilon_n(\hat{\delta})$$
$$=[\frac{\lambda_0}{\sigma_0^4}\frac{Y_n'Z_n'\varepsilon_n}{n}-\frac{\hat{\lambda}_n}{\hat{\sigma}_n^4}\frac{Y_n'Z_n'(\hat{\gamma}_n)\varepsilon_n}{n}]-\frac{\hat{\lambda}_n}{\hat{\sigma}_n^4}Y_n'Z_n'(\hat{\gamma}_n)X_n(\beta_0-\hat{\beta}_n)$$
$$-[\frac{\hat{\lambda}_n\lambda_0}{\hat{\sigma}_n^4}\frac{Y_n'Z_n'(\hat{\gamma}_n)W_nY_n}{n}-\frac{\hat{\lambda}_n^2}{\hat{\sigma}_n^4}\frac{Y_n'Z_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n}]$$

The same intuition as in (53) above applies here as well. By the mean value theorem, we first show that $\frac{1}{n}Y_n'Z_n'(\hat{\gamma}_n)\varepsilon_n \xrightarrow{p} \frac{1}{n}Y_n'Z_n'\varepsilon_n$.

$$\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)\varepsilon_n}{n}-\frac{Y_n'Z_n'\varepsilon_n}{n}\Big|\Big|\le\sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)\varepsilon_n}{n}\Big|\Big||\hat{\gamma}_n-\gamma_0|=o_p(1). \tag{63}$$

Then we show that $\frac{1}{n}Y_n'Z_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n$ and $\frac{1}{n}Y_n'Z_n'(\hat{\gamma}_n)W_nY_n$ converge in probability to $\frac{1}{n}Y_n'Z_n'W_nY_n$. By the mean value theorem,

$$\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)W_nY_n}{n}-\frac{Y_n'Z_n'W_nY_n}{n}\Big|\Big|\le\sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)W_nY_n}{n}\Big|\Big||\hat{\gamma}_n-\gamma_0|=o_p(1) \tag{64}$$

and

$$\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n}-\frac{Y_n'Z_n'W_nY_n}{n}\Big|\Big| \tag{65}$$
$$\le\sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)W_n(\bar{\gamma}_n)Y_n}{n}+\frac{Y_n'Z_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n-\gamma_0|=o_p(1),$$

where $\frac{1}{n}Y_n'A_n'(\hat{\gamma}_n)W_nY_n$ and $\frac{1}{n}Y_n'A_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n+\frac{1}{n}Y_n'Z_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n$ are $O_p(\frac{1}{h_n})$. Then, with (63) - (65) and (59), the convergence of (55) becomes

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\gamma\partial\sigma^2}-\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\gamma\partial\sigma^2}=(\frac{\lambda_0}{\sigma_0^4}-\frac{\hat{\lambda}_n}{\hat{\sigma}_n^4})\frac{Y_n'Z_n'\varepsilon_n}{n}-\frac{\hat{\lambda}_n}{\hat{\sigma}_n^4}Y_n'Z_n'X_n(\beta_0-\hat{\beta}_n)$$
$$-(\hat{\lambda}_n\lambda_0-\hat{\lambda}_n^2)\frac{Y_n'Z_n'W_nY_n}{\hat{\sigma}_n^4 n}+o_p(1)=o_p(1).$$

Note that by the continuous mapping theorem and $\hat{\theta}_n \xrightarrow{p} \theta_0$, we have $\frac{\hat{\lambda}_n}{\hat{\sigma}_n^4} \xrightarrow{p} \frac{\lambda_0}{\sigma_0^4}$ and $\hat{\lambda}_n^2 \xrightarrow{p} \hat{\lambda}_n\lambda_0$, and the above difference converges in probability to zero.

For (51), (52) and (54), the second-order derivatives involve the trace of matrices $G_n^2(\lambda,\gamma)$, $T_n(\lambda,\gamma)$, $G_n(\lambda,\gamma)T_n(\lambda,\gamma)$, $C_n(\lambda,\gamma)$, and $T_n^2(\lambda,\gamma)$. Note that $G_n(\lambda,\gamma)=W_n(\gamma)S_n^{-1}(\lambda,\gamma)$, $T_n(\lambda,\gamma)=Z_n(\gamma)S_n^{-1}(\lambda,\gamma)$, and $C_n(\lambda,\gamma)=A_n(\gamma)S_n^{-1}(\lambda,\gamma)$. The difference between the second-order derivatives in (51) at $\hat{\theta}_n$ and θ_0 is

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\lambda^2}-\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda^2}=\frac{1}{\sigma_0^2}\frac{Y_n'W_n'W_nY_n}{n}-\frac{1}{\hat{\sigma}_n^2}\frac{Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n}$$
$$+\frac{1}{n}tr(G_n^2)-\frac{1}{n}tr(G_n^2(\hat{\lambda}_n,\hat{\gamma}_n)).$$

As we have already shown in (61), $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n \xrightarrow{p} \frac{1}{n}Y_n'W_n'W_nY_n$. Next, we apply the mean value theorem to show that the differences between these traces at $\hat{\theta}_n$ and θ_0 are $o_p(1)$. Let $\bar{\lambda}_n$ lie between

$\hat{\lambda}_n$ and λ_0, and $\bar{\gamma}_n$ between $\hat{\gamma}_n$ and γ_0, respectively. By the mean value theorem,

$$tr(G_n^2(\hat{\lambda}_n,\hat{\gamma}_n)) - tr(G_n^2) = 2tr(G_n^3(\bar{\lambda}_n,\bar{\gamma}_n))[\hat{\lambda}_n - \lambda_0] + 2tr(G_n(\bar{\lambda}_n,\bar{\gamma}_n)T_n(\bar{\lambda}_n,\bar{\gamma}_n) + \bar{\lambda}_n G_n^2(\bar{\lambda}_n,\bar{\gamma}_n)T_n(\bar{\lambda}_n,\bar{\gamma}_n))[\hat{\gamma}_n - \gamma_0].$$

As $G_n(\bar{\lambda}_n,\bar{\gamma}_n)$ is uniformly bounded in both row and column sums uniformly in a neighbourhood of λ_0 and γ_0 by Assumption 8, then $tr(G_n^3(\bar{\lambda}_n,\bar{\gamma}_n)) = O(\frac{n}{h_n})$. Further, Lemma A.8 in Lee (2004b) implies that $tr(G_n(\bar{\lambda}_n,\bar{\gamma}_n)T_n(\bar{\lambda}_n,\bar{\gamma}_n) = O(\frac{n}{h_n})$ and $tr(G_n^2(\bar{\lambda}_n,\bar{\gamma}_n)T_n(\bar{\lambda}_n,\bar{\gamma}_n)) = O(\frac{n}{h_n})$. Since $\hat{\lambda}_n \xrightarrow{p} \lambda_0$ and $\hat{\gamma}_n \xrightarrow{p} \gamma_0$, all trace terms on the right hand side of the above equation become $o_p(1)$. Then, the difference of the second-order derivatives in (51) becomes

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\lambda^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda^2} = (\frac{1}{\sigma_0^2} - \frac{1}{\hat{\sigma}_n^2})\frac{Y_n'W_n'W_nY_n}{n} + o_p(1) = o_p(1).$$

For (52), the same technique applies. By the mean value theorem,

$$tr(T_n(\hat{\lambda}_n,\hat{\gamma}_n)) - tr(T_n) = tr(T_n(\bar{\lambda}_n,\bar{\gamma}_n)G_n(\bar{\lambda}_n,\bar{\gamma}_n))[\hat{\lambda}_n - \lambda_0] + tr(C_n(\bar{\lambda}_n,\bar{\gamma}_n) + \bar{\lambda}_n T_n^2(\bar{\lambda}_n,\bar{\gamma}_n))[\hat{\gamma}_n - \gamma_0]$$

and

$$tr(\hat{\lambda}_n G_n(\hat{\lambda}_n,\hat{\gamma}_n)T_n(\hat{\lambda}_n,\hat{\gamma}_n)) - tr(\lambda_0 G_n T_n) = tr(G_n(\bar{\lambda}_n,\bar{\gamma}_n)T_n(\bar{\lambda}_n,\bar{\gamma}_n) + 2\bar{\lambda}_n G_n^2(\bar{\lambda}_n,\bar{\gamma}_n)T_n(\bar{\lambda}_n,\bar{\gamma}_n))[\hat{\lambda}_n - \lambda_0] + \bar{\lambda}_n tr(T_n^2(\bar{\lambda}_n,\bar{\gamma}_n) + 2\bar{\lambda}_n G_n(\bar{\lambda}_n,\bar{\gamma}_n)T_n^2(\bar{\lambda}_n,\bar{\gamma}_n) + G_n(\bar{\lambda}_n,\bar{\gamma}_n)C_n(\bar{\lambda}_n,\bar{\gamma}_n))[\hat{\gamma}_n - \gamma_0].$$

Hence, the difference corresponding to (52) that must converge becomes

$$\begin{aligned}&\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\lambda\partial\gamma} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda\partial\gamma} = \frac{\lambda_0}{\sigma_0^2}\frac{Y_n'Z_n'W_nY_n}{n} - \frac{\hat{\lambda}_n}{\hat{\sigma}_n^2}\frac{Y_n'Z_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n}\\
&- \frac{1}{n}[tr(T_n(\bar{\lambda}_n,\bar{\gamma}_n)G_n(\bar{\lambda}_n,\bar{\gamma}_n))(\hat{\lambda}_n - \lambda_0) + tr(C_n(\bar{\lambda}_n,\bar{\gamma}_n) + \bar{\lambda}_n T_n^2(\bar{\lambda}_n,\bar{\gamma}_n))(\hat{\gamma}_n - \gamma_0)]\\
&- \frac{1}{n}[tr(G_n(\bar{\lambda}_n,\bar{\gamma}_n)T_n(\bar{\lambda}_n,\bar{\gamma}_n) + 2\bar{\lambda}_n G_n^2(\bar{\lambda}_n,\bar{\gamma}_n)T_n(\bar{\lambda}_n,\bar{\gamma}_n))(\hat{\lambda}_n - \lambda_0)\\
&+ \bar{\lambda}_n tr(T_n^2(\bar{\lambda}_n,\bar{\gamma}_n) + 2\bar{\lambda}_n G_n(\bar{\lambda}_n,\bar{\gamma}_n)T_n^2(\bar{\lambda}_n,\bar{\gamma}_n) + G_n(\bar{\lambda}_n,\bar{\gamma}_n)C_n(\bar{\lambda}_n,\bar{\gamma}_n))(\hat{\gamma}_n - \gamma_0)].\end{aligned}$$

Since $S_n^{-1}(\lambda,\gamma)$ is uniformly bounded in row and column sums uniformly in a neighbourhood of λ_0 and γ_0, then $tr(C_n(\bar{\lambda}_n,\bar{\gamma}_n)) = O(\frac{n}{h_n})$ by Lemma A.8 in Lee (2004b). Note that as $\hat{\lambda}_n \xrightarrow{p} \lambda_0$ and $\hat{\gamma}_n \xrightarrow{p} \gamma_0$, therefore, the trace terms become $o_p(1)$. As we have already shown in (65) that $\frac{1}{n}Y_n'Z_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n \xrightarrow{p} Y_n'Z_n'W_nY_n$, then

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\lambda\partial\gamma} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda\partial\gamma} = (\frac{\lambda_0}{\sigma_0^2} - \frac{\hat{\lambda}_n}{\hat{\sigma}_n^2})\frac{Y_n'Z_n'W_nY_n}{n} + o_p(1) = o_p(1).$$

Next, for equation (54), apply the mean value theorem to the traces as follows.

$$tr(\hat{\lambda}_n T_n^2(\hat{\lambda}_n,\hat{\gamma}_n)) - tr(\lambda_0 T_n^2) = 2\bar{\lambda}_n tr(T_n^2(\bar{\lambda}_n,\bar{\gamma}_n) + \bar{\lambda}_n T_n^2(\bar{\lambda}_n,\bar{\gamma}_n)G_n(\bar{\lambda}_n,\bar{\gamma}_n))[\hat{\lambda}_n - \lambda_0] + 2\bar{\lambda}_n^2 tr(T_n(\bar{\lambda}_n,\bar{\gamma}_n)C_n(\bar{\lambda}_n,\bar{\gamma}_n) + \bar{\lambda}_n T_n^3(\bar{\lambda}_n,\bar{\gamma}_n))[\hat{\gamma}_n - \gamma_0]$$

and

$$tr(\hat{\lambda}_n C_n(\hat{\lambda}_n, \hat{\gamma}_n)) - tr(\lambda_0 C_n) = tr(C_n(\bar{\lambda}_n, \bar{\gamma}_n) + \bar{\lambda}_n C_n(\bar{\lambda}_n, \bar{\gamma}_n) G_n(\bar{\lambda}_n, \bar{\gamma}_n))[\hat{\lambda}_n - \lambda_0]$$
$$+ \bar{\lambda}_n tr(V_n(\bar{\lambda}_n, \bar{\gamma}_n) + \bar{\lambda}_n C_n(\bar{\lambda}_n, \bar{\gamma}_n) T_n(\bar{\lambda}_n, \bar{\gamma}_n))[\hat{\gamma}_n - \gamma_0]$$

where $V_n(\lambda, \gamma) = B_n(\gamma) S_n^{-1}(\lambda, \gamma)$ and $B_n(\gamma) = \frac{\partial A_n(\gamma)}{\partial \gamma}$. The difference of (54) evaluated at $\hat{\theta}_n$ and θ_0 is

$$\begin{aligned}
&\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial \gamma^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial \gamma^2} = \frac{\lambda_0^2}{\sigma_0^2}\frac{Y_n' Z_n' Z_n Y_n}{n} - \frac{\hat{\lambda}_n^2}{\hat{\sigma}_n^2}\frac{Y_n' Z_n'(\hat{\gamma}_n) Z_n(\hat{\gamma}_n) Y_n}{n} \\
&- \frac{1}{n}[tr(C_n(\bar{\lambda}_n, \bar{\gamma}_n) + \bar{\lambda}_n C_n(\bar{\lambda}_n, \bar{\gamma}_n) G_n(\bar{\lambda}_n, \bar{\gamma}_n))(\hat{\lambda}_n - \lambda_0) \\
&+ \bar{\lambda}_n tr(V_n(\bar{\lambda}_n, \bar{\gamma}_n) + \bar{\lambda}_n C_n(\bar{\lambda}_n, \bar{\gamma}_n) T_n(\bar{\lambda}_n, \bar{\gamma}_n))(\hat{\gamma}_n - \gamma_0)] \\
&- \frac{1}{n}[2\bar{\lambda}_n tr(T_n^2(\bar{\lambda}_n, \bar{\gamma}_n) + 2\bar{\lambda}_n T_n^2(\bar{\lambda}_n, \bar{\gamma}_n) G_n(\bar{\lambda}_n, \bar{\gamma}_n))(\hat{\lambda}_n - \lambda_0) \\
&+ 2\bar{\lambda}_n^2 tr(T_n(\bar{\lambda}_n, \bar{\gamma}_n) C_n(\bar{\lambda}_n, \bar{\gamma}_n) + \bar{\lambda}_n T_n^3(\bar{\lambda}_n, \bar{\gamma}_n))(\hat{\gamma}_n - \gamma_0)].
\end{aligned}$$

Note that the elements of $B_n(\gamma)$ are uniformly bounded by Assumption 5. Next, we show that $\frac{1}{n} Y_n' Z_n'(\hat{\gamma}_n) Z_n(\hat{\gamma}_n) Y_n \xrightarrow{p} \frac{1}{n} Y_n' Z_n' Z_n Y_n$. By the mean value theorem,

$$\begin{aligned}
&\Big|\Big|\frac{Y_n' Z_n'(\hat{\gamma}_n) Z_n(\hat{\gamma}_n) Y_n}{n} - \frac{Y_n' Z_n' Z_n Y_n}{n}\Big|\Big| \\
&\leq \sup_{\gamma \in \Gamma}\Big|\Big|\frac{Y_n' A_n'(\bar{\gamma}_n) Z_n(\bar{\gamma}_n) Y_n}{n} + \frac{Y_n' Z_n'(\bar{\gamma}_n) A_n(\bar{\gamma}_n) Y_n}{n}\Big|\Big| |\hat{\gamma}_n - \gamma_0| = o_p(1)
\end{aligned} \tag{66}$$

where $\frac{1}{n} Y_n' A_n'(\bar{\gamma}_n) Z_n(\bar{\gamma}_n) Y_n + \frac{1}{n} Y_n' Z_n'(\bar{\gamma}_n) A_n(\bar{\gamma}_n) Y_n = O_p(\frac{1}{h_n})$. Hence, the difference of (54) becomes

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial \gamma^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial \gamma^2} = (\frac{\lambda_0^2}{\sigma_0^2} - \frac{\hat{\lambda}_n^2}{\hat{\sigma}_n^2})\frac{Y_n' Z_n' Z_n Y_n}{n} + o_p(1) = o_p(1).$$

Finally, for the last derivative (56), we have

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial (\sigma^2)^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial (\sigma^2)^2} = (\frac{1}{2\hat{\sigma}_n^4} - \frac{1}{2\sigma_0^4}) - [\frac{1}{\hat{\sigma}_n^6}\frac{\varepsilon_n'(\hat{\delta})\varepsilon_n(\hat{\delta})}{n} - \frac{1}{\sigma_0^6}\frac{\varepsilon_n' \varepsilon_n}{n}]$$

where

$$\begin{aligned}
\frac{1}{n}\varepsilon_n'(\hat{\delta}_n)\varepsilon_n(\hat{\delta}_n) &= \frac{\varepsilon_n' \varepsilon_n}{n} + (\hat{\beta}_n - \beta_0)'\frac{X_n' X_n}{n}(\hat{\beta}_n - \beta_0) - 2(\hat{\beta}_n - \beta_0)'\frac{X_n' \varepsilon_n}{n} \\
&+ 2(\hat{\beta}_n - \beta_0)'[\hat{\lambda}_n \frac{X_n' W_n(\hat{\gamma}_n) Y_n}{n} - \lambda_0 \frac{X_n' W_n Y_n}{n}] \\
&+ [\lambda_0^2 \frac{Y_n' W_n' W_n Y_n}{n} - \lambda_0 \hat{\lambda}_n \frac{Y_n' W_n' W_n(\hat{\gamma}_n) Y_n}{n}] \\
&- [\lambda_0 \hat{\lambda}_n \frac{Y_n' W_n'(\hat{\gamma}_n) W_n Y_n}{n} - \hat{\lambda}_n^2 \frac{Y_n' W_n'(\hat{\gamma}_n) W_n(\hat{\gamma}_n) Y_n}{n}] \\
&+ 2[\lambda_0 \frac{Y_n' W_n' \varepsilon_n}{n} - \hat{\lambda}_n \frac{Y_n' W_n'(\hat{\gamma}_n) \varepsilon_n}{n}].
\end{aligned}$$

As $\hat{\theta}_n \xrightarrow{p} \theta_0$ and by equations (58) and (60) - (62), the above equation can be written as

$$\begin{aligned}\frac{1}{n}\varepsilon_n'(\hat{\delta}_n)\varepsilon_n(\hat{\delta}_n) &= \frac{\varepsilon_n'\varepsilon_n}{n} + (\hat{\beta}_n - \beta_0)'\frac{X_n'X_n}{n}(\hat{\beta}_n - \beta_0) + 2(\beta_0 - \hat{\beta}_n)'\frac{X_n'\varepsilon_n}{n} \\ &+ 2(\hat{\lambda}_n - \lambda_0)(\hat{\beta}_n - \beta_0)'\frac{X_n'W_nY_n}{n} + (\lambda_0^2 - \lambda_0\hat{\lambda}_n)\frac{Y_n'W_n'W_nY_n}{n} \\ &- (\lambda_0\hat{\lambda}_n - \hat{\lambda}_n^2)\frac{Y_n'W_n'W_nY_n}{n} + 2(\lambda_0 - \hat{\lambda}_n)\frac{Y_n'W_n'\varepsilon_n}{n} + o_p(1) \\ &= \frac{\varepsilon_n'\varepsilon_n}{n} + o_p(1).\end{aligned}$$

Then the difference of (56) becomes

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial(\sigma^2)^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial(\sigma^2)^2} = (\frac{1}{2\hat{\sigma}_n^4} - \frac{1}{2\sigma_0^4}) + (\frac{1}{\sigma_0^6} - \frac{1}{\hat{\sigma}_n^6})\frac{\varepsilon_n'\varepsilon_n}{n} + o_p(1) = o_p(1).$$

We have now shown that all of the differences between the second-order derivatives at $\hat{\theta}_n$ and those at the true values converge in probability to zero uniformly on $\Lambda \otimes \Gamma$.

$\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) \xrightarrow{p} 0$ For the final step, we show that $\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})$ converges in probability to zero. By Lemma A.2 in Lee (2002), we have $\frac{1}{n}X_n'G_n\varepsilon_n = o_p(1)$, $\frac{1}{n}(G_nX_n\beta_0)'\varepsilon_n = o_p(1)$, $\frac{1}{n}(G_nX_n\beta_0)'G_n\varepsilon_n = o_p(1)$, $\frac{1}{n}X_n'T_n\varepsilon_n = o_p(1)$, $\frac{1}{n}(T_nX_n\beta_0)'G_n\varepsilon_n = o_p(1)$, $\frac{1}{n}\varepsilon_n'T_n'(G_nX_n\beta_0) = o_p(1)$, $\frac{1}{n}(T_nX_n\beta_0)'T_n\varepsilon_n = o_p(1)$, $\frac{1}{n}\varepsilon_n'T_n'(T_nX_n\beta_0) = o_p(1)$, and $\frac{1}{n}(T_nX_n\beta_0)'\varepsilon_n = o_p(1)$. It follows that,

$$\begin{aligned}\frac{1}{n}X_n'W_nY_n &= \frac{1}{n}[X_n'(G_nX_n\beta_0) + X_n'G_n\varepsilon_n] = \frac{1}{n}X_n'(G_nX_n\beta_0) + o_p(1), \\ \frac{1}{n}X_n'Z_nY_n &= \frac{1}{n}[X_n'(T_nX_n\beta_0) + X_n'T_n\varepsilon_n] = \frac{1}{n}X_n'(T_nX_n\beta_0) + o_p(1), \\ \frac{1}{n}Y_n'W_n'\varepsilon_n &= \frac{1}{n}[\varepsilon_n'G_n'\varepsilon_n + (G_nX_n\beta_0)'\varepsilon_n] = \frac{1}{n}\varepsilon_n'G_n'\varepsilon_n + o_p(1)\end{aligned}$$

where, by Lemmas A.8 and A.11 in Lee (2004b), and the Law of Large Numbers, $E(\varepsilon_n'G_n'\varepsilon_n) = \sigma_0^2 tr(G_n)$ and

$$var(\frac{1}{n}\varepsilon_n'G_n'\varepsilon_n) = (\frac{\mu_4 - 3\sigma_0^4}{n^2})\Sigma_{i=1}^n G_{n,ii}^2 + \frac{\sigma_0^4}{n^2}[tr(G_nG_n') + tr(G_n^2)] = O(\frac{1}{nh_n}).$$

Next,

$$\begin{aligned}\frac{1}{n}Y_n'W_n'W_nY_n &= \frac{1}{n}[(G_nX_n\beta_0)'(G_nX_n\beta_0) + \varepsilon_n'G_n'G_n\varepsilon_n + (G_nX_n\beta_0)'G_n\varepsilon_n] \\ &= \frac{1}{n}[(G_nX_n\beta_0)'(G_nX_n\beta_0) + \varepsilon_n'G_n'G_n\varepsilon_n] + o_p(1)\end{aligned}$$

with $E(\varepsilon_n'G_n'G_n\varepsilon_n) = \sigma_0^2 tr(G_n'G_n)$ and

$$var(\frac{1}{n}\varepsilon_n'G_n'G_n\varepsilon_n) = (\frac{\mu_4 - 3\sigma_0^4}{n^2})\Sigma_{i=1}^n (G_n'G_n)_{ii}^2 + \frac{2\sigma_0^4}{n^2}tr((G_n'G_n)^2) = O(\frac{1}{nh_n}).$$

Following,

$$\begin{aligned}\frac{1}{n}Y_n'Z_n'W_nY_n &= \frac{1}{n}[(T_nX_n\beta_0)'(G_nX_n\beta_0) + \varepsilon_n'T_n'G_n\varepsilon_n + (T_nX_n\beta_0)'G_n\varepsilon_n + \varepsilon_n'T_n'(G_nX_n\beta_0)] \\ &= \frac{1}{n}[(T_nX_n\beta_0)'(G_nX_n\beta_0) + \varepsilon_n'T_n'G_n\varepsilon_n] + o_p(1)\end{aligned}$$

where $E(\varepsilon_n' T_n' G_n \varepsilon_n) = \sigma_0^2 tr(T_n' G_n)$ and

$$
\begin{aligned}
var(\frac{1}{n}\varepsilon_n' T_n' G_n \varepsilon_n) &= (\frac{\mu_4 - 3\sigma_0^4}{n^2})\Sigma_{i=1}^n (T_n' G_n)_{ii}^2 + \frac{\sigma_0^4}{n^2}[tr((T_n' G_n)(T_n' G_n)') + tr((T_n' G_n)^2)] \\
&= O(\frac{1}{nh_n}).
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n} Y_n' Z_n' Z_n Y_n &= \frac{1}{n}(T_n X_n \beta_0)'(T_n X_n \beta_0) + \frac{1}{n}\varepsilon_n' T_n' T_n \varepsilon_n + \frac{1}{n}(T_n X_n \beta_0)' T_n \varepsilon_n + \frac{1}{n}\varepsilon_n' T_n'(T_n X_n \beta_0) \\
&= \frac{1}{n}(T_n X_n \beta_0)'(T_n X_n \beta_0) + \frac{1}{n}\varepsilon_n' T_n' T_n \varepsilon_n + o_p(1)
\end{aligned}
$$

where $E(\varepsilon_n' T_n' T_n \varepsilon_n) = \sigma_0^2 tr(T_n' T_n)$ and

$$
var(\frac{1}{n}\varepsilon_n' T_n' T_n \varepsilon_n) = (\frac{\mu_4 - 3\sigma_0^4}{n^2})\Sigma_{i=1}^n (T_n' T_n)_{ii}^2 + \frac{2\sigma_0^4}{n^2} tr((T_n' T_n)^2) = O(\frac{1}{nh_n})
$$

Finally,

$$
\frac{1}{n} Y_n' Z_n' \varepsilon_n = \frac{1}{n}\varepsilon_n' T_n' \varepsilon_n + \frac{1}{n}(T_n X_n \beta_0)' \varepsilon_n = \frac{1}{n}\varepsilon_n' T_n' \varepsilon_n + o_p(1)
$$

where $E(\varepsilon_n' T_n' \varepsilon_n) = \sigma_0^2 tr(T_n)$

$$
var(\frac{1}{n}\varepsilon_n' T_n \varepsilon_n) = (\frac{\mu_4 - 3\sigma_0^4}{n^2})\Sigma_{i=1}^n T_{n,ii}^2 + \frac{\sigma_0^4}{n^2}[tr(T_n T_n') + tr(T_n^2)] = O(\frac{1}{nh_n}).
$$

With the above results, we have shown that $\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) \xrightarrow{p} 0$. Hence, from $\sqrt{n}(\hat{\theta}_n - \theta_0) = -(\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta})}{\partial\theta\partial\theta'})^{-1} \cdot \frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}$, the asymptotic distribution of the QMLE $\hat{\theta}_n$ follows. Q.E.D.

7.3 List of Countries

Table 1 below presents a list of 91 countries and their isocodes.[2]

No	Country	Code	No	Country	Code	No	Country	Code
1	Angola	AGO	32	Greece	GRC	63	Pakistan	PAK
2	Argentina	ARG	33	Guatemala	GTM	64	Panama	PAN
3	Australia	AUS	34	Hong Kong	HKG	65	Peru	PER
4	Austria	AUT	35	Honduras	HND	66	Philippines	PHL
5	Burundi	BDI	36	Indonesia	IDN	67	Papua New Guinea	PNG
6	Belgium	BEL	37	India	IND	68	Portugal	PRT
7	Benin	BEN	38	Ireland	IRL	69	Paraguay	PRY
8	Burkina Faso	BFA	39	Israel	ISR	70	Rwanda	RWA
9	Bangladesh	BGD	40	Italy	ITA	71	Senegal	SEN
10	Bolivia	BOL	41	Jamaica	JAM	72	Singapore	SGP
11	Brazil	BRA	42	Jordan	JOR	73	Sierra Leone	SLE
12	Botswana	BWA	43	Japan	JPN	74	El Salvador	SLV
13	Cent. African Rep.	CAF	44	Kenya	KEN	75	Sweden	SWE
14	Canada	CAN	45	Korea, Rep. of	KOR	76	Syria	SYR
15	Congo, Rep. of	COG	46	Sri Lanka	LKA	77	Chad	TCD
16	Switzerland	CHE	47	Morocco	MAR	78	Togo	TGO
17	Chile	CHL	48	Madagascar	MDG	79	Thailand	THA
18	Cote d'Ivoire	CIV	49	Mexico	MEX	80	Trinidad & Tobago	TTO
19	Cameroon	CMR	50	Mali	MLI	81	Tunisia	TUN
20	Colombia	COL	51	Mozambique	MOZ	82	Turkey	TUR
21	Costa Rica	CRI	52	Mauritania	MRT	83	Tanzania	TZA
22	Denmark	DNK	53	Mauritius	MUS	84	Uganda	UGA
23	Dominican Rep.	DOM	54	Malawi	MWI	85	Uruguay	URY
24	Ecuador	ECU	55	Malaysia	MYS	86	USA	USA
25	Egypt	EGY	56	Niger	NER	87	Venezuela	VEN
26	Spain	ESP	57	Nigeria	NGA	88	South Africa	ZAF
27	Ethiopia	ETH	58	Nicaragua	NIC	89	Congo, Dem. Rep.	ZAR
28	Finland	FIN	59	Netherlands	NLD	90	Zambia	ZMB
29	France	FRA	60	Norway	NOR	91	Zimbabwe	ZWE
30	United Kingdom	GBR	61	Nepal	NPL			
31	Ghana	GHA	62	New Zealand	NZL			

Table 1: List of 91 countries and their isocodes.

[2] See http://qed.econ.queensu.ca/jae/2007-v22.6/ertur-koch/ for detail.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Anselin, L. (1988a). *Spatial Econometrics: Methods and Models (Studies in Operational Regional Science)*. Kluwer Academic Publishers.

Anselin, L. and Bera, A.K. (1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. *Handbook of Applied Economics and Statistics*, edited by Ullah, A. and Giles, D.E.A. New York: Marcel Dekker, 237-289.

Bavaud, F. (1998). Models for Spatial Weights: A Systematic Look. *Geographical Analysis*, 30, 153-171.

Billingsley, P. (1995). *Probability and Measure*, 3rd ed. New York: Wiley.

Case, A.C., Rosen, H.S. and Hines, J.R. (1993). Budget Spillovers and Fiscal Policy Interdependence. *Journal of Public Economics*, 52, 285-307.

Cliff, A. and Ord, J.K. (1973). *Spatial Autocorrelation*. London: Pion.

Cliff, A. and Ord, J.K. (1981). *Spatial Processes: Models and Applications*. London: Pion.

Ertur, C. and Koch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics*, 22, 1033-1062.

Getis, A. and Aldstadt, J. (2004). Constructing the Spatial Weights Matrix Using a Local Statistic. *Geographical Analysis*, 36(2), 90-104.

Heston, A., Summers, R. and Aten, B. (2002). *Penn World Tables Version 6.1*. Downloadable dataset. Centre for International Comparisons at the University of Pennsylvania.

Kelejian, H.H. and Prucha, I.R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, 40(2), 509-533.

Kelejian, H.H. and Prucha, I.R. (2001). On the Asymptotic Distribution of the Moran I Test Statistic with Applications. *Journal of Econometrics*, 104, 219-257.

Lee, L.F. (2002). Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models. *Econometric Theory*, 18, 252-277.

Lee, L.F. (2003). Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econometric Reviews*, 22(4), 307-335.

Lee, L.F. (2004a). Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica*, 72(6), 1899-1925.

Lee, L.F. (2004b). A supplement to "Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models". *http://economics.sbs.ohio-state.edu/lee/*.

Lee, L.F. (2007a). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137, 489-514.

Lee, L.F. (2007b). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140, 333-374.

Mankiw, N.G., Romer, D. and Weil, D.N. (1992). A Contribution to the Empirics of Economic Growth. *The Quarterly Journal of Economics*, 107(2), 407-437.

Ord, K. (1975). Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association*, 70(349), 120-126.

Paez, A., Scott, D.M. and Volz, E. (2008). Weight matrices for social influence analysis: An investigation of measurement errors and their effect on model identification and estimation quality. *Social Networks*, 30, 309-317.

Partridge, M.D., Boarnet, M., Brakman, S. and Ottaviano, G. (2012). Introduction: Whither Spatial Econometrics? *Journal of Regional Science*, 52(2), 167-171.

Pinkse, J. and Slade, M.E. (2010). The Future of Spatial Econometrics. *Journal of Regional Science*, 50(1), 103-117.

White, H. (1996). *Estimation, Inference and Specification Analysis*. Econometric Society Monographs No. 22, Cambridge University Press.