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

Saruta Benjanuvatra*

University of York, UK

June 2013

Preliminary version prepared for

VII World Conference of the Spatial Econometrics Association, Washington DC, USA

Abstract

In this paper, we introduce a sub-model for spatial weights and estimate a variable weight matrix for the mixed regressive, spatial autoregressive (MR-SAR) model by maximum Gaussian likelihood. We establish the identifiability of the parameter defining the weights, the consistency as well as the asymptotic normality of the QMLE under appropriate conditions that extend those given by Lee (2004a). Small sample properties are investigated in a Monte Carlo study and the estimator's performance is subsequently compared with other QML estimators using various pre-determined spatial weight matrices. Finally, the QML estimator using two types of sub-models for the spatial weights satisfying the identifiability, consistency and asymptotic normality is applied to cross-sectional dataset used in Ertur and Koch (2007), to study the impact of saving, population growth and neighbourhood on growth.

Keywords: Spatial autoregressive model, spatial weight matrix, maximum likelihood estimation, quasi-maximum likelihood estimator, Monte Carlo, growth

JEL Classification: C13, C15, C21, R15

---

*Correspondence: Saruta Benjanuvatra, Department of Economics and Related Studies, University of York, Heslington, York, YO10 5DD, UK; E-mail: sb652@york.ac.uk.

1 Introduction

One of the most important issues in spatial econometrics is the issue of spatial weight matrix and it has received much attention, particularly in the past few years. Spatial weight matrix captures dependence structure between spatial units (Anselin, 1988). Case et al. (1993) discuss that the distances between neighbours may be geographic, economic or demographic distance. Several forms of the spatial weight matrix have been introduced, see Anselin (1988) and Anselin and Bera (1998) for an overview of the spatial weight matrices. Some well-known weight matrices in the literature are such as the Cliff-Ord weight matrix introduced in Cliff and Ord (1973, 1981) where the weights are a combination of distances and relative border length between units; the weights that are in the form of Rook contiguity where the weights equal 1 if the two regions share common border and 0 otherwise; Queen contiguity where the weights equal 1 if the two regions share common side or vertex and 0 otherwise; inverse distance; or n-nearest neighbours. Others are geostatistical whose form is a function of values derived empirically (Getis and Aldstadt, 2004). Some of the weight matrices in this category are in the forms of Spherical Variogram, Gaussian Variogram, and Exponential Variogram.

Specification of the spatial weights is an important issue as different weight matrices yield different results and, hence, different interpretation of the results. Anselin (1988) discusses that the weights are generally chosen to be exogenous and the parameter values are determined a priori, which may cause spurious correlation if the pre-determined spatial structure is not correctly specified. Moreover, Anselin (1980, 1984) argue that spatial weights should be selected based on spatial interaction theory. Proper choice of the weights improves estimator's efficiency whereas inappropriate choice of the weights creates inefficiency of the estimator (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. Paez et al. (2008) show that errors in the weight matrix can lead to biased estimates. Several researchers have tried to construct the weight matrix using computer software, for example Can (1996) and Aldstadt and

Getis (2006). Others have tried to estimate the weight matrix using different techniques, for instance Geniaux (2012), Souza (2012), and Kelejian and Piras (2012).

In this paper we introduce a sub-model for the spatial weights and estimate a variable spatial weight matrix for the mixed regressive, spatial autoregressive (MR-SAR) model by the maximum Gaussian likelihood. The maximum likelihood estimator in spatial regression models is studied by Ord (1975), Anselin (1988) 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 approach carried out 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 performance of the estimator is subsequently compared with other QML estimators using various fixed spatial weight matrices. Our results show that our QML estimator using a freely estimated weight matrix is able to estimate the parameter defining the spatial weights reasonably well. It outperforms other competing estimators in many cases considered in this paper. Our results also show that using a wrong weight matrix strongly affects the estimation performance of the estimators, especially when estimating the spatial autoregressive parameter.

The applicability to 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, is also illustrated. We apply our QML estimator using these two sub-models for the weights to the cross-sectional data set of 91 countries used in Ertur and Koch (2007) in the framework of the MR-SAR model, to study the impact of saving, population growth and neighbourhood on growth. We evaluate and compare our estimator using freely-estimated spatial weight matrices with other QML estimators using pre-determined weight matrices. Our results show that our QML estimator with freely-estimated weight matrices in the framework of the MR-SAR model is able to capture positive spatial spillovers of growth among countries and provides significant estimates of the parameter defining the weights and other parameters with predicted signs.

This paper is constructed as follows. Section 2 describes the mixed regressive spatial autoregressive model. Assumptions are listed in Section 3. Section 4 analyses the identifiability of the parameters and the consistency of the QML estimator. The asymptotic normality of the QMLE is derived in Section 5. Section 6 explains how the Monte Carlo experiment is conducted and presents the corresponding results. Section 7 discusses the data set used in this paper and describes two sub-models for the spatial weights in the framework of the MR-SAR model. Empirical results for the QMLEs using pre-determined and freely-estimated weight matrices are also presented in this section. Section 8 concludes. Detailed proofs can be found in the Appendix.

2 Mixed Regressive, Spatial Autoregressive Model

The first-order mixed regressive, spatial autoregressive model (Ord, 1975 and Anselin, 1988) is described as follows

$$Y_n = X_n\beta + \lambda W_n(\gamma)Y_n + \varepsilon_n \tag{2.1}$$

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 only ones 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 degree of possible interaction of location j on location i (Ord, 1975). The weight elements are specified as

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

where $w^*_{n,ij}(\gamma) = f(\gamma, d_{ij})$ is a function of distances, d_{ij} is a fixed nonnegative distance between spatial units i and j, γ is a positive scalar parameter defining the weights, and $\sum_j w^*_{n,ij}(\gamma)$ is a row sum for all i. This spatial weight matrix is row-standardised such that $\sum_j w_{n,ij}(\gamma) = 1$ for all i, with zeros on the main diagonal, and the off-diagonal elements take values between 0 and 1. In the case of row-standardisation, the weights can be interpreted as an average of neighbouring values (Anselin and Bera, 1998) and they are perceived as relative values instead of absolute ones. Closer units are given relatively greater weights than farther units. Note that row-standardised matrices are usually asymmetric even though the original matrices, with elements $w^*_{n,ij}(\gamma)$, are symmetric.

The term $W_n(\gamma)Y_n$ in (2.1) is the spatially lagged dependent variable corresponding to the weight matrix $W_n(\gamma)$. A distinct characteristic of this model in spatial econometrics as opposed to time-series context is that $(W_n(\gamma)Y_n)_i$ is correlated not only with ε_i, but also with the error terms at all other locations. The subscript n indicates that each component of the model depends on n, which is the total number of spatial units.

The objective is to estimate $\theta = (\beta', \lambda, \gamma, \sigma^2)'$. Our approach follows the approach in Lee (2004a) and Lee (2002), and we extend their notations as follows. Let $S_n(\lambda, \gamma) = I_n - \lambda W_n(\gamma)$, equation (2.1) becomes

$$\begin{aligned} S_n(\lambda,\gamma)Y_n &= X_n\beta + \varepsilon_n \\ Y_n &= S_n^{-1}(\lambda,\gamma)(X_n\beta + \varepsilon_n) \end{aligned} \tag{2.2}$$

where $S_n(\lambda,\gamma)Y_n$ is a spatially filtered dependent variable. Denote $\theta_0 = (\beta_0', \lambda_0, \gamma_0, \sigma_0^2)'$ 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. The log-likelihood function of equation (2.1) is given by

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \ln|det(S_n(\lambda,\gamma))| - \frac{1}{2\sigma^2}\varepsilon_n'(\delta)\varepsilon_n(\delta) \tag{2.3}$$

where $\varepsilon_n(\delta) = Y_n - X_n\beta - \lambda W_n(\gamma)Y_n$, with $\delta = (\beta', \lambda, \gamma)'$ and $\theta = (\beta', \lambda, \gamma, \sigma^2)'$. Note that the term $\ln|det(S_n(\lambda,\gamma))|$ stands for the natural logarithm of the absolute value of the determinant of $S_n(\lambda,\gamma)$. We take the absolute value of the determinant of $S_n(\lambda,\gamma)$ before taking the logarithm.

The quasi-maximum likelihood estimator is obtained by maximising (2.3) with respect to the parameters. To obtain the concentrated log-likelihood function, we first concentrate out β and σ^2. Then, for given λ and γ, the QMLE of β is

$$\hat{\beta}_n(\lambda,\gamma) = (X_n'X_n)^{-1}(X_n'Y_n - \lambda X_n'W_n(\gamma)Y_n) = (X_n'X_n)^{-1}X_n'S_n(\lambda,\gamma)Y_n. \tag{2.4}$$

Insert this $\hat{\beta}_n(\lambda,\gamma)$ into the first-order derivative of the log-likelihood function with respect to σ_n^2, then the QMLE of σ^2 is given by

$$\begin{aligned}\hat{\sigma}_n^2(\lambda,\gamma) &= \frac{1}{n}[(S_n(\lambda,\gamma)Y_n - X_n\hat{\beta}(\lambda,\gamma))'(S_n(\lambda,\gamma)Y_n - X_n\hat{\beta}(\lambda,\gamma)] \\ &= \frac{1}{n}[Y_n'S_n'(\lambda,\gamma)M_nS_n(\lambda,\gamma)Y_n] \end{aligned} \tag{2.5}$$

where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. Insert (2.4) and (2.5) back into the log-likelihood function and obtain the following concentrated log-likelihood function of λ and γ.

$$\ln L_n(\lambda,\gamma) = -\frac{n}{2}(\ln(2\pi)+1) + \ln|det(S_n(\lambda,\gamma))| - \frac{n}{2}\ln\hat{\sigma}_n^2(\lambda,\gamma). \tag{2.6}$$

To obtain the QMLEs $\hat{\lambda}_n$ and $\hat{\gamma}_n$, maximise (2.6) with respect to λ and γ. Then, the QMLEs of β and σ^2 become $\hat{\beta}_n(\hat{\lambda}_n,\hat{\gamma}_n)$ and $\hat{\sigma}_n^2(\hat{\lambda}_n,\hat{\gamma}_n)$.

3 Assumptions

Before we proceed, we list the assumptions necessary for analysing asymptotic properties of the QML estimator $\hat{\theta}_n$ below.

Assumption 1. *$\varepsilon_1,\ldots,\varepsilon_n$ are independently and identically distributed with mean 0 and finite variance σ^2 for all n. The third and fourth moments of ε_n exist and 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. *Ratio $\frac{h_n}{n} \to 0$ 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 $w_{n,ij}(\gamma) = \frac{e^{-\gamma d_{ij}}}{\sum_j e^{-\gamma d_{ij}}}$ 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.*[1]

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 a basic assumption of the disturbances. 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 (2.4) and (2.5), and their identifiable uniqueness follows 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 (2.1) 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

[1] See Horn and Johnson (1985)

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).

4 Consistency of the QMLE

In this section we 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 (2.2) can be rewritten as

$$Y_n = (I_n + \lambda_0 G_n)(X_n\beta_0 + \varepsilon_n) = X_n\beta_0 + \lambda_0 G_n X_n \beta_0 + S_n^{-1}\varepsilon_n. \tag{4.1}$$

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 will be close to the expected log-likelihood, so we may expect the QML estimator to be close to the maximum of the expected log-likelihood as well.

As already mentioned in Section 3, 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 maintains that if $\ln L_n(\lambda, \gamma)$ is continuous on the parameter space, then the limit function $Q_n(\lambda, \gamma)$ is continuous on the parameter space as well.

Before we continue, we make the following additional assumption.

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

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

where $T_n = Z_nS_n^{-1}$, and $Z_n = \frac{\partial W_n(\gamma_0)}{\partial\gamma}$ is the first-order derivative of the W-matrix at γ_0, the true value of γ. This assumption ensures that $G_nX_n\beta_0$ in (4.1) 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

$$\beta_n^*(\lambda,\gamma) = (X_n'X_n)^{-1}X_n'S_n(\lambda,\gamma)S_n^{-1}X_n\beta_0 \tag{4.2}$$

and

$$\sigma_n^{2*}(\lambda,\gamma) = \frac{1}{n}[(\lambda_0-\lambda)^2(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \sigma_0^2 tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})] \tag{4.3}$$

Substitute (4.2) and (4.3) into the log-likelihood, then we get

$$Q_n(\lambda,\gamma) = -\frac{n}{2}(\ln(2\pi)+1) + \ln|det(S_n(\lambda,\gamma))| - \frac{n}{2}\ln\sigma_n^{2*}(\lambda,\gamma) \tag{4.4}$$

and it is concave and continuous in $\theta\in\Theta$. We establish our theorems below. See the Appendix for detailed proofs of these theorems.

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, $Q_n(\lambda,\gamma)$ is uniquely maximised at θ_0.

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

5 Asymptotic Normality of the QMLE

In this section we analyse the issue of asymptotic normality of the QML estimator $\hat{\theta}_n$. In other words, 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 derived below.

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \beta} = \frac{1}{\sigma_0^2 \sqrt{n}} X_n' \varepsilon_n \tag{5.1}$$
$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \lambda} = \frac{1}{\sigma_0^2 \sqrt{n}} [(G_n X_n \beta_0)' \varepsilon_n + \varepsilon_n' G_n \varepsilon_n - \sigma_0^2 tr(G_n)] \tag{5.2}$$
$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \gamma} = \frac{\lambda_0}{\sigma_0^2 \sqrt{n}} [(T_n X_n \beta_0)' \varepsilon_n + \varepsilon_n' T_n \varepsilon_n - \sigma_0^2 tr(T_n)] \tag{5.3}$$
$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \sigma^2} = \frac{1}{2\sigma_0^4 \sqrt{n}} (\varepsilon_n' \varepsilon_n - n\sigma_0^2) \tag{5.4}$$

These first-order derivatives appear in linear and quadratic forms of ε_n. 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 $G_n X_n \beta_0$ and $T_n X_n \beta_0$ for all n are uniformly bounded by Lemma A.6 in Lee (2004b).

If $\{h_n\}$ is a bounded process, then 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 a divergent process, then we can apply the Kolmogorov's 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

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

where

$$\Omega_{\theta,n} = E(\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta'}) + E(\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'}). \tag{5.5}$$

Introduce $P_n = G_n X_n \beta_0$ and $R_n = T_n X_n \beta_0$, then we have

$$-E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) = \tag{5.6}$$

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

with $G_n^s = G_n + G_n'$ and $T_n^s = T_n + T_n'$, and the matrix $\Omega_{\theta,n}$ is derived as follows.

$$\Omega_{\theta,n} = \tag{5.7}$$

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

The matrix $\Omega_{\theta,n}$ above is symmetric and the asterisks (*) above the main diagonal stand for their symmetric entries with respect to the main diagonal. Note that μ_3 and μ_4 are the third and fourth moments of ε_n, respectively. $G_{n,ij}$ and $T_{n,ij}$ are the (i,j) entries of

G_n and T_n, and $G_{n,i}$ and $x_{n,i}$ are the i-th rows of G_n and X_n, respectively.

If ε_i's are i.i.d.,

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} \xrightarrow{D} N[0, -E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) + \Omega_{\theta,n}] \tag{5.8}$$

and, subsequently,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_{\theta,n}\Sigma_\theta^{-1}], \tag{5.9}$$

with $\Sigma_\theta = -\lim_{n\to\infty} E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})$. Note that Assumption 10 ensures that Σ_θ is nonsingular. For normally distributed ε_i's, $\Omega_{\theta,n}$ disappears and we get

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

and, hence,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1}]. \tag{5.10}$$

Given the above results and assumptions, we state the following theorem.

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

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1}] \tag{5.11}$$

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 ε_i's are normally distributed, then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1}]. \tag{5.12}$$

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 (5.6) and (5.7) can be simplified to

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

where $P_n = G_nX_n\beta_0$ and $R_n = T_nX_n\beta_0$, and

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

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.

6 Monte Carlo Results

6.1 Experiment Design

We investigate small sample properties of our estimator using the freely estimated weight matrix and compare its performance with several QML estimators using a randomly generated weight matrix, and weight matrices with the same weight structure based on pre-determined γ values including the true γ value in a Monte Carlo study. So far we have not specified a functional form of the sub-model for the spatial weights and have kept it general instead. To carry out our Monte Carlo experiment in this section, we now specify a functional form of the sub-model for the spatial weights. Recall that the elements of the row-standardised weight matrix $W_n(\gamma)$ is

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

with $w^*_{n,ij}(\gamma) = f(\gamma, d_{ij})$ a function of distances. We specify this function of distances as follows.

$$w^*_{n,ij}(\gamma) = \begin{cases} 0 & \text{if } i = j \\ e^{-\gamma d_{ij}} & \text{if } i \neq j \end{cases}$$

where γ is a positive scalar parameter specifying the weights and d_{ij} is a fixed nonnegative distance between spatial units i and j. Then, the elements $w_{n,ij}(\gamma)$ of the weight matrix $W_n(\gamma)$ become

$$w_{n,ij}(\gamma) = \begin{cases} 0 & \text{if } i = j \\ \frac{e^{-\gamma d_{ij}}}{\sum_j e^{-\gamma d_{ij}}} \geq 0 & \text{if } i \neq j \end{cases} \tag{6.1}$$

where $\sum_j e^{-\gamma d_{ij}}$ is a row sum for all i. This matrix $W_n(\gamma)$ is row-standardised so the weight elements on the main diagonal are zero whereas all other elements are nonnegative.

We perform Monte Carlo experiments for n = 200, 400 and 800 for 1000 replications. The row-standardised weight matrix is created following equation (6.1), with associated γ values equal to 3, 5, and 7. The distances d_{ij} between units i and j are generated by randomly drawing n pairs of coordinates from a standard uniform distribution from which the Euclidean distances are produced. For each weight matrix, we generate the matrix X_n which consists of 3 columns with associated coefficients; $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$. The first column of the matrix X_n is a vector of ones and the other 2 columns are 1000 independent draws from the standard $n-$ variate Normal distribution. For each W_n and X_n we draw a further independent standard Normal vector of disturbances, of which the variance σ_n^2 is fixed at 1. The matrix $S_n(\lambda, \gamma)$ is generated for each W_n, λ and γ, with $\lambda \in (0.1, 0.3, 0.5, 0.7, 0.9)$.

In the simulations, we impose bounds on λ estimates to be $|\hat{\lambda}| \leq 0.99$, and on γ estimates to be $\hat{\gamma} \geq 0.01$. In the following subsections, we only report the results for a selection of cases. All other results not reported here are available on request.

6.2 Estimates of λ

Table 1 shows the mean, median, standard deviation and root mean squares error of estimates obtained from different estimators of λ for n = 400. The estimators are our QML estimator using the freely estimated weight matrix and 4 competing QML estimators using fixed weight matrices obtained from different values of γ. The results on the left panel of Table 1 are obtained from the DGP based on $\gamma = 5$, $\lambda \in (0.1, 0.3, 0.5, 0.7, 0.9)$ and $\sigma^2 = 1$ for n = 400, while the results on the right panel of the table are obtained from the DGP based on $\gamma = 7$. The first 2 columns list the true values of λ and the weight matrices used for each estimator. The next 4 columns show the mean, median, standard deviation and the RMSE of estimates obtained for each estimator. The structure of the right panel of the table is the same as that of the left panel.

For each true value of λ, the first row shows results for the QML estimator with a fixed and correctly chosen weight matrix, which we use as a benchmark estimator. The second row gives results for our QML estimator with the weight matrix from equation (6.1) which freely estimates the parameter γ that defines the weights. This weight matrix is denoted by W($\hat{\gamma}$) in the table below. The third to fifth rows give results for competing QML estimators using wrongly chosen weight matrices. W(3) and W(7) in the second column of the left panel stand for weight matrices obtained from equation (6.1) with associated values of $\gamma = 3$ and 7, respectively. For each true value of λ, the last competing estimator in the last row uses a fixed weight matrix 'Wrand', which is generated by randomly drawing n pairs of coordinates from a standard bivariate Normal distribution to which the Delaunay routine is applied to produce Voronoi polygons, and subsequently row-standardised. The right panel is constructed in the same way as the left panel, with the benchmark estimator using the weight matrix W(7) and other competing estimators using the weight matrices W(3), W(5) and Wrand, respectively.

Table 1 shows that our QML estimator performs well, producing estimates for λ with smaller bias than any other estimators for most cases across true values of λ. The associated estimates are close to the true values of λ and to the results obtained for the benchmark estimator which uses fixed and correctly chosen weight matrix. Especially for a large true value of λ, $\lambda = 7$ here, our QML estimator is able to estimate λ clearly better than other estimators. Even though the standard deviation and the RMSE are quite large for smaller λ and true values of γ, they decrease significantly when λ increases. Looking at the last row associated with each true value of λ, we see that the estimates obtained for the QML estimator using a randomly generated weight matrix Wrand have the largest bias in most cases. These results show that using a wrongly chosen weight matrix strongly affects the estimates of the spatial autoregressive parameter, λ.

True λ	W	Mean	Med.	St.D.	RMSE	W	Mean	Med.	St.D.	RMSE
0.1	W(5)	-0.040	-0.034	0.523	0.541	W(7)	0.078	0.092	0.274	0.274
	W($\hat{\gamma}$)	-0.204	0.030	0.648	0.715	W($\hat{\gamma}$)	0.018	0.092	0.414	0.422
	W(3)	-0.346	-0.631	0.714	0.842	W(3)	-0.292	-0.527	0.730	0.829
	W(7)	0.023	0.034	0.289	0.299	W(5)	0.039	0.044	0.503	0.506
	Wrand	-0.016	-0.015	0.105	0.156	Wrand	-0.014	-0.010	0.104	0.154
0.3	W(5)	0.196	0.223	0.513	0.523	W(7)	0.344	0.356	0.272	0.275
	W($\hat{\gamma}$)	0.043	0.222	0.627	0.677	W($\hat{\gamma}$)	0.320	0.299	0.342	0.343
	W(3)	-0.065	-0.111	0.761	0.844	W(3)	0.233	0.382	0.738	0.741
	W(7)	0.155	0.165	0.290	0.324	W(5)	0.476	0.532	0.446	0.479
	Wrand	-0.007	-0.001	0.106	0.325	Wrand	-0.008	-0.004	0.105	0.326
0.5	W(5)	0.427	0.479	0.463	0.469	W(7)	0.618	0.625	0.249	0.276
	W($\hat{\gamma}$)	0.303	0.389	0.547	0.582	W($\hat{\gamma}$)	0.575	0.542	0.279	0.289
	W(3)	0.257	0.432	0.725	0.764	W(3)	0.720	0.990	0.512	0.557
	W(7)	0.280	0.287	0.279	0.356	W(5)	0.822	0.990	0.286	0.431
	Wrand	-0.010	-0.004	0.104	0.521	Wrand	-0.012	-0.006	0.102	0.522
0.7	W(5)	0.649	0.793	0.396	0.399	W(7)	0.818	0.872	0.188	0.222
	W($\hat{\gamma}$)	0.536	0.606	0.454	0.482	W($\hat{\gamma}$)	0.734	0.744	0.231	0.233
	W(3)	0.569	0.990	0.619	0.632	W(3)	0.907	0.990	0.282	0.350
	W(7)	0.421	0.422	0.272	0.390	W(5)	0.950	0.990	0.129	0.281
	Wrand	-0.013	-0.009	0.107	0.721	Wrand	-0.011	-0.010	0.102	0.719
0.9	W(5)	0.811	0.990	0.294	0.307	W(7)	0.949	0.990	0.092	0.104
	W($\hat{\gamma}$)	0.708	0.779	0.334	0.385	W($\hat{\gamma}$)	0.874	0.990	0.160	0.162
	W(3)	0.781	0.990	0.450	0.465	W(3)	0.981	0.990	0.085	0.118
	W(7)	0.568	0.581	0.259	0.421	W(5)	0.988	0.990	0.026	0.091
	Wrand	-0.013	-0.011	0.101	0.919	Wrand	-0.014	-0.007	0.106	0.920

Table 1: Estimation of λ for $n = 400$, $\sigma^2 = 1$, and $\lambda \in (0.1, 0.3, 0.5, 0.7, 0.9)$. For the left panel, the true value of $\gamma = 5$ and for the competing estimators, $\gamma = 3$, 7. For the right panel, the true value of $\gamma = 7$ and for the competing estimators, $\gamma = 3$, 5.

6.3 Estimates of γ

Another parameter of interest is γ which defines the spatial weights according to equation (6.1). In Table 2 we report the mean, median, standard deviation and root mean square error of our QML estimator of γ for $n \in (200, 400, 800)$, $\sigma^2 = 1$ and $\lambda \in (0.1, 0.3, 0.5, 0.7, 0.9)$ for true value of $\gamma = 5$.

It is shown that our QML estimator performs reasonably well in estimating γ. For each value of n, bias of the mean and median of the estimates decreases as λ increases. The standard deviation and RMSE also decrease when λ becomes larger. When we compare the results associated with each value of λ across all n, we see that the performance of our QML estimator slightly improves for some values of λ as n becomes larger.

True γ	n	λ	Mean	Med.	St.D.	RMSE
5	200	0.1	5.369	5.349	0.647	0.745
		0.3	5.267	5.320	0.702	0.750
		0.5	5.131	5.169	0.682	0.694
		0.7	5.112	5.127	0.612	0.622
		0.9	5.142	4.989	0.500	0.520
	400	0.1	5.274	5.266	0.635	0.691
		0.3	5.212	5.242	0.627	0.662
		0.5	5.105	5.182	0.616	0.624
		0.7	5.065	5.112	0.578	0.582
		0.9	5.095	4.951	0.468	0.477
	800	0.1	5.341	5.347	0.615	0.703
		0.3	5.230	5.257	0.615	0.657
		0.5	5.099	5.145	0.604	0.612
		0.7	5.004	5.037	0.585	0.585
		0.9	5.115	4.950	0.467	0.480

Table 2: Estimation of γ for the true value of $\gamma = 5$, $n = 200$, 400, 800, $\sigma^2 = 1$, and $\lambda \in (0.1, 0.3, 0.5, 0.7, 0.9)$.

6.4 Estimates of β and σ^2

As β's and σ^2 are not our main parameters of interest, we only discuss a summary of the results below. Detailed results for these 2 parameters are available on request.

For β's, we carry out Monte Carlo experiments for $n = 200$, 400, and 800, $\sigma^2 = 1$, and $\lambda = 0.5$. The true values of β's are $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$, and the true values of γ used in the experiments are 5 and 7, respectively. Our results show that our QML estimator, the benchmark estimator and other competing estimators perform equally well in estimating β_1, β_2 and β_3, except the QML estimator using randomly generated weight matrix Wrand reported on the last row for the true value of β_1, the intercept. The standard deviation and the RMSE are small, except for the estimates for β_1. Note that when the true values of γ and λ are large, or when both are small, all estimators including the benchmark estimator tend to produce estimates for β_1 with larger bias in the mean and median. However, the estimates produced for β_2 and β_3 are still robust across different values of λ and n. For $\gamma = 7$, the results show that there is a larger bias in the mean and median of the estimates for β_1 even when λ is moderate and n is large, whereas the estimates for β_2 and β_3 seem to improve, especially when n is large.

For σ^2, we carry out Monte Carlo experiments for $n = 200$, 400, and 800, $\sigma^2 = 1$, and $\lambda = 0.5$ as well. The true values of γ used are 5 and 7, respectively. The results show that all estimators perform equally well in estimating σ^2 across all values of λ and n, regardless of the weight matrices used in the experiments. The performance of the estimators are comparable, producing estimates with small bias in the mean and median and small standard deviation and RMSE. The standard deviation and the RMSE of the estimates also become smaller when n increases.

7 Empirical Evidence

In this section, we illustrate the applicability of our QML estimator to a real spatial data set using two forms of the sub-models for the spatial weights that satisfy the identifiability, consistency and asymptotic normality established earlier, to study the impact of saving, population growth and neighbourhood on growth. We first discuss the data set used in this paper.

7.1 Data Analysis

The data set used in this paper is obtained from Ertur and Koch (2007)[2], of which the data are originally acquired from the Penn World Tables version 6.1 (Heston et al., 2002). It consists of cross-sectional data of 7 variables for 91 countries for the period of 1960-1995. These countries are from the non-oil sample in Mankiw et al. (1992), see Table 10 in the Appendix for a list of these countries and their ISO codes. Table 3 below presents the variables and their abbreviations used in this paper.

No	**Variable**	**Code**
1	initial level of per worker income (in 1960)	$lny60$
2	level of per worker income in 1995	$lny95$
3	average rate of growth between 1960 and 1995	gy
4	average investment rate of the period 1960-1995	lns
5	average rate of growth of working-age population (n_p) plus ($g + \delta$)	$lnngd$
6	longitude of capital	$xlong$
7	lattitude of capital	$ylat$

Table 3: List of variables and their acronyms.

The first five variables in Table 3 are used to evaluate the impact of saving, population growth and neighbourhood on growth. The last two variables, i.e. longitude of capital and lattitude of capital, are used to construct the distance matrix of which the elements d_{ij} are great-circle, geographical distances between country capitals. This distance matrix is subsequently used to build weight matrices W1(γ_1) and W2(γ_2) to be introduced in the next subsection. We discuss each of the main variables below.

We first look at the first three variables in Table 3. Logarithms of real income in 1960 and 1995 for 91 countries are illustrated in Figure 1. Countries' ISO codes are listed on the horizontal axis in the order according to Table 10 in the Appendix. The dotted and solid bars represent the initial level of real income (in 1960) and the level of real income in 1995, respectively. The figure shows that the levels of real income differ very strongly

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

across countries. However, within each country, the levels of real income in 1960 and 1995 stay close to each other with those in 1995 are usually higher for most countries.

Figure 2 shows the average rates of growth between 1960 and 1995 for 91 countries. These values are computed as (lny95- lny60)/35. The figure shows that the average rates of growth indeed differ strongly across countries. Out of these 91 countries considered in this paper, 17 countries have negative average rates of growth. Hong Kong is the country with the highest rate of growth between 1960 and 1995, with the average rate of growth of 6.24%. The country with the lowest rate of growth between 1960 and 1995 is Democratic Republic of the Congo, with the average rate of growth of -3.43%.

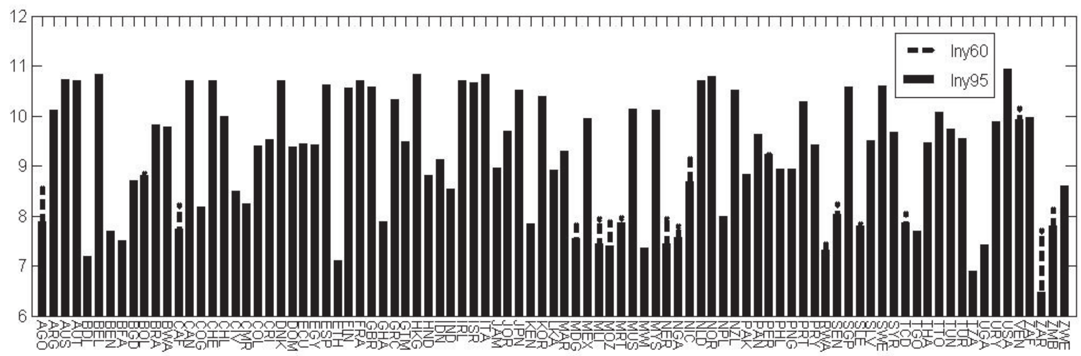

Figure 1: Logarithms of the levels of per worker income in 1960 and 1995 for 91 countries.

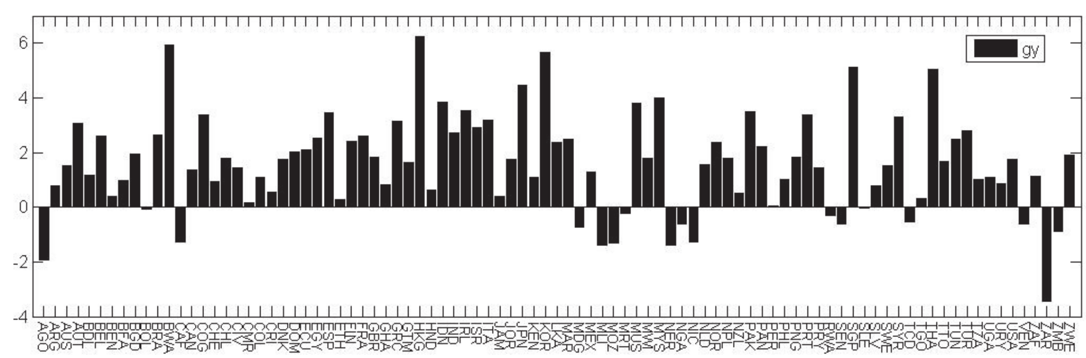

Figure 2: Average rate of growth between 1960 and 1995 for 91 countries.

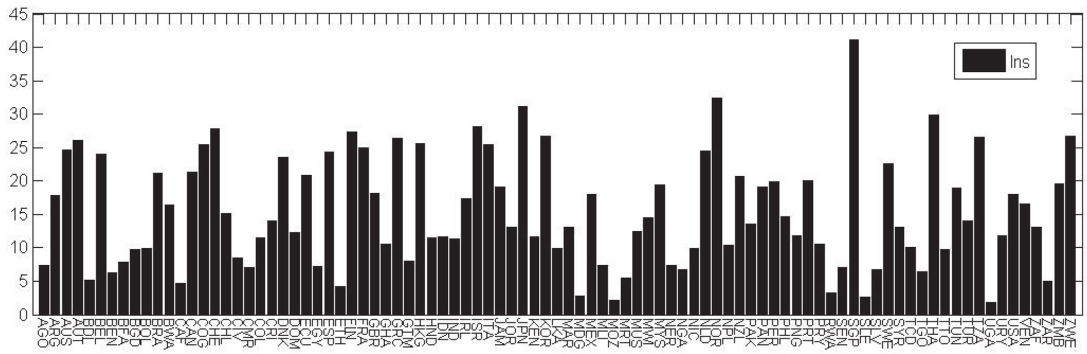

Figure 3: Average investment rates of the period 1960-1995 for 91 countries.

Next, we look at the fourth variable; the average investment rates of the period 1960-1995. These are measured as the average shares of real investment, including government investment, in real GDP. Note that the average investment rates of the period 1960-1995 for 91 countries are shown in Figure 3 while we use the logarithm values of this variable in our empirical study. We can see that the average investment rates vary sharply across countries. Singapore has the highest average investment rate with its share of real investment in real GDP of 41%, whereas Uganda has the lowest average investment rate with its share of real investment of 1.9%.

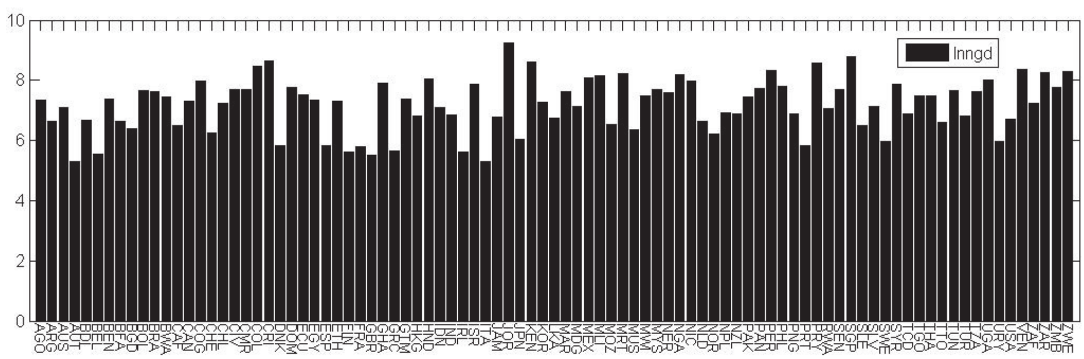

Figure 4: Average rates of growth of working-age population plus 0.05 for 91 countries.

Finally, the average rates of growth of the working-age population (n_p) plus ($g+\delta$) are shown in Figure 4. Note that this figure shows the average rates of growth while we use their logarithm values in our empirical study. The working age is restricted to 15-64 years old and $(g + \delta)$ is assumed to equal 0.05 as in Mankiw et al. (1992). The figure shows that there are large differences in the population growth rates among countries considered here. The countries with the highest growth rate of working-age population plus 0.05 is Jordan (9.3%) and the lowest rate is Austria (5.3%).

7.2 Empirical Results

We apply our QML estimator using two types of sub-models for the spatial weights with fixed and freely-estimated parameters defining the weights, γ, to the data set from Ertur and Koch (2007) discussed in the previous section. Then, we evaluate the impact of saving, population growth and neighbourhood on growth for each type of these weight matrices. The corresponding empirical results are reported in the following subsections. The hypothesis test results will be reported elsewhere.

We first explain how the model is constructed. As the MR-SAR model is a special case of the spatial durbin model (SDM), we modify the spatial durbin model used in Ertur and Koch (2007) to suit our MR-SAR case. The extension of our work to the SDM model is for the future and is non-trivial. Here, for country i with $i = 1, \ldots, 91$, our MR-SAR model is described as follows

$$gy_i = \beta_1 + \beta_2 lny60_i + \beta_3 lns_i + \beta_4 lnngd_i + \lambda \sum_{j \neq i}^{n} w_{ij}(\gamma_l) gy_j + \varepsilon_i. \tag{7.1}$$

The dependent variable is the average rate of growth between the year 1960 and 1995 for country i, computed as $(lny95_i - lny60_i)/35$, where $lny95_i$ is logarithm of the level of per worker income in 1995 for country i, $lny60_i$ is logarithm of the initial level of per worker income (in 1960) for country i, and 35 is the number of years.

For the explanatory variables for country i, $x_{1,i}$ consists of ones, $x_{2,i}$ is logarithm of the initial level of per worker income ($lny60_i$), $x_{3,i}$ is logarithm of the average investment rate of the period 1960-1995 (lns_i), and $x_{4,i}$ is logarithm of the average rate of growth of working-age population (n_p) plus 0.05 ($lnngd_i$). ε_i is country i's shock. $\beta_1, \ldots, \beta_4$ are parameters associated with $X_1, \ldots, X_4$, respectively, and assumed to be the same for all countries. λ is the spatial autoregressive parameter which is also assumed to be the same for all countries. $w_{ij}(\gamma_l)$ for $l = 1$ and 2 is the spatial weight between countries i and j with γ_l fixed across countries for each type of the weight matrices.

The first weight matrix $W1_n(\gamma_1)$ considered in this section has its weight elements of the following function form

$$w1_{n,ij}(\gamma_1) = \begin{cases} 0 & \text{if } i = j \\ \frac{e^{-\gamma_1 d_{ij}}}{\sum_j e^{-\gamma_1 d_{ij}}} \geq 0 & \text{if } i \neq j \end{cases} \tag{7.2}$$

Note that this is the same as in (6.1). Next, we specify the second sub-model for the spatial weights and the weight matrix $W2_n(\gamma_2)$ has the following weight elements.

$$w2_{n,ij}(\gamma_2) = \begin{cases} 0 & \text{if } i = j \\ \frac{d_{ij}^{-\gamma_2}}{\sum_j d_{ij}^{-\gamma_2}} \geq 0 & \text{if } i \neq j \end{cases} \tag{7.3}$$

where γ_2 is a positive scalar parameter specifying the weights, d_{ij} is a fixed nonnegative distance between spatial units i and j, and $\sum_j d_{ij}^{-\gamma_2}$ is a row sum for all i. Matrix $W2_n(\gamma_2)$ is also row-standardised so the weight elements on the main diagonal are zero whereas all other elements are nonnegative.

In the next subsections we report the empirical results obtained from fixed and freely estimated weight matrices.

7.2.1 With fixed Spatial Weight Matrices

In this subsection we present the results obtained by evaluating the log-likelihood function derived from equation (7.1) above based on two types of sub-models for the spatial weights. The evaluation is carried out using a one-dimensional grid search. Table 4 reports the QML estimates of parameters β, λ, and σ^2 for two weight matrices W1(γ_1) and W2(γ_2) with γ_1 and γ_2 fixed at 2 as in Ertur and Koch (2007). The variables are listed in the first column. The second and third columns show the QML estimates obtained based on weight matrices W1(γ_1) and W2(γ_2), respectively.

From the results, we can see that the coefficients of the initial level of per worker income (lny60) and the average rate of growth of working-age population (lnngd) are both negative. The negative coefficient of the initial level of income indicates that there exists conditional $\beta-$convergence, i.e. a country's growth rate declines as it approaches its steady state. On the other hand, the coefficient of the average investment rate of the period 1960-1995 (lns) and the spatial autoregressive parameter are both positive as expected. The average investment rate has positive effect on growth, so higher investment rate leads to higher growth and positive coefficient of the spatial autoregressive parameter suggests positive spillovers of growth across countries.

Variable	W1(2)	W2(2)
constant	0.0322	0.0348
lny60	-0.0066	-0.0070
lns	0.0181	0.0192
lnngd	-0.0277	-0.0291
W(2) gy	0.3	0.28
σ^2	0.000148	0.000154
log-likelihood	271.1240	269.4306

Table 4: QML estimates for the MR-SAR model based on weight matrices W1(γ_1) and W2(γ_2), with γ_1 and γ_2 fixed at 2.

Variable	constant	lny60	lns	lnngd	W1(2) gy	σ^2
constant	0.886	-0.012	0.027	0.275	0.364	-0.000
lny60	-0.012	0.003	-0.002	0.006	0.014	-0.000
lns	0.027	-0.002	0.006	-0.002	-0.068	0.000
lnngd	0.275	0.006	-0.002	0.127	0.295	-0.000
W1(2) gy	0.364	0.014	-0.068	0.295	9.776	-0.000
σ^2	-0.000	-0.000	0.000	-0.000	-0.000	0.000

Table 5: Estimated asymptotic variance matrix for all coefficients based on weight matrix W1(γ_1), with γ_1 fixed at 2.

In Tables 5 and 6 we report the estimated asymptotic variances for the coefficients, based on weight matrices W1(γ_1) and W2(γ_2) with γ_1 and γ_2 fixed at 2, respectively. These variance matrices are obtained from taking the inverse of the average Hessian matrix in equation (5.6) and dividing by n. Then, we multiply these variances by 10^3 and round them to the nearest 3th decimal before reporting them in these tables to improve the readability.

Variable	constant	lny60	lns	lnngd	W2(2) gy	σ^2
constant	0.926	-0.013	0.029	0.287	0.420	-0.000
lny60	-0.013	0.003	-0.002	0.006	0.011	-0.000
lns	0.029	-0.002	0.006	-0.001	-0.070	0.000
lnngd	0.287	0.006	-0.001	0.132	0.325	-0.000
W2(2) gy	0.420	0.011	-0.070	0.325	12.894	-0.000
σ^2	-0.000	-0.000	0.000	-0.000	-0.000	0.000

Table 6: Estimated asymptotic variance matrix for all coefficients based on weight matrix W2(γ_2), with γ_2 fixed at 2.

7.2.2 With Freely Estimated Spatial Weight Matrices

In this subsection we present empirical results obtained by evaluating the log-likelihood function based on two types of sub-models for the spatial weights, where the weight parameters γ are freely estimated. The evaluation is carried out using a two-dimensional grid search. Table 7 shows the QML estimates of parameters γ, β, λ, and σ^2 for two weight matrices W1(γ_1) and W2(γ_2) with freely-estimated γ_1 and γ_2.

Variable	W1(γ_1)	W2(γ_2)
γ	0.81	2.49
constant	0.0363	0.0336
lny60	-0.0063	-0.0069
lns	0.0169	0.0191
lnngd	-0.0231	-0.0292
W(γ) gy	0.47	0.25
σ^2	0.000140	0.000154
log-likelihood	273.3922	269.6053

Table 7: QML estimates for the MR-SAR model based on weight matrices W1(γ_1) and W2(γ_2), with freely-estimated γ_1 and γ_2.

The first row reports estimates of γ_1 and γ_2 and we can see that they both have positive signs as predicted. The rest of the results in Table 7 are similar to those in

Table 4. As expected, the coefficients of the initial level of per worker income (in 1960) and the average rate of growth of working-age population are negative for both types of the weight matrices used. The negative coefficient of the initial level of income again confirms the conditional $\beta-$convergence. For the average investment rate of the period 1960-1995 and the spatial autoregressive parameter, their coefficients are both positive and there are positive spillovers of growth across countries.

Variable	constant	lny60	lns	lnngd	W1(γ_1) gy	γ_1	σ^2
constant	0.864	-0.012	0.026	0.266	-0.078	2.150	-0.000
lny60	-0.012	0.003	-0.002	0.006	0.046	-0.078	-0.000
lns	0.026	-0.002	0.006	-0.002	-0.099	0.051	0.000
lnngd	0.266	0.006	-0.002	0.124	0.385	0.141	-0.000
W1(γ_1) gy	-0.078	0.046	-0.099	0.385	30.106	-59.157	-0.000
γ_1	2.150	-0.078	0.051	0.141	-59.157	229.510	0.000
σ^2	-0.000	-0.000	0.000	-0.000	-0.000	0.000	0.000

Table 8: Estimated asymptotic variance matrix for all coefficients based on weight matrix W1(γ_1), with freely-estimated $\gamma_1 = 0.81$.

Variable	constant	lny60	lns	lnngd	W2(γ_2) gy	γ_2	σ^2
constant	0.939	-0.014	0.030	0.291	0.790	-8.452	-0.000
lny60	-0.014	0.003	-0.002	0.006	-0.004	0.284	- 0.000
lns	0.030	-0.002	0.006	-0.001	-0.047	-0.321	0.000
lnngd	0.291	0.006	-0.001	0.133	0.436	-3.025	-0.000
W2(γ_2) gy	0.790	-0.004	-0.047	0.436	20.173	-183.883	-0.000
γ_2	-8.452	0.284	-0.321	-3.025	-183.883	3307.032	-0.000
σ^2	-0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000

Table 9: Estimated asymptotic variance matrix for all coefficients based on weight matrix W2(γ_2), with freely-estimated $\gamma_2 = 2.49$.

In Tables 8 and 9 we report the estimated asymptotic variances for the coefficients based on weight matrices W1(γ_1) and W2(γ_2) with freely-estimated $\gamma_1 = 0.81$ and $\gamma_2 = 2.49$, respectively. Similarly to the results reported in Tables 5 and 6, we multiply the variances by 10^3 and round them to the nearest 3th decimal to improve the readability of the tables.

7.3 Hypothesis Tests

[RESULTS WILL BE REPORTED LATER]

8 Conclusion

Specification of the spatial weight matrix is one of the most important issues in spatial econometrics and it has received much attention, especially in the past few years. As the weight matrix captures the dependence structure between spatial units, it is crucial to specify the elements of the weights properly. Different weight matrices lead to different results and different interpretations of the results. In this paper we introduce a sub-model for the spatial weights and estimate a variable spatial weight matrix in the mixed regressive, spatial autoregressive (MR-SAR) model by the maximum Gaussian likelihood. We establish the identifiability of the parameter defining the weights as well as 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 subsequently compared with other QML estimators using various fixed spatial weight matrices.

The Monte Carlo results show that our QML estimator using a freely estimated weight matrix is able to estimate the parameter defining the spatial weights, γ, reasonably well. It outperforms other competing estimators in many cases considered in this paper. Moreover, our results show that using a wrong weight matrix strongly affects the estimation performance of the estimators, especially when estimating the spatial autoregressive parameter λ.

To illustrate the applicability of our QML estimator, we apply our QML estimator using two functional forms of the weight matrices to a real data set to study the impact of saving, population growth and neighbourhood on growth in the framework of the mixed regressive, spatial autoregressive (MR-SAR) model. We evaluate and compare our estimator using freely-estimated spatial weight matrices with other QML estimators using weight matrices with the parameter defining the weights adopted in previous work. Asymptotic variances are evaluated and the hypothesis test results will be reported elsewhere.

The empirical results show that our QML estimator with freely-estimated weight matrices in the framework of the MR-SAR model is applicable to a real data set. It is able to capture positive spatial spillovers of growth among countries and provides significant estimates of the parameter defining the weights and other parameters with predicted signs.

9 Appendix

9.1 List of Notations

The list below presents the notations frequently used in this paper, most of which are extensions of Lee (2004a)'s notations.

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

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

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

$$S_n = I_n - \lambda_0 W_n$$

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

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

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

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

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

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

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

$$Q_n(\lambda,\gamma) = max_{\beta,\sigma^2} E[\ln L_n(\theta)]$$

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

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

9.2 Useful Properties

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 $W_n(\gamma_0)$.

9.2.1 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 mean value theorem,

$$\begin{aligned}
&\frac{1}{n}(\ln|det(S_n(\lambda_2,\gamma_2))| - \ln|det(S_n(\lambda_1,\gamma_1))|)\\
&= -\frac{1}{n}tr(W_n(\bar{\gamma}_n)S_n^{-1}(\bar{\lambda}_n,\bar{\gamma}_n))[\lambda_2-\lambda_1] - \frac{\bar{\lambda}_n}{n}tr(Z_n(\bar{\gamma}_n)S_n^{-1}(\bar{\lambda}_n,\bar{\gamma}_n))[\gamma_2-\gamma_1]\\
&= -\frac{1}{n}tr(G_n(\bar{\lambda}_n,\bar{\gamma}_n))[\lambda_2-\lambda_1] - \frac{\bar{\lambda}_n}{n}tr(T_n(\bar{\lambda}_n,\bar{\gamma}_n))[\gamma_2-\gamma_1] \qquad (9.1)
\end{aligned}$$

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 6, 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 (9.1), 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 6, 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 $\Lambda \otimes \Gamma$.

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

We describe the following auxiliary model as follows

$$Q_{p,n}(\lambda, \gamma) = -\frac{n}{2}(\ln 2\pi + 1) - \frac{n}{2}\ln \sigma_n^2(\lambda, \gamma) + \ln|det(S_n(\lambda, \gamma))| \tag{9.2}$$

and the log likelihood of a SAR model $Y_n = \lambda W_n Y_n + \varepsilon_n$, where $\varepsilon_n \sim N(0, \sigma_0^2 I_n)$, is as follows

$$\begin{aligned}\ln L_{p,n}(\lambda, \gamma, \sigma^2) = & -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|det(S_n(\lambda, \gamma))| \\ & - \frac{1}{2\sigma^2} Y_n' S_n'(\lambda, \gamma) S_n(\lambda, \gamma) Y_n.\end{aligned}$$

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

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

For $\sigma_n^2(\lambda, \gamma)$, note that

$$\begin{aligned}\sigma_n^2(\lambda, \gamma) &= \frac{\sigma_0^2}{n} tr(S_n^{-1\prime} S_n'(\lambda, \gamma) S_n(\lambda, \gamma) S_n^{-1}) \\ &= \sigma_0^2[1 + 2(\lambda_0 - \lambda)\frac{1}{n}tr(G_n) + (\lambda_0 - \lambda)^2 \frac{1}{n} tr(G_n G_n')].\end{aligned} \tag{9.3}$$

We show that $\sigma_n^2(\lambda, \gamma)$ 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

$$-\frac{1}{2}\ln\sigma_n^2(\lambda, \gamma) \leq -\frac{1}{2}\ln\sigma_0^2 + \frac{1}{n}(\ln|det(S_n)| - \ln|det(S_n(\lambda, \gamma))|). \tag{9.4}$$

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 $\lim_{n\to\infty} \sigma_n^2(\lambda_n, \gamma_n) = 0$, which implies that $-\lim_{n\to\infty} \ln\sigma_n^2(\lambda_n, \gamma_n) = \infty$. Therefore, $\sigma_n^2(\lambda, \gamma)$ must be bounded away from zero uniformly on $\Lambda \otimes \Gamma$.

9.2.4 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 $\frac{1}{n}Q_n(\lambda, \gamma) = -\frac{1}{2}(\ln(2\pi) + 1) - \frac{1}{2}\ln \sigma_n^{*2}(\lambda, \gamma) + \frac{1}{n}\ln |det(S_n(\lambda, \gamma))|$. Substitute (9.3) into σ_n^{*2}, we have

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

It is quadratic in λ and its coefficients $\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$, $\frac{1}{n}tr(G_n)$, and $\frac{1}{n}tr(G_nG_n')$ 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 show detailed proofs of the identifiable uniqueness, consistency and asymptotic normality of $\hat{\theta}_n$.

9.3 Proof of Theorem 1: Identifiable Uniqueness

We show that

$$\lim_{n\to\infty} \sup[\max_{\theta \in \eta_n^c(\nu)} \{\frac{1}{n}Q_n(\lambda, \gamma) - \frac{1}{n}Q_n\}] < 0 \tag{9.5}$$

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

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

Note that $\sigma_n^{*2}(\lambda_0, \gamma_0) = \frac{\sigma_0^2}{n}tr[S_n^{-1\prime}S_n(\lambda_0, \gamma_0)'S_n(\lambda_0, \gamma_0)S_n^{-1}] = \frac{\sigma_0^2}{n}tr[S_n^{-1\prime}S_n'S_nS_n^{-1}] = \sigma_0^2$. Add $\frac{1}{2}\ln \sigma_n^2(\lambda, \gamma)$ to both sides of the above equation and rearrange the terms, then it becomes

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

We prove this theorem by a counter example. Suppose that the condition of identifiable uniqueness would 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 $\lim_{n\to\infty}\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ exists and is positive. Hence, the identifiable uniqueness must hold. Q.E.D.

9.4 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

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

and

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

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

$$H_{1n}(\lambda,\gamma)=\frac{1}{n}(G_nX_n\beta_0)'M_nS_n(\lambda,\gamma)S_n^{-1}\varepsilon_n \tag{9.6}$$

and

$$H_{2n}(\lambda,\gamma)=\frac{1}{n}\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)M_nS_n(\lambda,\gamma)S_n^{-1}\varepsilon_n. \tag{9.7}$$

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 (9.6) as follows

$$H_{1n}(\lambda,\gamma)=\frac{1}{n}(G_nX_n\beta_0)'M_n\varepsilon_n+(\lambda_0-\lambda)\frac{1}{n}(G_nX_n\beta_0)'M_nG_n\varepsilon_n \tag{9.8}$$

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,

$$\begin{aligned}H_{2n}(\lambda,\gamma)-\sigma_n^2(\lambda,\gamma)&=\frac{1}{n}\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)M_nS_n(\lambda,\gamma)S^{-1}\varepsilon_n-\sigma_n^2(\lambda,\gamma)\\&=\frac{1}{n}\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S^{-1}\varepsilon_n\\&-\frac{\sigma_0^2}{n}tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})-H_{3n}(\lambda,\gamma)\end{aligned} \tag{9.9}$$

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

$$\frac{1}{\sqrt{n}}X_n'S_n(\lambda,\gamma)S_n^{-1}\varepsilon_n=\frac{1}{\sqrt{n}}X_n'S_n^{-1}\varepsilon_n-\frac{\lambda}{\sqrt{n}}X_n'G_n\varepsilon_n=O_p(1) \tag{9.10}$$

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),

$$\frac{1}{n}[\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1}\varepsilon_n-\sigma_0^2tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})]=o_p(1) \tag{9.11}$$

uniformly in $(\lambda,\gamma)\in\Lambda\otimes\Gamma$. Subsequently, 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 $\sigma_n^{*2}(\lambda,\gamma)=(\lambda_0-\lambda)^2\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ $+\frac{\sigma_0^2}{n}tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})=(\lambda_0-\lambda)^2\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)+\sigma_n^2(\lambda,\gamma)$. 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 the 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.

9.5 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 $\frac{1}{n}\frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'}-E(\frac{1}{n}\frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'})\xrightarrow{p}0$.

9.5.1 Nonsingularity of Σ_θ

[BEING WRITTEN UP]

9.5.2 $\frac{1}{n}\frac{\partial^2\ln L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'}-\frac{1}{n}\frac{\partial^2\ln L_n(\theta_0)}{\partial\theta\partial\theta'}\xrightarrow{p}0$

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.

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2}X_n'X_n, \tag{9.12}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\lambda} = -\frac{1}{\sigma^2}X_n'W_n(\gamma)Y_n, \tag{9.13}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\gamma} = -\frac{\lambda}{\sigma^2}X_n'Z_n(\gamma)Y_n, \tag{9.14}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}X_n'\varepsilon_n(\delta), \tag{9.15}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda^2} = -tr(G_n^2(\lambda,\gamma)) - \frac{1}{\sigma^2}Y_n'W_n'(\gamma)W_n(\gamma)Y_n, \tag{9.16}$$

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\gamma} =& - tr(T_n(\lambda,\gamma)) - \lambda tr(G_n(\lambda,\gamma)T_n(\lambda,\gamma)) \\ &- \frac{\lambda}{\sigma^2}Y_n'Z_n'(\gamma)W_n(\gamma)Y_n,\end{aligned} \tag{9.17}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\sigma^2} = -\frac{1}{\sigma^4}Y_n'W_n'(\gamma)\varepsilon_n(\delta), \tag{9.18}$$

$$\begin{aligned}\frac{\partial^2 \ln L_n(\theta)}{\partial\gamma^2} =& - \lambda tr(C_n(\lambda,\gamma)) - \lambda^2 tr(T_n^2(\lambda,\gamma)) \\ &- \frac{\lambda^2}{\sigma^2}Y_n'Z_n'(\gamma)Z_n(\gamma)Y_n,\end{aligned} \tag{9.19}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\gamma\partial\sigma^2} = -\frac{\lambda}{\sigma^4}Y_n'Z_n'(\gamma)\varepsilon_n(\delta), \tag{9.20}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\varepsilon_n'(\delta)\varepsilon_n(\delta) \tag{9.21}$$

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 (9.12) 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 (9.13) at $\hat\theta_n$ and at θ_0 is

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat\theta_n)}{\partial\beta\partial\lambda} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\beta\partial\lambda} = \frac{1}{\sigma_0^2}\frac{X_n'W_nY_n}{n} - \frac{1}{\hat\sigma_n^2}\frac{X_n'W_n(\hat\gamma_n)Y_n}{n} \tag{9.22}$$

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 vector-valued function. Then, for $\bar{\gamma}_n$ that lies between $\hat{\gamma}_n$ and γ_0, we have

$$\Big|\Big|\frac{X_n'W_n(\hat{\gamma}_n)Y_n}{n} - \frac{X_n'W_nY_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{X_n'Z_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| \tag{9.23}$$

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 $\hat{\gamma}_n \xrightarrow{p} \gamma_0$, $\Big|\Big|\frac{X_n'W_n(\hat{\gamma}_n)Y_n}{n} - \frac{X_n'W_nY_n}{n}\Big|\Big| \xrightarrow{p} 0$. 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, (9.22) 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 (9.14), we first show that

$$\Big|\Big|\frac{X_n'Z_n(\hat{\gamma}_n)Y_n}{n} - \frac{X_n'Z_nY_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{X_n'A_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1) \tag{9.24}$$

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 (9.15), 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 (9.15) and as we have shown in (9.23) 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 (9.15) evaluated at $\hat{\theta}_n$ and θ_0 becomes

$$\begin{aligned}&\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\beta\partial\sigma^2} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\beta\partial\sigma^2}\\ &= (\frac{1}{\sigma_0^4} - \frac{1}{\hat{\sigma}_n^4})\frac{X_n'\varepsilon_n}{n} + \frac{X_n'X_n}{\hat{\sigma}_n^4 n}(\hat{\beta}_n - \beta_0) + \frac{1}{\hat{\sigma}_n^4 n}[\hat{\lambda}_nX_n'W_n(\hat{\gamma}_n)Y_n - \lambda_0X_n'W_nY_n]\\ &= (\frac{1}{\sigma_0^4} - \frac{1}{\hat{\sigma}_n^4})\frac{X_n'\varepsilon_n}{n} + \frac{X_n'X_n}{\hat{\sigma}_n^4 n}(\hat{\beta}_n - \beta_0) + (\hat{\lambda}_n - \lambda_0)\frac{X_n'W_nY_n}{\hat{\sigma}_n^4 n} + o_p(1) = o_p(1)\end{aligned}$$

for $\hat{\theta}_n \xrightarrow{p} \theta_0$. Next, for (9.18), we have

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

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

$$\Big|\Big|\frac{Y_n'W_n'(\hat{\gamma}_n)\varepsilon_n}{n} - \frac{Y_n'W_n'\varepsilon_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'Z_n'(\bar{\gamma}_n)\varepsilon_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1). \quad (9.25)$$

For $\hat{\gamma}_n \xrightarrow{p} \gamma_0$ and $\bar{\gamma}_n$ lies between $\hat{\gamma}_n$ and γ_0, (9.25) 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

$$\Big|\Big|\frac{Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n} - \frac{Y_n'W_n'W_nY_n}{n}\Big|\Big| \quad (9.26)$$
$$\leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'Z_n'(\bar{\gamma}_n)W_n(\bar{\gamma}_n)Y_n}{n} + \frac{Y_n'W_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1)$$

and

$$\Big|\Big|\frac{Y_n'W_n'(\hat{\gamma}_n)W_nY_n}{n} - \frac{Y_n'W_n'W_nY_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'Z_n'(\bar{\gamma}_n)W_nY_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1), \quad (9.27)$$

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 (9.23) and $\hat{\theta}_n \xrightarrow{p} \theta_0$, the difference of (9.18) evaluated at $\hat{\theta}_n$ and θ_0 becomes

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

For (9.20), 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 (9.18) above applies here as well. By the mean value theorem, we first show that $\frac{1}{n}Y_n'Z_n'(\hat{\gamma}_n)\varepsilon_n \xrightarrow{p} \frac{1}{n}Y_n'Z_n'\varepsilon_n$.

$$\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)\varepsilon_n}{n} - \frac{Y_n'Z_n'\varepsilon_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)\varepsilon_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1). \tag{9.28}$$

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,

$$\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)W_nY_n}{n} - \frac{Y_n'Z_n'W_nY_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)W_nY_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1) \tag{9.29}$$

and

$$\begin{aligned}&\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n} - \frac{Y_n'Z_n'W_nY_n}{n}\Big|\Big| \\ &\leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)W_n(\bar{\gamma}_n)Y_n}{n} + \frac{Y_n'Z_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1),\end{aligned} \tag{9.30}$$

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 (9.28) - (9.30) and (9.24), the convergence of (9.20) becomes

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

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 (9.16), (9.17) and (9.19), 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 (9.16) at $\hat{\theta}_n$ and θ_0 is

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

As we have already shown in (9.26), $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n \xrightarrow{p} \frac{1}{n}Y_n'W_n'W_nY_n$. 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 (9.16) becomes

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

For (9.17), the same technique applies. By 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 convergence of (9.17) becomes

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

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 (9.30) 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 (9.19), apply the mean value theorem to the traces as follows.

$$\begin{aligned} tr(\hat\lambda_n T_n^2(\hat\lambda_n,\hat\gamma_n)) - tr(\lambda_0 T_n^2) &= 2\bar\lambda_n tr(T_n^2(\bar\lambda_n,\bar\gamma_n) + \bar\lambda_n T_n^2(\bar\lambda_n,\bar\gamma_n)G_n(\bar\lambda_n,\bar\gamma_n))[\hat\lambda_n - \lambda_0] \\ &+ 2\bar\lambda_n^2 tr(T_n(\bar\lambda_n,\bar\gamma_n)C_n(\bar\lambda_n,\bar\gamma_n) + \bar\lambda_n T_n^3(\bar\lambda_n,\bar\gamma_n))[\hat\gamma_n - \gamma_0] \end{aligned}$$

and

$$\begin{aligned} tr(\hat\lambda_n C_n(\hat\lambda_n,\hat\gamma_n)) - tr(\lambda_0 C_n) &= tr(C_n(\bar\lambda_n,\bar\gamma_n) + \bar\lambda_n C_n(\bar\lambda_n,\bar\gamma_n)G_n(\bar\lambda_n,\bar\gamma_n))[\hat\lambda_n - \lambda_0] \\ &+ \bar\lambda_n tr(V_n(\bar\lambda_n,\bar\gamma_n) + \bar\lambda_n C_n(\bar\lambda_n,\bar\gamma_n)T_n(\bar\lambda_n,\bar\gamma_n))[\hat\gamma_n - \gamma_0] \end{aligned}$$

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 (9.19) evaluated at $\hat\theta_n$ and θ_0 is

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

Note that the elements of $B_n(\gamma)$ are uniformly bounded by Assumption 6. Next, we show that $\frac{1}{n}Y_n'Z_n'(\hat\gamma_n)Z_n(\hat\gamma_n)Y_n \xrightarrow{p} \frac{1}{n}Y_n'Z_n'Z_nY_n$. By the mean value theorem,

$$\begin{aligned} &\Big|\Big|\frac{Y_n'Z_n'(\hat\gamma_n)Z_n(\hat\gamma_n)Y_n}{n} - \frac{Y_n'Z_n'Z_nY_n}{n}\Big|\Big| \\ &\le \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar\gamma_n)Z_n(\bar\gamma_n)Y_n}{n} + \frac{Y_n'Z_n'(\bar\gamma_n)A_n(\bar\gamma_n)Y_n}{n}\Big|\Big||\hat\gamma_n - \gamma_0| = o_p(1) \end{aligned} \tag{9.31}$$

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 (9.19) becomes

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

Finally, for the last derivative (9.21), 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

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

As $\hat{\theta}_n \xrightarrow{p} \theta_0$ and by equations (9.23) and (9.25) - (9.27), 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 (9.21) 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 $\Lambda \otimes \Gamma$.

9.5.3 $\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) \xrightarrow{p} 0$

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,

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

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

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

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

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,

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

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

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

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.

9.6 List of Countries

Table 10 below presents a list of 91 countries and their ISO codes.[3]

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 10: List of 91 countries and their ISO codes.

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

References

[1] Aldstadt, J. and Getis, A. (2006). Using AMOEBA to Create a Spatial Weights Matrix and Identify Spatial Clusters. *Geographical Analysis*, 38, 327-343.

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

[3] Anselin, L. (1980). *Estimation Methods for Spatial Autoregressive Structures*. Ithaca. NY: Cornell University, Regional Science Dissertation and Monograph Series 8.

[4] Anselin, L. (1984). Specification Tests on the Structure of Interaction in Spatial Econometric Models. *Papers, Regional Science Association*, 54, 165-182.

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

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

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

[8] Blonigen, B.A., Davies, R.B., Waddell, G.R. and Naughton, H.T. (2007). FDI in space: Spatial autoregressive relationships in foreign direct investment. *European Economic Review*, 51, 1303-1325.

[9] Can, A. (1996). Weight matrices and spatial autocorrelation statistics using a topological vector data model. *International Journal Geographical Information Systems*, 10(8), 1009-1017.

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

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

[12] Cliff, A. and Ord, J.K. (1981). *Spatial Processes: Models and Application*. London: Pion.

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

[14] Geniaux, G. (2012). In search of W for the spatial lag model. mimeo.

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

[16] 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.

[17] Horn, R. and Johnson, C. (1985). *Matrix Analysis*. Cambridge University Press.

[18] Kelejian, H.H. and Piras, G. (2012). Estimating of Spatial Models with Endogenous Weighting Matrices, and an Application to a Demand Model for Cigarettes. mimeo.

[19] 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.

[20] 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.

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

[22] 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.

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

[24] 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/*.

[25] 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.

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

[27] 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.

[28] Souza, P.C.L de (2012). Estimating Networks: Lasso for Spatial Weights. mimeo.

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