Essays on Spatial Econometrics

Saruta Benjanuvatra

PhD

University of York

Economics

October 2012

Abstract

A bias-adjusted estimator for small samples and a hybrid estimator, which combines the guaranteed invertibility of the MLE with original non-hybridised estimators, are introduced in Chapter 2. Their performance is extensively compared with that of the Maximum Gaussian Likelihood and several Instrumental Variable-type estimators in the context of the spatial error model (SEM). We show that the bias-adjusted estimator is effective across various sample sizes and the hybridised forms of the estimators outperform even the best of the IV methods across a majority of the cases examined. Chapter 3 introduces a sub-model for spatial weights and estimates a variable weight matrix for the mixed regressive, spatial autoregressive (MR-SAR) model by maximum Gaussian likelihood. We establish the identifiability of the weight parameter, the consistency and the asymptotic normality of the QMLE under appropriate conditions that extend those given by Lee (2004a). Finite properties of our estimator are investigated in a Monte Carlo study and we show that it outperforms other competing estimators in many cases considered. Its applicability is illustrated in Chapter 4, where the estimator using two types of sub-models for the spatial weights is applied to the cross-sectional data set used in Ertur and Koch (2007) in the framework of the MR-SAR model to study the impact of saving, population growth and interdependence among countries on growth. It is shown that our QML estimator is able to capture positive spatial spillovers of growth among countries and provide significant estimates of other parameters of the model including the parameter defining the spatial weights.

Contents

List of Tables

List of Figures

Acknowledgements

Several people have significantly contributed to my PhD life while working on this thesis. The most important person in my PhD life cannot be anyone else but my supervisor, Professor Peter Burridge. I am deeply grateful to his great supervision, kind support, and valuable advice not only on my thesis but also on other aspects of being a PhD student. I would also like to thank Doctor Francesco Bravo and Professor Takashi Yamagata for their helpful comments on my work and kind support on my future career.

I am grateful to the Royal Thai Government for the financial support throughout my PhD, and to the Department of Economics and Related Studies, University of York, for the financial support for me to attend a number of conferences.

I want to thank all my friends and fellow PhD students for their friendship. In particular, I thank Vivien Burrows for all our relaxing 'breaks', Mai Farid for her kind support from the first day I walked into the PhD study area, Yuhsuan Lin for being such a great friend and fantastic chef, and Ryota Nakamura for countless interesting discussions. I am also thankful to Massaporn Kannasoot and Neungrutai Saesaengseerung for always being there for me.

Finally, I would like to dedicate this thesis to my family; Pa, Mah, P' A and P' Ohm. I am profoundly grateful to them for their unconditional love and encouragement. I would not have made it this far without them.

Author's Declaration

This thesis represents the work I have carried out towards my PhD degree in Economics from the Department of Economics and Related Studies at the University of York.

Chapter 2 is joint work with Peter Burridge. Earlier versions of this chapter were presented at the Far East and South Asia Meeting of the Econometric Society held in Tokyo, Japan, in August 2009; the Research Student Workshop at the University of York in February 2010; and the 9th International Workshop on Spatial Econometrics and Statistics held in Orleans, France, in June 2010.

Chapters 3 and 4 are my own work and have not been co-authored. Chapter 3 was presented at the VI World Conference of the Spatial Econometrics Association held in Salvador, Brazil, in July 2012, and the 5th Seminar of Spatial Econometrics in honour of Doctor J.H.P. Paelinck held in Coimbra, Portugal, in October 2012.

A combined version of Chapters 3 and 4 was presented at the VII World Conference of the Spatial Econometrics Association held in Washington, DC, USA, in July 2013.

Chapter 1

Introduction

Spatial econometrics is a sub-field of econometrics that combines econometrics with spatial analysis. The term 'spatial econometrics' was originated by Jean Paelinck. It deals with estimation and specification of models that involve interactions between units, or with data that have spatial autocorrelation or neighbourhood effects. Paelinck and Klaassen (1979) discuss the following distinct features that separate spatial econometrics from (standard) econometrics; spatial interdependence in spatial models, asymmetry among observations, space-distant explanatory factors, ex ante and ex post distance interaction, and space in spatial models. Standard econometric techniques are not always applicable for dealing with these features so they were often ignored or assumed away in the literature of econometrics.

Anselin and Rey (1997) provide a collection of papers on spatial econometrics to emphasise the importance of this field. An overview of development of spatial econometrics in the past three decades and challenging directions of future research can be found in Anselin (2010). Pinkse and Slade (2010) provide an overview of the direction of spatial econometrics and recognise problems still unsolved in this field. They recommend that researchers should begin with concrete empirical problems when trying to establish a new methodology. Partridge et al. (2012) discuss three papers that address problems in spatial econometrics when estimating geographic spillovers and propose alternative approaches to deal with these problems.

1.1 Spatial Effects

Spatial effects contained in spatial data can be divided into two types; spatial dependence and spatial heterogeneity. Spatial dependence is the dependence among data observations in cross-section and can be either positive or negative. Comparing to time series with dependence in time dimension only, spatial dependence can be multidimensional with dependence both in time and space dimensions. Anselin (1988a) argues that this dependence may be caused by (i) measurement problems and (ii) complex patterns of spatial interactions. As spatial dependence can be multidirectional, standard econometric techniques are not applicable and results obtained from these techniques are often not valid. Spatial econometric techniques, therefore, need to be developed.

The second type of spatial effects is spatial heterogeneity, which can be seen as observations being distributed unevenly in the area. It can lead to heteroskedasticity if it is reflected in measurement errors. However, this aspect of spatial effects can often be dealt with by standard econometric techniques and a separate estimation method is not always necessary (Anselin, 1988a and 2010).

1.2 Spatial Weight Matrix

The specification of a spatial weight matrix is one of the most important issues in the analysis of spatial econometric models of the type described briefly in the following sections. It is a square matrix with weight elements capturing dependence or interaction between spatial units. Earlier forms of spatial weight matrix in the literature are based on binary contiguity between spatial units

(Moran, 1948 and Geary, 1954), which use values $0-1$ to capture the interactions. Value 1 represents two spatial units having a common border, and 0 otherwise. Moran (1948) first assigns B ("black") to a county if an event has occurred in that particular county, and W ("white") otherwise. Then, if two contiguous counties are both "black", the value 1 is assigned, and 0 otherwise. Weight matrices based on the binary contiguity are of, for example, Rook contiguity where the weights equal 1 if the two regions share common border, and 0 otherwise. Another form is of Queen contiguity where the weights equal 1 if the two regions share common side or vertex, and 0 otherwise. If cities or points are considered as spatial units, then two cities are neighbours if they are within a chosen distance from each other.

As the binary contiguity is sometimes not sufficient to represent a more complex spatial interaction, general weight matrices have been proposed to include (relative) distances between spatial units. One of the well-known weight matrices is the Cliff-Ord weight matrix introduced in Cliff and Ord (1973, 1981), where the weight elements are a combination of distances and relative border length of common border between units. Distances in these contexts are geographic. However, Case et al. (1993) discuss that the distances between neighbours are not limited to only geographic distances but can represent economic or demographic distances as well. Other forms of the weights are of, for instance, inverse distance, n-nearest neighbours, or 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. A nice overview of the spatial weight matrices can be found in Anselin (1988a) and Anselin and Bera (1998).

Specification of the spatial weights is an important issue as different weight matrices yield different results and, hence, different interpretation of the results. Anselin (1988a) 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) argues that spatial weights should be selected based on spatial interaction theory. Proper choice of the weights improves an 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. Besides, Plumper and Neumayer (2010) study specification issues relating to the spatial weight matrix and argue that row-standardising and changes in a functional form of the weight matrix can lead to significantly different estimated results of the spatial effect.

Researchers have tried to construct the weight matrix using computer software. For example, Can (1996) develops software to construct the weight matrix in C programming language, and Aldstadt and Getis (2006) suggest an algorithm called 'A Multidirectional Optimal Ecotope-Based Algorithm' (AMOEBA) using empirical data that can distinguish clusters of weighted spatial units. GeoDa has also become a useful tool for constructing the weight matrix. Other studies attempt to find a proper weight matrix using different techniques and approaches. Bavaud (1998) gives a theoretical overview of general properties of spatial weight models and discusses several examples depicting these properties. Leenders (2002) discusses the four steps that should be taken when constructing a weight matrix and provides specification tests for choosing the most appropriate models for network autocorrelation. Mur et al. (2012) provide an overview of literature on the criteria for specifying a weight matrix and propose a simple nonparametric approach for selecting the appropriate weight matrix from a set of matrices.

Estimating the weight matrix has recently become a challenging alternative for specifying the weight matrix properly. Souza (2012) proposes an estimation technique for estimating networks using the Least Absolute Shrinkage and Selection Operator (Lasso), and shows that the estimator is consistent under sparsity requirements. Geniaux (2012) introduces a parametric approach to endogenously estimate the spatial weight matrix based on geographical distances in the spatial lag model using the iterated IV estimation method. Kelejian and Piras (2012) provide an estimator for regression parameters in the spatial panel data model that incorporates an endogenous weight matrix as a spatially lagged dependent variable, and show that the estimator is consistent and asymptotically normal.

1.3 Regression Models in Spatial Econometrics

Several regression models have been introduced in the literature to deal with the spatial effects. For the spatial dependence, two groups of regression models have been introduced. The first group consists of the regression models that include the spatial lag dependence, which is the dependence in variables associated with different spatial units. An example of models in this group is the spatial autoregressive (SAR) model (Anselin, 1988a):

$$Y = \lambda WY + \varepsilon \tag{1.1}$$

where Y is an $n \times 1$ vector of observations of the dependent variable, ε is an $n \times 1$ vector of disturbances, λ is the spatial autoregressive parameter, and W is an $n \times n$ weight matrix of fixed non-negative constants. This model is also called the spatial lag model. If the model includes regressors X, then it is called the mixed regressive, spatial autoregressive (MR-SAR) model (Ord, 1975 and Anselin, 1988a). The MR-SAR model is described as

$$Y = X\beta + \lambda WY + \varepsilon \tag{1.2}$$

where X is an $n \times k$ matrix of values of k exogenous explanatory variables, and β is a $k \times 1$ vector of parameters.

The second group of the regression models deals with the spatial error dependence, which is the dependence in the error terms. The most common model is the spatial error model (SEM) (Cliff and Ord, 1973):

$$Y = X\beta + U \tag{1.3}$$

with U an $n \times 1$ disturbance vector defined as

$$U = \rho MU + \varepsilon \tag{1.4}$$

where M is an $n \times n$ weight matrix of fixed non-negative constants, ρ is a scalar parameter, and ε is an $n \times 1$ vector of innovations that are homoskedastic and independently distributed.

A model that contains both the spatial lag dependence and the spatial error dependence is the spatial autoregressive model with spatial autoregressive disturbance (SARAR) described below.

$$Y = X\beta + \lambda WY + U \tag{1.5}$$

with

$$U = \rho MU + \varepsilon \tag{1.6}$$

where W and M can be the same or different. This model can also be extended to capture higher order spatial processes (Lee et al., 2010).

1.4 Spatial Panel Data Model

Panel data models have recently received much attention in spatial econometrics and several studies, both theoretical and applied, have been carried out in this framework. For example, Kapoor et al. (2007) suggest generalisations of the GM estimator introduced in Kelejian and Prucha (1999) to estimate the spatial autoregressive parameter and variances of the disturbance process in the framework of panel data models. Elhorst (2003) investigates eight panel data models with fixed effects, random effects, fixed coefficients, and random coefficients extended for the spatial error models (SEM) as well as the spatial lag models (SAR) respectively. This survey provides a nice overview of the spatial panel data models and concentrates on the model specification and the comparison between the estimation methods. Lee and Yu (2010) discuss recent developments in the spatial panel data models for static and dynamic cases. Finite sample properties are studied in Monte Carlo experiments and effects of misspecification are provided. Elhorst (2011) surveys the literature on the static and dynamic spatial panel data models, and shows that incorporating lags of the dependent and independent variables into spatial econometric models can be useful to assess direct and indirect effects.

1.5 Estimation Methods

The least squares estimator is generally an inconsistent estimator for the MR-SAR model whether or not the disturbances are spatially correlated, because the spatial lagged variables are correlated with the disturbances (Ord, 1975 and Anselin, 1988a). Incorporating the spatial dependence in the error terms, on the other hand, results in the OLS estimates being unbiased but inefficient (Anselin and Bera, 1998). However, Lee (2002) shows that the OLS estimator can be a consistent and asymptotically efficient estimator for the MR-SAR models when each spatial unit is aggregately influenced by a significant number of other spatial units. In this situation, the OLS estimator possesses advantage over the ML and IV estimators as it is computationally easier. Nevertheless, the OLS estimator is still inconsistent for the SAR model without exogenous regressors.

Consequently, several estimation methods have been proposed in the literature as alternatives to the OLS estimator. We discuss some of them in the following subsections.

1.5.1 Maximum Likelihood Estimation

As the least squares estimators are not suitable for estimating a spatial process with spatial dependence, the maximum likelihood estimation has widely been used as an alternative. The ML estimator in spatial regression models with Gaussian shocks is studied by Ord (1975), Anselin (1988a) and Anselin and Bera (1998). Ord (1975) also presents a computational scheme extended to the MR-SAR model and compares the MLE with other alternative estimators. Dubin (1988) simultaneously estimates regression coefficients and parameters of the correlation function by the maximum likelihood estimation. Asymptotic properties of the MLE are developed by Lee (2004a) for spatial autoregressive models with fixed sequences of weights. He also argues that the MLE method is still applicable when applied on the pure SAR model, while alternatives such as the IV estimation method will break down. Exact properties of the MLE in the spatial autoregressive models are derived by Hillier and Martellosio (2012). Lee, Liu and Lin (2010) suggest a QML estimation approach for social interaction models with network structures as well as endogenous and correlated effects. Asymptotic distribution of the estimator is derived and its small sample performance is investigated in a Monte Carlo study.

Nevertheless, the ML method may run into numerical problems associated with matrix inversion and eigenvalue calculations at least in the Gaussian case and especially with large numbers of observations. To avoid this problem, several alternative methods including the GMM estimation have been proposed. We discuss the GMM/IV estimation in the next subsection.

1.5.2 GMM/IV Estimation

The GMM estimation method has been introduced as an alternative to avoid numerical problems of the MLE method, which are due to the matrix inversion and especially when the number of observations is large. Various GMM estimators found in the literature are generally computationally feasible and consistent under appropriate conditions. A summary of these alternatives is presented below.

Among the alternatives are the GMM estimators introduced by Kelejian and Prucha (1998, 1999). Kelejian and Prucha (1998) introduce a generalised spatial two-stage least squares (GS2SLS) procedure for estimating the spatial autoregressive model with autoregressive disturbances, and show that their feasible estimator is consistent and asymptotically normal. However, Lee (2003) argues that this estimator may not be asymptotically optimal. He proposes a best spatial two-stage least squares estimator for this model and provides a three-step procedure similar to that given by Kelejian and Prucha (1998). Fingleton and Le Gallo (2008b) propose an estimation method which extends the GMM/IV estimators introduced by Kelejian and Prucha (1998) and Fingleton and Le Gallo (2008a) to include an endogenous spatial lag, other endogenous variables and a spatial error process, and investigate its finite sample properties in a Monte Carlo study. Drukker et al. (2011) extend the work by Kelejian and Prucha (1998, 1999) and propose a two-step GMM and IV estimation methods for the spatial autoregressive model with spatial autoregressive error terms and endogenous variables. The joint asymptotic distribution for the estimators is also derived.

Kelejian and Prucha (1999) propose a generalised moments (GM) estimator for the spatial autoregressive parameter in the SAR model and prove the consistency of the estimator under a set of conditions. Bell and Bockstael (2000) apply this method to microlevel data, whose feature concerns large numbers of observations scattered irregularly on the landscape that can cause problems with the ML estimation. They also compare its performance with that of the ML estimator and find that the GM estimator performs relatively well. A small-sample adjustment to the method of Kelejian and Prucha (1999) is introduced by Arnold and Wied (2010a).

Kelejian and Prucha (2002) show that the 2SLS and OLS estimators of linear Cliff-Ord type spatial models are not consistent in single cross section data if the weight matrix is row-normalised and has equal weights, whereas these estimators are consistent and efficient if two or more cross-sections of data are used. Liu et al. (2006) propose the best GMM estimator for the MR-SAR model and model with spatial autoregressive disturbances, and include potential skewness and kurtosis of the disturbances into the moment conditions. They show that this estimator is asymptotically as efficient as the ML estimator with normal disturbances and more efficient otherwise. Lee (2007a) proposes a GMM estimator that is superior to the 2SLS estimator for estimating the MR-SAR model. This GMM estimator is a combination of the moments in the 2SLS estimator and those obtained from the pure SAR model. He shows that this GMM estimator can be asymptotically more efficient than the 2SLS estimator and as efficient as the ML estimator. Lee and Liu (2010) expand the GMM estimator proposed by Lee (2007a) for the MR-SAR model to estimate a high order MR-SAR model with spatial autoregressive disturbances and show that this estimator is consistent and asymptotically normal. They also derive the best GMM estimator based on linear and quadratic moment conditions of the error terms and show that the best GMM estimator is asymptotically as efficient as the ML estimator when the disturbances are normally distributed, more efficient than the MLE in other cases, and efficient relative to the G2SLS estimator.

The ML estimator of the spatial autoregressive parameter for the SAR model can be inconsistent if the disturbances are heteroskedastic and several alternative estimators have been suggested. Kelejian and Prucha (2007) propose a robust spatial HAC estimator of a variance-covariance matrix and show that this estimator is consistent. Lin and Lee (2010) introduce a GMM estimator acquired from certain moment conditions that take into account the heteroskedasticity. They show that this estimator is consistent, asymptotically normal and robust, and the efficiency of the estimator can be improved by including an optimal weight matrix. Kelejian and Prucha (2010) introduce a GM estimator for the autoregressive parameter in the Cliff-Ord model (SARAR(1,1)) with heteroskedastic innovations of unknown form in the disturbance process and show that their GM estimator is consistent. They also specify IV estimators for the regression parameters in the model and provide the joint asymptotic distribution of the GM estimator for the spatial autoregressive parameter in the disturbance process and of the IV estimator for the model regression parameters. Badinger and Egger (2011) extend the two-step GM estimation procedure introduced in Kelejian and Prucha (2010) to the case of higher order (SARAR(R,S)) and establish the consistency of the estimator and provide the joint asymptotic distribution of the GM and the TSLS estimator as well. Arraiz et al. (2010) describes a multi-step GMM/IV type estimation procedure for estimating the linear Cliff-Ord-type model with spatial lagged dependent variable and heteroskedastic innovations of unknown form in the disturbance process. Their results also show that the ML estimator of the autoregressive parameter can be biased when the disturbances are heteroskedastic. Arnold and Wied (2010b) propose a two-step GMM estimation approach to estimate parameters in a spatial model with three kinds of spatial dependence as well as heteroskedastic innovations and apply their approach to daily stock returns of the Euro Stoxx 50 members.

Pinkse et al. (2002) introduce an IV estimator for the price response coefficients and provide the consistency and asymptotic distribution of this IV estimator. It is then applied to data of wholesales gasoline markets in the

United States. Das et al. (2003) investigate finite sample properties of several estimators in the SARAR (1,1) model. The estimators they consider are the maximum likelihood estimator, least squares estimator, two-stage least squares (2SLS) estimator, generalised spatial two-stage least squares (GS2SLS) estimator, feasible generalised spatial two-stage least squares (FGS2SLS) estimator, and iterated FGS2SLS estimator. Their results suggest that there is small difference in finite sample efficiency between the ML and FGS2SLS estimators so the latter can be considered with small cost. Moreover, for the autoregressive parameter in the disturbance process, there is also small difference in finite sample efficiency between the ML and GM estimators.

1.5.3 Bayesian Approach

Bayesian approaches have been applied in spatial econometrics to help researchers make choice between models. LeSage (1997) suggests a Bayesian approach based on Gibbs sampling for the spatial autoregressive models. Mur et al. (2012) use the Bayesian as one of the approaches to study performance of different weight matrices in their research. As least squares estimator may be biased and inconsistent when spatial dependence is present, the Markov Chain Monte Carlo model composition (MC^3) procedure and the Bayesian Model Averaging (BMA) method using least squares estimates will also be invalid in such cases. Therefore, LeSage and Parent (2007) introduce a MC^3 procedure and extend the Bayesian estimation for the spatial autoregressive (SAR) model and the spatial error model (SEM), focusing on comparing models with different matrices of explanatory variables. Efficient computational algorithms are also provided.

1.6 Hypothesis Tests

Well-known hypothesis tests on the parameters of the spatial models based on Maximum Likelihood are the Wald (W), Likelihood Ratio (LR) and Lagrange Multiplier (LM) tests. Another test also frequently used in spatial econometrics is Moran's I test. We discuss in brief these tests below. See Anselin (1988a) for an overview of these tests.

1.6.1 Wald Test

The Wald test can be applied to test the significance of an individual parameter or the joint significance of the parameter vector. Suppose that we want to test the significance of the spatial autoregressive parameter λ, the test statistic is described as

$$Wald = \hat{\lambda}^2 / \hat{v}_\lambda,$$

where $\hat{\lambda}$ is the ML estimate for λ, and $\hat{v}_\lambda$ is the diagonal element corresponding to λ in the variance matrix obtained from the unrestricted model. This test only uses information obtained from the unrestricted model and is, under $H_0 : g(\theta) = 0$, asymptotically distributed as χ^2 with 1 degree of freedom.

When testing for the joint significance of all parameters, the test statistic becomes

$$Wald = g'[G'VG]^{-1}g,$$

where g is a $q \times 1$ vector of ML estimates, G is a $z \times q$ matrix of partial derivatives evaluated for the parameter estimates with z the number of total parameters in the model, and V is the variance matrix obtained from the unrestricted model. This test is, under H_0, asymptotically distributed as χ^2 with q degrees of freedom. See Anselin (1988a) for more details.

1.6.2 Likelihood Ratio Test

The Likelihood Ratio (LR) test statistic is described as

$$LR = 2[\ln L(\theta) - \ln L(\theta_r)]$$

where $\ln L(\theta)$ is the log-likelihood for the unrestricted model with parameter vector θ and $\ln L(\theta_r)$ is the log-likelihood for the restricted model with parameter vector θ_r. Note that the LR test uses log-likelihood values for both the restricted and unrestricted models. It is, under H_0, asymptotically distributed as χ^2_q, where q is the degree of freedom corresponding to the number of constraints.

1.6.3 Lagrange Multiplier Test

The Lagrange Multiplier (LM) test, or the Rao Score (RS) test, only uses information from the restricted model. The test statistic is described as

$$LM = s_r' I(\theta_r)^{-1} s_r$$

where s_r is the score vector of the model evaluated at the null and $I(\theta_r)$ is a consistent estimator for the information matrix evaluated at the null. This test is, under H_0, asymptotically distributed as χ^2 with q degrees of freedom.

Note that, asymptotically, the Wald, LR and LM tests are equivalent (Engle (1984)). However, these tests yield different test statistics in finite samples. In particular, Berndt and Savin (1977) show that when the model is linear, these test statistics follow the following inequalities,

$$LM \leq LR \leq Wald.$$

The Wald and LR tests have received more attention in the literature than the LM test. However, the likelihood approach has been exploited to create a battery of LM-type specification tests for spatial regression models with

Gaussian shocks, examples of which can be found in Burridge (1980), Anselin (1988b), and Anselin et al. (1996). Burridge (1980) shows, in particular, that the test for the spatial autoregressive parameter in the disturbance process can be derived by the application of Silvey's (1959) LM method. Anselin et al. (1996) apply the LM test introduced by Bera and Yoon (1993) to the spatial models and provide simple tests based on the OLS residuals for the spatial dependence. They claim that these tests are robust and computationally simple. Debarsy and Ertur (2010) suggest a number of LM and LR test statistics to distinguish between models with endogenous spatial lag and those with spatially autocorrelated errors in a fixed effects panel data model. Finite sample performance of these tests is investigated in Monte Carlo experiments. Based on this work and Bera and Yoon (1993), He (2011) introduces locally adjusted LM tests for spatially lagged dependent variable with spatially correlated error and for spatially correlated error with spatially lagged dependent variable, respectively, and investigates the tests' finite sample performance in Monte Carlo experiments.

Anselin (2001) delivers an overview of Rao's score test applied on the spatial autoregressive and moving average processes, spatial error components and direct representation models, and introduces new Rao's Score tests for the last two spatial processes. Monte Carlo experiments are carried out and he finds that the test does not have standard asymptotic properties for the spatial error components models. For the direct representation models, in which the error covariance between two observations is a direct function of the distance between them, the nuisance parameter is identified only under the alternative. Baltagi et al. (2003) derive the Breusch and Pagan (1980)'s LM test for the panel data models that incorporate the spatial error correlation and random region effects, and test for their joint significance. Two conditional LM tests are also given for the existence of spatial error correlation while random region effects are present, and vice versa. They show that when testing for the existence of random regional effects in panel data, one should not ignore the spatial error correlation.

1.6.4 Moran's I Test

Moran's I test is one of the most popular tests for spatial correlation in a linear regression model. Applying Moran's I to regression residuals can be used to test spatial autocorrelation. The test statistic is described as

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{e'We}{e'e}$$

where n is the number of observations, e is a vector of OLS residuals, W is a weight matrix, and w_{ij} is the element of the weight matrix. Note that if the weight matrix is standardised with row sums equal to 1, the test statistic becomes $I = \frac{e'We}{e'e}$. Moran's I can yield values ranging from -1 to 1, where negative (positive) I indicates negative (positive) spatial autocorrelation. When I is equal to zero, there is no spatial dependence.

The asymptotic distribution of Moran's I statistic is developed by Cliff and Ord (1972, 1973, 1981). Anselin and Rey (1991) compare Moran's I and LM tests for spatial error autocorrelation and for a spatially lagged dependent variable based on several sample sizes, weight matrices and error distributions. They provide sample sizes for which the asymptotic properties of the tests would provide good approximations to the sampling distributions and power of the LM tests to discriminate between spatial lag and error autocorrelation. For the case when endogenous variables are included in the regression specification, the asymptotic distribution of Moran's I statistic is derived by Anselin and Kelejian (1997). This test statistic is based on residuals obtained from an IV procedure, and its small sample performance is evaluated in Monte Carlo experiments and compared with several other approaches. This test is the only acceptable test among all tests considered when spatially lagged dependent variables are present. Kelejian and Prucha (2001) provide large sample distribution of Moran's I test statistics in general and for specific spatial models. A new central limit theorem for linear-quadratic forms is also provided. Finite sample properties of Moran's I test for spatial autocorrelation in Tobit models are studied by Amaral and Anselin (2013) using Monte Carlo simulations. They find that the test statistic is unbiased and approximately normally distributed confirming the results obtained by Kelejian and Prucha (2001). They also find that the test statistic is, however, sensitive to misspecification of heteroskedasticity.

Saavedra (2003) introduces the Wald, LR and LM tests based on the work of Newey and West (1987) and the GMM estimator suggested by Kelejian and Robinson (1993), for spatial lag dependence in the spatial lag model with autocorrelated errors. The finite sample performance of these tests is investigated in a Monte Carlo experiment and compared with tests based on least squares and generalised least squares estimation.

Kelejian and Robinson (1992) introduce a test for the spatial correlation of the disturbances in large sample. They suggest that the test is computationally simple and does not need linearity of the model nor normality of the disturbances. Martellosio (2012) investigates the Cliff-Ord test and point optimal invariant tests in the framework of the spatial error model. Results show that for any fixed sample size, any fixed size of the tests, and almost any fixed weight matrix, there exists a positive measure set of regression spaces such that the limiting power disappears. In other words, it is always possible that the tests will not detect large autocorrelation in practice.

1.7 Applied Work

Spatial econometrics has been used in applied work in many fields, especially in regional science as it is able to account for spillovers while traditional econometrics is not. Some of the applied studies in spatial econometrics are summarised below.

Can (1990) extends an econometric model to include spatial neighbourhood dynamics in the hedonic housing price models. Case (1991) investigates spatial patterns in data and suggests an estimation method that allows for spatial random effects. The model is applied to demand for rice in Indonesia. Anselin et al. (1997) empirically study the degree of spatial spillovers between university research and high technology innovations. This work is broadened to the disaggregated level and the measures of local geographic spillovers are suggested in Anselin et al. (2000). Bolduc et al. (1992) propose an efficient estimation procedure using the maximum likelihood estimation to deal with the spatial autocorrelation in the error terms in travel flow models. Case et al. (1993) examine whether or not a state's spending depends on spending of its neighbouring states and find that it is positively and significantly affected by its neighbours' spending levels. Brett and Pinkse (1997) suggest a test for spatial independence in the local tax rates in British Columbia, Canada. Overmars et al. (2003) use the MR-SAR model to deal with the spatial autocorrelation in land use data. Moreno et al. (2003) investigate the spreading of innovative activity in Europe and provide a structure of this activity at the regional level. Le Gallo (2004) uses spatial markov chains approach to study the GDP disparities in the European regions. Holly et al. (2011) use generalised spatio-temporal impulse responses to study the effect of shocks on house prices in the UK.

Spatial econometric models have also been used in estimating social interactions and social network, foreign direct investment (FDI) as well as in political economy. For instance, Lee (2007b) uses the SAR model as a group effect model and estimates structural interaction effects in a social interaction model by the conditional maximum likelihood and instrumental variables methods. Bramoulle et al. (2009) define network interaction, with and without the presence of correlated effects, which yield the identified endogenous and exogenous effects. Coughlin and Segev (2000) study the FDI of the United States in Chinese provinces. Baltagi et al. (2007) include spatially weighted third-country determinants when estimating the FDI, and Blonigen et al. (2007) examine the spatial interactions in the FDI models using outbound FDI of the United States. Beck et al. (2006) discuss the use of spatial econometric models in political science and suggest that economic distances such as relative trade should be considered.

1.8 Outline of the Thesis

The thesis is organised as follows. In Chapter 2 we introduce a bias-adjusted estimator for small samples and extensively compare its performance with that of the Maximum Gaussian Likelihood and several Instrumental Variable-type estimators in the context of the spatial error model. The bias-adjusted estimator for small samples is effective across various sample sizes, being virtually mean and median unbiased. This improvement, however, comes at the cost of increasing the frequency of non-invertible estimates, which is our motivation to develop a hybrid estimator that combines the guaranteed invertibility of the MLE with the original non-hybridised estimators. We show that the hybridised forms of the estimators outperform even the best of the IV methods across a majority of the cases examined.

In Chapter 3 we introduce a sub-model for spatial weights and estimate a variable spatial weight matrix for the mixed regressive, spatial autoregressive (MR-SAR) model using maximum likelihood estimation. We establish the identifiability of the parameter defining the weights as well as the consistency and the asymptotic normality of the QMLE of the MR-SAR model under appropriate conditions that extend those given by Lee (2004a). Finite sample properties of the QMLE are investigated in a Monte Carlo study. The performance of the estimator is subsequently compared with that of other QML estimators using various fixed spatial weight matrices.

In Chapter 4 we apply our QML estimator with freely-estimated weight matrix using two types of sub-models for the spatial weights satisfying the identifiability, consistency and asymptotic normality conditions to the cross-sectional data set used in Ertur and Koch (2007) in the framework of the MR-SAR model to study the impact of saving, population growth and interdependence among countries on growth. Our QML estimator using freely-estimated weight matrices is also compared with other QML estimators using weight matrices with weight parameter values adopted in previous work. Asymptotic variances are evaluated and Wald test for our estimator is carried out.

Chapter 5 concludes. Additional results as well as detailed proofs to Chapters 2, 3 and 4 are presented in Appendices A, B and C, respectively.

Chapter 2

Improved Estimators for the Spatial Error Model

2.1 Introduction

The maximum likelihood estimator in spatial regression models with Gaussian shocks is studied by Ord (1975), Anselin (1988a) and Anselin and Bera (1998). Its asymptotic properties are developed by Lee (2004a), and those of alternative estimators have been explored in recent papers by a number of authors, examples being the spatial two-stage least squares and GMM estimators of Kelejian and Prucha (1998, 1999), a small-sample adjustment to the method of Kelejian and Prucha (1999) introduced by Arnold and Wied (2010a), the optimal instrumental variable estimator of Lee (2003), the robust HAC estimator and weighted GMM estimator of Kelejian and Prucha (2007, 2010), and the efficient GMM estimators of Lee and Liu (2010). Over the same period, the likelihood approach has been exploited to create a battery of LM-type specification tests for such models, examples of which can be found in Burridge (1980), Anselin (1988b), and Anselin, Bera, Florax and Yoon (1996). At least in the Gaussian case, and especially with large numbers of observations, the maximum likelihood method may run into numerical problems associated with matrix inversion and eigenvalue calculations, the avoidance of which is a major motivation for most of the alternative methods that have been proposed.[1]

The various GMM/IV-type estimators found in the literature are generally computationally feasible and consistent under appropriate conditions; however, they may all give non-invertible estimates of the spatial autocorrelation parameter(s) and exhibit bias in finite samples. Moreover, there is no obvious ranking of their performance in such a case. In this chapter we introduce a bias-adjusted estimator for small samples, designated [BB], and provide evidence on the small sample performance of the leading methods in a simple spatial error model.

We initially consider seven different estimators of the parameters in a spatial error model: Maximum Gaussian Likelihood [ML], the method of Kelejian and Prucha (1999) [KP], the Kelejian and Prucha method with weighting matrix (Kelejian and Prucha, 2009) [KPW], a small-sample adjustment to the KPW method [BB], the Lee and Liu (2010) method [LL], a small-sample adjustment to the KP method introduced by Arnold and Wied (2010a) [AW], and the Arnold and Wied method with an optimal weighting matrix included [AWW].[2] We define and then compare these approaches in the context of the widely employed spatial error model, SEM, an important special case of the general SARAR model. As we shall see, the adjustment to the KP and KPW estimators introduced in BB, AW, and AWW is very effective in reducing the small-sample bias. However, it tends to increase the number of samples that lead to a non-invertible estimate of the spatial error parameter. For this reason

[1]It is worth noting, however, that the dramatic reduction in computing costs over the past two decades, together with development of efficient numerical methods have reduced the importance of such obstacles: see LeSage and Pace (2009, Ch3) and Bivand (2010) for evidence on the current state of the art.

[2]We became aware of the related work of Arnold and Wied (2010a) at a late stage. In this chapter we have included two methods based on their work in our numerical experiments.

we find that a further improvement is possible by switching to an invertible estimator in such cases. This motivates a hybrid estimator that exploits the guaranteed invertibility of the parameter estimates that maximise the Gaussian likelihood to improve small-sample efficiency. The hybridisation method, which can be applied to all estimators considered, combines the guaranteed invertibility of the MLE with the original non-hybridised estimator, in which all the parameters' estimates are replaced with the MLE estimates when the original non-hybridised estimator produces a non-invertible estimate of the spatial error parameter. We find that the hybridised forms of the BB, AW and AWW estimators are superior to the other estimators considered in reducing the small-sample bias.

Reducing the mean square errors of parameter estimates is of course an important objective, but since in practice such parameter estimates will usually be reported with their associated standard errors or in the form of $t-$ statistics, it is equally important to assess the reliability of the inferences which may then be drawn. We therefore provide evidence on this, finding the hybridised forms of BB, AW, and AWW estimators perform extremely well.

The chapter is organised as follows. Section 2.2 introduces the model and defines the estimators, Section 2.3 describes the experimental results, and Section 2.4 concludes.

2.2 The Model and Estimators

2.2.1 The Spatial Error Regression Model

The spatial error model, SEM, is described as follows:

$$Y = X\beta + U \tag{2.1}$$

where Y is an $n \times 1$ vector of observations of the dependent variable, X is an $n \times k$ matrix of values of k exogenous explanatory variables, β is a $k \times 1$ vector of parameters, and U is the $n \times 1$ disturbance vector defined as

$$U = \rho M U + \varepsilon \tag{2.2}$$

with M an $n \times n$ matrix of fixed non-negative constants, $m_{i,j}$, where for each row i, $\sum_{j=1}^{n} m_{i,j} = 1$ and $m_{i,i} = 0$, ρ is a scalar parameter, and ε an $n \times 1$ vector of innovations that are independently distributed with mean 0 and variance $\sigma^2 I$, independent of X. The objective is to estimate β, ρ and σ^2 from a single sample of n observations taken at the spatial units indexed by $i = 1, ..., n$. If the innovations, ε are Gaussian, the log-likelihood function is given by

$$\ln(L) = -\frac{n}{2}[\ln(\sigma^2) + \ln(2\pi)] - \frac{1}{2\sigma^2} U'(\beta) B'(\rho) B(\rho) U(\beta) + \ln|det(B(\rho))|, \tag{2.3}$$

where $U(\beta) = Y - X\beta$, $B(\rho) = I - \rho M$, and $|det(B(\rho))|$ denotes the absolute value of the determinant of $B(\rho)$. To avoid degeneracy it is necessary that the matrix, $B(\rho)$ be non-singular. Commonly, M will have been constructed from a symmetric matrix of positive elements, by row-standardisation; as is well known, (see Ord, 1975, p.125) in such a case all the eigenvalues of M are real, and so will be those of $B(\rho)$; taking $[-\frac{1}{\omega_{min}} < \rho < 1]$, where ρ lies in the 'invertible region' and ω_{min} is the largest negative eigenvalue in absolute value of matrix M,[3] will then ensure that $B(\rho)$ is non-singular, as required; see for example Lee and Liu (2010, Endnote #6).

For the case of ρ lying outside the invertible region, i.e. outside $(-\frac{1}{\omega_{min}}, 1)$, it is then called 'non-invertible' and can result in $B(\rho)$ being singular.

[3] See Anselin (1988a, p. 78-79, Endnote #10).

2.2.2 The Maximum Likelihood Estimator

The maximum likelihood estimator is obtained by maximising (2.3) with respect to β, ρ and σ^2. The score vector is:

$$\begin{aligned}
\frac{\partial \ln(L)}{\partial \beta} &= \frac{1}{\sigma^2} X'B'BU \\
\frac{\partial \ln(L)}{\partial \rho} &= -\frac{1}{2\sigma^2} U'[2\rho M'M - M - M']U - tr[MB^{-1}] \\
\frac{\partial \ln(L)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} U'B'BU
\end{aligned} \tag{2.4}$$

where we have suppressed the dependence of U and B on the parameters to enhance legibility. A little manipulation and rearrangement yields the more transparent forms for the f.o.c.:

$$X'B'BU = 0 \tag{2.5}$$

$$U'B'[MB^{-1} - \frac{tr[MB^{-1}]}{n}.I_n]BU = 0 \tag{2.6}$$

$$\hat{\sigma}^2_{ML} = \frac{1}{n} U'B'BU. \tag{2.7}$$

Introducing the notation, $\hat{\theta}_{ML} = (\hat{\beta}_{ML}, \hat{\rho}_{ML}, \hat{\sigma}^2_{ML})'$ would enable (2.7) to be written as $\hat{\sigma}^2_{ML} = \frac{1}{n}\varepsilon(\hat{\theta}_{ML})'\varepsilon(\hat{\theta}_{ML})$ and so on. Finding the solution to these moment conditions is a numerical problem that in practice may be solved to whatever degree of accuracy can be obtained in the evaluation of $\ln|det(B(\rho))|$. Notice that the MLE of ρ, $\hat{\rho}_{ML}$, will be forced to lie in the invertible region by the behaviour of $\ln|det(B(\rho))|$ near the invertibility boundary.[4] We return to these conditions when we discuss the LL estimator, below.

The asymptotic covariance matrix of $(\hat{\rho}_{ML}, \hat{\sigma}^2_{ML})$ is given by

$$AsyVar(\hat{\rho}_{ML}, \hat{\sigma}^2_{ML}) = \sigma^4 \begin{bmatrix} \sigma^4 tr((MB^{-1})^2 + B^{-1\prime}M'MB^{-1}) & \sigma^2 tr(MB^{-1}) \\ \sigma^2 tr(MB^{-1}) & n/2 \end{bmatrix}^{-1},$$

and the estimator of β is asymptotically uncorrelated with $(\hat{\rho}_{ML}, \hat{\sigma}^2_{ML})$ with

[4] At the invertibility boundary, $\ln|det(B(\rho))| = -\infty$.

asymptotic variance given by

$$AsyVar(\hat{\beta}_{ML}) = \sigma^2(X'B'BX)^{-1}, \tag{2.8}$$

see Ord (1975, p.125, eq. B3, after correcting the sign on his α). Computational aspects of maximising (2.3) are discussed in LeSage and Pace (2009, Ch.3).

2.2.3 The KP and KPW Estimators

Kelejian and Prucha (1999) suggest a multi-step estimation procedure: an initial consistent estimator of β is used to obtain a vector of residuals; a method-of-moments estimator is applied to the first step residuals to estimate σ^2 and the autoregressive parameter ρ, and in the final step a feasible GLS estimator is used to re-estimate β.

To estimate ρ and σ^2 Kelejian and Prucha (1999) use three moment conditions:

$$E[\frac{1}{n}\varepsilon'\varepsilon] = \sigma^2 \tag{2.9}$$

$$E[\frac{1}{n}\varepsilon' M' M\varepsilon] = \frac{\sigma^2}{n} tr(M'M) \tag{2.10}$$

$$E[\frac{1}{n}\varepsilon' M'\varepsilon] = \frac{\sigma^2}{n} tr(M') = 0 \tag{2.11}$$

These moment conditions imply that

$$E\{\Gamma_n \begin{bmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{bmatrix} - \gamma_n\} = 0 \tag{2.12}$$

where

$$\Gamma_n = \begin{bmatrix} \frac{2}{n}(U'MU) & -\frac{1}{n}(U'M'MU) & 1 \\ \frac{2}{n}(U'M'M'MU) & -\frac{1}{n}(U'M'M'MMU) & \frac{1}{n}tr(M'M) \\ \frac{1}{n}(U'[MM + M'M]U) & -\frac{1}{n}(U'M'MMU) & 0 \end{bmatrix}$$

$$\gamma_n = \begin{bmatrix} \frac{1}{n}(U'U) \\ \frac{1}{n}(U'M'MU) \\ \frac{1}{n}(U'MU) \end{bmatrix}.$$

For use later, suppose $\lim_{n\to\infty} E(\Gamma_n) = \Gamma$, a finite matrix of constants.

To implement these moment conditions U is required; however, U is not directly observed, and so in the [KP] estimator it is replaced by $\hat{U}$ the vector of ordinary least squares residuals from the equation

$$Y = X\hat{\beta}_{OLS} + \hat{U}.$$

Writing G_n and g_n for Γ_n and γ_n with $\hat{U}$ in place of U define the 3×1 vector of moment residuals,

$$\nu_n(\rho, \sigma^2) = G_n \begin{bmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{bmatrix} - g_n \tag{2.13}$$

Then the [KP] generalised moments estimators for ρ and σ^2 are obtained by minimising the sum of squares of ν_n:

$$(\hat{\rho}_{KP}, \hat{\sigma}^2_{KP}) = argmin[\nu_n(\rho, \sigma^2)'\nu_n(\rho, \sigma^2)]. \tag{2.14}$$

The final step is a feasible GLS estimator of β obtained by performing an OLS regression of $B(\hat{\rho}_{KP})Y$ on $B(\hat{\rho}_{KP})X$.

Kelejian and Prucha (1999) give conditions under which the above method provides consistent estimates for ρ, σ^2 and β. However, their method is not efficient. In particular, note that the sum of squares in (2.14) is unweighted, so that one direct way to improve efficiency is to include an optimal weighting matrix in the method of moments procedure. This is done in Kelejian and Prucha (2010).

The required optimal weighting matrix, Ψ_n^{opt} say, has probability limit equal to the inverse of the asymptotic variance of the moment residuals.

So, in our finite sample implementation, we choose the weighting matrix $\Psi_n = \{NVar[\nu_n(\rho, \sigma^2, \beta)]\}^{-1}$, which corresponds to that given in Kelejian and Prucha (2010) after specialising to the present SEM case:

$$\Psi_n = \frac{1}{\sigma^4} \begin{bmatrix} 2 & \frac{2}{n}tr(M'M) & 0 \\ \frac{2}{n}tr(M'M) & \frac{2}{n}tr(M'MM'M) & \frac{2}{n}tr(M'MM') \\ 0 & \frac{2}{n}tr(M'MM') & \frac{1}{n}tr((M+M')M) \end{bmatrix}^{-1}.$$

Suppose $\lim_{n\to\infty} \Psi_n = \Psi$ a finite positive definite matrix. Observe that Ψ_n is parameter free, apart from a scalar division by σ^4 which is irrelevant to the minimisation. The estimators for ρ and σ^2 for the [KPW] approach are obtained as

$$(\hat{\rho}_{KPW}, \hat{\sigma}^2_{KPW}) = argmin[\nu_n(\rho, \sigma^2)'\Psi_n \nu_n(\rho, \sigma^2)] \tag{2.15}$$

and an estimate of β is again obtained by performing an OLS regression of $B(\hat{\rho}_{KPW})Y$ on $B(\hat{\rho}_{KPW})X$. We denote this second estimator as [KPW].

The asymptotic distribution of generalisations of the estimator of Kelejian and Prucha (1999) has recently been derived under appropriate conditions in Kelejian and Prucha (2010). In the case of the two estimators described above $\hat{\theta}_{KP} = (\hat{\beta}_{KP}, \hat{\rho}_{KP}, \hat{\sigma}^2_{KP})'$ and $\hat{\theta}_{KPW} = (\hat{\beta}_{KPW}, \hat{\rho}_{KPW}, \hat{\sigma}^2_{KPW})'$ the asymptotic distribution is as follows.

Define

$$J_n = \frac{\partial \nu_n(\rho, \sigma^2)}{\partial(\rho, \sigma^2)} = \Gamma_n \begin{bmatrix} 1 & 0 \\ 2\rho & 0 \\ 0 & 1 \end{bmatrix} \tag{2.16}$$

with $\lim_{n\to\infty} J_n = J$.

Then the asymptotic distribution of $(\hat{\rho}_{KP}, \hat{\sigma}^2_{KP})'$ is

$$n^{1/2} \left(\begin{bmatrix} \hat{\rho}_{KP} \\ \hat{\sigma}^2_{KP} \end{bmatrix} - \begin{bmatrix} \rho \\ \sigma^2 \end{bmatrix} \right) \overset{a}{\sim} N(0, \Omega_{KP}) \tag{2.17}$$

with

$$\Omega_{KP} = (J'J)^{-1}J'\Psi J(J'J)^{-1} \tag{2.18}$$

whereas the asymptotic distribution of $(\hat{\rho}_{KPW}, \hat{\sigma}^2_{KPW})'$ is

$$n^{1/2}\left(\begin{bmatrix} \hat{\rho}_{KPW} \\ \hat{\sigma}^2_{KPW} \end{bmatrix} - \begin{bmatrix} \rho \\ \sigma^2 \end{bmatrix}\right) \overset{a}{\sim} N(0, \Omega_{KPW}) \tag{2.19}$$

with

$$\Omega_{KPW} = (J'\Psi^{-1}J)^{-1}. \tag{2.20}$$

Estimates of β obtained from an OLS regression of $B(\hat{\rho}_{KP})Y$ on $B(\hat{\rho}_{KP})X$ or from an OLS regression of $B(\hat{\rho}_{KPW})Y$ on $B(\hat{\rho}_{KPW})X$ are asymptotically uncorrelated with $(\hat{\rho}_{KP}, \hat{\sigma}^2_{KP})$ or $(\hat{\rho}_{KPW}, \hat{\sigma}^2_{KPW})$, and the asymptotic variance is given by

$$AsyVar(\tilde{\beta}) = \sigma^2(X'B'(\rho)B(\rho)X)^{-1} \tag{2.21}$$

which is the same as that of the MLE (2.8); see Kelejian and Prucha (2010) for details.

2.2.4 The Bias-Adjusted Estimators, BB, AW, and AWW

Both the KP and KPW estimators use an initial consistent estimate of β in order to construct residuals, $\hat{U}$, that are substituted for the unobservable U in the evaluation of the moment conditions used to estimate ρ and σ^2. Although this leads to a consistent estimate of ρ and σ^2, there may be a substantial bias in small samples, because, as we show below, the expectation of ν_n in (2.13) is not zero. Therefore, we propose a simple bias adjustment, and designate the modified estimator, BB.

Recall that ε has been assumed independent of X with $E\{\varepsilon\} = 0$ and $Var\{\varepsilon\} = \sigma^2 I_n$. Estimating β using OLS we obtain

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'(X\beta + U) \tag{2.22}$$

giving the residual

$$\begin{aligned} \hat{U} &= (I - X(X'X)^{-1}X')U \\ &= AU = AB^{-1}\varepsilon, \text{ say} \end{aligned} \tag{2.23}$$

where we drop the explicit dependence of B on ρ for readability.

Substituting from (2.23) into G and g and taking expectations we find, with a little routine algebra, and writing $C = AB^{-1}$ that

$$E\{G_n \begin{bmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{bmatrix} - g_n\} = \frac{\sigma^2}{n} \begin{bmatrix} n - tr(C'B'BC) \\ tr(M'M - C'B'M'MBC) \\ -tr(C'B'MBC) \end{bmatrix} \tag{2.24}$$

$$= \frac{\sigma^2}{n} \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}, \text{ say.} \tag{2.25}$$

We may adjust for this bias in the moment conditions by replacing G_n by

$$\hat{G}_n(\rho) = \frac{1}{n} \begin{bmatrix} 2\hat{U}'M\hat{U} & -\hat{U}'M'M\hat{U} & tr(C'B'BC) \\ 2\hat{U}'M'M'M\hat{U} & -\hat{U}'M'M'MM\hat{U} & tr(C'B'M'MBC) \\ \hat{U}'[MM + M'M]\hat{U} & -\hat{U}'M'MM\hat{U} & tr(C'B'MBC) \end{bmatrix} \tag{2.26}$$

in which B and C are functions of ρ. Writing

$$\hat{\nu}_n(\rho, \sigma^2) = \hat{G}_n(\rho) \begin{bmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{bmatrix} - g_n \tag{2.27}$$

and equation (2.15) then becomes

$$(\hat{\rho}_{BB}, \hat{\sigma}^2_{BB}) = argmin\{\hat{\nu}_n(\rho, \sigma^2)'\Psi_n \hat{\nu}_n(\rho, \sigma^2)\} \tag{2.28}$$

The estimate of β is obtained as before by performing an OLS regression of $B(\hat{\rho}_{BB})Y$ on $B(\hat{\rho}_{BB})X$. The asymptotic distribution of $(\hat{\rho}_{BB}, \hat{\sigma}^2_{BB})'$ is the same as that of $(\hat{\rho}_{KPW}, \hat{\sigma}^2_{KPW})'$ as in (2.19).

Our BB estimator is developed independently of a similar method introduced by Arnold and Wied (2010a) [AW]. This estimator may be considered to lie between the KP and BB estimators. The KP and BB estimators start with using ε in the moment conditions (2.9) - (2.11), which imply (2.12). Then the KP estimator replaces U with $\hat{U}$ to obtain the moment residuals in (2.13), whose expectation is not zero in finite sample. The BB estimator, instead, substitutes $\hat{U} = AU = AB^{-1}\varepsilon$ into G_n and g to obtain the moment residuals in (2.27).

The AW estimator, on the other hand, starts with $\hat{\epsilon} = A\epsilon = AU - \rho AMU$, instead of ϵ, in the moment conditions such that

$$E\{\check{\Gamma}_n \begin{bmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{bmatrix} - g_n\} = 0 \tag{2.29}$$

where

$$\check{\Gamma}_n = \frac{1}{n} \begin{bmatrix} 2U'AMU & -U'M'AMU & tr(A) \\ 2U'M'MAMAU & -U'M'AM'MAMU & tr(AM'M) \\ (U'A[M+M']AMU) & -(U'M'AMAMU) & tr(MA) \end{bmatrix}$$

and calculates the expectation in (2.29) above based on $A\varepsilon$ while, for the KP estimator, the expectation in (2.12) is calculated based on ε. Then the AW estimator replaces U with $\hat{U} = AU$ and applies $tr(A) = \frac{n-k}{n}$ to obtain the following moment residuals

$$\check{\nu}_n(\rho, \sigma^2) = \check{G}_n(\rho) \begin{bmatrix} \rho \\ \rho^2 \\ \sigma^2 \end{bmatrix} - g_n \tag{2.30}$$

where

$$\check{G}_n(\rho) = \frac{1}{n} \begin{bmatrix} 2\hat{U}'M\hat{U} & -\hat{U}'M'AM\hat{U} & n-k \\ 2\hat{U}'M'MAM\hat{U} & -\hat{U}'M'AM'MAM\hat{U} & tr(AM'M) \\ \hat{U}'[MAM+M'AM]\hat{U} & -\hat{U}'M'AMAM\hat{U} & tr(MA) \end{bmatrix}. \tag{2.31}$$

So equation (2.15) becomes

$$(\hat{\rho}_{AW}, \hat{\sigma}^2_{AW}) = argmin\{\check{\nu}_n(\rho, \sigma^2)'\check{\nu}_n(\rho, \sigma^2)\} \tag{2.32}$$

and the asymptotic distribution of $(\hat{\rho}_{AW}, \hat{\sigma}^2_{AW})'$ is the same as that of $(\hat{\rho}_{KP}, \hat{\sigma}^2_{KP})'$ as in equation (2.17).

While the expectation in (2.29) is zero, the expectation of the moment residuals in (2.30) may not be zero in finite sample as the AW method uses $\hat{\varepsilon}$ in the moment conditions in the first step and subsequently replaces U with $\hat{U}$, which could lead to a bias in small samples.

Note that Arnold and Wied (2010a) do not include the optimal weighting matrix Ψ_n in their estimator AW, which also makes this method not efficient. To make the AW estimator efficient, and to be able to compare its performance with other estimators, we also include the optimal weighting matrix Ψ_n in their AW estimator, designated [AWW], and compare both the AW and the AWW estimators separately with the other estimators in Section 2.3.

The asymptotic distribution of $(\hat{\rho}_{AWW}, \hat{\sigma}^2_{AWW})'$ is the same as that of $(\hat{\rho}_{KPW}, \hat{\sigma}^2_{KPW})'$ as in equation (2.19).

2.2.5 The Lee and Liu (2010) Estimator, LL

The main difference between the approaches of Kelejian and Prucha (1999), and Lee and Liu (2010) lies in the nature of the moment conditions. The model studied by Lee and Liu (2010) is a mixed regression-spatial autoregression with spatially autoregressive disturbances, of which the SEM in this chapter is a special case. Lee and Liu introduce an infeasible GMM estimator that is the best in the class determined by linear and quadratic moment conditions as in their Proposition 3 (Lee and Liu, 2010, p.196) and surrounding discussion. That is, the infeasible estimator delivers the smallest asymptotic variance-covariance matrix. The motivation for considering this class of moment conditions is to imitate the structure of the score function of the Gaussian likelihood (see Lee and Liu 2010, p.192). To implement their "best GMM" estimator in the mixed regressive, spatially autoregressive model, in a numerical experiment, Lee and

Liu define a feasible two-step approximation to it in which the first step uses the KP estimator for that model, the so-called "generalised two stage least squares estimator" of Kelejian and Prucha (1998), to generate the parameter-dependent matrices appearing in the moment conditions (Lee and Liu 2010 p.200 and Endnote 27).

Specialising their analysis to the Gaussian SEM case, the relevant moment conditions are

$$E[\frac{1}{n}\varepsilon(\rho,\beta)'\varepsilon(\rho,\beta)] = \sigma^2 \tag{2.33}$$

$$E[\frac{1}{n}\varepsilon(\rho,\beta)'P_1\varepsilon(\rho,\beta)] = \frac{\sigma^2}{n}tr(P_1) \tag{2.34}$$

$$E[\frac{1}{n}Q'\varepsilon(\rho,\beta)] = 0 \tag{2.35}$$

where

$$\varepsilon(\rho,\beta) = BU = (I-\rho M)(Y - X\beta)$$

$$P_1 = MB^{-1}$$

and

$$Q = BX$$

Notice there are now $2+k$ moment conditions and $2+k$ parameters, so the model is exactly identified under the conditions specified in Lee and Liu (2010) and the system of moment conditions (2.33) - (2.35) could be solved exactly. However, Q and P_1 both depend on the unknown true parameter ρ. To obtain a simple feasible estimator in the spirit of Lee and Liu, an initial consistent estimator of ρ, say $(\tilde{\rho})$, is used in order to obtain an estimate of Q and P_1, that is, $\tilde{Q} = B(\tilde{\rho})X$ and $\tilde{P}_1 = MB(\tilde{\rho})^{-1}$. In our simulations we have used the KPW estimator for ρ as the initial consistent estimator $\tilde{\rho}$.

Then, the following nonlinear system of equations is solved to obtain the final estimators for parameters ρ, σ^2, and β. That is, define $\theta \equiv (\beta', \rho, \sigma^2)'$ as above and $g(\theta) \equiv (g_1(\theta), g_2(\theta), g_3(\theta))'$ with

$$g_1(\theta) = \frac{1}{n}\varepsilon(\rho,\beta)'\varepsilon(\rho,\beta) - \sigma^2 \tag{2.36}$$

$$g_2(\theta) = \frac{1}{n}\varepsilon(\rho,\beta)'\tilde{P}_1\varepsilon(\rho,\beta) - \frac{\sigma^2}{n}tr(\tilde{P}_1) \tag{2.37}$$

$$g_3(\theta) = \frac{1}{n}\tilde{Q}'\varepsilon(\rho,\beta). \tag{2.38}$$

Then the LL estimator for θ is obtained by solving $g(\theta) = 0$. Observe that by making the substitution, $\varepsilon = BU$ in (2.5) to (2.7) we recover equations with the same structure as (2.33) to (2.35), thus the LL moment conditions, if solved exactly, would yield the Gaussian MLE for the present model. The asymptotic efficiency of the LL estimator is thus the same as that of the MLE in the Gaussian case, as proved in general in LL (2010, Propositions 5 and 6).

2.2.6 The Hybrid Estimator

In the numerical experiments to be reported in the next section we find that the bias-adjusted estimators BB, AW, and AWW reduce the bias of the KP and KPW estimators, but at the cost of increasing the frequency of non-invertible estimates of the ρ parameter. This situation may be explained as follows. When small-sample bias is not corrected for, it pushes the estimator's distribution to the left, causing the estimates to be biased downward. Correcting for small-sample bias by a bias-adjusted estimator, on the other hand, shifts the estimator's distribution back to the right, resulting in more estimates of ρ becoming non-invertible.

The only estimator that is guaranteed to yield invertible $\hat{\rho}$ values is the MLE; this motivates the hybrid estimator, which will be applied to all estimators considered in the next section. The hybrid estimator is equal to the original non-hybridised estimator when that gives an invertible estimate of ρ, but is equal to the MLE otherwise. Note that when the original non-hybridised estimator produces non-invertible estimates of ρ, we not only replace the estimates of ρ, but also estimates of all other parameters with the MLE estimates. In the experiments, the hybridised forms of the BB, KP, KPW, LL, AW, and

AWW estimators are designated BB2, KP2, KPW2, LL2, AW2, and AWW2, respectively.

2.3 Simulation Results

2.3.1 Experiment Design

We perform experiments for two sets of spatial weight matrices: the first set is for $n = 20, 50, 100, 245$, and 490, and the second set is for $n = 49$. For the first set, the matrix M is generated by randomly drawing n pairs of coordinates from a standard bivariate Normal distribution to which the Delaunay routine is then applied to produce Voronoi polygons. The contiguity weights are then based on the set of nearest neighbouring polygons, and subsequently row-standardised; for $n = 49$ the matrix M is a row-standardised form of the Columbus weights used by Anselin (1988a). For each n in the first set we draw three different M using each one in 1000 replications which are then pooled. For $n = 49$ a single set of 3000 replications is used.

For each M the matrix X consists of 3 columns with associated coefficients, $\beta_1 = 1, \beta_2 = 0, \beta_3 = -1$. The first column of X is the constant vector, 1, and the other two columns are 1000 independent draws from the standard $n-$variate Normal distribution. For each M and X we draw a further independent standard Normal vector of disturbances, ε. For each M, X and ε draw for $n \in (20, 50, 100, 245, 490)$ the variance of ε is fixed at 1; for $n = 49$ ε is scaled to have variance equal to 0.25, 0.5, 1.0, or 2.0, and in each case we form $U = B(\rho)^{-1}\varepsilon$ using values of $\rho \in (0.0, 0.1, 0.2, ..., 0.9)$.

As we look at the problem in the region where $[-\frac{1}{\omega_{min}} < \rho < 1]$, with ω_{min} the largest negative eigenvalue in absolute value of matrix M, we impose bounds on ρ to be $|\rho| \leq 0.99$ in the simulations to ensure that $B(\rho)^{-1}$ exists and to speed up the simulation.[5] Note that for the MLE, the likelihood is heavily penalised as the boundary is approached, and the estimator never hits the upper bound. The results obtained for the randomly generated spatial weight matrices were compared for consistency across weight structures; we found no significant variation and so the three samples of 1000 replications for each n are combined into a single sample of 3000 in the tables that follow.

In this chapter we only report the simulation results for a selection of cases. In particular, t statistics and estimates of ρ from all non-hybridised and hybridised estimators for various values of n are reported in the next subsection. Estimates of other parameters obtained for σ^2 fixed at unity and various values of n are also reported in this chapter. For the results for these parameters obtained for other values of σ^2 for the case of $n = 49$, see Appendix A. All other results not reported in this thesis are available on request.

2.3.2 Estimates of ρ

Tables 2.1 - 2.6 give the mean, median, standard deviation and root mean square error of the estimators of ρ for $n \in (20, 100, 245)$. Tables 2.1, 2.3, 2.5 give results for the non-hybridised estimators, while Tables 2.2, 2.4, 2.6 show the effects of hybridising each of the estimators with the MLE whenever the boundary constraint $|\hat{\rho}| \leq 0.99$ is binding. Note that BB1, KP1, KPW1, LL1, AW1, and AWW1 stand for the non-hybridised BB, KP, KPW, LL, AW, and AWW estimators, whereas BB2, KP2, KPW2, LL2, AW2, and AWW2 stand for the hybridised BB, KP, KPW, LL, AW, and AWW estimators. The results given in bold show the lowest bias among all estimators for each measurement category. The summary tables, 2.7 and 2.8 show the best non-hybridised and hybridised estimators, respectively, with the relative efficiencies of the BB1 or

[5] The purpose was to speed up the simulations, which would crash numerically at search points very close to the boundary. This modification is costless given that the final estimate of ρ cannot lie close to the boundary.

BB2 estimators given in parentheses; in these tables, a * indicates that the mean squared error of the best method is significantly lower than that of the second-best method as measured by a z-test at 5%.

Looking first at Tables 2.1 and 2.2, notice that for $n = 20$ the bias adjustment implemented in BB1, AW1, and AWW1 is highly effective, producing an estimator with very much smaller bias than any of the others across all the ρ values considered. However, all the estimators, except the MLE, produce a significant proportion of non-invertible estimates, especially for $\rho = 0.9$ as evidenced in Figure 2.1, where the histograms in the left column show that this proportion is particularly high for BB1, AW1, and AWW1. For practical use, therefore, we introduce the hybrid estimators in Table 2.2, where all estimators are hybridised if they produce non-invertible estimates of ρ. Here we see a similar pattern repeated, though with, obviously, a less striking reduction in bias. Notice that although the hybrid BB2, AW2, and AWW2 have larger bias than their non-hybridised ones, they have lower bias than any of the other hybridised estimators. Moreover, as evidenced in Figure 2.1, the histograms in the right column show that all hybridised estimators no longer produce non-invertible estimates. When the sample size is increased to 100 the BB1 and AWW1 estimators are almost unbiased, and retain their clear advantage over the other non-hybrid estimators. After hybridising, as shown in Table 2.4, the BB2, AW2, and AWW2 estimators have very small increase in bias and remain much superior to the other hybridised estimators. Figure 2.2 shows that the non-hybridised bias-adjusted estimators hit the invertibility constraint less often than the case for $n = 20$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB1	0.032	0.034	0.499	0.500	0.5	BB1	0.471	0.526	0.415	**0.416**
	ML	-0.251	-0.233	0.446	0.511		ML	0.212	0.299	0.418	0.508
	KP1	-0.237	-0.215	**0.423**	0.485		KP1	0.205	0.262	0.417	0.511
	KPW1	-0.092	-0.091	0.452	**0.462**		KPW1	0.332	0.366	**0.409**	0.443
	LL1	-0.276	-0.271	0.455	0.532		LL1	0.175	0.244	0.473	0.574
	AW1	-0.055	**-0.020**	0.474	0.477		AW1	0.421	**0.497**	0.429	0.436
	AWW1	**0.021**	0.027	0.494	0.494		AWW1	**0.477**	0.545	0.421	0.421
0.1	BB1	0.121	0.133	0.489	0.490	0.7	BB1	0.636	0.718	**0.354**	**0.359**
	ML	-0.163	-0.122	0.449	0.521		ML	0.410	0.511	0.375	0.475
	KP1	-0.155	-0.124	**0.429**	0.499		KP1	0.401	0.468	0.386	0.489
	KPW1	-0.009	-0.003	0.451	**0.464**		KPW1	0.504	0.553	0.366	0.415
	LL1	-0.191	-0.172	0.464	0.548		LL1	0.376	0.472	0.459	0.561
	AW1	0.038	**0.080**	0.473	0.477		AW1	0.617	**0.714**	0.376	0.385
	AWW1	**0.112**	0.131	0.487	0.487		AWW1	**0.656**	0.753	0.361	0.364
0.3	BB1	**0.298**	0.333	0.460	0.460	0.9	BB1	0.803	**0.917**	**0.260**	**0.278**
	ML	0.021	0.088	0.442	0.522		ML	0.611	0.703	0.309	0.423
	KP1	0.019	0.063	**0.431**	0.514		KP1	0.606	0.681	0.330	0.443
	KPW1	0.160	0.184	0.438	**0.460**		KPW1	0.674	0.741	0.303	0.378
	LL1	-0.014	0.028	0.475	0.570		LL1	0.591	0.697	0.410	0.513
	AW1	0.227	**0.286**	0.460	0.465		AW1	0.799	0.944	0.290	0.306
	AWW1	0.295	0.338	0.462	0.462		AWW1	**0.818**	0.973	0.275	0.287

Table 2.1: Non-hybridised estimators of ρ for $n = 20$, $\sigma^2 = 1$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB2	**-0.004**	0.022	0.459	0.459	0.5	BB2	0.414	0.477	0.381	0.391
	ML	-0.251	-0.233	0.446	0.511		ML	0.212	0.299	0.418	0.508
	KP2	-0.232	-0.215	**0.416**	0.477		KP2	0.203	0.262	0.409	0.506
	KPW2	-0.104	-0.091	0.430	**0.443**		KPW2	0.303	0.354	0.380	0.428
	LL2	-0.271	-0.258	0.441	0.517		LL2	0.169	0.249	0.437	0.548
	AW2	-0.058	**-0.020**	0.466	0.470		AW2	0.397	0.489	0.404	0.417
	AWW2	-0.006	0.023	0.459	0.459		AWW2	**0.418**	**0.493**	**0.379**	**0.388**
0.1	BB2	**0.082**	**0.117**	0.450	0.450	0.7	BB2	0.566	0.644	**0.325**	0.351
	ML	-0.163	-0.122	0.449	0.521		ML	0.410	0.511	0.375	0.475
	KP2	-0.151	-0.124	**0.422**	0.491		KP2	0.395	0.468	0.378	0.485
	KPW2	-0.023	-0.004	0.428	**0.446**		KPW2	0.468	0.529	0.337	0.409
	LL2	-0.186	-0.162	0.446	0.529		LL2	0.366	0.470	0.406	0.526
	AW2	0.031	0.080	0.461	0.466		AW2	0.574	**0.682**	0.346	0.368
	AWW2	0.081	0.120	0.451	0.452		AWW2	**0.575**	0.662	0.325	**0.348**
0.3	BB2	0.250	**0.300**	0.422	0.425	0.9	BB2	0.710	0.790	0.262	0.324
	ML	0.021	0.088	0.442	0.522		ML	0.611	0.703	0.309	0.423
	KP2	0.020	0.063	0.424	0.508		KP2	0.594	0.678	0.320	0.443
	KPW2	0.140	0.178	**0.412**	0.442		KPW2	0.634	0.707	0.281	0.387
	LL2	-0.011	0.041	0.444	0.542		LL2	0.569	0.673	0.356	0.486
	AW2	0.213	0.285	0.442	0.450		AW2	**0.725**	**0.809**	0.259	0.313
	AWW2	**0.251**	0.312	0.422	**0.424**		AWW2	0.716	0.800	**0.251**	**0.312**

Table 2.2: Hybridised estimators of ρ for $n = 20$, $\sigma^2 = 1$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB1	**-0.000**	0.006	0.183	0.183	0.5	BB1	0.494	0.504	0.143	0.143
	ML	-0.049	-0.040	0.179	0.186		ML	0.450	0.465	**0.136**	0.144
	KP1	-0.053	-0.041	0.179	0.187		KP1	0.446	0.458	0.139	0.149
	KPW1	-0.028	-0.022	**0.178**	**0.180**		KPW1	0.459	0.469	0.140	0.146
	LL1	-0.050	-0.039	0.179	0.185		LL1	0.447	0.462	0.137	0.146
	AW1	-0.019	-0.006	0.181	0.182		AW1	0.484	**0.497**	0.140	**0.141**
	AWW1	-0.000	**0.005**	0.182	0.182		AWW1	**0.496**	0.508	0.142	0.142
0.1	BB1	**0.099**	0.106	0.177	0.177	0.7	BB1	0.693	0.702	0.115	0.115
	ML	0.051	0.063	0.174	0.181		ML	0.651	0.666	**0.107**	0.117
	KP1	0.045	0.058	0.174	0.182		KP1	0.650	0.662	0.113	0.123
	KPW1	0.069	0.076	0.173	**0.176**		KPW1	0.658	0.667	0.113	0.121
	LL1	0.049	0.062	**0.173**	0.180		LL1	0.646	0.662	0.110	0.123
	AW1	0.081	**0.094**	0.175	0.176		AW1	0.690	**0.701**	0.113	**0.113**
	AWW1	0.098	0.106	0.177	0.177		AWW1	**0.698**	0.709	0.114	0.114
0.3	BB1	0.296	0.306	0.163	0.163	0.9	BB1	0.891	**0.900**	0.072	0.073
	ML	0.250	0.264	0.158	0.165		ML	0.854	0.869	**0.065**	0.080
	KP1	0.244	0.257	0.159	0.169		KP1	0.857	0.866	0.074	0.086
	KPW1	0.263	0.273	0.159	0.163		KPW1	0.860	0.869	0.073	0.083
	LL1	0.248	0.263	**0.157**	0.166		LL1	0.844	0.860	0.080	0.098
	AW1	0.281	**0.295**	0.160	**0.161**		AW1	**0.899**	0.908	0.072	0.072
	AWW1	**0.297**	0.306	0.162	0.162		AWW1	0.902	0.913	0.070	**0.070**

Table 2.3: Non-hybridised estimators of ρ for $n = 100$, $\sigma^2 = 1$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB2	**-0.000**	0.006	0.183	0.183	0.5	BB2	0.493	0.504	0.141	0.141
	ML	-0.049	-0.040	0.179	0.186		ML	0.450	0.465	**0.136**	0.144
	KP2	-0.053	-0.041	0.179	0.187		KP2	0.446	0.458	0.139	0.149
	KPW2	-0.028	-0.022	**0.178**	**0.180**		KPW2	0.459	0.469	0.140	0.145
	LL2	-0.050	-0.039	0.179	0.185		LL2	0.447	0.462	0.137	0.146
	AW2	-0.019	-0.006	0.181	0.182		AW2	0.484	**0.497**	0.140	**0.141**
	AWW2	-0.000	**0.005**	0.182	0.182		AWW2	**0.496**	0.508	0.142	0.142
0.1	BB2	**0.099**	0.106	0.177	0.177	0.7	BB2	0.691	**0.701**	0.112	**0.112**
	ML	0.051	0.063	0.174	0.181		ML	0.651	0.666	**0.107**	0.117
	KP2	0.045	0.058	0.174	0.182		KP2	0.650	0.662	0.113	0.123
	KPW2	0.069	0.076	0.173	**0.176**		KPW2	0.658	0.667	0.112	0.120
	LL2	0.049	0.062	**0.173**	0.180		LL2	0.646	0.662	0.110	0.123
	AW2	0.081	**0.094**	0.175	0.176		AW2	0.690	0.701	0.113	0.113
	AWW2	0.098	0.106	0.177	0.177		AWW2	**0.697**	0.709	0.113	0.113
0.3	BB2	0.296	0.306	0.162	0.162	0.9	BB2	0.884	0.896	0.066	0.068
	ML	0.250	0.264	0.158	0.165		ML	0.854	0.869	0.065	0.080
	KP2	0.244	0.257	0.159	0.169		KP2	0.855	0.866	0.072	0.085
	KPW2	0.263	0.273	0.159	0.163		KPW2	0.858	0.869	0.071	0.083
	LL2	0.248	0.263	**0.157**	0.166		LL2	0.844	0.860	0.080	0.098
	AW2	0.281	**0.295**	0.160	**0.161**		AW2	0.891	**0.904**	0.066	0.067
	AWW2	**0.297**	0.306	0.162	0.162		AWW2	**0.894**	0.907	**0.065**	**0.065**

Table 2.4: Hybridised estimators of ρ for $n = 100$, $\sigma^2 = 1$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB1	**-0.000**	0.002	0.113	0.113	0.5	BB1	0.496	**0.501**	0.085	**0.085**
	ML	-0.020	-0.016	0.114	0.115		ML	0.480	0.487	**0.084**	0.086
	KP1	-0.022	-0.018	0.114	0.116		KP1	0.478	0.482	0.086	0.088
	KPW1	-0.012	-0.009	**0.112**	**0.112**		KPW1	0.482	0.486	0.085	0.087
	LL1	-0.019	-0.016	0.113	0.115		LL1	0.480	0.485	0.084	0.086
	AW1	-0.008	-0.005	0.114	0.114		AW1	0.494	0.497	0.086	0.086
	AWW1	-0.000	**0.001**	0.113	0.113		AWW1	**0.497**	0.502	0.085	0.085
0.1	BB1	0.099	0.101	0.109	0.109	0.7	BB1	0.695	**0.700**	0.067	**0.067**
	ML	0.080	0.083	0.110	0.111		ML	0.680	0.689	0.065	0.068
	KP1	0.078	0.082	0.110	0.112		KP1	0.680	0.684	0.067	0.070
	KPW1	0.086	0.088	**0.108**	**0.109**		KPW1	0.682	0.687	0.067	0.069
	LL1	0.081	0.084	0.109	0.111		LL1	0.680	0.686	**0.065**	0.068
	AW1	0.092	0.095	0.110	0.110		AW1	0.696	0.700	0.068	0.068
	AWW1	**0.099**	**0.101**	0.109	0.109		AWW1	**0.698**	0.703	0.067	0.067
0.3	BB1	0.298	**0.301**	0.099	**0.099**	0.9	BB1	0.896	**0.900**	0.041	0.041
	ML	0.280	0.283	**0.098**	0.100		ML	0.881	0.884	**0.037**	0.041
	KP1	0.278	0.281	0.099	0.102		KP1	0.883	0.886	0.042	0.045
	KPW1	0.284	0.288	0.099	0.100		KPW1	0.884	0.888	0.041	0.044
	LL1	0.280	0.284	0.099	0.101		LL1	0.881	0.886	0.037	0.041
	AW1	0.292	0.296	0.100	0.100		AW1	**0.901**	0.904	0.042	0.042
	AWW1	**0.298**	0.302	0.099	0.099		AWW1	0.902	0.905	0.041	**0.041**

Table 2.5: Non-hybridised estimators of ρ for $n = 245$, $\sigma^2 = 1$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB2	**-0.000**	0.002	0.113	0.113	0.5	BB2	0.496	**0.501**	0.085	**0.085**
	ML	-0.020	-0.016	0.114	0.115		ML	0.480	0.487	**0.084**	0.086
	KP2	-0.022	-0.018	0.114	0.116		KP2	0.478	0.482	0.086	0.088
	KPW2	-0.012	-0.009	**0.112**	**0.112**		KPW2	0.482	0.486	0.085	0.087
	LL2	-0.019	-0.016	0.113	0.115		LL2	0.480	0.485	0.084	0.086
	AW2	-0.008	-0.005	0.114	0.114		AW2	0.494	0.497	0.086	0.086
	AWW2	-0.000	**0.001**	0.113	0.113		AWW2	**0.497**	0.502	0.085	0.085
0.1	BB2	0.099	0.101	0.109	0.109	0.7	BB2	0.695	**0.700**	0.067	**0.067**
	ML	0.080	0.083	0.110	0.111		ML	0.680	0.689	0.065	0.068
	KP2	0.078	0.082	0.110	0.112		KP2	0.680	0.684	0.067	0.070
	KPW2	0.086	0.088	**0.108**	**0.109**		KPW2	0.682	0.687	0.067	0.069
	LL2	0.081	0.084	0.109	0.111		LL2	0.680	0.686	**0.065**	0.068
	AW2	0.092	0.095	0.110	0.110		AW2	0.696	0.700	0.068	0.068
	AWW2	**0.099**	**0.101**	0.109	0.109		AWW2	**0.698**	0.703	0.067	0.067
0.3	BB2	0.298	**0.301**	0.099	**0.099**	0.9	BB2	0.895	**0.900**	0.040	0.040
	ML	0.280	0.283	**0.098**	0.100		ML	0.881	0.884	**0.037**	0.041
	KP2	0.278	0.281	0.099	0.102		KP2	0.883	0.886	0.042	0.045
	KPW2	0.284	0.288	0.099	0.100		KPW2	0.884	0.888	0.040	0.044
	LL2	0.280	0.284	0.099	0.101		LL2	0.881	0.886	0.037	0.041
	AW2	0.292	0.296	0.100	0.100		AW2	**0.900**	0.904	0.041	0.041
	AWW2	**0.298**	0.302	0.099	0.099		AWW2	0.901	0.905	0.040	**0.040**

Table 2.6: Hybridised estimators of ρ for $n = 245$, $\sigma^2 = 1$.

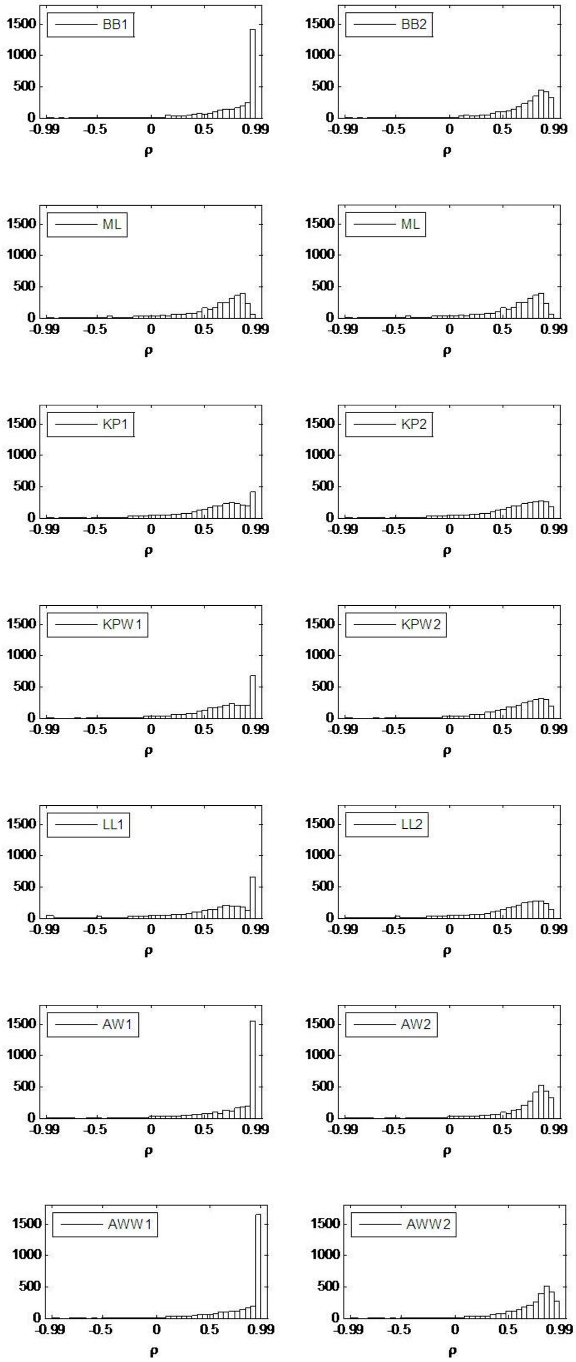


Figure 2.1: Histograms of non-hybridised and hybridised estimators of ρ for $n = 20$, $\sigma^2 = 1$ and the true value of $\rho = 0.9$.

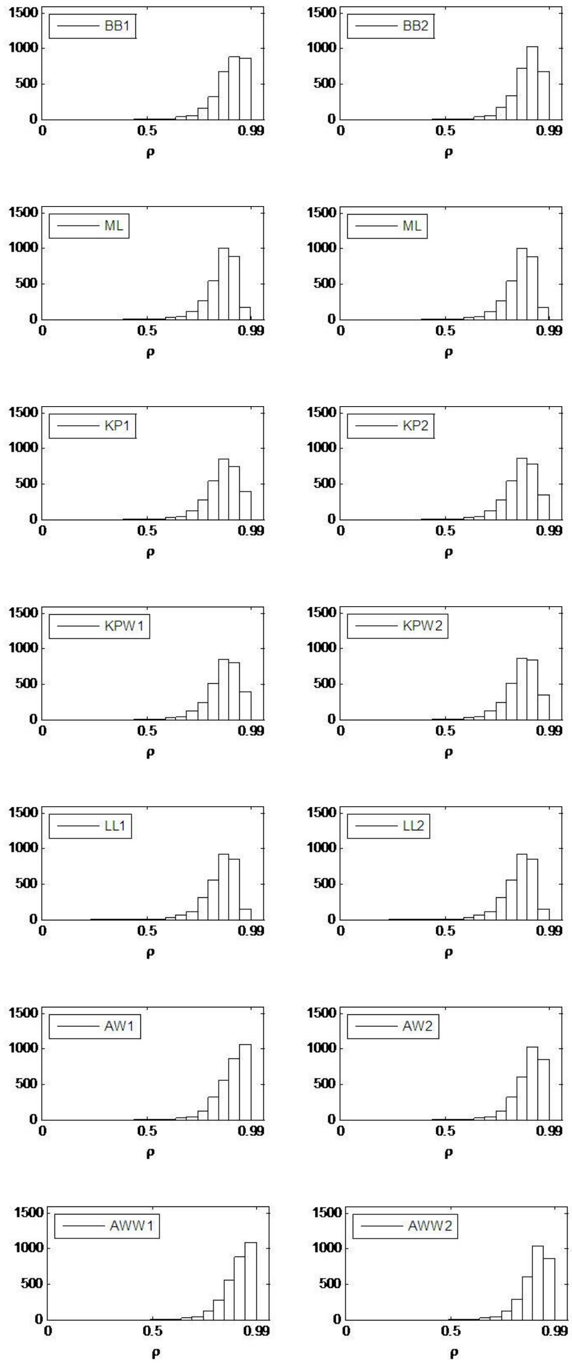


Figure 2.2: Histograms of non-hybridised and hybridised estimators of ρ for $n = 100$, $\sigma^2 = 1$ and the true value of $\rho = 0.9$.

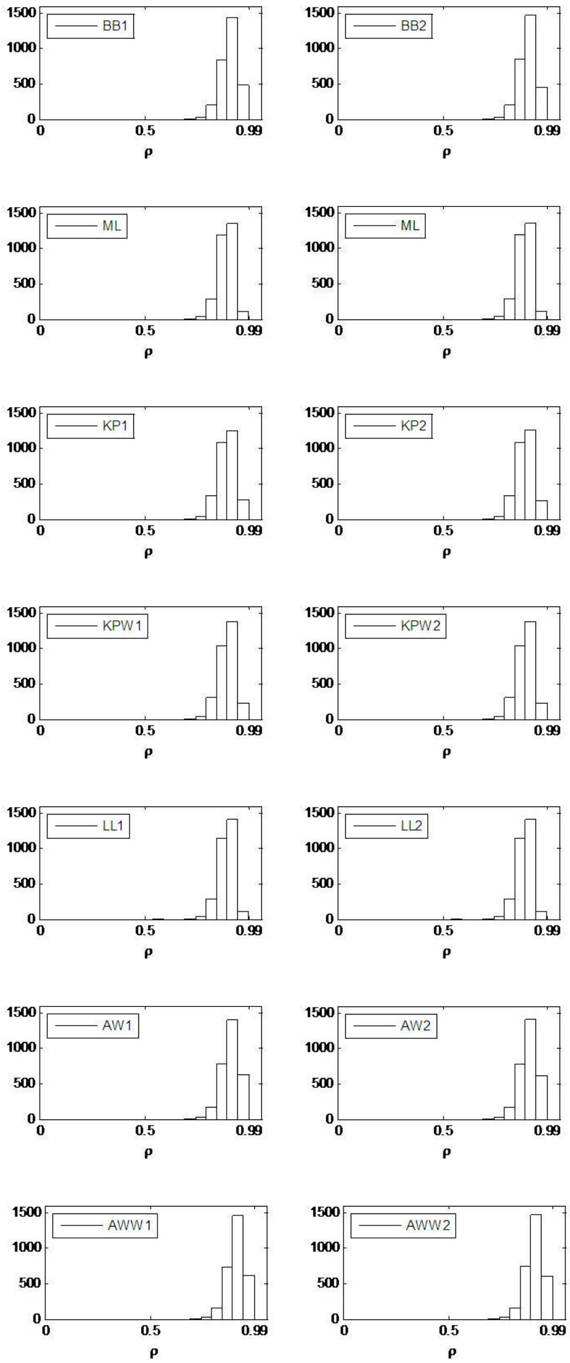


Figure 2.3: Histograms of non-hybridised and hybridised estimators of ρ for $n = 245$, $\sigma^2 = 1$ and the true value of $\rho = 0.9$.

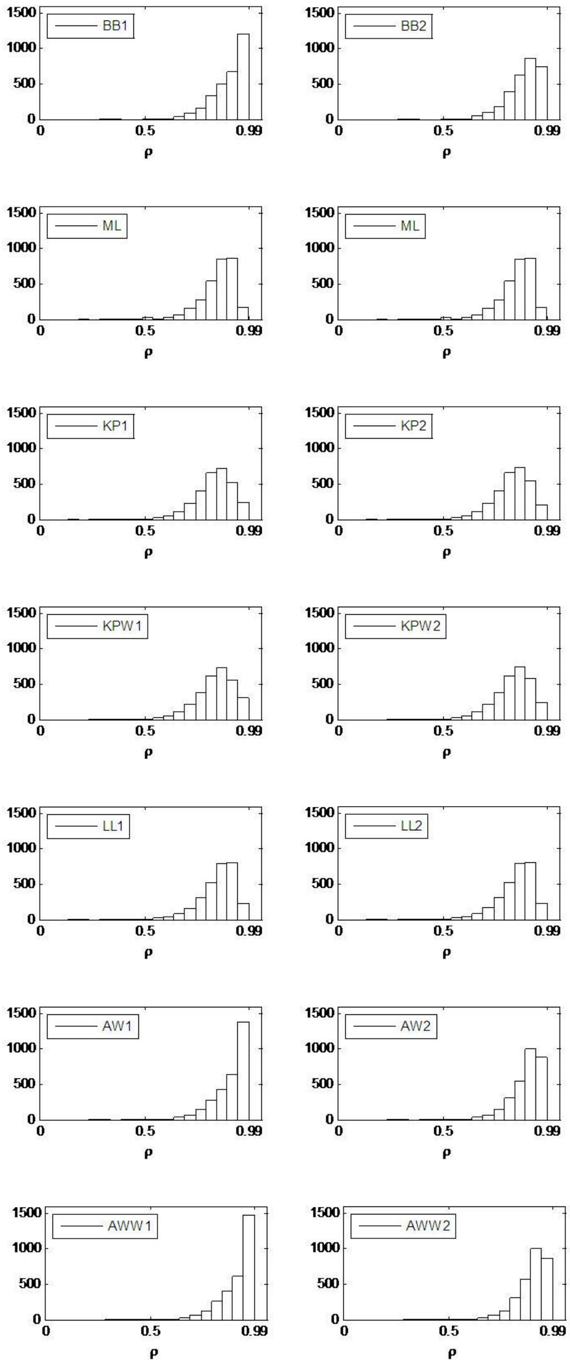


Figure 2.4: Histograms of non-hybridised and hybridised estimators of ρ for $n = 49$, $\sigma^2 = 1$ and the true value of $\rho = 0.9$.

For $n = 245$ a similar pattern emerges in Tables 2.5 and 2.6, with BB and AWW performing equally well in most cases. Though the various estimators' performance now much closer together, the BB, AW, and AWW are still ahead in their classes as shown in summary form in Tables 2.7 and 2.8. Notice that as non-hybridised estimators rarely produce non-invertible estimates for large n, results shown in Tables 2.5 and 2.6, and Figure 2.3 are more-or-less identical. Moreover, differences in estimators' performance for $n = 490$ are less significant, as shown in Tables 2.7 and 2.8, where only LL1 and LL2 are significantly better than the second-best when $\rho \geq 0.7$.

From the results for non-hybridised estimators, we can see that the MLE does not seem to perform well especially for small n. This situation may be explained as follows. As we mentioned earlier, the MLE and estimators that are not bias-adjusted for small samples are biased downward. The estimates obtained for these estimators are therefore much lower than those obtained for the bias-adjusted estimators such as the BB, AW and AWW, as well as lower than the true parameter values. Note that the differences between the estimates obtained for the MLE and bias-adjusted estimators are less striking for the hybridised cases, especially for small n and large true values of ρ.

	n									
True ρ	**20**		**50**		**100**		**245**		**490**	
0.0	KPW1*	(0.923)	KPW1*	(0.944)	KPW1*	(0.986)	KPW1*	(0.996)	KPW1	(0.998)
0.1	KPW1*	(0.947)	KPW1*	(0.956)	KPW1*	(0.993)	KPW1	(0.999)	BB1	(1.000)
0.3	KPW1	(0.999)	AW1*	(0.963)	AW1	(0.990)	BB1*	(1.000)	ML	(0.998)
0.5	BB1*	(1.000)	AW1*	(0.972)	AW1*	(0.983)	BB1*	(1.000)	ML	(0.993)
0.7	BB1*	(1.000)	AWW1	(0.998)	AW1*	(0.987)	BB1	(1.000)	LL1*	(0.990)
0.9	BB1*	(1.000)	AWW1*	(0.952)	AWW1*	(0.971)	AWW1	(0.987)	LL1*	(0.940)

Table 2.7: Estimation of ρ. For each ρ and n combination, the table entry is the non-hybridised estimator of ρ giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB1.

Tables 2.1 - 2.8 record results for a model in which σ^2, the variance of ε, is fixed at unity and in which the spatial weight matrices were created from randomly generated Voronoi polygons. To safeguard against unanticipated effects of using such random weight matrices and to reveal any sensitivity to variation in R^2 we conducted the experiment for $n = 49$ using the Columbus weights with $\sigma^2 = 0.25, 0.5, 1.0$, and 2.0.

Tables A.1 and A.2 in Appendix A show the mean, median, standard deviation and root mean square error of the non-hybridised and hybridised estimators, respectively, of ρ for $n = 49$ and $\sigma^2 = 1.0$. The results are reassuring, being little different from the corresponding tables for $n = 50$ obtained using the random weights. For the case $\rho = 0$ the R^2 values corresponding to the innovation variances are: $\frac{1}{1+\sigma^2} = 0.5, 0.8, 0.66$ and 0.33 respectively. Inspection of Tables A.3 - A.8 in Appendix A confirms the previous patterns where the bias-adjusted estimators take turn in having the smallest bias among all estimators in both the non-hybridised and hybridised forms.

The summaries, Tables 2.9 and 2.10, show that the bias-adjusted estimators are the most efficient estimators across different values of σ^2, except for when $\rho \leq 0.2$ where the KPW estimator is most efficient.

	n									
True ρ	**20**		**50**		**100**		**245**		**490**	
0.0	KPW2*	(0.964)	KPW2*	(0.974)	KPW2*	(0.986)	KPW2*	(0.996)	KPW2	(0.998)
0.1	KPW2	(0.990)	KPW2	(0.992)	KPW2*	(0.993)	KPW2	(0.999)	BB2	(1.000)
0.3	AWW2	(0.999)	BB2*	(1.000)	AW2	(0.995)	BB2*	(1.000)	ML	(0.998)
0.5	AWW2	(0.993)	BB2*	(1.000)	AW2	(0.999)	BB2*	(1.000)	ML	(0.993)
0.7	AWW2	(0.992)	AWW2	(0.995)	BB2*	(1.000)	BB2	(1.000)	LL2*	(0.990)
0.9	AWW2	(0.963)	AWW2*	(0.924)	AWW2*	(0.953)	AWW2	(0.990)	LL2*	(0.940)

Table 2.8: Estimation of ρ. For each ρ and n combination, the table entry is the hybridised estimator of ρ giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB2.

True ρ	σ^2 0.25		0.5		1		2	
0.0	KPW1*	(0.958)	KPW1*	(0.966)	KPW1*	(0.956)	KPW1*	(0.959)
0.1	KPW1*	(0.971)	KPW1*	(0.963)	KPW1*	(0.964)	KPW1*	(0.976)
0.3	AW1	(0.968)	AW1*	(0.964)	AW1	(0.984)	AWW1	(0.982)
0.5	AW1*	(0.956)	AW1*	(0.952)	AW1*	(0.961)	AW1*	(0.973)
0.7	AW1*	(0.962)	AW1*	(0.952)	AW1*	(0.968)	AW1*	(0.961)
0.9	AWW1	(0.992)	BB1*	(1.000)	BB1*	(1.000)	AWW1	(0.995)

Table 2.9: Estimation of ρ. For each ρ and σ^2 combination, the table entry is the non-hybridised estimator of ρ giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB1.

True ρ	σ^2 0.25		0.5		1		2	
0.0	KPW2*	(0.966)	KPW2*	(0.971)	KPW2*	(0.969)	KPW2*	(0.971)
0.1	KPW2*	(0.980)	KPW2*	(0.979)	KPW2*	(0.982)	KPW2*	(0.983)
0.3	AW2	(0.989)	AW2	(0.989)	AWW2	(0.995)	BB2	(1.000)
0.5	AW2*	(0.978)	AWW2	(0.981)	AWW2	(0.977)	AWW2	(0.993)
0.7	AWW2*	(0.975)	AWW2*	(0.966)	AWW2*	(0.976)	AWW2*	(0.984)
0.9	AWW2	(0.897)	AWW2	(0.899)	AWW2*	(0.911)	AWW2	(0.909)

Table 2.10: Estimation of ρ. For each ρ and σ^2 combination, the table entry is the hybridised estimator of ρ giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB2.

2.3.3 Estimates of β_1

The intercept is rarely a parameter of interest in spatial regression models, but of course good estimates are preferable to bad ones. Somewhat surprisingly, none of the bias-adjusted estimators does well on the mean square error criterion; as shown in Tables A.9 and A.10 in Appendix A where the ML and the LL estimators are most efficient in the majority of cases across all values of σ^2 considered. Note that the most efficient estimator is, however, not significantly better than the second-best estimator. For $n = 20$ and 50, as shown in Tables A.11 and A.12, MLE is again the most efficient estimator of β_1 in most cases. For $n = 100$, the AW and the AWW estimators are most efficient, even though not significantly better than the second-best, only when $\rho \leq 0.5$. For $n = 245$ and 490, the differences in estimators' performance are less significant, and the relative efficiency of the BB is at least 0.97.

The news is much better for the BB, AW, and AWW estimators when the focus is inference about β_1. The t statistics are calculated using the asymptotic standard errors described in Section 2.2; for the hybrid estimators, the standard error is calculated from either the MLE or the alternative method as appropriate. For $n = 20$, Table 2.11 reveals that rejection rates for the two-sided true null hypothesis, $\beta_1 = 1$ are much closer to nominal significance levels for the bias-adjusted estimators than for the others, although all estimators give badly over-sized t statistics when ρ is large. Note that the OLS estimator performs better than the other estimators when $\rho = 0.0$, as in that case the OLS assumption that U_i is independently distributed, is indeed correct. For $n = 100$ as seen in Table 2.12 the bias-adjusted estimators give more-or-less correctly sized two-sided t tests while the other estimators are clearly less satisfactory on this criterion. For $n = 490$, as seen in Table 2.13, they also give more-or-less correctly sized two-sided t tests, while the other estimators, even the OLS estimator in some cases when $\rho = 0.0$, are less satisfactory although with smaller differences.

		Rejection rates											
		Non-hybridised						Hybridised					
		β_1		β_2		β_3		β_1		β_2		β_3	
True ρ	**Method**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
0.0	BB	7.8	12.2	7.1	12.5	8.3	14.0	8.0	12.5	7.0	12.3	8.2	14.1
	ML	14.9	20.9	10.2	16.7	11.8	18.7	14.9	20.9	10.2	16.7	11.8	18.7
	KP	13.8	20.1	9.0	15.0	10.2	16.9	13.8	20.0	9.0	15.1	10.2	16.9
	KPW	11.5	17.0	8.6	14.5	10.3	16.4	11.5	17.0	8.6	14.5	10.2	16.2
	LL	15.2	21.6	9.3	15.5	11.0	17.5	15.1	21.6	9.5	16.0	11.1	17.6
	AW	9.3	13.4	7.2	12.1	8.1	14.2	9.4	13.6	7.4	12.3	8.2	14.2
	AWW	7.7	12.3	7.0	12.5	8.3	13.9	7.9	12.5	7.0	12.5	8.3	13.9
	OLS	4.4	9.0	4.7	9.0	5.5	10.6	4.4	9.0	4.7	9.0	5.5	10.6
0.5	BB	11.4	15.7	6.7	12.3	7.7	13.6	12.4	17.2	6.6	11.7	7.5	13.3
	ML	21.1	26.8	8.8	14.3	9.8	16.0	21.1	26.8	8.8	14.3	9.8	16.0
	KP	21.7	27.9	8.3	13.5	9.0	15.3	21.6	27.9	8.2	13.6	9.0	15.3
	KPW	17.7	23.4	7.9	13.1	8.9	15.0	18.0	23.9	7.5	12.8	8.6	14.8
	LL	22.8	29.1	8.5	13.5	9.6	15.5	22.7	29.1	8.2	13.3	9.3	15.2
	AW	12.6	17.0	6.7	11.6	6.9	12.9	12.8	17.3	6.5	11.5	6.9	12.8
	AWW	11.2	15.2	6.3	11.5	7.2	12.8	11.8	16.3	6.2	11.3	7.2	12.9
	OLS	25.4	34.4	4.6	9.0	5.3	10.7	25.4	34.4	4.6	9.0	5.3	10.7
0.9	BB	23.2	26.6	5.4	10.3	6.1	11.5	35.6	42.0	5.2	10.4	5.7	11.5
	ML	47.9	54.7	6.5	11.5	6.7	12.6	47.9	54.7	6.5	11.5	6.7	12.6
	KP	46.1	51.4	5.6	10.7	6.3	11.5	47.6	53.5	5.5	10.6	6.2	11.4
	KPW	38.8	43.8	5.4	10.3	6.2	11.5	44.6	50.8	5.3	10.3	6.0	11.5
	LL	44.9	49.6	6.6	11.6	7.0	12.5	49.7	55.7	6.4	11.2	6.6	12.2
	AW	21.8	25.0	5.0	9.1	5.4	10.0	33.0	39.6	5.4	10.1	6.0	10.9
	AWW	20.7	23.3	5.0	9.3	5.2	10.2	35.5	41.7	5.4	10.2	6.0	11.1
	OLS	78.0	81.4	4.7	9.3	4.9	10.2	78.0	81.4	4.7	9.3	4.9	10.2

Table 2.11: Size of non-hybridised and hybridised t-statistics for $n = 20$, $\sigma^2 = 1$. The table entry is the rejection percentage of the two-sided t test.

		Rejection rates											
		Non-hybridised						**Hybridised**					
		β_1		β_2		β_3		β_1		β_2		β_3	
True ρ	**Method**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
0.0	BB	5.8	11.4	4.6	10.2	5.3	10.4	5.8	11.4	4.6	10.2	5.3	10.4
	ML	7.6	12.9	5.2	10.6	5.9	10.9	7.6	12.9	5.2	10.6	5.9	10.9
	KP	7.8	12.9	5.2	10.6	6.0	10.8	7.8	12.9	5.2	10.6	6.0	10.8
	KPW	7.2	12.4	5.1	10.6	5.9	10.9	7.2	12.4	5.1	10.6	5.9	10.9
	LL	7.6	13.0	5.2	10.6	6.0	10.9	7.6	13.0	5.2	10.6	6.0	10.9
	AW	6.3	11.8	4.7	10.1	5.4	10.3	6.3	11.8	4.7	10.1	5.4	10.3
	AWW	6.0	11.4	4.6	10.2	5.3	10.4	6.0	11.4	4.6	10.2	5.3	10.4
	OLS	5.2	10.4	4.1	9.1	4.8	9.9	5.2	10.4	4.1	9.1	4.8	9.9
0.5	BB	6.7	12.0	4.5	9.9	4.8	10.1	6.7	12.0	4.5	9.9	4.8	10.0
	ML	8.9	14.4	4.8	10.3	4.9	10.3	8.9	14.4	4.8	10.3	4.9	10.3
	KP	9.1	14.5	4.7	9.9	4.9	10.3	9.1	14.5	4.7	9.9	4.9	10.3
	KPW	8.6	13.9	4.6	10.1	5.0	10.3	8.6	13.9	4.6	10.1	5.0	10.3
	LL	9.1	14.6	4.8	10.3	4.9	10.3	9.1	14.6	4.8	10.3	4.9	10.3
	AW	7.0	12.1	4.5	9.8	4.5	9.8	7.0	12.1	4.5	9.8	4.5	9.8
	AWW	6.5	11.6	4.4	9.9	4.7	9.9	6.5	11.6	4.4	9.9	4.7	9.9
	OLS	28.4	36.8	4.8	9.6	5.0	9.9	28.4	36.8	4.8	9.6	5.0	9.9
0.9	BB	11.3	15.3	4.7	9.9	5.0	9.0	11.7	16.0	4.5	9.5	4.8	8.8
	ML	17.1	23.7	4.4	9.5	4.4	9.0	17.1	23.7	4.4	9.5	4.4	9.0
	KP	16.7	22.9	3.9	8.6	4.4	8.3	16.8	23.1	3.9	8.6	4.3	8.3
	KPW	16.4	22.1	4.0	8.6	4.4	8.4	16.5	22.3	4.0	8.6	4.4	8.4
	LL	18.9	26.1	4.3	9.3	4.4	8.8	18.9	26.1	4.3	9.3	4.4	8.8
	AW	10.0	13.8	4.3	9.3	4.5	8.6	10.3	14.3	4.2	9.3	4.5	8.6
	AWW	9.4	12.6	4.4	9.5	4.5	8.7	9.8	13.3	4.3	9.5	4.5	8.7
	OLS	72.7	76.9	4.9	10.1	5.0	9.9	72.7	76.9	4.9	10.1	5.0	9.9

Table 2.12: Size of non-hybridised and hybridised t-statistics for $n = 100$, $\sigma^2 = 1$. The table entry is the rejection percentage of the two-sided t test.

		Rejection rates											
		Non-hybridised						Hybridised					
		β_1		β_2		β_3		β_1		β_2		β_3	
True ρ	**Method**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**	**5%**	**10%**
0.0	BB	5.1	10.7	5.1	9.7	5.5	11.3	5.1	10.7	5.1	9.7	5.5	11.3
	ML	5.5	11.4	5.2	9.8	5.6	11.5	5.5	11.4	5.2	9.8	5.6	11.5
	KP	5.5	11.4	5.2	9.8	5.6	11.4	5.5	11.4	5.2	9.8	5.6	11.4
	KPW	5.4	11.3	5.2	9.8	5.5	11.4	5.4	11.3	5.2	9.8	5.5	11.4
	LL	5.4	11.5	5.2	9.8	5.6	11.5	5.4	11.5	5.2	9.8	5.6	11.5
	AW	5.3	11.1	5.2	9.8	5.6	11.3	5.3	11.1	5.2	9.8	5.6	11.3
	AWW	5.1	10.6	5.1	9.7	5.5	11.3	5.1	10.6	5.1	9.7	5.5	11.3
	OLS	4.9	10.8	5.1	9.6	5.6	11.2	4.9	10.8	5.1	9.6	5.6	11.2
0.5	BB	5.3	11.0	5.2	9.8	5.5	11.3	5.3	11.0	5.2	9.8	5.5	11.3
	ML	5.7	11.4	5.3	9.9	5.5	11.5	5.7	11.4	5.3	9.9	5.5	11.5
	KP	5.7	11.7	5.3	10.0	5.5	11.3	5.7	11.7	5.3	10.0	5.5	11.3
	KPW	5.5	11.7	5.3	9.9	5.5	11.3	5.5	11.7	5.3	9.9	5.5	11.3
	LL	5.7	11.5	5.3	10.0	5.5	11.3	5.7	11.5	5.3	10.0	5.5	11.3
	AW	5.4	11.0	5.2	9.8	5.4	11.3	5.4	11.0	5.2	9.8	5.4	11.3
	AWW	5.3	10.8	5.2	9.8	5.5	11.3	5.3	10.8	5.2	9.8	5.5	11.3
	OLS	28.9	37.7	4.8	9.4	5.3	10.9	28.9	37.7	4.8	9.4	5.3	10.9
0.9	BB	6.7	11.4	5.6	9.9	5.4	11.0	6.7	11.4	5.6	9.9	5.4	11.0
	ML	8.1	13.7	5.5	9.8	5.3	11.0	8.1	13.7	5.5	9.8	5.3	11.0
	KP	8.0	13.3	5.5	9.7	5.3	10.6	8.0	13.3	5.5	9.7	5.3	10.6
	KPW	7.9	12.9	5.4	9.7	5.3	10.6	7.9	12.9	5.4	9.7	5.3	10.6
	LL	8.1	13.2	5.5	9.8	5.3	11.0	8.1	13.2	5.5	9.8	5.3	11.0
	AW	6.3	10.6	5.5	9.9	5.3	10.9	6.3	10.6	5.5	9.9	5.3	10.9
	AWW	6.2	10.5	5.5	9.9	5.4	11.0	6.2	10.5	5.5	9.9	5.4	11.0
	OLS	72.1	76.8	5.4	9.6	4.9	10.5	72.1	76.8	5.4	9.6	4.9	10.5

Table 2.13: Size of non-hybridised and hybridised t-statistics for $n = 490$, $\sigma^2 = 1$. The table entry is the rejection percentage of the two-sided t test.

2.3.4 Estimates of β_2 and β_3

Consider β_2 first; the true value is 0.0 so the associated explanatory variable is redundant. The empirical significance levels for t tests reported in Table 2.11 suggests there is little to choose between the estimators, most resulting in slightly liberal inferences, while in Tables 2.12 - 2.13, for $n = 100$ and $n =$

490 respectively, we see that all the estimators are approximately correctly sized. For β_3 the pattern is similar, with BB, AW, and AWW essentially indistinguishable from their competitors.

The summary tables 2.14 and 2.15 show that BB1 and BB2 perform very well, especially for $n = 20$ and 50 where they are (significantly) the most efficient estimators in several cases. For larger n, estimators' performance is much closer together as expected and the most efficient estimator is not significantly better than the second-best. Notice that the relative efficiency of BB1 and BB2 is at least 0.99 for $n = 20$ and 50, and 1.0 for $n \geq 100$. Tables A.13 and A.14 in Appendix A show that for $n = 49$, in most cases, the KPW performs the best for small ρ values, the MLE and the AWW for large ρ values, and the BB when ρ values are between the two extremes. Note that the differences between the best and the second-best methods disappear with the relative efficiency of BB very close to the best estimators.

	n									
True ρ	**20**		**50**		**100**		**245**		**490**	
0.0	KPW1	(0.990)	KPW1	(0.997)	KPW1	(1.000)	AW1	(1.000)	KPW1	(1.000)
0.1	KPW1	(0.993)	KPW1	(0.998)	AW1	(1.000)	AW1	(1.000)	KPW1	(1.000)
0.3	KPW1	(1.000)	BB1	(1.000)	AW1	(1.000)	AW1	(1.000)	KPW1	(1.000)
0.5	BB1*	(1.000)	BB1	(1.000)	AW1	(1.000)	AW1	(1.000)	LL1	(1.000)
0.7	BB1*	(1.000)	AWW1	(1.000)	ML	(1.000)	AWW1	(1.000)	ML	(1.000)
0.9	BB1*	(1.000)	AWW1	(0.997)	LL1	(1.000)	AWW1	(1.000)	ML	(1.000)

Table 2.14: Estimation of β_2. For each ρ and n combination, the table entry is the non-hybridised estimator of β_2 giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB1.

For β_3 as shown in Tables 2.16 and 2.17, BB1/2 are significantly most efficient in several cases for $n = 20$. For $n \geq 50$, the bias-adjusted estimators still outperform other estimators for the majority of cases, though not significantly better than the second-best. Tables A.15 and A.16 in Appendix A confirm the similarity of pattern of β_3 with that of β_2 where the bias-adjusted estimators

True ρ	n									
	20		50		100		245		490	
0.0	KPW2	(0.998)	KPW2	(0.999)	KPW2	(1.000)	AW2	(1.000)	KPW2	(1.000)
0.1	KPW2	(0.999)	KPW2	(1.000)	AW2	(1.000)	AW2	(1.000)	KPW2	(1.000)
0.3	BB2	(1.000)	BB2	(1.000)	AW2	(1.000)	AW2	(1.000)	KPW2	(1.000)
0.5	BB2*	(1.000)	BB2*	(1.000)	AW2	(1.000)	AW2	(1.000)	LL2	(1.000)
0.7	BB2*	(1.000)	AWW2	(1.000)	AWW2	(1.000)	AWW2	(1.000)	ML	(1.000)
0.9	AW2*	(0.990)	AWW2*	(0.996)	AWW2	(1.000)	AWW2	(1.000)	ML	(1.000)

Table 2.15: Estimation of β_2. For each ρ and n combination, the table entry is the hybridised estimator of β_2 giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB2.

take it in turn to perform the best in most cases, except for small ρ values.

True ρ	n									
	20		50		100		245		490	
0.0	KPW1	(0.993)	KPW1	(0.997)	BB1	(1.000)	BB1	(1.000)	KPW1	(1.000)
0.1	KPW1	(0.997)	KPW1	(0.998)	BB1	(1.000)	BB1	(1.000)	KPW1	(1.000)
0.3	BB1	(1.000)	KPW1	(0.999)	AW1	(1.000)	BB1	(1.000)	KPW1	(1.000)
0.5	BB1*	(1.000)	BB1	(1.000)	AW1	(1.000)	AW1	(1.000)	ML	(1.000)
0.7	AWW1	(0.998)	BB1	(1.000)	AW1	(0.999)	AW1	(1.000)	AW1	(1.000)
0.9	AWW1	(0.993)	AWW1	(0.999)	AWW1	(0.999)	ML	(1.000)	AW1	(1.000)

Table 2.16: Estimation of β_3. For each ρ and n combination, the table entry is the non-hybridised estimator of β_3 giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB1.

2.3.5 Estimates of σ^2

Like β_1 in this model, the innovation variance is seldom a key parameter. However, the relative performance of the various estimators is still of some interest. Tables 2.18 and 2.19 below show that AW and AWW are most efficient for small ρ values, KP and KPW for moderate ρ, and MLE for $\rho \geq 0.7$. Note that for the larger sample sizes the LL estimator makes an appearance. The

True ρ	n 20		50		100		245		490	
0.0	KPW2*	(0.996)	KPW2	(0.998)	BB2	(1.000)	BB2	(1.000)	KPW2	(1.000)
0.1	KPW2	(1.000)	KPW2	(1.000)	BB2	(1.000)	BB2	(1.000)	KPW2	(1.000)
0.3	BB2*	(1.000)	BB2	(1.000)	AW2	(1.000)	BB2	(1.000)	KPW2	(1.000)
0.5	BB2*	(1.000)	BB2	(1.000)	AW2	(1.000)	AW2	(1.000)	ML	(1.000)
0.7	AWW2	(0.995)	AWW2	(0.999)	AWW2	(0.999)	AW2	(1.000)	AW2	(1.000)
0.9	AW2*	(0.988)	AW2	(0.998)	AWW2	(0.999)	ML	(1.000)	AW2	(1.000)

Table 2.17: Estimation of β_3. For each ρ and n combination, the table entry is the hybridised estimator of β_3 giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB2.

relative efficiencies suggest that except perhaps at the largest ρ values and for small n, there is little to choose between the estimators.

Tables A.17 and A.18 in Appendix A repeat similar pattern for the Columbus weight matrix with $n = 49$ and various σ^2. Detailed results are available on request.

True ρ	n 20		50		100		245		490	
0.0	AW1*	(0.987)	AW1*	(0.995)	AW1	(0.999)	AWW1*	(1.000)	AWW1	(1.000)
0.1	AW1*	(0.988)	AW1*	(0.995)	AW1	(0.999)	AWW1	(1.000)	AWW1	(1.000)
0.3	AW1*	(0.991)	AW1*	(0.989)	AW1	(0.998)	AWW1	(1.000)	AWW1	(1.000)
0.5	AW1	(0.990)	KP1	(0.976)	KP1	(0.990)	KPW1	(0.999)	ML	(0.999)
0.7	ML	(0.971)	ML	(0.948)	ML	(0.969)	ML	(0.990)	LL1	(0.994)
0.9	ML*	(0.877)	ML*	(0.871)	ML*	(0.859)	ML	(0.915)	LL1	(0.951)

Table 2.18: Estimation of σ^2. For each ρ and n combination, the table entry is the non-hybridised estimator of σ^2 giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB1.

True ρ	n									
	20		50		100		245		490	
0.0	AW2	(0.997)	AW2	(0.997)	AW2	(0.999)	AWW2*	(1.000)	AWW2	(1.000)
0.1	AW2	(0.996)	AW2	(0.997)	AW2	(0.999)	AWW2	(1.000)	AWW2	(1.000)
0.3	BB2	(1.000)	AW2*	(0.994)	AW2	(0.999)	AWW2	(1.000)	AWW2	(1.000)
0.5	KPW2	(0.994)	KP2	(0.988)	KP2	(0.993)	KPW2	(0.999)	ML	(0.999)
0.7	KPW2	(0.981)	KPW2	(0.966)	ML	(0.974)	ML	(0.990)	LL2	(0.994)
0.9	ML*	(0.911)	ML*	(0.897)	ML*	(0.898)	ML	(0.932)	LL2	(0.951)

Table 2.19: Estimation of σ^2. For each ρ and n combination, the table entry is the hybridised estimator of σ^2 giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB2.

2.4 Conclusion

Small sample performance of six existing estimators and the bias-adjusted estimator BB introduced in this chapter are compared in the spatial error model framework. The existing estimators considered are the Maximum Gaussian Likelihood [ML], the method of Kelejian and Prucha (1999) [KP], the Kelejian and Prucha method with weighting matrix (Kelejian and Prucha, 2009) [KPW], the Lee and Liu (2010) method [LL], the small-sample adjustment to the KP method introduced by Arnold and Wied (2010a) [AW], and the Arnold and Wied method with weighting matrix included [AWW]. We show that the BB estimator is robust and its performance does not depend on a particular spatial weighting matrix M. An optimal weighting matrix W should also be incorporated in the method of moments procedure to improve the efficiency of the estimators. Furthermore, the bias-adjusted estimators; the BB, AW, and AWW, perform extremely well in reducing the small-sample bias, being virtually mean and median unbiased. Nevertheless, all estimators except the MLE produce a significant proportion of non-invertible estimates.

This motivates us to develop the hybrid estimator for the spatial autocorrelation parameter to improve the small-sample efficiency. This method combines the original non-hybridised estimator with the maximum likelihood estimator when the former delivers non-invertible parameter estimates. The hybridised forms of the BB, AW and AWW estimators are clearly superior to other estimators in small samples and, in our experiments, the use of the hybrid estimator in the first step of a feasible GLS estimator leads to inferences about the regression coefficients in the second stage that are at least as robust as those of competing estimators.

Chapter 3

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

3.1 Introduction

In this chapter 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 (1988a) and Anselin and Bera (1998). Ord (1975) also presents a computational scheme extended to the MR-SAR models. Asymptotic properties of the MLE and QMLE are developed by Lee (2004a) for the spatial autoregressive models with fixed sequences of weights. Our approach relies heavily on the 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 chapter. 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.

This chapter is constructed as follows. Section 3.2 describes the mixed regressive spatial autoregressive model and introduces a sub-model for spatial weights. Assumptions are listed in Section 3.3. Section 3.6 provides figures of the shape of the concentrated log-likelihood. Section 3.4 analyses the identifiability of the parameters and the consistency of the QML estimator. The asymptotic normality of the QMLE is derived in Section 3.5. Section 3.7 explains how the Monte Carlo experiment is conducted and presents the corresponding results. Section 3.8 concludes. Detailed proofs can be found in Appendix B.

3.2 Mixed Regressive, Spatial Autoregressive Model

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

$$Y_n = X_n\beta + \lambda W_n(\gamma)Y_n + \varepsilon_n \tag{3.2.1}$$

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, which represent the degree of possible interaction of location j on location i (Ord, 1975). The elements are specified as

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

with

$$w^*_{n,ij}(\gamma) = \begin{cases} 0 & \text{for } i = j \\ f(\gamma, d_{ij}) & \text{for } i \neq j \end{cases} \tag{3.2.2}$$

where $f(\gamma, d_{ij})$ is a function of distances, d_{ij} is a fixed nonnegative distance between spatial units i and j, γ is a positive scalar parameter, 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 (3.2.1) 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 $(W_n(\gamma)Y_n)_i$ may be 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 (3.2.1) becomes

$$\begin{aligned} S_n(\lambda, \gamma)Y_n &= X_n\beta + \varepsilon_n \\ Y_n &= S_n^{-1}(\lambda, \gamma)(X_n\beta + \varepsilon_n) \end{aligned} \tag{3.2.3}$$

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 (3.2.1) is given by

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \ln|det(S_n(\lambda, \gamma))| - \frac{1}{2\sigma^2}\varepsilon_n'(\delta)\varepsilon_n(\delta) \tag{3.2.4}$$

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 (3.2.4) 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

$$\hat{\beta}_n(\lambda, \gamma) = (X_n'X_n)^{-1}(X_n'Y_n - \lambda X_n'W_n(\gamma)Y_n) = (X_n'X_n)^{-1}X_n'S_n(\lambda, \gamma)Y_n \tag{3.2.5}$$

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

$$\begin{aligned} \hat{\sigma}_n^2(\lambda, \gamma) &= \frac{1}{n}[(S_n(\lambda, \gamma)Y_n - X_n\hat{\beta}(\lambda, \gamma))'(S_n(\lambda, \gamma)Y_n - X_n\hat{\beta}(\lambda, \gamma)] \\ &= \frac{1}{n}[Y_n'S_n'(\lambda, \gamma)M_nS_n(\lambda, \gamma)Y_n] \end{aligned} \tag{3.2.6}$$

where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. Insert (3.2.5) and (3.2.6) back into the log-likelihood function and obtain the following concentrated log-likelihood function of λ and γ.

$$\ln L_n(\lambda, \gamma) = -\frac{n}{2}(\ln(2\pi) + 1) + \ln|det(S_n(\lambda, \gamma))| - \frac{n}{2}\ln\hat{\sigma}_n^2(\lambda, \gamma). \tag{3.2.7}$$

To obtain the QMLEs $\hat{\lambda}_n$ and $\hat{\gamma}_n$, maximise (3.2.7) 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 $\hat{\sigma}_n^2(\hat{\lambda}_n, \hat{\gamma}_n)$.

3.3 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 concave. The true values of λ and γ denoted by λ_0 and γ_0 respectively, are in the interior of Θ.*

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

Assumption 4. *The 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 5. *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 is the rate whose sequence, $\{h_n\}$, is nonrandom and can be bounded or divergent. There exists an open neighbourhood $\eta_n(\gamma_0)$ of γ_0 such that $w_{n,ij}(\gamma)$ 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 6. *Ratio $\frac{h_n}{n} \to 0$ as $n \to \infty$, where n is the total number of spatial units.*

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 (3.2.5) and (3.2.6), 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 4 and 5 provide the characteristics of the spatial weight matrix and the functional form of its elements. Assumption 6 rules out the case of $\sum_j w_{n,ij}^*$ diverging to infinity at a rate equal to or faster than the rate of sample size n.

Assumption 7 is sufficient to ensure that S_n is nonsingular such that (3.2.1) 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 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

[1]See Horn and Johnson (1985)

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). See Appendix B.2 for more detail of Lemmas used in this chapter.

3.4 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 (3.2.3) can be rewritten as

$$Y_n = (I_n + \lambda_0 G_n)(X_n\beta_0 + \varepsilon_n) = X_n\beta_0 + \lambda_0 G_n X_n \beta_0 + S_n^{-1}\varepsilon_n \qquad (3.4.1)$$

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 3.3, 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.

We make the following additional assumption.

Assumption 10. *The following limits exist and are nonsingular.*
$\lim_{n\to\infty}\dfrac{1}{n}(X_n, G_nX_n\beta_0)'(X_n, G_nX_n\beta_0)$, $\lim_{n\to\infty}\dfrac{1}{n}(X_n, T_nX_n\beta_0)'(X_n, T_nX_n\beta_0)$ *and*
$\lim_{n\to\infty}\dfrac{1}{n}(X_n, G_nX_n\beta_0)'(X_n, T_nX_n\beta_0)$
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 (3.4.1) 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

$$\beta_n^*(\lambda,\gamma) = (X_n'X_n)^{-1}X_n'S_n(\lambda,\gamma)S_n^{-1}X_n\beta_0 \tag{3.4.2}$$

and

$$\sigma_n^{2*}(\lambda,\gamma) = \frac{1}{n}[(\lambda_0-\lambda)^2(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)+\sigma_0^2tr(S_n^{-1\prime}S_n'(\lambda,\gamma)S_n(\lambda,\gamma)S_n^{-1})] \tag{3.4.3}$$

Substitute (3.4.2) and (3.4.3) into the log-likelihood, then we get

$$Q_n(\lambda,\gamma) = -\frac{n}{2}(\ln(2\pi)+1) + \ln|det(S_n(\lambda,\gamma))| - \frac{n}{2}\ln\sigma_n^{2*}(\lambda,\gamma) \tag{3.4.4}$$

and it is concave and continuous in $\theta \in \Theta$. We establish our theorems below. See Appendix B 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. *Under Assumptions 1 - 10, $\hat{\theta}_n$ is a consistent estimator of θ_0.*

3.5 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.

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \beta} = \frac{1}{\sigma_0^2 \sqrt{n}} X_n' \varepsilon_n \tag{3.5.1}$$

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \lambda} = \frac{1}{\sigma_0^2 \sqrt{n}} [(G_n X_n \beta_0)' \varepsilon_n + \varepsilon_n' G_n \varepsilon_n - \sigma_0^2 tr(G_n)] \tag{3.5.2}$$

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \gamma} = \frac{\lambda_0}{\sigma_0^2 \sqrt{n}} [(T_n X_n \beta_0)' \varepsilon_n + \varepsilon_n' T_n \varepsilon_n - \sigma_0^2 tr(T_n)] \tag{3.5.3}$$

$$\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \sigma^2} = \frac{1}{2\sigma_0^4 \sqrt{n}} (\varepsilon_n' \varepsilon_n - n\sigma_0^2) \tag{3.5.4}$$

where G_n and T_n are as defined in Section 3.4.

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). See Appendix B.2 for more detail of Lemmas used in this chapter.

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 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

$$\Omega_{\theta,n} = E(\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta} \cdot \frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta'}) + E(\frac{1}{n} \frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'}). \tag{3.5.5}$$

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

$$-E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) = \tag{3.5.6}$$

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

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.

$$\Omega_{\theta,n} = \tag{3.5.7}$$

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

The matrix $\Omega_{\theta,n}$ above is symmetric 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.,

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} \xrightarrow{D} N[0, -E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) + \Omega_{\theta,n}]. \tag{3.5.8}$$

and, consequently,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_{\theta,n}\Sigma_\theta^{-1}] \tag{3.5.9}$$

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,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1}]. \tag{3.5.10}$$

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

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

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1}] \tag{3.5.11}$$

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

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N[0, \Sigma_\theta^{-1}]. \tag{3.5.12}$$

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 (3.5.6) and (3.5.7) 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 $P_n = G_n X_n \beta_0$ and $R_n = T_n X_n \beta_0$, 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.

3.6 Shape of the Concentrated Log-Likelihood

This section shows three-dimensional shape of the concentrated log-likelihood evaluated at several values of (λ, γ) coordinates for different numbers of observations. So far we have not specified a functional form of the sub-model for the spatial weights and, instead, have kept it general. To illustrate a shape of the concentrated log-likelihood, we now specify a functional form of the sub-model for the weights below. Recall first that the elements of the row-standardised weight matrix $W_n(\gamma)$ are

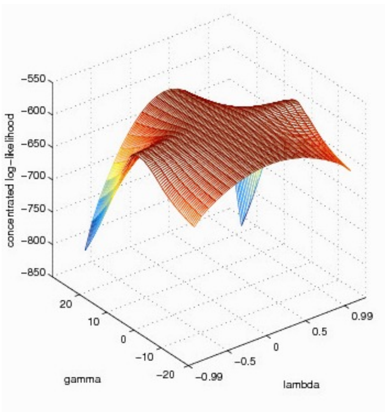
$$w_{n,ij}(\gamma) = \frac{w^*_{n,ij}(\gamma)}{\Sigma_j w^*_{n,ij}(\gamma)} \tag{3.6.1}$$

where $w^*_{n,ij}(\gamma) = 0$ for $i = j$ and $w^*_{n,ij}(\gamma) = f(\gamma, d_{ij})$ for $i \neq j$ with γ a positive scalar parameter specifying the weights and d_{ij} a fixed nonnegative distance between spatial units i and j. Then if $f(\gamma, d_{ij}) = e^{-\gamma d_{ij}}$ the elements $w_{n,ij}(\gamma)$ of the weight matrix $W_n(\gamma)$ become

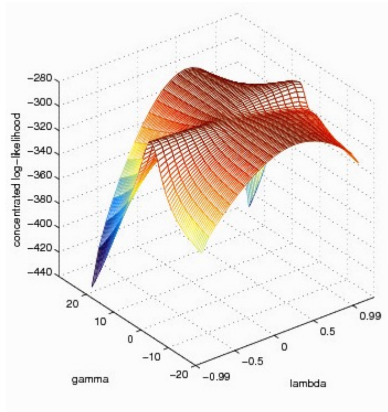
$$w_{n,ij}(\gamma) = \begin{cases} 0 & \text{if } i = j \\ \frac{e^{-\gamma d_{ij}}}{\sum_j e^{-\gamma d_{ij}}} \geq 0 & \text{if } i \neq j \end{cases} \tag{3.6.2}$$

where $\sum_j e^{-\gamma d_{ij}}$ is a row sum for all i. 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. Then we draw an independent standard Normal vector of disturbances, of which the variance σ_n^2 is fixed at 1.0, and 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 is the vector of ones and the other 2 columns are draws from the standard $n-$variate Normal distribution. The row-standardised weight matrix is created based on equation (3.6.2) with true values of $\gamma = 0$, 0.5, 1, 2, 5, and 10, and the matrix $S_n(\lambda, \gamma)$ is generated using this weight matrix and true value of $\lambda = 0$ and 0.5.

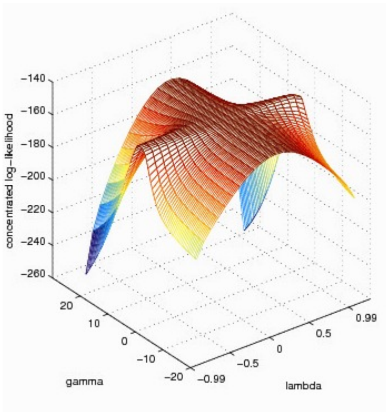
Figures 3.1 - 3.6 below show three-dimensional shape of the concentrated log-likelihood evaluated at λ ranging from -0.99 to 0.99 and γ from -20 to 20 for n = 400, 200, 100, and 50. We can see that the shape of the concentrated log-likelihood is concave and continuous, even though it is generally flat, especially when (λ, γ) coordinates are around zero or around the true value of γ. This feature may be due to the chosen functional form of the sub-model for the weights and how the distances d_{ij} are generated. On the other hand, the shape of the concentrated log-likelihood becomes much steeper towards the extreme values of γ and λ. Differences between the maximum and minimum values of the likelihood in each graph vary considerably, with larger differences for large n and smaller differences for small n. Peaks are generally at or close to the DGP values of γ and λ.



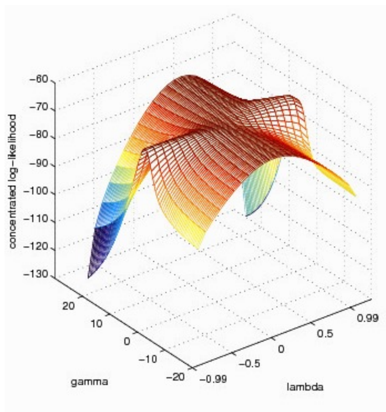
n = 400



n = 200

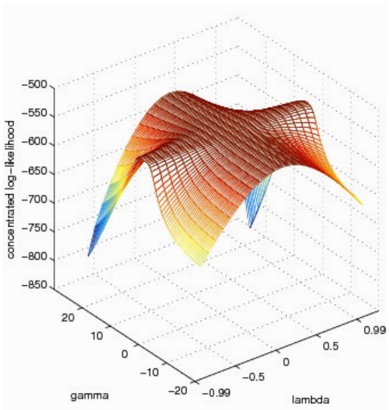


n = 100

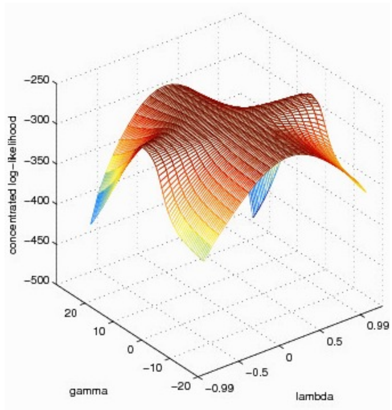


n = 50

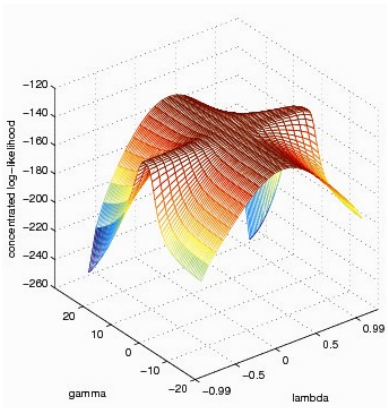
Figure 3.1: Shape of the concentrated log-likelihood based on DGP $\gamma = 2$ and $\lambda = 0.5$, evaluated at λ ranging from -0.99 to 0.99 and γ from -20 to 20 for n = 400, 200, 100 and 50, and $\sigma^2 = 1$.



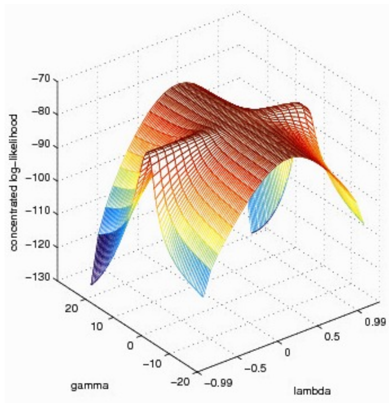
n = 400



n = 200

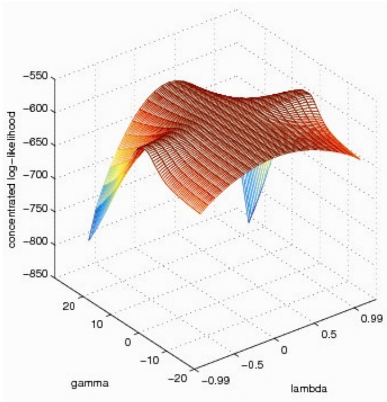


n = 100

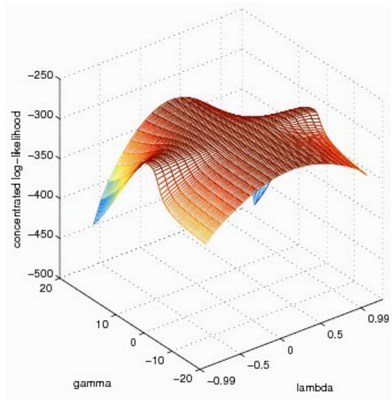


n = 50

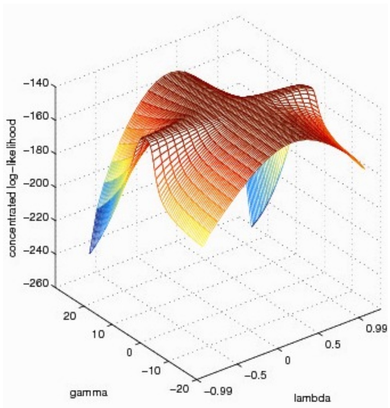
Figure 3.2: Shape of the concentrated log-likelihood based on DGP $\gamma = 0$ and $\lambda = 0$, evaluated at λ ranging from -0.99 to 0.99 and γ from -20 to 20 for n = 400, 200, 100 and 50, and $\sigma^2 = 1$.



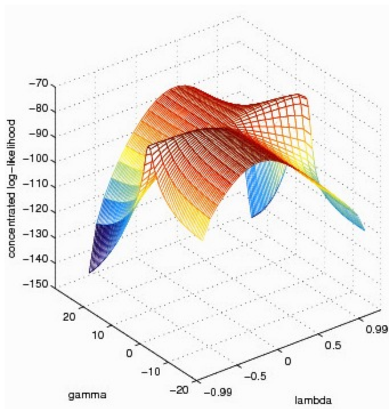
n = 400



n = 200

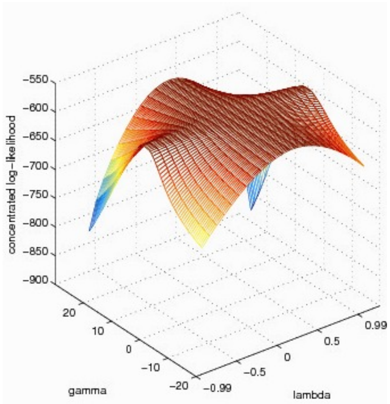


n = 100

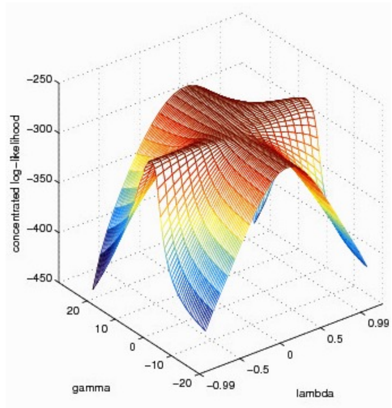


n = 50

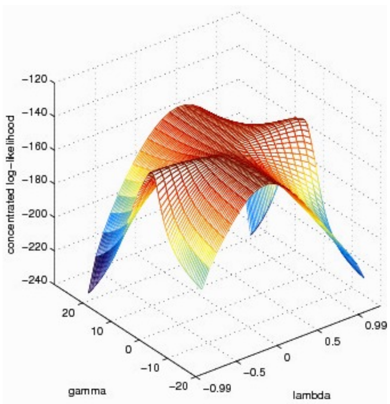
Figure 3.3: Shape of the concentrated log-likelihood based on DGP $\gamma = 0.5$ and $\lambda = 0.5$, evaluated at λ ranging from -0.99 to 0.99 and γ from -20 to 20 for n = 400, 200, 100 and 50, and $\sigma^2 = 1$.



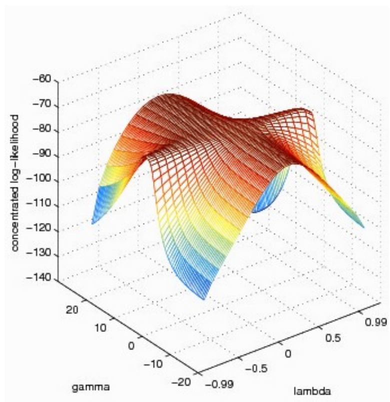
n = 400



n = 200

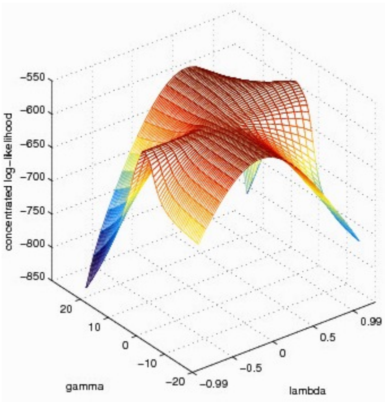


n = 100

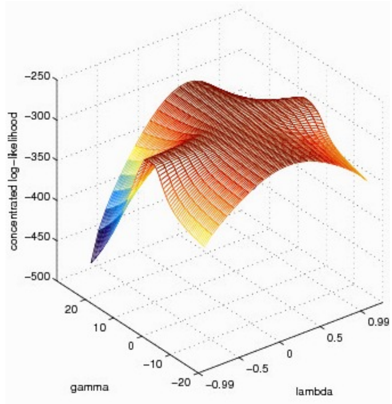


n = 50

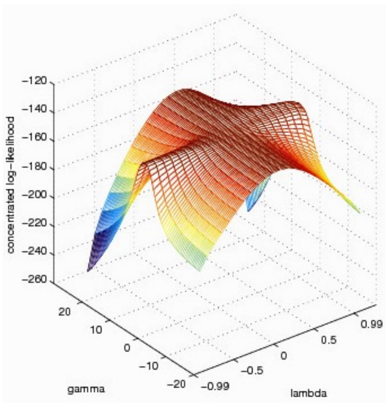
Figure 3.4: Shape of the concentrated log-likelihood based on DGP $\gamma = 1$ and $\lambda = 0.5$, evaluated at λ ranging from -0.99 to 0.99 and γ from -20 to 20 for n = 400, 200, 100 and 50, and $\sigma^2 = 1$.



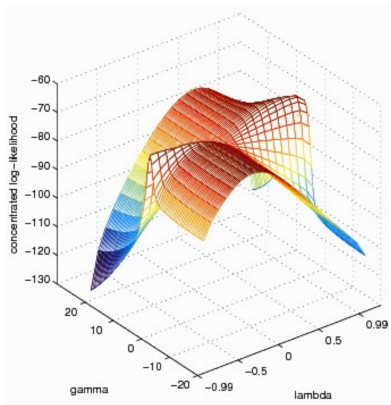
n = 400



n = 200

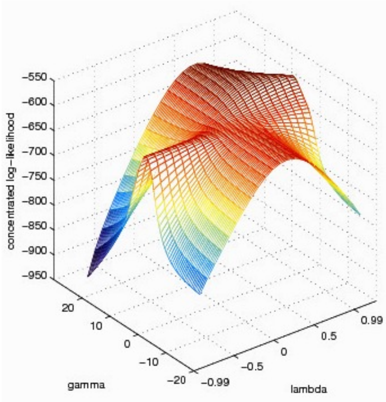


n = 100

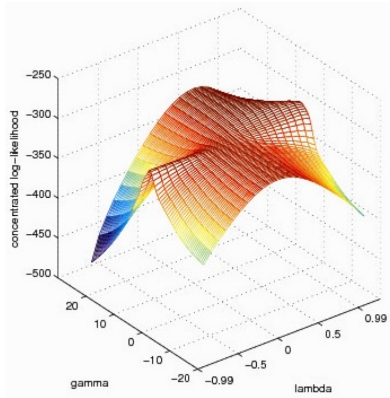


n = 50

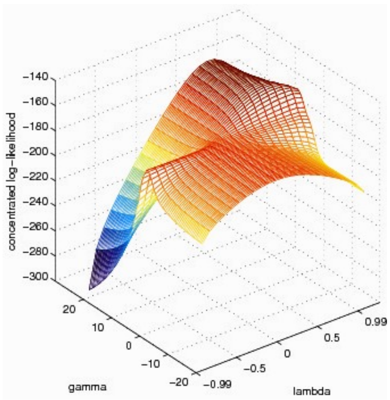
Figure 3.5: Shape of the concentrated log-likelihood based on DGP $\gamma = 5$ and $\lambda = 0.5$, evaluated at λ ranging from -0.99 to 0.99 and γ from -20 to 20 for n = 400, 200, 100 and 50, and $\sigma^2 = 1$.



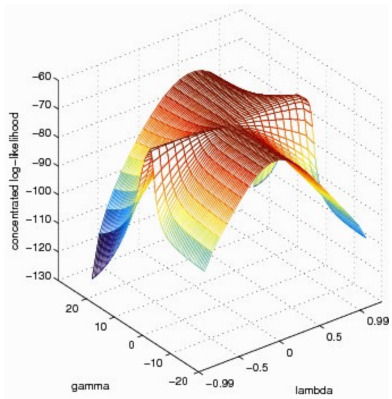
n = 400



n = 200



n = 100



n = 50

Figure 3.6: Shape of the concentrated log-likelihood based on DGP $\gamma = 10$ and $\lambda = 0.5$, evaluated at λ ranging from -0.99 to 0.99 and γ from -20 to 20 for n = 400, 200, 100 and 50, and $\sigma^2 = 1$.

3.7 Monte Carlo Results

3.7.1 Experiment Design

We investigate small sample properties of our estimator using the freely estimated weight matrix as in (3.6.2) 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.

We perform experiments for n = 200, 400 and 800 for 1000 replications. The row-standardised weight matrix is created following equation (3.6.2), with associated γ values = 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 is the 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 1.0. The matrix $S_n(\lambda, \gamma)$ is generated for each W, 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 $|\hat{\lambda}| \leq 0.99$, to ensure that the matrix $S_n(\lambda, \gamma)$ is nonsingular and to speed up the simulation, and on γ estimates to be $\hat{\gamma} \geq 0.01$. The simulation results are reported in the following subsections.

3.7.2 Estimates of λ

Tables 3.1 - 3.3 show the mean, median, standard deviation and root mean squares error of estimates obtained from different estimators of λ for n = 200, 400, and 800. 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. Looking first at Table 3.1, there are 2 panels of results. The results on the left panel of the table 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 = 200, 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 (3.6.2) which freely estimates the parameter γ that defines the weight matrix. This weight matrix is denoted by W($\hat{\gamma}$) in the tables 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 (3.6.2) 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 with competing estimators using the weight matrices W(3), W(5) and Wrand.

Table 3.1 below 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 have the largest bias in most cases. Moreover, these estimates suggest that the randomly generated weight matrix does not seem to be able to properly capture the dependence between spatial units. All estimates obtained in the last row are negative and close to zero regardless of the true values of λ. For small λ, this situation may seem to suggest that the estimator based on the random weight matrix is better than other methods. However, this could be merely a coincidence for small λ since all estimates obtained for this estimator are negative and close to zero.

For $n = 400$ and 800, similar patterns emerge in Tables 3.2 and 3.3 respectively, with our QML estimator performing better than other competing estimators in most cases across the true values of λ. The QML estimator using the randomly generated weight matrix, again, has the largest bias in the mean and median in most cases. Even though the other competing QML estimators using the wrong weight matrices have smaller bias than the estimator in the last row, they produce larger bias in the mean and median than our QML estimator in most cases. These results clearly 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)	-0.067	-0.065	0.518	0.544	W(7)	0.078	0.092	0.304	0.305
	W($\hat{\gamma}$)	-0.244	-0.155	0.648	0.733	W($\hat{\gamma}$)	0.015	0.107	0.425	0.433
	W(3)	-0.376	-0.663	0.702	0.848	W(3)	-0.293	-0.481	0.712	0.813
	W(7)	0.009	0.006	0.297	0.310	W(5)	0.033	0.064	0.511	0.515
	Wrand	-0.028	-0.027	0.142	0.191	Wrand	-0.027	-0.025	0.147	0.195
0.3	W(5)	0.186	0.213	0.505	0.518	W(7)	0.351	0.376	0.298	0.302
	W($\hat{\gamma}$)	0.042	0.216	0.624	0.675	W($\hat{\gamma}$)	0.332	0.330	0.364	0.365
	W(3)	-0.049	-0.092	0.753	0.830	W(3)	0.243	0.424	0.741	0.742
	W(7)	0.148	0.157	0.295	0.332	W(5)	0.467	0.549	0.469	0.497
	Wrand	-0.022	-0.019	0.143	0.352	Wrand	-0.033	-0.025	0.151	0.366
0.5	W(5)	0.454	0.522	0.467	0.469	W(7)	0.605	0.624	0.275	0.294
	W($\hat{\gamma}$)	0.323	0.414	0.558	0.585	W($\hat{\gamma}$)	0.557	0.543	0.302	0.307
	W(3)	0.299	0.541	0.737	0.763	W(3)	0.660	0.990	0.573	0.595
	W(7)	0.315	0.327	0.292	0.346	W(5)	0.773	0.990	0.337	0.434
	Wrand	-0.021	-0.018	0.145	0.541	Wrand	-0.031	-0.029	0.151	0.552
0.7	W(5)	0.639	0.774	0.400	0.404	W(7)	0.814	0.868	0.197	0.228
	W($\hat{\gamma}$)	0.525	0.596	0.467	0.499	W($\hat{\gamma}$)	0.739	0.752	0.223	0.226
	W(3)	0.549	0.990	0.637	0.654	W(3)	0.915	0.990	0.266	0.342
	W(7)	0.443	0.451	0.285	0.384	W(5)	0.942	0.990	0.145	0.282
	Wrand	-0.021	-0.012	0.152	0.737	Wrand	-0.015	-0.010	0.153	0.731
0.9	W(5)	0.800	0.990	0.307	0.323	W(7)	0.940	0.990	0.118	0.124
	W($\hat{\gamma}$)	0.699	0.806	0.369	0.420	W($\hat{\gamma}$)	0.872	0.984	0.167	0.169
	W(3)	0.778	0.990	0.467	0.483	W(3)	0.971	0.990	0.151	0.167
	W(7)	0.576	0.589	0.270	0.422	W(5)	0.979	0.990	0.077	0.111
	Wrand	-0.019	-0.015	0.143	0.930	Wrand	-0.018	-0.006	0.155	0.931

Table 3.1: Estimation of λ for $n = 200$, $\sigma^2 = 1$, 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.

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 3.2: 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.

True λ	W	Mean	Med.	St.D.	RMSE	W	Mean	Med.	St.D.	RMSE
0.1	W(5)	-0.087	-0.085	0.524	0.556	W(7)	0.071	0.081	0.268	0.269
	W($\hat{\gamma}$)	-0.256	-0.176	0.645	0.737	W($\hat{\gamma}$)	0.003	0.092	0.414	0.426
	W(3)	-0.399	-0.723	0.690	0.852	W(3)	-0.328	-0.562	0.705	0.825
	W(7)	0.002	0.008	0.272	0.289	W(5)	0.015	0.018	0.511	0.518
	Wrand	-0.002	-0.001	0.072	0.125	Wrand	-0.005	-0.001	0.071	0.127
0.3	W(5)	0.187	0.206	0.518	0.530	W(7)	0.348	0.347	0.271	0.275
	W($\hat{\gamma}$)	0.048	0.206	0.630	0.678	W($\hat{\gamma}$)	0.324	0.296	0.358	0.358
	W(3)	-0.069	-0.129	0.767	0.850	W(3)	0.246	0.453	0.743	0.745
	W(7)	0.143	0.146	0.276	0.317	W(5)	0.487	0.571	0.461	0.497
	Wrand	-0.004	-0.003	0.073	0.312	Wrand	-0.007	-0.006	0.072	0.315
0.5	W(5)	0.431	0.475	0.472	0.476	W(7)	0.610	0.619	0.244	0.267
	W($\hat{\gamma}$)	0.304	0.377	0.553	0.587	W($\hat{\gamma}$)	0.571	0.511	0.291	0.300
	W(3)	0.264	0.441	0.731	0.768	W(3)	0.685	0.990	0.527	0.559
	W(7)	0.276	0.279	0.275	0.355	W(5)	0.814	0.990	0.283	0.423
	Wrand	-0.009	-0.010	0.076	0.514	Wrand	-0.004	-0.003	0.076	0.510
0.7	W(5)	0.644	0.794	0.399	0.402	W(7)	0.829	0.890	0.187	0.228
	W($\hat{\gamma}$)	0.531	0.590	0.466	0.496	W($\hat{\gamma}$)	0.746	0.784	0.242	0.246
	W(3)	0.548	0.990	0.627	0.645	W(3)	0.902	0.990	0.299	0.361
	W(7)	0.411	0.418	0.270	0.396	W(5)	0.951	0.990	0.134	0.284
	Wrand	-0.005	-0.001	0.072	0.709	Wrand	-0.010	-0.006	0.075	0.714
0.9	W(5)	0.799	0.990	0.297	0.314	W(7)	0.950	0.990	0.097	0.109
	W($\hat{\gamma}$)	0.696	0.778	0.342	0.398	W($\hat{\gamma}$)	0.872	0.990	0.173	0.176
	W(3)	0.762	0.990	0.456	0.477	W(3)	0.978	0.990	0.093	0.121
	W(7)	0.542	0.543	0.260	0.442	W(5)	0.986	0.990	0.033	0.092
	Wrand	-0.005	-0.006	0.073	0.908	Wrand	-0.007	-0.005	0.074	0.910

Table 3.3: Estimation of λ for $n = 800$, $\sigma^2 = 1$, 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.

3.7.3 Estimates of γ

Another parameter of interest is γ which defines the spatial weights according to equation (3.6.2). In Tables 3.4 - 3.6 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 values of $\gamma = 3$, 5, and 7.

For a small true value of γ; $\gamma = 3$, Table 3.4 shows 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
3	200	0.1	3.807	4.135	0.875	1.189
		0.3	3.815	4.091	0.920	1.229
		0.5	3.768	3.912	0.892	1.177
		0.7	3.745	3.878	0.912	1.177
		0.9	3.786	3.807	0.813	1.131
	400	0.1	3.865	4.219	0.818	1.190
		0.3	3.803	4.043	0.890	1.199
		0.5	3.758	3.900	0.887	1.166
		0.7	3.653	3.748	0.906	1.117
		0.9	3.798	3.867	0.802	1.131
	800	0.1	3.812	4.181	0.899	1.211
		0.3	3.809	4.034	0.898	1.208
		0.5	3.761	3.847	0.870	1.156
		0.7	3.662	3.760	0.911	1.126
		0.9	3.721	3.703	0.791	1.070

Table 3.4: Estimation of γ for the true value of $\gamma = 3$, $n = 200$, 400, 800, $\sigma^2 = 1$, and $\lambda \in (0.1, 0.3, 0.5, 0.7, 0.9)$.

Looking at results in Table 3.5 obtained for the true value of γ equals 5, the same pattern appears with the standard deviation and RMSE of the estimates for each n clearly reduce when λ becomes larger. Looking at the results across different n, we see that the bias of the mean and median as well as the errors of the estimates associated with each λ become smaller when n increases.

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

For results corresponding to a larger true value of γ in Table 3.6, we see a similar pattern repeated for the true value of $\gamma = 7$. Note that the performance of the QML estimator in this case strongly improves compared to those obtained for smaller true values of γ we report earlier. Here there is a clear reduction in bias of the mean and median and smaller standard deviation and RMSE.

True γ	n	λ	Mean	Med.	St.D.	RMSE
7	200	0.1	7.221	7.230	0.400	0.456
		0.3	7.084	7.115	0.442	0.450
		0.5	6.976	7.019	0.475	0.475
		0.7	6.860	6.816	0.444	0.465
		0.9	6.942	6.951	0.321	0.326
	400	0.1	7.177	7.194	0.341	0.384
		0.3	7.104	7.107	0.383	0.397
		0.5	7.002	7.033	0.448	0.447
		0.7	6.874	6.864	0.422	0.440
		0.9	6.904	6.944	0.287	0.303
	800	0.1	7.165	7.163	0.329	0.368
		0.3	7.053	7.076	0.369	0.372
		0.5	6.973	7.028	0.414	0.415
		0.7	6.824	6.772	0.428	0.463
		0.9	6.889	6.950	0.288	0.308

Table 3.6: Estimation of γ for the true value of $\gamma = 7$, $n = 200,\ 400,\ 800$, $\sigma^2 = 1$, and $\lambda \in (0.1, 0.3, 0.5, 0.7, 0.9)$.

3.7.4 Estimates of β

In this subsection we report the mean, median, standard deviation and the RMSE of the benchmark estimator, our QML estimator, and three competing estimators for estimating β's for $n = 200$, 400, and 800, $\sigma^2 = 1$, and $\lambda = 0.5$. The true values of β's are $\beta_1 = 1.0$, $\beta_2 = 0.0$, and $\beta_3 = -1.0$, and the true values of γ used in this experiment are 5 and 7, respectively. Note that the results shown below are rounded to the nearest third decimal. All other results in our Monte Carlo experiment not reported here are available on request.

True λ	n	True β	W	Mean	Med.	St.D.	RMSE
0.5	200	$\beta_1 = 1$	W(5)	1.089	0.954	0.941	0.945
			W($\hat{\gamma}$)	1.350	1.144	1.136	1.188
			W(3)	1.403	0.925	1.489	1.542
			W(7)	1.365	1.335	0.597	0.700
			Wrand	2.039	2.019	0.337	1.093
		$\beta_2 = 0$	W(5)	0.002	-0.000	0.070	0.070
			W($\hat{\gamma}$)	0.002	0.000	0.070	0.070
			W(3)	0.002	0.000	0.070	0.070
			W(7)	0.002	0.001	0.070	0.070
			Wrand	0.002	-0.000	0.071	0.071
		$\beta_3 = -1$	W(5)	-0.997	-0.997	0.071	0.071
			W($\hat{\gamma}$)	-0.997	-0.998	0.072	0.072
			W(3)	-0.998	-0.998	0.072	0.072
			W(7)	-0.996	-0.998	0.071	0.071
			Wrand	-0.997	-0.998	0.073	0.073

Table 3.7: Estimation of β_1, β_2, and β_3 for the true value of $\gamma = 5$, $n = 200$, $\sigma^2 = 1$, $\lambda = 0.5$, and the true values of $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$.

Tables 3.7 - 3.9 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, reported on the last row for the true value of β_1. The standard deviation and the RMSE are small, except for the estimates for β_1, the intercept. 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 larger value of γ, $\gamma = 7$ here, Tables 3.10 - 3.12 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.

True λ	n	True β	W	Mean	Med.	St.D.	RMSE
0.5	400	$\beta_1 = 1$	W(5)	1.139	1.031	0.923	0.933
			W($\hat{\gamma}$)	1.385	1.201	1.089	1.155
			W(3)	1.475	1.144	1.443	1.518
			W(7)	1.433	1.417	0.560	0.707
			Wrand	2.015	1.998	0.242	1.043
		$\beta_2 = 0$	W(5)	0.001	0.002	0.050	0.050
			W($\hat{\gamma}$)	0.001	0.003	0.050	0.050
			W(3)	0.001	0.002	0.050	0.050
			W(7)	0.001	0.003	0.050	0.050
			Wrand	0.001	0.001	0.050	0.050
		$\beta_3 = -1$	W(5)	-0.999	-0.999	0.049	0.049
			W($\hat{\gamma}$)	-0.999	-0.999	0.049	0.049
			W(3)	-1.000	-1.000	0.049	0.049
			W(7)	-0.999	-0.998	0.049	0.049
			Wrand	-0.999	-0.999	0.049	0.049

Table 3.8: Estimation of β_1, β_2, and β_3 for the true value of $\gamma = 5$, $n = 400$, $\sigma^2 = 1$, $\lambda = 0.5$, and the true values of $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$.

True λ	n	True β	W	Mean	Med.	St.D.	RMSE
0.5	800	$\beta_1 = 1$	W(5)	1.136	1.038	0.944	0.954
			W($\hat{\gamma}$)	1.393	1.232	1.114	1.181
			W(3)	1.472	1.132	1.466	1.540
			W(7)	1.446	1.419	0.552	0.710
			Wrand	2.015	2.004	0.176	1.031
		$\beta_2 = 0$	W(5)	0.001	-0.000	0.036	0.036
			W($\hat{\gamma}$)	0.001	-0.000	0.036	0.036
			W(3)	0.001	0.000	0.036	0.036
			W(7)	0.001	-0.000	0.036	0.036
			Wrand	0.001	0.000	0.036	0.036
		$\beta_3 = -1$	W(5)	-0.999	-0.998	0.036	0.036
			W($\hat{\gamma}$)	-0.998	-0.998	0.036	0.036
			W(3)	-0.999	-0.998	0.036	0.036
			W(7)	-0.998	-0.998	0.036	0.036
			Wrand	-0.998	-0.998	0.036	0.036

Table 3.9: Estimation of β_1, β_2, and β_3 for the true value of $\gamma = 5$, $n = 800$, $\sigma^2 = 1$, $\lambda = 0.5$, and the true values of $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$.

True λ	n	True β	W	Mean	Med.	St.D.	RMSE
0.5	200	$\beta_1 = 1$	W(7)	0.792	0.753	0.558	0.595
			W($\hat{\gamma}$)	0.889	0.910	0.613	0.622
			W(3)	0.678	0.075	1.149	1.193
			W(5)	0.453	0.114	0.680	0.873
			Wrand	2.069	2.042	0.353	1.126
		$\beta_2 = 0$	W(7)	0.000	-0.000	0.073	0.073
			W($\hat{\gamma}$)	0.000	-0.000	0.073	0.073
			W(3)	0.002	0.000	0.070	0.070
			W(5)	0.000	0.001	0.073	0.073
			Wrand	-0.000	-0.001	0.074	0.073
		$\beta_3 = -1$	W(7)	-0.996	-0.993	0.072	0.072
			W($\hat{\gamma}$)	-0.996	-0.995	0.071	0.071
			W(3)	-1.004	-1.002	0.072	0.072
			W(5)	-1.000	-0.998	0.071	0.071
			Wrand	-1.001	-1.000	0.073	0.073

Table 3.10: Estimation of β_1, β_2, and β_3 for the true value of $\gamma = 7$, $n = 200$, $\sigma^2 = 1$, $\lambda = 0.5$, and the true values of $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$.

True λ	n	True β	W	Mean	Med.	St.D.	RMSE
0.5	400	$\beta_1 = 1$	W(7)	0.763	0.744	0.503	0.556
			W($\hat{\gamma}$)	0.848	0.910	0.563	0.583
			W(3)	0.559	0.053	1.030	1.120
			W(5)	0.355	0.077	0.576	0.865
			Wrand	2.022	2.010	0.236	1.049
		$\beta_2 = 0$	W(7)	-0.001	-0.001	0.051	0.051
			W($\hat{\gamma}$)	-0.001	-0.001	0.051	0.051
			W(3)	-0.001	-0.001	0.051	0.051
			W(5)	-0.001	-0.001	0.051	0.051
			Wrand	-0.001	-0.001	0.051	0.051
		$\beta_3 = -1$	W(7)	-1.000	-1.001	0.051	0.051
			W($\hat{\gamma}$)	-1.000	-1.000	0.051	0.051
			W(3)	-1.004	-1.005	0.051	0.052
			W(5)	-1.002	-1.003	0.051	0.051
			Wrand	-1.003	-1.003	0.052	0.052

Table 3.11: Estimation of β_1, β_2, and β_3 for the true value of $\gamma = 7$, $n = 400$, $\sigma^2 = 1$, $\lambda = 0.5$, and the true values of $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$.

True λ	n	True β	W	Mean	Med.	St.D.	RMSE
0.5	800	$\beta_1 = 1$	W(7)	0.779	0.756	0.489	0.536
			W($\hat{\gamma}$)	0.855	0.977	0.583	0.601
			W(3)	0.626	0.048	1.051	1.115
			W(5)	0.372	0.062	0.566	0.845
			Wrand	2.004	2.001	0.176	1.019
		$\beta_2 = 0$	W(7)	0.001	0.000	0.035	0.035
			W($\hat{\gamma}$)	-0.001	-0.001	0.051	0.051
			W(3)	0.001	0.001	0.035	0.035
			W(5)	0.001	0.000	0.035	0.035
			Wrand	0.001	-0.000	0.035	0.035
		$\beta_3 = -1$	W(7)	-0.998	-0.999	0.035	0.035
			W($\hat{\gamma}$)	-0.998	-0.999	0.035	0.035
			W(3)	-1.000	-1.001	0.035	0.035
			W(5)	-1.000	-1.001	0.035	0.035
			Wrand	-1.000	-1.000	0.035	0.035

Table 3.12: Estimation of β_1, β_2, and β_3 for the true value of $\gamma = 7$, $n = 800$, $\sigma^2 = 1$, $\lambda = 0.5$, and the true values of $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = -1$.

3.7.5 Estimates of σ^2

As σ^2 is not a key parameter and all estimators perform equally well in estimating σ^2 across all values of λ and n, regardless of the weight matrices used in the experiment, we only report the results for a selection of cases we have carried out in this chapter. Tables 3.13 and 3.14 below show the mean, median, standard deviation and the RMSE of the benchmark estimator, our QML estimator, and three competing estimators for estimating σ^2 for $n = 200$, 400, and 800, $\sigma^2 = 1$, and $\lambda = 0.5$. The true values of γ used are 5 and 7, respectively. Note that the results shown below are rounded to the nearest third decimal.

From the results we see that 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.

True σ^2	True λ	n	W	Mean	Med.	St.D.	RMSE
1.0	0.5	200	W(5)	0.979	0.978	0.100	0.102
			W($\hat{\gamma}$)	0.981	0.980	0.100	0.101
			W(3)	0.985	0.986	0.100	0.101
			W(7)	0.977	0.977	0.100	0.102
			Wrand	0.985	0.985	0.100	0.101
		400	W(5)	0.992	0.991	0.068	0.069
			W($\hat{\gamma}$)	0.993	0.993	0.069	0.069
			W(3)	0.995	0.995	0.069	0.069
			W(7)	0.991	0.990	0.069	0.069
			Wrand	0.995	0.995	0.069	0.070
		800	W(5)	0.994	0.992	0.051	0.051
			W($\hat{\gamma}$)	0.995	0.993	0.051	0.051
			W(3)	0.995	0.994	0.051	0.051
			W(7)	0.994	0.992	0.051	0.051
			Wrand	0.995	0.994	0.051	0.051

Table 3.13: Estimation of σ^2 for the true value of $\gamma = 5$, $n = 200$, 400, 800, $\sigma^2 = 1$, and $\lambda = 0.5$.

True σ^2	True λ	n	W	Mean	Med.	St.D.	RMSE
1.0	0.5	200	W(7)	0.982	0.976	0.098	0.100
			W($\hat{\gamma}$)	0.985	0.979	0.098	0.100
			W(3)	1.009	1.004	0.102	0.102
			W(5)	0.991	0.985	0.099	0.100
			Wrand	1.013	1.006	0.104	0.105
		400	W(7)	0.988	0.985	0.068	0.069
			W($\hat{\gamma}$)	0.990	0.988	0.069	0.069
			W(3)	1.003	1.002	0.070	0.070
			W(5)	0.993	0.991	0.069	0.069
			Wrand	1.006	1.005	0.070	0.071
		800	W(7)	0.994	0.993	0.051	0.051
			W($\hat{\gamma}$)	0.995	0.993	0.051	0.051
			W(3)	1.002	1.001	0.052	0.052
			W(5)	0.997	0.996	0.051	0.051
			Wrand	1.003	1.002	0.052	0.052

Table 3.14: Estimation of σ^2 for the true value of $\gamma = 7$, $n = 200, 400, 800$, $\sigma^2 = 1$, and $\lambda = 0.5$.

3.8 Conclusion

In this chapter 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 chapter. 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, λ.

Chapter 4

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

4.1 Introduction

Spatial econometrics has been used in applied work in many fields including the field of economic growth as it is able to account for spillovers between spatial units. Some of the applied studies related to economic growth are summarised below.

Abreu, De Groot and Florax (2005) present a survey of the empirical literature on growth and space. They differentiate models into models of absolute and relative location and concentrate their survey on regression techniques applied to growth processes, and suggest that models in spatial econometrics and results should relate more closely to theory. An overview of the literature on regional economic growth and convergence is given by Bode and Rey (2006). They particularly discuss papers that involve open-economy models, the role of space in convergence dynamics, innovative framework towards regional interactions, and new spatial tool-kits for applied work on regional growth. Fingleton (2004) also provides a survey of the literature on growth and introduces an iterative approach for the stochastic equilibrium. He uses a spatial econometric model to study the productivity growth variations and computes the steady-states and stochastic equilibrium for the manufacturing productivity ratios of the EU regions.

Henry, Schmitt and Piguet (2001) compare several spatial econometric models of small region growth applied to data on French rural community to investigate the determinants of population and employment change in the rural areas. They test for the impacts of urban growth on rural communities and find robust evidence of dispersion of population from neighbouring communities. Lundberg (2006) examines the determinants of average income growth and net migration in Swedish municipalities and tests the hypothesis if growth and net migration rates of one municipality depend on growth rates of nearby municipalities. He finds spatial spillovers of net migration as well as spatial dependence in the error terms for the average income growth rates. Ying (2003) investigates China's growth from a spatial econometric perspective and presents new insights of the Chinese economy. The author analyses the determinants of growth and takes into account spatial effects to provide a better understanding of the spatial process underlying the Chinese economy.

Ertur and Koch (2007) extend the Solow model to include technological interdependence among countries and investigate the impact of spillover effects. Dall'erba and Le Gallo (2008) also use a neoclassical growth model to study the impact of structural funds on the convergence process among the European regions while taking into account the presence of spillover effects and possible risk of endogeneity of the funds.

Next, we discuss some studies related to regional income and foreign direct investment (FDI). Rey and Montouri (1999) use exploratory spatial data analysis and spatial econometric methods to study the US's regional income convergence patterns. They take into account spatial effects and geographical aspects of the income growth, and find strong spatial autocorrelation in the regional income convergence. Their results also indicate that omitting spatial error dependence can result in model misspecification. LeSage and Fischer (2008) demonstrate that long-run regional income level not only depends on its own characteristics, but also on neighbouring's characteristics, connectivity structure and spatial dependence, and suggest the use of spatial econometric methods that take into account these spatial aspects. For the FDI, Madariaga and Poncet (2007) include spatial effects and spatially lagged levels of the FDI and per capita GDP in their cross-section, pooling, panel and GMM estimations to study the impact of FDI on China's economic growth. They find spillover effects of the FDI inflows and income per capita among 180 Chinese cities considered and conclude that economic growth of one city is affected positively by its own as well as neighbours' FDI.

The following studies by Fingleton involve increasing returns to scale. Fingleton and McCombie (1998) investigate the effect of increasing returns to scale on economic growth rate disparities among the EU regions. They estimate the effect of spatial spillovers of technical change and find large increasing returns to scale. Fingleton (2001a) develops a model that assumes increasing returns and spatially varying technical progress, and applies this model to data on the EU regions. The results indicate spillover effects of productivity and growth rates among the EU regions. Fingleton (2001b) uses 3SLS estimation method to manufacturing productivity growth data for the EU regions. The results report increasing returns, which support the new economic geography theory, and indicate that across-region spillovers are, among others, ones of the determinants of regional productivity growth variations.

Our focus in this chapter is to illustrate the applicability of our QML estimator developed in Chapter 3 to a real spatial data set. To do this, we first specify two forms of sub-models for the spatial weights that satisfy the identifiability, consistency and asymptotic normality conditions established in Chapter 3. Then 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 mixed regressive, spatial autoregressive (MR-SAR) model, to study the impact of saving, population growth and interdependence among countries on growth. We evaluate and compare our estimator using freely-estimated spatial weight matrices with other QML estimators using fixed weight matrices. Asymptotic variances are evaluated and the Wald test for our estimator is carried out. Other hypothesis tests for our estimator are for future work, as the nuisance parameter problem is present.

The results show that our QML estimator with freely-estimated weight matrices in the framework of the MR-SAR model introduced in Chapter 3 is applicable to a real data set. It is able to capture positive spatial spillovers of growth among countries and provides significant estimates of other parameters, including the parameter defining the weights, with predicted signs. Moreover, our estimator yields an estimate which is significantly different from its fixed counterpart for the weight matrix with exponential distances. We conclude that our QML estimator with freely-estimated weight matrix is able to provide an estimate of the weight parameter that is comparable to, and in one case testably different from, the value previously assumed.

This chapter is constructed as follows. Section 4.2 describes two sub-models for the spatial weights in the framework of the MR-SAR model. Section 4.3 discusses the data set used in this chapter. Section 4.4 presents the empirical results for the QMLEs using fixed and freely-estimated weight matrices. Section 4.5 concludes. A list of the countries considered in this chapter can be found in Appendix C.

4.2 MR-SAR Model and Spatial Weight Matrices

Recall that the first-order mixed regressive, spatial autoregressive (MR-SAR) model in (3.2.1) is described as

$$Y_n = X_n\beta + \lambda W_n(\gamma)Y_n + \varepsilon_n$$

and the elements of the row-standardised weight matrix $W_n(\gamma)$ is as follows

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

All elements in the equations above are described as in Chapter 3. We specify two sub-models for the spatial weights as follows.

$$w1^*_{n,ij}(\gamma_1) = \begin{cases} 0 & \text{if } i = j \\ e^{-\gamma_1 d_{ij}} & \text{if } i \neq j \end{cases}$$

and

$$w2^*_{n,ij}(\gamma_2) = \begin{cases} 0 & \text{if } i = j \\ d_{ij}^{-\gamma_2} & \text{if } i \neq j \end{cases}$$

where γ_1 and γ_2 are positive scalar parameters specifying the weights, and d_{ij} is a fixed nonnegative distance between spatial units i and j. Then, the elements $w1_{n,ij}(\gamma_1)$ of the weight matrix $W1_n(\gamma_1)$ become

$$w1_{n,ij}(\gamma_1) = \begin{cases} 0 & \text{if } i = j \\ \frac{e^{-\gamma_1 d_{ij}}}{\sum_j e^{-\gamma_1 d_{ij}}} \geq 0 & \text{if } i \neq j \end{cases} \tag{4.2.1}$$

where $\sum_j e^{-\gamma_1 d_{ij}}$ is a row sum for all i. For the weight matrix $W2_n(\gamma_2)$, its elements $w2_{n,ij}(\gamma_2)$ become

$$w2_{n,ij}(\gamma_2) = \begin{cases} 0 & \text{if } i = j \\ \frac{d_{ij}^{-\gamma_2}}{\sum_j d_{ij}^{-\gamma_2}} \geq 0 & \text{if } i \neq j \end{cases} \tag{4.2.2}$$

where $\sum_j d_{ij}^{-\gamma_2}$ is a row sum for all i. These matrices $W1_n(\gamma_1)$ and $W2_n(\gamma_2)$ are row-standardised so weight elements on the main diagonal are zero whereas all other elements are nonnegative.

Recall that the log-likelihood function of equation (3.2.1) is given by

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

where $\varepsilon_n(\delta) = Y_n - X_n\beta - \lambda W_n(\gamma)Y_n$, with $\delta = (\beta', \lambda, \gamma)'$ and $\theta = (\beta', \lambda, \gamma, \sigma^2)'$. Finally, the concentrated log-likelihood function of λ and γ is described as

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

Maximising the above equation with respect to λ and γ yields the QMLEs $\hat{\lambda}_n$ and $\hat{\gamma}_n$. Then, the QMLEs of β and σ^2 become $\hat{\beta}_n(\hat{\lambda}_n,\hat{\gamma}_n)$ and $\hat{\sigma}_n^2(\hat{\lambda}_n,\hat{\gamma}_n)$, respectively.

4.3 Data Analysis

Data set used in this chapter is obtained from Ertur and Koch (2007)[1], 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 C.1 in Appendix C for a list of these countries and their ISO codes. Table 4.1 below presents the variables and their abbreviations used in this chapter.

The first five variables in Table 4.1 are used to evaluate the impact of saving, population growth and interdependence among countries on growth. The last two variables, i.e. longitude of capital and latitude 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

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

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	latitude of capital	ylat

Table 4.1: List of variables and their acronyms.

is subsequently used to build weight matrices W1(γ_1) and W2(γ_2) described in Section 4.2. We discuss each of the main variables below.

Logarithms of real income in 1960 and 1995 for 91 countries are illustrated in Figure 4.1. Countries' ISO codes are listed on the horizontal axis in the order according to Table C.1 in Appendix C. 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 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 4.2 shows the average rates of growth between 1960 and 1995 for 91 countries. The figure shows that the average rates of growth indeed differ strongly across countries. Out of 91 countries considered in this chapter, 17 countries have negative average rates of growth. Hong Kong has the highest rate of growth between 1960 and 1995, with the average rate of growth of 6.24%, and Democratic Republic of the Congo has the lowest rate of growth between 1960 and 1995, with the average rate of −3.43%.

Next, we look at the average investment rates of the period 1960-1995. These are measured as the average shares of real investment, including gov-

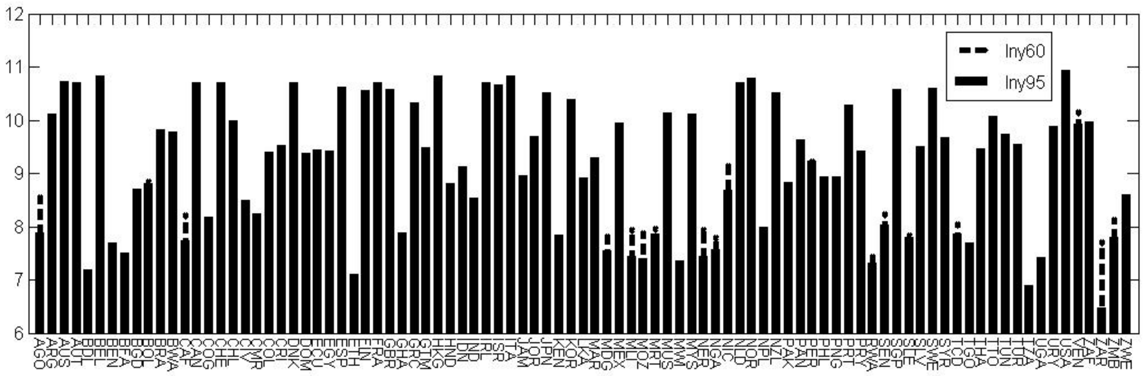


Figure 4.1: Logarithms of the levels of per worker income in 1960 and 1995 for 91 countries.

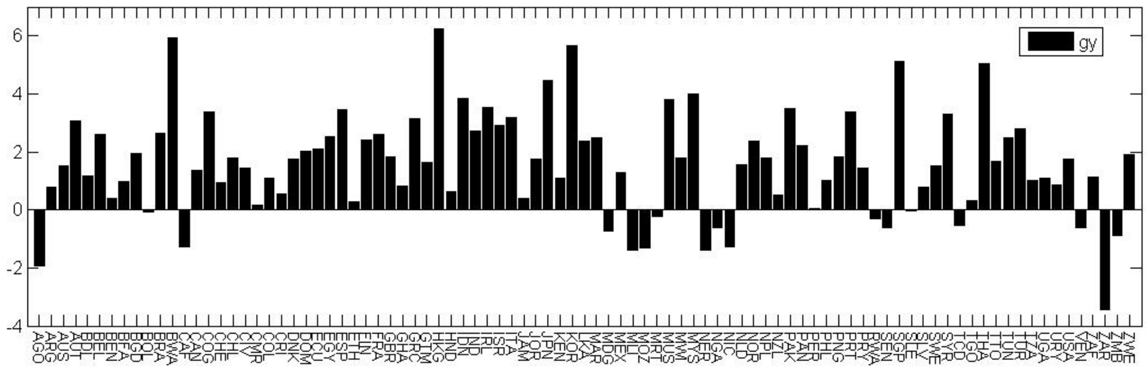


Figure 4.2: Average rates of growth between 1960 and 1995 for 91 countries.

ernment investment, in real GDP. Note that the average investment rates of the period 1960-1995 for 91 countries are shown in Figure 4.3 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%.

Finally, the average rates of growth of the working-age population n_p plus

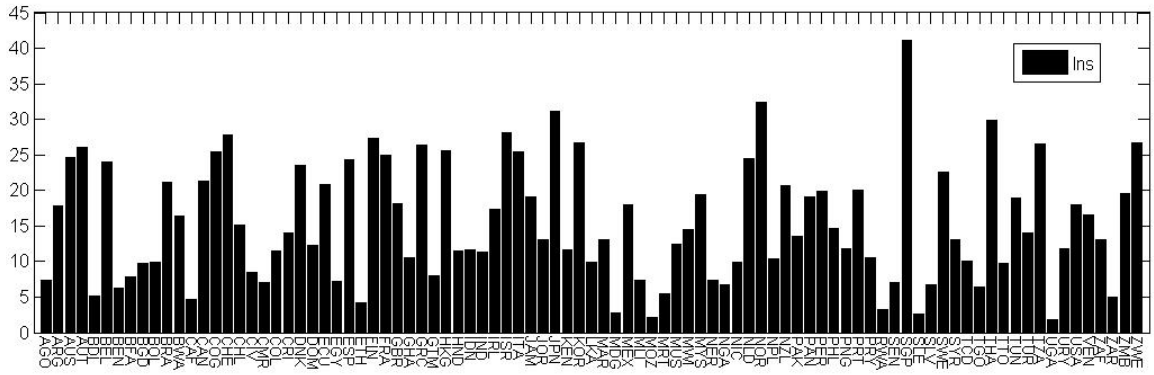


Figure 4.3: Average investment rates of the period 1960-1995 for 91 countries.

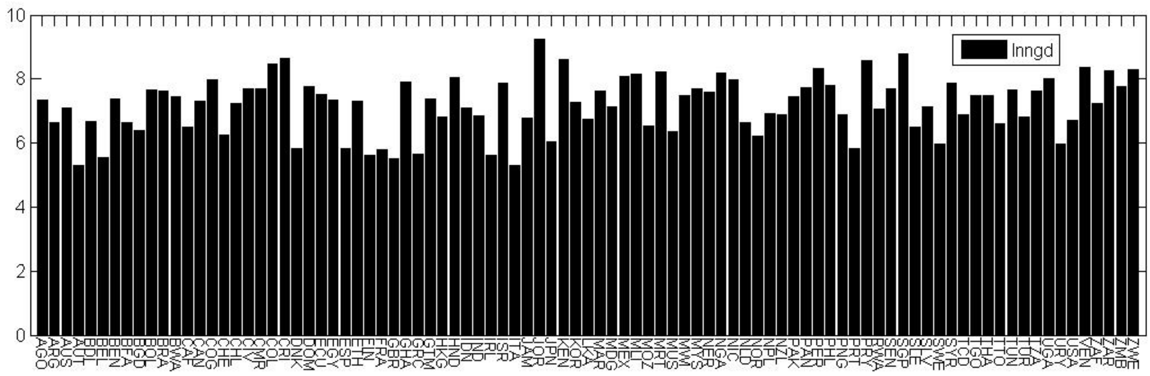


Figure 4.4: Average rates of growth of working-age population plus 0.05 for 91 countries.

0.05 ($g + \delta$) are shown in Figure 4.4. 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. The figure shows that there are large differences in the population growth rates among countries considered here. Countries with the highest and the lowest growth rates of working-age population plus 0.05 are Jordan (9.3%) and Austria (5.3%), respectively.

4.4 Empirical Results

In this section, 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 interdependence among countries on growth for each type of these weight matrices. The corresponding empirical results including the Wald test results are reported in the following subsections.

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 future work and is non-trivial. Here, for country i, with $i = 1, \ldots, 91$, our MR-SAR model is described as follows

$$gy_i = \beta_1 + \beta_2 lny60_i + \beta_3 lns_i + \beta_4 lnngd_i + \lambda \sum_{j \neq i}^{n} w_{ij}(\gamma_l) gy_j + \varepsilon_i. \qquad (4.4.1)$$

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 assumed to be the same for all countries. $w_{ij}(\gamma_l)$ is the spatial weight of countries i and j following two parametric sub-models for the weights as in equations (4.2.1) and (4.2.2), with γ_l, for $l = 1$ and 2, fixed across countries for each type of the weight matrices.

Note that in our model, the parameter defining the weights, γ, is a nuisance parameter. It is not identified under the null hypothesis. Work on inference about γ and λ based on hypothesis tests other than the Wald test is for future work. See Davies (1977 and 1987), Andrews and Ploberger (1994), and Hansen (1996), among others, for more details about hypothesis tests when a nuisance parameter is present only under the alternative.

We first report empirical results obtained from fixed weight matrices below.

4.4.1 With Fixed Spatial Weight Matrices

In this subsection we present the results obtained by evaluating the log-likelihood function derived from equation (4.4.1) above based on two types of sub-models for the spatial weights. The evaluation is carried out using a one-dimensional grid search. Table 4.2 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.

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	0.0001	0.0002
log-likelihood	271.1240	269.4306

Table 4.2: 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.

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

In Tables 4.3 and 4.4 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 (3.5.6) 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. Note that there may be a computing error in calculating these variances. This doubt will be re-checked and removed later.

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

4.4.2 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 4.5 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	0.0001	0.0002
log-likelihood	273.3922	269.6053

Table 4.5: 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 4.5 are similar to those in Table 4.2. 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 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.

In Tables 4.6 and 4.7 we report the estimated asymptotic variances for the coefficients based on weight matrices $W1(\gamma_1)$ and $W2(\gamma_2)$ with freely-estimated $\gamma_1 = 0.81$ and $\gamma_2 = 2.49$, respectively. Similarly to the results reported in Tables 4.3 and 4.4, we multiply the variances by 10^3 and round them to the nearest 3th decimal to improve the readability of the tables. Note also that there may be a computing error in calculating these variances and the doubt will be re-checked and removed later.

Variable	constant	lny60	lns	lnngd	$W1(\gamma_1)$ 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
$W1(\gamma_1)$ 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 4.6: Estimated asymptotic variance matrix for all coefficients based on weight matrix $W1(\gamma_1)$, with freely-estimated $\gamma_1 = 0.81$.

4.4.3 Wald Test

For each form of the weight matrices, we now carry out the Wald test to test whether the spatial autoregressive parameter λ is significantly greater than zero for fixed and freely estimated parameter defining the weights, γ. We report the results of Wald tests on the significance of λ for each type of the

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

weight matrices in Table 4.8. The first row of the table reports the QML estimates of λ obtained using different weight matrices. The second row lists the diagonal elements corresponding to λ in the estimated asymptotic variance matrices obtained from Tables 4.3 - 4.4 and 4.6 - 4.7. The last two rows report the Wald statistics and their associated p-values for each case considered here.

	QML with			
	W1(2)	W1(γ_1)	W2(2)	W2(γ_2)
λ	0.3	0.47	0.28	0.25
v_λ	0.0098	0.0301	0.0129	0.0202
Wald	9.2064	7.3373	6.0804	3.0983
p-value	0.0024	0.0068	0.0137	0.0784

Table 4.8: Wald tests on significance of the spatial autoregressive parameter λ, based on two different weight matrices with pre-determined and freely-estimated weight parameter γ.

As the critical value at 5% significance level for a one-sided test is 3.841, the results show that we can reject the null hypotheses for all cases except for the case of W2(γ_2) with freely-estimated γ_2. The associated p-values for the first three cases also confirm the significance of λ and we conclude that, the first three QML estimators can account for spatial spillover effects in the data, while the data cannot reject the null hypothesis of the estimated value of λ at

0.25 for the case of freely-estimated W2(γ_2).

Next, we use Wald test to test whether freely-estimated γ_l is significantly different from the pre-determined value of γ_l fixed at 2 for both forms of the weight matrices. Results of Wald tests on γ_l for each type of the weight matrices are presented in Table 4.9 below.

The first row lists the pre-determined values and QML estimates of γ for both forms of the weight matrices. The second row reports the diagonal elements corresponding to γ in the estimated asymptotic variance matrices obtained from Tables 4.3 - 4.4 and 4.6 - 4.7. The last two rows report the Wald statistics and their associated p-values, respectively.

	QML with			
	W1(2)	W1(γ_1)	W2(2)	W2(γ_2)
γ	2	0.81	2	2.49
v_γ	-	0.2295	-	3.3070
Wald	-	6.1701	-	0.0726
p-value	-	0.0260	-	0.4248

Table 4.9: Wald tests on restrictions on the parameters defining the weights, γ_1 and γ_2.

As the critical values at 5% significance level for a two-sided test are 5.024 and 0.001, the results suggest that we can reject the null hypothesis in the case based on weight matrix W1(γ_1). The Wald statistic is 6.1701, greater than the critical value of 5.024. The associated p-value also confirms this outcome. For the case based on weight matrix W2, the null hypothesis cannot be rejected. We conclude that at 5% significance level, the freely-estimated parameter defining the weights for the QMLE with weight matrix W1 is significantly different from its pre-determined counterpart fixed at 2, whereas it seems that we cannot choose between fixed and freely-estimated weight parameter for the QML estimators with weight matrix W2.

4.5 Conclusion

Spatial econometrics has been used in applied work in many fields including the field of economic growth as it is able to account for spillovers between spatial units. To illustrate the applicability of our QML estimator, we apply our QML estimator introduced in Chapter 3, using two functional forms of the weight matrices, to a real data set to study the impact of saving, population growth and interdependence among countries 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 Wald test is carried out. Hypothesis tests other than the Wald test will be carried out in future work as the nuisance parameter problem is present.

The empirical results show that our QML estimator with freely-estimated weight matrices in the framework of the MR-SAR model introduced in Chapter 3 is applicable to a real data set. It is able to capture positive spatial spillovers of growth among countries and provides significant estimates of other parameters including the parameter defining the weights, with predicted signs. Moreover, our estimator yields an estimate which is significantly different from its pre-determined counterpart for weight matrix W1 with exponential distances. We conclude that our QML estimator with freely-estimated weight matrix is able to provide an estimate of the weight parameter that is comparable to, and in one case testably different from, the value previously assumed.

Chapter 5

Conclusion

This thesis explores two issues in spatial econometrics. The first issue involves a bias-adjusted estimator for small samples and the second issue is in regard to the spatial weight matrix. Chapter 1 provides an introduction to and a review of the literature in the field of spatial econometrics.

The maximum likelihood has widely been used as an estimation method in spatial econometrics since it is consistent and asymptotically efficient. However, this method involves matrix inversion and eigenvalue calculations, which may cause numerical problems when the sample size is large. Several alternative methods have been proposed, among which are the GMM/IV-type estimators that are consistent and computationally feasible under appropriate conditions. Even though the GMM estimators are consistent, they may have large bias in finite samples. Besides, it is not clear which estimator performs better in such a case.

Chapter 2 introduces a bias-adjusted estimator [BB] for small samples and extensively compares its performance with that of six existing estimators in the context of a spatial error model. We show that the BB estimator is robust and its performance does not depend on a particular spatial weight matrix M. An optimal weight matrix W should also be incorporated in the method of moments procedure to improve the efficiency of the estimators. Furthermore, the bias-adjusted estimators; BB, AW, and AWW, perform extremely well in reducing small-sample bias, being virtually mean and median unbiased. Nevertheless, all estimators except the MLE produce a significant proportion of non-invertible estimates. This motivates us to develop the hybrid estimator for the spatial autocorrelation parameter to improve the small-sample efficiency. The hybridised forms of the BB, AW and AWW estimators are clearly superior to other estimators in small samples and, in our experiments, the use of the hybrid estimator in the first stage of a feasible GLS estimator leads to inferences about the regression coefficients in the second stage that are at least as robust as those of competing estimators.

For future research based on Chapter 2, the bias-adjusted estimator for small samples may be extended to the general spatial process model that includes both the spatial lag dependence as well as the spatial error dependence.

Chapter 3 deals with the second focal issue of this thesis; the spatial weight matrix. This issue 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 matrix leads to different results and different interpretations of the results. While spatial weights should be chosen based on spatial interaction theory (Anselin, 1980 and 1984), practitioners sometimes choose a weight matrix based on empirical convenience, that may not properly capture the dependence structure.

In Chapter 3 we introduce a sub-model for the spatial weights and estimate the variable spatial weight matrix in the mixed regressive, spatial autoregressive (MR-SAR) model by the quasi-maximum 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). Its small sample properties are studied in a Monte Carlo experiment. The performance of the estimator is subsequently compared with that of other QML estimators using various fixed spatial weight matrices. The 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 Chapter 3. The results also show that using a wrong weight matrix strongly affects the estimation performance of the estimators, especially when estimating the spatial autoregressive parameter, λ.

Extending the QML estimator to a panel/dynamic setting and/or to the framework of the spatial autoregressive model with spatial autoregressive disturbance is an interesting path for future research.

Finally, in Chapter 4, we apply our QML estimator using freely-estimated weight matrix based on two functional forms of sub-models for the weights to cross-sectional data set to study the impact of saving, population growth and interdependence among countries on growth in the framework of the MR-SAR model. Our QML estimator using freely-estimated weight matrices is compared with other QML estimators using weight matrices with the weight parameter values adopted in previous work. Asymptotic variances are evaluated and the Wald test is carried out. Other hypothesis tests for our estimator are for future research as the nuisance parameter problem is present. The 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 other parameters of the model including the parameter defining the spatial weights. The QML estimator with freely-estimated weight matrix is able to provide an estimate of the weight parameter that is comparable to, and in one case testably different from, the value previously assumed.

For future research based on Chapter 4, we may apply our QML estimator using variable weight matrices to data sets in other fields of economics such as social interactions, foreign direct investment, migration or trade. Developing hypothesis tests for our estimator when a nuisance parameter is present under the alternative is also an interesting topic for future work.

Appendix A

Appendix to Chapter 2

A.1 Simulation Results

This section presents the simulation results obtained for Chapter 2 - Improved Estimators for the Spatial Error Model. The results listed in this section show the performance of the non-hybridised and hybridised estimators considered in Chapter 2, obtained for various values of σ^2 for the case of $n = 49$. All other results not reported in this thesis are available on request.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB1	0.006	0.011	0.237	0.237	0.5	BB1	**0.501**	0.508	0.193	0.193
	ML	-0.071	-0.062	0.228	0.239		ML	0.429	0.450	**0.175**	0.189
	KP1	-0.077	-0.064	0.227	0.240		KP1	0.412	0.432	0.181	0.201
	KPW1	-0.045	-0.040	**0.222**	**0.226**		KPW1	0.435	0.447	0.180	0.192
	LL1	-0.072	-0.063	0.226	0.237		LL1	0.421	0.443	0.181	0.197
	AW1	-0.025	-0.011	0.235	0.236		AW1	0.481	**0.500**	0.184	**0.185**
	AWW1	**0.002**	**0.009**	0.234	0.234		AWW1	0.502	0.513	0.187	0.187
0.1	BB1	0.104	0.113	0.231	0.231	0.7	BB1	**0.697**	**0.707**	0.153	0.153
	ML	0.028	0.041	0.222	0.233		ML	0.633	0.655	**0.137**	0.152
	KP1	0.018	0.032	0.222	0.237		KP1	0.616	0.633	0.145	0.167
	KPW1	0.050	0.055	**0.217**	**0.223**		KPW1	0.632	0.646	0.146	0.161
	LL1	0.026	0.038	0.220	0.233		LL1	0.623	0.648	0.155	0.173
	AW1	0.074	**0.090**	0.229	0.230		AW1	0.692	0.711	0.148	**0.148**
	AWW1	**0.101**	0.111	0.228	0.228		AWW1	0.707	0.722	0.150	0.150
0.3	BB1	0.301	0.310	0.215	0.215	0.9	BB1	**0.898**	**0.913**	0.084	**0.084**
	ML	0.228	0.245	**0.203**	0.216		ML	0.845	0.867	**0.081**	0.098
	KP1	0.212	0.229	0.206	0.224		KP1	0.822	0.833	0.093	0.121
	KPW1	0.241	0.250	0.203	0.212		KPW1	0.830	0.841	0.093	0.116
	LL1	0.222	0.240	0.204	0.218		LL1	0.835	0.858	0.110	0.128
	AW1	0.275	**0.293**	0.210	**0.212**		AW1	0.906	0.926	0.087	0.087
	AWW1	**0.300**	0.311	0.212	0.212		AWW1	0.912	0.934	0.085	0.086

Table A.1: Non-hybridised estimators of ρ for $n = 49$, $\sigma^2 = 1$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB2	0.004	0.011	0.233	0.233	0.5	BB2	0.496	0.506	0.188	0.188
	ML	-0.071	-0.062	0.228	0.239		ML	0.429	0.450	**0.175**	0.189
	KP2	-0.077	-0.064	0.226	0.239		KP2	0.412	0.432	0.181	0.201
	KPW2	-0.045	-0.040	**0.222**	**0.226**		KPW2	0.434	0.447	0.179	0.191
	LL2	-0.072	-0.063	0.226	0.237		LL2	0.421	0.443	0.179	0.195
	AW2	-0.025	-0.011	0.235	0.236		AW2	0.481	**0.500**	0.184	0.185
	AWW2	**0.002**	**0.009**	0.234	0.234		AWW2	**0.499**	0.513	0.183	**0.183**
0.1	BB2	0.102	0.113	0.227	0.227	0.7	BB2	0.690	**0.704**	0.147	0.147
	ML	0.028	0.041	0.222	0.233		ML	0.633	0.655	**0.137**	0.152
	KP2	0.018	0.032	0.221	0.236		KP2	0.616	0.633	0.145	0.167
	KPW2	0.050	0.055	**0.217**	**0.223**		KPW2	0.630	0.645	0.143	0.159
	LL2	0.026	0.038	0.220	0.233		LL2	0.624	0.648	0.149	0.167
	AW2	0.075	**0.090**	0.228	0.230		AW2	0.691	0.711	0.146	0.147
	AWW2	**0.101**	0.111	0.228	0.228		AWW2	**0.701**	0.719	0.144	**0.144**
0.3	BB2	**0.299**	0.309	0.212	0.212	0.9	BB2	0.877	0.892	0.081	0.084
	ML	0.228	0.245	0.203	0.216		ML	0.845	0.867	0.081	0.098
	KP2	0.212	0.229	0.206	0.224		KP2	0.821	0.833	0.091	0.121
	KPW2	0.240	0.250	0.203	0.211		KPW2	0.827	0.840	0.090	0.116
	LL2	0.223	0.240	**0.203**	0.217		LL2	0.836	0.858	0.105	0.123
	AW2	0.275	**0.293**	0.210	0.212		AW2	0.889	0.906	0.078	0.079
	AWW2	0.299	0.310	0.211	**0.211**		AWW2	**0.889**	**0.906**	**0.076**	**0.077**

Table A.2: Hybridised estimators of ρ for $n = 49$, $\sigma^2 = 1$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB1	0.001	0.009	0.241	0.241	0.5	BB1	0.495	0.506	0.197	0.197
	ML	-0.075	-0.064	0.234	0.246		ML	0.426	0.452	**0.179**	0.194
	KP1	-0.081	-0.063	0.229	0.243		KP1	0.408	0.427	0.183	0.205
	KPW1	-0.048	-0.039	**0.226**	**0.231**		KPW1	0.430	0.446	0.184	0.197
	LL1	-0.077	-0.064	0.232	0.244		LL1	0.419	0.444	0.184	0.201
	AW1	-0.029	-0.011	0.238	0.240		AW1	0.477	**0.500**	0.187	**0.189**
	AWW1	**-0.001**	**0.008**	0.239	0.239		AWW1	**0.497**	0.514	0.191	0.191
0.1	BB1	**0.099**	**0.109**	0.235	0.235	0.7	BB1	0.692	**0.701**	0.157	0.157
	ML	0.024	0.041	0.228	0.241		ML	0.631	0.656	**0.140**	0.156
	KP1	0.014	0.032	0.224	0.240		KP1	0.613	0.630	0.148	0.172
	KPW1	0.046	0.057	**0.222**	**0.228**		KPW1	0.628	0.641	0.149	0.166
	LL1	0.021	0.038	0.226	0.240		LL1	0.622	0.650	0.151	0.170
	AW1	0.071	0.089	0.232	0.234		AW1	0.689	0.708	0.151	**0.151**
	AWW1	0.097	0.110	0.234	0.234		AWW1	**0.703**	0.721	0.154	0.154
0.3	BB1	**0.298**	0.308	0.223	0.223	0.9	BB1	0.891	**0.906**	0.088	0.088
	ML	0.224	0.246	0.209	0.222		ML	0.845	0.867	**0.084**	0.101
	KP1	0.208	0.227	0.208	0.227		KP1	0.820	0.830	0.096	0.125
	KPW1	0.236	0.252	**0.207**	0.217		KPW1	0.827	0.837	0.095	0.120
	LL1	0.218	0.244	0.210	0.225		LL1	0.837	0.860	0.098	0.117
	AW1	0.271	**0.293**	0.214	**0.215**		AW1	**0.904**	0.923	0.089	0.089
	AWW1	0.296	0.312	0.217	0.217		AWW1	0.909	0.928	0.087	**0.088**

Table A.3: Non-hybridised estimators of ρ for $n = 49$, $\sigma^2 = 0.25$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB2	**0.001**	0.009	0.240	0.239	0.5	BB2	0.491	0.504	0.193	0.193
	ML	-0.075	-0.064	0.234	0.246		ML	0.426	0.452	**0.179**	0.194
	KP2	-0.081	-0.063	0.229	0.243		KP2	0.408	0.427	0.183	0.205
	KPW2	-0.048	-0.039	**0.226**	**0.231**		KPW2	0.429	0.446	0.183	0.196
	LL2	-0.077	-0.064	0.232	0.244		LL2	0.419	0.444	0.184	0.201
	AW2	-0.029	-0.011	0.238	0.240		AW2	0.477	**0.500**	0.187	**0.188**
	AWW2	-0.001	**0.008**	0.239	0.239		AWW2	**0.495**	0.514	0.189	0.189
0.1	BB2	**0.098**	**0.109**	0.233	0.233	0.7	BB2	0.685	**0.698**	0.151	0.152
	ML	0.024	0.041	0.228	0.241		ML	0.631	0.656	**0.140**	0.156
	KP2	0.014	0.032	0.224	0.240		KP2	0.613	0.630	0.148	0.172
	KPW2	0.046	0.057	**0.222**	**0.228**		KPW2	0.626	0.641	0.147	0.164
	LL2	0.021	0.038	0.226	0.240		LL2	0.622	0.650	0.151	0.170
	AW2	0.071	0.089	0.232	0.234		AW2	0.688	0.708	0.149	0.150
	AWW2	0.097	0.110	0.234	0.234		AWW2	**0.697**	0.717	0.148	**0.148**
0.3	BB2	0.295	0.307	0.218	0.218	0.9	BB2	0.874	0.890	0.085	0.089
	ML	0.224	0.246	0.209	0.222		ML	0.845	0.867	0.084	0.101
	KP2	0.208	0.227	0.208	0.227		KP2	0.819	0.830	0.094	0.124
	KPW2	0.235	0.252	**0.206**	0.216		KPW2	0.824	0.836	0.093	0.120
	LL2	0.218	0.244	0.210	0.225		LL2	0.837	0.860	0.098	0.116
	AW2	0.271	**0.293**	0.214	**0.215**		AW2	0.887	**0.906**	0.079	0.080
	AWW2	**0.295**	0.312	0.215	0.216		AWW2	**0.887**	0.907	**0.079**	**0.080**

Table A.4: Hybridised estimators of ρ for $n = 49$, $\sigma^2 = 0.25$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB1	**-0.001**	0.006	0.239	0.239	0.5	BB1	0.496	**0.502**	0.197	0.197
	ML	-0.076	-0.064	0.232	0.244		ML	0.424	0.446	**0.178**	0.193
	KP1	-0.081	-0.067	0.228	0.242		KP1	0.408	0.426	0.183	0.204
	KPW1	-0.050	-0.042	**0.226**	**0.231**		KPW1	0.431	0.443	0.184	0.196
	LL1	-0.078	-0.066	0.229	0.242		LL1	0.415	0.439	0.189	0.207
	AW1	-0.029	-0.015	0.237	0.238		AW1	0.476	0.496	0.186	**0.187**
	AWW1	-0.003	**0.003**	0.238	0.238		AWW1	**0.496**	0.509	0.190	0.190
0.1	BB1	**0.099**	0.106	0.237	0.237	0.7	BB1	0.695	**0.702**	0.158	0.158
	ML	0.023	0.040	0.226	0.238		ML	0.629	0.653	**0.139**	0.156
	KP1	0.014	0.027	0.223	0.239		KP1	0.613	0.631	0.147	0.171
	KPW1	0.045	0.055	**0.221**	**0.228**		KPW1	0.629	0.645	0.148	0.165
	LL1	0.020	0.035	0.223	0.237		LL1	0.617	0.645	0.165	0.185
	AW1	0.070	0.087	0.230	0.232		AW1	0.688	0.706	0.150	**0.150**
	AWW1	0.095	**0.104**	0.232	0.232		AWW1	**0.702**	0.718	0.152	0.152
0.3	BB1	**0.297**	**0.304**	0.222	0.222	0.9	BB1	0.896	**0.912**	0.088	**0.088**
	ML	0.222	0.242	**0.206**	0.220		ML	0.843	0.862	**0.085**	0.102
	KP1	0.208	0.223	0.207	0.227		KP1	0.821	0.836	0.095	0.124
	KPW1	0.236	0.247	0.207	0.216		KPW1	0.829	0.842	0.095	0.119
	LL1	0.216	0.237	0.207	0.224		LL1	0.831	0.856	0.115	0.134
	AW1	0.271	0.290	0.212	**0.214**		AW1	**0.903**	0.922	0.089	0.089
	AWW1	0.294	0.307	0.216	0.216		AWW1	0.909	0.929	0.088	0.088

Table A.5: Non-hybridised estimators of ρ for $n = 49$, $\sigma^2 = 0.5$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB2	**-0.001**	0.006	0.238	0.238	0.5	BB2	0.490	**0.500**	0.190	0.190
	ML	-0.076	-0.064	0.232	0.244		ML	0.424	0.446	**0.178**	0.193
	KP2	-0.081	-0.067	0.228	0.242		KP2	0.408	0.426	0.183	0.204
	KPW2	-0.050	-0.042	**0.226**	**0.231**		KPW2	0.430	0.443	0.182	0.195
	LL2	-0.078	-0.066	0.229	0.242		LL2	0.417	0.439	0.184	0.202
	AW2	-0.029	-0.015	0.236	0.238		AW2	0.476	0.496	0.185	0.187
	AWW2	-0.003	**0.003**	0.238	0.238		AWW2	**0.494**	0.508	0.186	**0.186**
0.1	BB2	**0.097**	0.106	0.233	0.233	0.7	BB2	0.688	**0.699**	0.151	0.151
	ML	0.023	0.040	0.226	0.238		ML	0.629	0.653	**0.139**	0.156
	KP2	0.014	0.027	0.223	0.239		KP2	0.613	0.631	0.147	0.171
	KPW2	0.045	0.055	**0.221**	**0.228**		KPW2	0.626	0.644	0.145	0.162
	LL2	0.020	0.035	0.223	0.237		LL2	0.620	0.645	0.152	0.172
	AW2	0.070	0.087	0.230	0.232		AW2	0.686	0.706	0.148	0.149
	AWW2	0.095	**0.104**	0.232	0.232		AWW2	**0.697**	0.715	0.146	**0.146**
0.3	BB2	0.294	**0.304**	0.216	0.216	0.9	BB2	0.876	0.892	0.086	0.090
	ML	0.222	0.242	**0.206**	0.220		ML	0.843	0.862	0.085	0.102
	KP2	0.208	0.223	0.207	0.227		KP2	0.820	0.836	0.094	0.124
	KPW2	0.236	0.247	0.206	0.216		KPW2	0.826	0.841	0.093	0.119
	LL2	0.217	0.237	0.206	0.222		LL2	0.832	0.856	0.110	0.129
	AW2	0.271	0.290	0.211	**0.213**		AW2	0.886	**0.903**	0.080	0.081
	AWW2	**0.294**	0.307	0.215	0.215		AWW2	**0.887**	0.904	**0.079**	**0.081**

Table A.6: Hybridised estimators of ρ for $n = 49$, $\sigma^2 = 0.5$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB1	**0.000**	0.007	0.243	0.243	0.5	BB1	0.494	0.507	0.196	0.196
	ML	-0.074	-0.063	0.234	0.245		ML	0.426	0.449	**0.181**	0.196
	KP1	-0.081	-0.063	0.232	0.246		KP1	0.408	0.431	0.186	0.208
	KPW1	-0.050	-0.041	**0.227**	**0.233**		KPW1	0.430	0.447	0.185	0.198
	LL1	-0.076	-0.063	0.232	0.244		LL1	0.417	0.444	0.191	0.208
	AW1	-0.029	-0.010	0.241	0.243		AW1	0.476	**0.499**	0.189	**0.191**
	AWW1	-0.003	**0.004**	0.240	0.240		AWW1	**0.496**	0.515	0.192	0.192
0.1	BB1	**0.098**	0.106	0.235	0.235	0.7	BB1	0.694	**0.705**	0.159	0.159
	ML	0.025	0.042	0.228	0.240		ML	0.631	0.656	**0.142**	0.158
	KP1	0.014	0.033	0.227	0.243		KP1	0.613	0.633	0.150	0.174
	KPW1	0.045	0.054	**0.223**	**0.230**		KPW1	0.628	0.645	0.150	0.167
	LL1	0.021	0.038	0.227	0.241		LL1	0.620	0.648	0.160	0.179
	AW1	0.070	0.090	0.234	0.236		AW1	0.688	0.709	0.153	**0.153**
	AWW1	0.096	**0.105**	0.234	0.234		AWW1	**0.702**	0.721	0.154	0.154
0.3	BB1	**0.296**	0.308	0.221	0.221	0.9	BB1	0.894	**0.913**	0.090	0.090
	ML	0.224	0.245	0.209	0.223		ML	0.843	0.866	**0.086**	0.103
	KP1	0.208	0.231	0.211	0.230		KP1	0.820	0.834	0.097	0.126
	KPW1	0.235	0.249	**0.208**	0.218		KPW1	0.828	0.841	0.097	0.121
	LL1	0.218	0.240	0.213	0.228		LL1	0.833	0.859	0.118	0.136
	AW1	0.271	**0.294**	0.216	0.218		AW1	**0.903**	0.925	0.090	0.090
	AWW1	0.294	0.310	0.217	**0.217**		AWW1	0.909	0.932	0.089	**0.089**

Table A.7: Non-hybridised estimators of ρ for $n = 49$, $\sigma^2 = 2$.

ρ	Method	Mean	Med.	St.D.	RMSE	ρ	Method	Mean	Med.	St.D.	RMSE
0.0	BB2	**-0.001**	0.007	0.239	0.239	0.5	BB2	0.489	0.504	0.190	0.190
	ML	-0.074	-0.063	0.234	0.245		ML	0.426	0.449	**0.181**	0.196
	KP2	-0.081	-0.063	0.232	0.246		KP2	0.408	0.431	0.186	0.208
	KPW2	-0.050	-0.041	**0.227**	**0.232**		KPW2	0.429	0.446	0.183	0.196
	LL2	-0.076	-0.063	0.232	0.244		LL2	0.419	0.444	0.184	0.201
	AW2	-0.029	-0.010	0.240	0.242		AW2	0.476	**0.499**	0.189	0.191
	AWW2	-0.003	**0.004**	0.240	0.240		AWW2	**0.494**	0.514	0.189	**0.189**
0.1	BB2	**0.097**	0.106	0.234	0.234	0.7	BB2	0.685	**0.698**	0.151	0.151
	ML	0.025	0.042	0.228	0.240		ML	0.631	0.656	**0.142**	0.158
	KP2	0.014	0.033	0.227	0.242		KP2	0.613	0.633	0.150	0.174
	KPW2	0.045	0.054	**0.223**	**0.230**		KPW2	0.626	0.645	0.147	0.165
	LL2	0.021	0.038	0.227	0.241		LL2	0.619	0.648	0.160	0.179
	AW2	0.070	0.090	0.234	0.236		AW2	0.687	0.709	0.151	0.151
	AWW2	0.095	**0.105**	0.233	0.233		AWW2	**0.696**	0.718	0.149	**0.149**
0.3	BB2	0.293	0.307	0.216	**0.216**	0.9	BB2	0.875	0.891	0.086	0.090
	ML	0.224	0.245	0.209	0.223		ML	0.843	0.866	0.086	0.103
	KP2	0.208	0.231	0.211	0.230		KP2	0.819	0.834	0.095	0.125
	KPW2	0.235	0.249	**0.207**	0.217		KPW2	0.824	0.839	0.094	0.121
	LL2	0.219	0.240	0.211	0.225		LL2	0.832	0.858	0.118	0.136
	AW2	0.271	**0.294**	0.216	0.218		AW2	0.887	0.907	0.081	0.082
	AWW2	**0.293**	0.310	0.216	0.216		AWW2	**0.887**	**0.905**	**0.081**	**0.082**

Table A.8: Hybridised estimators of ρ for $n = 49$, $\sigma^2 = 2$.

True ρ	σ^2 0.25		0.5		1		2	
0.0	KPW1	(0.741)	KPW1	(0.940)	LL1	(0.471)	KP1	(0.551)
0.1	KPW1	(0.705)	KPW1	(0.716)	LL1	(0.580)	KP1	(0.871)
0.3	KP1	(0.570)	ML	(0.598)	LL1	(0.486)	LL1	(0.584)
0.5	ML	(0.579)	ML	(0.525)	LL1	(0.537)	LL1	(0.573)
0.7	ML	(0.540)	ML	(0.464)	ML	(0.493)	ML	(0.472)
0.9	ML	(0.613)	ML*	(0.596)	ML*	(0.566)	ML	(0.580)

Table A.9: Estimation of β_1. For each ρ and σ^2 combination, the table entry is the non-hybridised estimator of β_1 giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB1.

True ρ	σ^2 0.25		0.5		1		2	
0.0	KPW2	(0.984)	KPW2	(0.999)	LL2	(0.965)	KP2	(0.959)
0.1	KPW2	(0.993)	KPW2	(0.964)	LL2	(0.975)	KP2	(0.951)
0.3	KP2	(0.775)	ML	(0.932)	LL2	(0.761)	LL2	(0.888)
0.5	ML	(0.779)	ML	(0.851)	LL2	(0.764)	ML	(0.868)
0.7	ML	(0.847)	ML	(0.786)	ML	(0.807)	ML	(0.863)
0.9	ML	(0.905)	ML	(0.900)	ML	(0.885)	ML	(0.910)

Table A.10: Estimation of β_1. For each ρ and σ^2 combination, the table entry is the hybridised estimator of β_1 giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB2.

True ρ	n 20		50		100		245		490	
0.0	ML	(0.244)	AW1	(0.471)	AWW1	(1.000)	KPW1	(1.000)	KP1	(1.000)
0.1	ML	(0.249)	ML	(0.457)	AW1	(1.000)	KPW1	(1.000)	KP1	(1.000)
0.3	ML*	(0.258)	ML	(0.510)	AW1	(0.903)	BB1	(1.000)	ML	(1.000)
0.5	ML*	(0.284)	ML	(0.482)	AW1	(0.849)	KP1	(1.000)	LL1	(1.000)
0.7	ML*	(0.346)	ML*	(0.526)	ML	(0.870)	KP1	(0.994)	ML	(1.000)
0.9	ML*	(0.647)	ML*	(0.709)	ML	(0.860)	LL1	(0.966)	LL1	(0.995)

Table A.11: Estimation of β_1. For each ρ and n combination, the table entry is the non-hybridised estimator of β_1 giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB1.

True ρ	n 20		50		100		245		490	
0.0	KP2	(0.835)	AW2	(0.885)	AWW2	(1.000)	KPW2	(1.000)	KP2	(1.000)
0.1	ML	(0.771)	ML	(0.809)	AW2	(1.000)	KPW2	(1.000)	KP2	(1.000)
0.3	ML*	(0.814)	ML	(0.962)	AW2	(1.000)	BB2	(1.000)	ML	(1.000)
0.5	ML*	(0.794)	ML	(0.969)	AW2	(0.991)	KP2	(1.000)	LL2	(1.000)
0.7	ML*	(0.891)	LL2	(0.912)	ML	(0.971)	KP2	(0.994)	ML	(1.000)
0.9	ML	(0.982)	ML	(0.981)	ML	(0.981)	LL2	(0.995)	LL2	(0.995)

Table A.12: Estimation of β_1. For each ρ and n combination, the table entry is the hybridised estimator of β_1 giving the smallest RMSE with $\sigma^2 = 1$. The figure in parentheses is the relative efficiency of BB2.

True ρ	σ^2 0.25		0.5		1		2	
0.0	KPW1	(0.999)	KPW1	(0.998)	KPW1	(0.999)	KP1	(0.997)
0.1	KPW1	(1.000)	KPW1	(0.999)	BB1	(1.000)	KP1	(0.998)
0.3	BB1	(1.000)	BB1	(1.000)	BB1	(1.000)	AW1	(0.999)
0.5	BB1	(1.000)	BB1	(1.000)	BB1	(1.000)	AW1	(0.998)
0.7	ML	(1.000)	AW1	(0.999)	AWW1	(1.000)	ML*	(0.996)
0.9	ML	(0.999)	AW1	(0.998)	LL1	(0.998)	ML	(0.998)

Table A.13: Estimation of β_2. For each ρ and σ^2 combination, the table entry is the non-hybridised estimator of β_2 giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB1.

True ρ	σ^2 0.25		0.5		1		2	
0.0	KPW2	(0.999)	KPW2	(0.998)	KPW2	(0.999)	KP2	(0.997)
0.1	KPW2	(1.000)	KPW2	(0.999)	KPW2	(0.999)	KP2	(0.998)
0.3	BB2	(1.000)	BB2	(1.000)	BB2	(1.000)	AW2	(0.999)
0.5	BB2	(1.000)	BB2	(1.000)	BB2	(1.000)	AW2	(0.998)
0.7	BB2	(1.000)	AWW2	(0.999)	AWW2	(1.000)	ML*	(0.996)
0.9	BB2	(1.000)	AW2	(0.997)	LL2	(1.000)	AWW2	(0.997)

Table A.14: Estimation of β_2. For each ρ and σ^2 combination, the table entry is the hybridised estimator of β_2 giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB2.

	σ^2							
True ρ	**0.25**		**0.5**		**1**		**2**	
0.0	KPW1	(0.999)	KP1	(0.998)	KPW1	(0.998)	KPW1	(0.999)
0.1	BB1	(1.000)	KPW1	(0.999)	KPW1	(0.999)	BB1	(1.000)
0.3	BB1	(1.000)	AW1	(1.000)	BB1	(1.000)	BB1	(1.000)
0.5	AWW1	(1.000)	AW1	(0.999)	BB1	(1.000)	BB1	(1.000)
0.7	AWW1	(0.999)	AW1	(0.999)	AWW1	(0.999)	AWW1	(0.998)
0.9	AW1	(0.998)	AWW1	(1.000)	ML	(0.999)	AWW1	(0.996)

Table A.15: Estimation of β_3. For each ρ and σ^2 combination, the table entry is the non-hybridised estimator of β_3 giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB1.

	σ^2							
True ρ	**0.25**		**0.5**		**1**		**2**	
0.0	KPW2	(0.999)	KP2	(0.999)	KPW2	(0.999)	KPW2	(1.000)
0.1	BB2	(1.000)	KPW2	(0.999)	KPW2	(1.000)	BB2	(1.000)
0.3	BB2	(1.000)	BB2	(1.000)	BB2	(1.000)	BB2	(1.000)
0.5	AWW2	(0.999)	AW2	(1.000)	BB2	(1.000)	AWW2	(1.000)
0.7	AW2	(0.999)	AWW2	(0.999)	AWW2	(0.999)	AWW2	(0.999)
0.9	BB2	(1.000)	AW2	(0.999)	AWW2	(0.998)	AWW2	(0.998)

Table A.16: Estimation of β_3. For each ρ and σ^2 combination, the table entry is the hybridised estimator of β_3 giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB2.

	σ^2							
True ρ	**0.25**		**0.5**		**1**		**2**	
0.0	AW1*	(0.993)	AWW1	(0.996)	AWW1	(0.994)	AWW1	(0.996)
0.1	AW1*	(0.993)	AWW1	(0.987)	AW1*	(0.994)	AWW1	(0.993)
0.3	AW1*	(0.971)	AWW1	(0.970)	AW1	(0.985)	AWW1	(0.977)
0.5	KP1	(0.936)	KPW1	(0.935)	KP1	(0.943)	KPW1	(0.945)
0.7	ML*	(0.849)	ML*	(0.845)	ML*	(0.853)	ML	(0.847)
0.9	ML*	(0.627)	ML*	(0.648)	ML*	(0.630)	ML*	(0.629)

Table A.17: Estimation of σ^2. For each ρ and σ^2 combination, the table entry is the non-hybridised estimator of σ^2 giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB1.

	σ^2							
True ρ	**0.25**		**0.5**		**1**		**2**	
0.0	AW2	(0.995)	AWW2	(0.998)	AWW2	(0.998)	AWW2	(0.997)
0.1	AW2*	(0.995)	AWW2	(0.992)	AW2	(0.998)	AWW2	(0.995)
0.3	AW2*	(0.984)	AWW2	(0.985)	AW2	(0.988)	AWW2	(0.987)
0.5	KP2	(0.953)	KPW2	(0.959)	KPW2	(0.959)	KPW2	(0.963)
0.7	ML*	(0.888)	ML*	(0.889)	ML*	(0.889)	ML	(0.892)
0.9	ML*	(0.683)	ML*	(0.712)	ML*	(0.686)	ML*	(0.685)

Table A.18: Estimation of σ^2. For each ρ and σ^2 combination, the table entry is the hybridised estimator of σ^2 giving the smallest RMSE for $n = 49$. The figure in parentheses is the relative efficiency of BB2.

Appendix B

Appendix to Chapter 3

B.1 List of Notations

The list below presents the notations frequently used in Chapter 3 and Appendix B, most of which are extensions of Lee (2004a)'s notations.

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

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

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

$$S_n = I_n - \lambda_0 W_n$$

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

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

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

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

$$Z_n = \frac{\partial W_n}{\partial \gamma}$$
$$A_n = \frac{\partial Z_n}{\partial \gamma}$$
$$B_n = \frac{\partial A_n}{\partial \gamma}$$
$$Q_n(\lambda,\gamma) = max_{\beta,\sigma^2} E[\ln L_n(\theta)] = -\frac{n}{2}(\ln(2\pi)+1) + \ln|det(S_n(\lambda,\gamma))| - \frac{n}{2}\ln \sigma_n^{2*}(\lambda,\gamma)$$
$$\hat{\sigma}_n^2(\lambda,\gamma) = \frac{1}{n}[Y_n' S_n'(\lambda,\gamma) M_n S_n(\lambda,\gamma) Y_n]$$
$$\sigma_n^{2*}(\lambda,\gamma) = \frac{1}{n}[(\lambda_0-\lambda)^2(G_nX_n\beta_0)' M_n (G_nX_n\beta_0) + \sigma_0^2 tr(S_n^{-1\prime} S_n'(\lambda,\gamma) S_n(\lambda,\gamma) S_n^{-1})]$$
$$\sigma_n^2(\lambda,\gamma) = \frac{\sigma_0^2}{n} tr[S_n^{-1\prime} S_n'(\lambda,\gamma) S_n(\lambda,\gamma) S_n^{-1}]$$
$$M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$$

B.2 List of Lemmas, Theorem and Definition

For convenience, we gather and list the existing Lemmas, Definition and Theorem that are used frequently in Chapter 3 and Appendix B in this section. Note that these Lemmas, Definition and Theorem below are written exactly as the originals appearing in the references.

B.2.1 Lemmas in Lee (2002, 2003, 2004b)

Lee (2002) - Lemma A.2: Suppose that A_n is a square matrix with its column sums being uniformly bounded and elements of the $n \times k$ matrix C_n are uniformly bounded. Then, $(1/\sqrt{n})C_n'A_nV_n = O(1)$. Furthermore, if the limit of $(1/n)C_n'A_nA_n'C_n$ exists and it is positive definite, then $(1/\sqrt{n})C_n'A_nV_n \xrightarrow{D} N(0, \sigma^2 \lim_{n\to\infty}(1/n)C_n'A_nA_n'C_n)$.

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

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

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

1. If $\{A_n\}$ are uniformly bounded in either row or column sums, then $|Q_n' A_n P_n| = O(n)$.
2. If the row sums of $\{A_n\}$ and $\{Z_n\}$ are uniformly bounded, $|z_{i,n} A_n P_n| = O(1)$ uniformly in i, where $z_{i,n}$ is the ith row of Z_n.

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

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

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

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

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

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

Lee (2004b) - Lemma A.11: Let $A_n = [a_{ij}]$ be an n-dimensional square matrix. Then

1. $E(V_n' A_n V_n) = \sigma^2 tr(A_n)$,

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

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

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

Proof: The result in 1) is trivial. For the second moment,

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

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

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

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

Lee (2004b) - Lemma A.12: Suppose that $\{A_n\}$ are uniformly bounded in either row and column sums, and the elements $a_{n,ij}$ of A_n are $O(\frac{1}{h_n})$ uniformly in all i and j. Then, $E(V_n'A_nV_n) = O(\frac{n}{h_n})$, $var(V_n'A_nV_n) = O(\frac{n}{h_n})$ and $V_n'A_nV_n = O_p(\frac{n}{h_n})$. Furthermore, if $\lim_{n\to\infty}\frac{h_n}{n} = 0$, $\frac{h_n}{n}V_n'A_nV_n - \frac{h_n}{n}E(V_n'A_nV_n) = o_p(1)$.

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

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

B.2.2 Definition in White (1996)

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

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

□

B.2.3 Theorem in White (1996)

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

B.3 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 $W_n(\gamma_0)$.

B.3.1 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,

$$
\begin{aligned}
&\frac{1}{n}(\ln|det(S_n(\lambda_2,\gamma_2))| - \ln|det(S_n(\lambda_1,\gamma_1))|) \\
&= -\frac{1}{n}tr(W_n(\bar{\gamma}_n)S_n^{-1}(\bar{\lambda}_n,\bar{\gamma}_n))[\lambda_2-\lambda_1] - \frac{\bar{\lambda}_n}{n}tr(Z_n(\bar{\gamma}_n)S_n^{-1}(\bar{\lambda}_n,\bar{\gamma}_n))[\gamma_2-\gamma_1] \\
&= -\frac{1}{n}tr(G_n(\bar{\lambda}_n,\bar{\gamma}_n))[\lambda_2-\lambda_1] - \frac{\bar{\lambda}_n}{n}tr(T_n(\bar{\lambda}_n,\bar{\gamma}_n))[\gamma_2-\gamma_1] \qquad \text{(B.3.1)}
\end{aligned}
$$

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 5, then Lemma A.8 in Lee (2004b) implies that $\frac{1}{n}tr(G_n(\bar{\lambda},\bar{\gamma})) = O(\frac{1}{h_n})$. See Appendix B.2 for more detail of Lemmas frequently used in this Appendix. The term $Z_n(\bar{\gamma}_n)$ on the right hand side of (B.3.1), 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 5, then $\frac{\bar{\lambda}_n}{n}tr(T_n(\bar{\lambda},\bar{\gamma})) = O(\frac{1}{h_n})$ as well. Hence, $\frac{1}{n}\ln|det(S_n(\lambda,\gamma))|$ is uniformly equicontinuous in λ and γ in $\Lambda\otimes\Gamma$. Because $\Lambda\otimes\Gamma$ is a compact set, $\frac{1}{n}(\ln|det(S_n(\lambda_2,\gamma_2))| - \ln|det(S_n(\lambda_1,\gamma_1))|) = O(1)$ uniformly in λ_1 and λ_2, and γ_1 and γ_2 in $\Lambda\otimes\Gamma$.

B.3.2 Auxiliary Model $Q_{p,n}(\lambda, \gamma)$

We describe the following auxiliary model as follows

$$Q_{p,n}(\lambda, \gamma) = -\frac{n}{2}(\ln 2\pi + 1) - \frac{n}{2} \ln \sigma_n^2(\lambda, \gamma) + \ln |det(S_n(\lambda, \gamma))| \qquad \text{(B.3.2)}$$

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.

B.3.3 Properties of $\sigma_n^2(\lambda, \gamma)$

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

$$\begin{aligned} \sigma_n^2(\lambda, \gamma) &= \frac{\sigma_0^2}{n} tr(S_n^{-1\prime} S_n'(\lambda, \gamma) S_n(\lambda, \gamma) S_n^{-1}) \\ &= \sigma_0^2[1 + 2(\lambda_0 - \lambda)\frac{1}{n} tr(G_n) + (\lambda_0 - \lambda)^2 \frac{1}{n} tr(G_n G_n')]. \end{aligned} \qquad \text{(B.3.3)}$$

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

$$-\frac{1}{2} \ln \sigma_n^2(\lambda, \gamma) \leq -\frac{1}{2} \ln \sigma_0^2 + \frac{1}{n}(\ln |det(S_n)| - \ln |det(S_n(\lambda, \gamma))|). \qquad \text{(B.3.4)}$$

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 $\Lambda \otimes \Gamma$.

B.3.4 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 (B.3.3) 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 $\hat{\theta}_n$.

B.4 Proof of Theorem 1: Identifiable Uniqueness

We show that

$$\lim_{n\to\infty} \sup[\max_{\theta \in \eta_n^c(\nu)}\{\frac{1}{n}Q_n(\lambda, \gamma) - \frac{1}{n}Q_n\}] < 0 \tag{B.4.1}$$

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.

B.5 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

$$H_{1n}(\lambda,\gamma) = \frac{1}{n}(G_nX_n\beta_0)'M_nS_n(\lambda,\gamma)S_n^{-1}\varepsilon_n \tag{B.5.1}$$

and

$$H_{2n}(\lambda,\gamma) = \frac{1}{n}\varepsilon_n'S_n^{-1\prime}S_n'(\lambda,\gamma)M_nS_n(\lambda,\gamma)S_n^{-1}\varepsilon_n. \tag{B.5.2}$$

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 (B.5.1) as follows

$$H_{1n}(\lambda,\gamma) = \frac{1}{n}(G_nX_n\beta_0)'M_n\varepsilon_n + (\lambda_0-\lambda)\frac{1}{n}(G_nX_n\beta_0)'M_nG_n\varepsilon_n \tag{B.5.3}$$

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 $(\lambda,\gamma)\in\Lambda\otimes\Gamma$. See Appendix B.2 for more detail of Lemmas used in this Appendix. Next,

$$\begin{aligned}H_{2n}(\lambda,\gamma)-\sigma_n^2(\lambda,\gamma) &= \frac{1}{n}\varepsilon_n' S_n^{-1\prime} S_n'(\lambda,\gamma) M_n S_n(\lambda,\gamma) S^{-1}\varepsilon_n - \sigma_n^2(\lambda,\gamma)\\ &= \frac{1}{n}\varepsilon_n' S_n^{-1\prime} S_n'(\lambda,\gamma) S_n(\lambda,\gamma) S^{-1}\varepsilon_n\\ &- \frac{\sigma_0^2}{n} tr(S_n^{-1\prime} S_n'(\lambda,\gamma) S_n(\lambda,\gamma) S_n^{-1}) - H_{3n}(\lambda,\gamma) \quad \text{(B.5.4)}\end{aligned}$$

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

$$\frac{1}{\sqrt{n}} X_n' S_n(\lambda,\gamma) S_n^{-1}\varepsilon_n = \frac{1}{\sqrt{n}} X_n' S_n^{-1}\varepsilon_n - \frac{\lambda}{\sqrt{n}} X_n' G_n \varepsilon_n = O_p(1) \quad \text{(B.5.5)}$$

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),

$$\frac{1}{n}[\varepsilon_n' S_n^{-1\prime} S_n'(\lambda,\gamma) S_n(\lambda,\gamma) S_n^{-1}\varepsilon_n - \sigma_0^2 tr(S_n^{-1\prime} S_n'(\lambda,\gamma) S_n(\lambda,\gamma) S_n^{-1})] = o_p(1) \quad \text{(B.5.6)}$$

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.

B.6 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 $\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) \xrightarrow{p} 0$.

B.6.1 Nonsingularity of Σ_θ

First we show that Σ_θ is nonsingular. Let $\alpha = (\alpha_1,\alpha_2,\alpha_3,\alpha_4)'$ be a column vector of constants such that $\Sigma_\theta\alpha = 0$. Here we need to show that $\alpha = 0$. From the first row block of the linear equation system $\Sigma_\theta\alpha = 0$ based on (3.5.6), we have

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

Consequently,

$$\alpha_1 = -\lim_{n\to\infty}(X_n'X_n)^{-1}X_n'(G_nX_n\beta_0)\alpha_2 - \lim_{n\to\infty}\lambda_0(X_n'X_n)^{-1}X_n'(T_nX_n\beta_0)\alpha_3 \tag{B.6.1}$$

From the fourth equation of the linear system, we have

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

Rearrange the terms and solve for α_4, we get

$$\alpha_4 = -\lim_{n\to\infty}\frac{2\sigma_0^2}{n}tr(G_n)\alpha_2 - \lim_{n\to\infty}\frac{2\lambda_0\sigma_0^2}{n}tr(T_n)\alpha_3 \tag{B.6.2}$$

From the second equation of the linear system, we have

$$0 = \lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'X_n\alpha_1 + \lim_{n\to\infty}[\frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'(G_nX_n\beta_0) + \frac{1}{n}tr(G_n^SG_n)]\alpha_2$$
$$+ \lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}[(G_nX_n\beta_0)'(T_nX_n\beta_0) + \sigma_0^2 tr(G_n^ST_n)]\alpha_3 + \lim_{n\to\infty}\frac{1}{\sigma_0^2 n}tr(G_n)\alpha_4$$

for $\lambda_0 \neq 0$ and $G_n^S = G_n + G_n'$. Substitute α_1 in (B.6.1) and α_4 in (B.6.2) into the above equation, we get

$$0 = \Big[\lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty}\frac{1}{n}[tr(G_n^SG_n) - \frac{2}{n}tr^2(G_n)]\Big]\alpha_2$$
$$+ \Big[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty}\frac{\lambda_0}{n}[tr(G_n^ST_n) - \frac{2}{n}tr(G_n)tr(T_n)]\Big]\alpha_3 \tag{B.6.3}$$

Rearrange the terms and solve for α_2

$$\alpha_2 = -\Big[\lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty}\frac{1}{n}[tr(G_n^SG_n) - \frac{2}{n}tr^2(G_n)]\Big]^{-1}$$
$$\times \Big[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty}\frac{\lambda_0}{n}[tr(G_n^ST_n) - \frac{2}{n}tr(G_n)tr(T_n)]\Big]\alpha_3 \tag{B.6.4}$$

Note that the inverse in equation (B.6.4) above exists as Assumption 10 implies that $\lim_{n\to\infty}\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ is positive and that $[tr(G_n^SG_n) - \frac{2}{n}tr^2(G_n)] = \frac{1}{2}tr[(\Psi_n' + \Psi_n)(\Psi_n' + \Psi_n)'] \geq 0$, where $\Psi_n = G_n - (tr(G_n)/n)I_n$ (Lee, 2004a).

From the third equation of the linear system, we have

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

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

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

Rearrange the terms and solve for α_2

$$\begin{aligned}\alpha_2 = &-\Big[\lim_{n\to\infty}\frac{\lambda_0}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0)+\lim_{n\to\infty}\frac{\lambda_0}{n}[tr(T_n^SG_n)-\frac{2}{n}tr(T_n)tr(G_n)]\Big]^{-1}\\ &\times\Big[\lim_{n\to\infty}\frac{\lambda_0^2}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0)+\lim_{n\to\infty}\frac{\lambda_0^2}{n}[tr(T_n^ST_n)-\frac{2}{n}tr^2(T_n)]\Big]\alpha_3\end{aligned} \tag{B.6.5}$$

The inverse in equation (B.6.5) above exists for $\lambda_0 \neq 0$ as Assumption 10 implies that $\lim_{n\to\infty}\frac{1}{n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ exists. If this limit is positive, then the sum of the terms in the inverse will exist and be positive whereas if this limit is negative, then the sum of the terms in the inverse will exist and be either negative or positive. Finally, combine equations (B.6.4) with (B.6.5), we get

$$\begin{aligned}0 = &\Big\{\Big[\lim_{n\to\infty}\frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)+\lim_{n\to\infty}\frac{1}{n}[tr(G_n^SG_n)-\frac{2}{n}tr^2(G_n)]\Big]^{-1}\\ &\times\Big[\lim_{n\to\infty}\frac{\lambda_0}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(T_nX_n\beta_0)+\lim_{n\to\infty}\frac{\lambda_0}{n}[tr(G_n^ST_n)-\frac{2}{n}tr(G_n)tr(T_n)]\Big]\\ &-\Big[\lim_{n\to\infty}\frac{\lambda_0}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0)+\lim_{n\to\infty}\frac{\lambda_0}{n}[tr(T_n^SG_n)-\frac{2}{n}tr(T_n)tr(G_n)]\Big]^{-1}\\ &\times\Big[\lim_{n\to\infty}\frac{\lambda_0^2}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0)+\lim_{n\to\infty}\frac{\lambda_0^2}{n}[tr(T_n^ST_n)-\frac{2}{n}tr^2(T_n)]\Big]\Big\}\alpha_3\end{aligned} \tag{B.6.6}$$

We show that the products of the above equation are nonzero. First of all, Assumption 10 implies that $\lim_{n\to\infty}\frac{1}{n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ exists, and $\lim_{n\to\infty}\frac{1}{n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0)$ and $\lim_{n\to\infty}\frac{1}{n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0)$ are positive. As stated earlier, because $[tr(G_n^SG_n)-\frac{2}{n}tr^2(G_n)]=\frac{1}{2}tr[(\Psi_n'+\Psi_n)(\Psi_n'+\Psi_n)'] \geq 0$, then the first and fourth lines of equation (B.6.7) above are positive while the second and third lines exist and can be either positive or negative.

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

$$
\begin{aligned}
0 =& \Bigg\{ \Big[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(G_n^ST_n) - \frac{2}{n}tr(G_n)tr(T_n)] \Big] \\
& \times \Big[\lim_{n\to\infty} \frac{\lambda_0}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0}{n}[tr(T_n^SG_n) - \frac{2}{n}tr(T_n)tr(G_n)] \Big] \\
& - \Big[\lim_{n\to\infty} \frac{1}{\sigma_0^2 n}(G_nX_n\beta_0)'M_n(G_nX_n\beta_0) + \lim_{n\to\infty} \frac{1}{n}[tr(G_n^SG_n) - \frac{2}{n}tr^2(G_n)] \Big] \\
& \times \Big[\lim_{n\to\infty} \frac{\lambda_0^2}{\sigma_0^2 n}(T_nX_n\beta_0)'M_n(T_nX_n\beta_0) + \lim_{n\to\infty} \frac{\lambda_0^2}{n}[tr(T_n^ST_n) - \frac{2}{n}tr^2(T_n)] \Big] \Bigg\} \alpha_3
\end{aligned}
\tag{B.6.7}
$$

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

B.6.2 $\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\theta\partial\theta'} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} 0$

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.

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2}X_n'X_n, \tag{B.6.8}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\lambda} = -\frac{1}{\sigma^2}X_n'W_n(\gamma)Y_n, \tag{B.6.9}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\gamma} = -\frac{\lambda}{\sigma^2}X_n'Z_n(\gamma)Y_n, \tag{B.6.10}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}X_n'\varepsilon_n(\delta), \tag{B.6.11}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda^2} = -tr(G_n^2(\lambda,\gamma)) - \frac{1}{\sigma^2}Y_n'W_n'(\gamma)W_n(\gamma)Y_n, \tag{B.6.12}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\gamma} = - tr(T_n(\lambda,\gamma)) - \lambda tr(G_n(\lambda,\gamma)T_n(\lambda,\gamma)) - \frac{\lambda}{\sigma^2}Y_n' Z_n'(\gamma)W_n(\gamma)Y_n, \tag{B.6.13}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\sigma^2} = -\frac{1}{\sigma^4}Y_n' W_n'(\gamma)\varepsilon_n(\delta), \tag{B.6.14}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\gamma^2} = - \lambda tr(C_n(\lambda,\gamma)) - \lambda^2 tr(T_n^2(\lambda,\gamma)) - \frac{\lambda^2}{\sigma^2}Y_n' Z_n'(\gamma)Z_n(\gamma)Y_n, \tag{B.6.15}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\gamma\partial\sigma^2} = -\frac{\lambda}{\sigma^4}Y_n' Z_n'(\gamma)\varepsilon_n(\delta), \tag{B.6.16}$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial(\sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\varepsilon_n'(\delta)\varepsilon_n(\delta) \tag{B.6.17}$$

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 (B.6.8) 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 (B.6.9) at $\hat{\theta}_n$ and at θ_0 is

$$\frac{1}{n}\frac{\partial^2 \ln L_n(\hat{\theta}_n)}{\partial\beta\partial\lambda} - \frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\beta\partial\lambda} = \frac{1}{\sigma_0^2}\frac{X_n'W_nY_n}{n} - \frac{1}{\hat{\sigma}_n^2}\frac{X_n'W_n(\hat{\gamma}_n)Y_n}{n} \tag{B.6.18}$$

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

$$\Big|\Big|\frac{X_n'W_n(\hat{\gamma}_n)Y_n}{n} - \frac{X_n'W_nY_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{X_n'Z_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| \tag{B.6.19}$$

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, (B.6.18) 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 (B.6.10), we first show that

$$\Big|\Big|\frac{X_n' Z_n(\hat{\gamma}_n) Y_n}{n} - \frac{X_n' Z_n Y_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{X_n' A_n(\bar{\gamma}_n) Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1) \quad \text{(B.6.20)}$$

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,

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

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 (B.6.11), we first look at the following equation.

$$\varepsilon_n(\delta_n) = Y_n - X_n\beta_n - \lambda_n W_n(\gamma_n) Y_n = X_n(\beta_0 - \beta_n) + [\lambda_0 W_n - \lambda_n W_n(\gamma_n)]Y_n + \varepsilon_n,$$

where $\delta_n = (\beta_n', \lambda_n, \gamma_n)'$. Substitute this equation into (B.6.11) and as we have shown in (B.6.19) that $\frac{1}{n}X_n' W_n(\hat{\gamma}_n) Y_n \xrightarrow{p} \frac{1}{n}X_n' W_n Y_n$, the difference of (B.6.11) evaluated at $\hat{\theta}_n$ and θ_0 becomes

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

for $\hat{\theta}_n \xrightarrow{p} \theta_0$. Next, for (B.6.14), we have

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

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

$$\Big|\Big|\frac{Y_n' W_n'(\hat{\gamma}_n)\varepsilon_n}{n} - \frac{Y_n' W_n'\varepsilon_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n' Z_n'(\bar{\gamma}_n)\varepsilon_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1). \quad \text{(B.6.21)}$$

For $\hat{\gamma}_n \xrightarrow{p} \gamma_0$ and $\bar{\gamma}_n$ lies between $\hat{\gamma}_n$ and γ_0, (B.6.21) above implies that $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)\varepsilon_n \xrightarrow{p} \frac{1}{n}Y_n'W_n'\varepsilon_n$. Next, we show that $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n$ and $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)W_nY_n$ converge in probability to $\frac{1}{n}Y_n'W_n'W_nY_n$. Apply the mean value theorem for vector-valued function, we have

$$\begin{aligned}
&\Big|\Big|\frac{Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n} - \frac{Y_n'W_n'W_nY_n}{n}\Big|\Big| \\
&\leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'Z_n'(\bar{\gamma}_n)W_n(\bar{\gamma}_n)Y_n}{n} + \frac{Y_n'W_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1)
\end{aligned} \tag{B.6.22}$$

and

$$\Big|\Big|\frac{Y_n'W_n'(\hat{\gamma}_n)W_nY_n}{n} - \frac{Y_n'W_n'W_nY_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'Z_n'(\bar{\gamma}_n)W_nY_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1), \tag{B.6.23}$$

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 (B.6.19) and $\hat{\theta}_n \xrightarrow{p} \theta_0$, the difference of (B.6.14) evaluated at $\hat{\theta}_n$ and θ_0 becomes

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

For (B.6.16), the convergence is as follows

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

The same intuition as in (B.6.14) above applies here as well. By the mean value theorem, we first show that $\frac{1}{n}Y_n'Z_n'(\hat{\gamma}_n)\varepsilon_n \xrightarrow{p} \frac{1}{n}Y_n'Z_n'\varepsilon_n$.

$$\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)\varepsilon_n}{n} - \frac{Y_n'Z_n'\varepsilon_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)\varepsilon_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1). \tag{B.6.24}$$

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,

$$\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)W_nY_n}{n} - \frac{Y_n'Z_n'W_nY_n}{n}\Big|\Big| \leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)W_nY_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1) \tag{B.6.25}$$

and

$$\begin{aligned}&\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n}{n} - \frac{Y_n'Z_n'W_nY_n}{n}\Big|\Big| \qquad\qquad (B.6.26)\\ &\leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)W_n(\bar{\gamma}_n)Y_n}{n} + \frac{Y_n'Z_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n - \gamma_0| = o_p(1),\end{aligned}$$

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 (B.6.24) - (B.6.26) and (B.6.20), the convergence of (B.6.16) 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 (B.6.12), (B.6.13) and (B.6.15), 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 (B.6.12) 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 (B.6.22), $\frac{1}{n}Y_n'W_n'(\hat{\gamma}_n)W_n(\hat{\gamma}_n)Y_n \xrightarrow{p} \frac{1}{n}Y_n'W_n'W_nY_n$. 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,

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

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 (B.6.12) 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 (B.6.13), the same technique applies. By mean value theorem,

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

and

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

Hence, the convergence of (B.6.13) becomes

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

Since $S_n^{-1}(\lambda, \gamma)$ is uniformly bounded in row and column sums uniformly in a neighbourhood of λ_0 and γ_0, then $tr(C_n(\bar{\lambda}_n, \bar{\gamma}_n)) = O(\frac{n}{h_n})$ by Lemma A.8 in Lee (2004b). Note that as $\hat{\lambda}_n \xrightarrow{p} \lambda_0$ and $\hat{\gamma}_n \xrightarrow{p} \gamma_0$, therefore, the trace terms become $o_p(1)$. As we have already shown in (B.6.26) that $\frac{1}{n}Y_n' Z_n'(\hat{\gamma}_n) W_n(\hat{\gamma}_n) Y_n \xrightarrow{p} Y_n' Z_n' W_n Y_n$, then

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

Next, for equation (B.6.15), apply the mean value theorem to the traces as follows.

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

and

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

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 (B.6.15) evaluated at $\hat{\theta}_n$ and θ_0 is

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

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

$$\begin{aligned}
&\Big|\Big|\frac{Y_n'Z_n'(\hat{\gamma}_n)Z_n(\hat{\gamma}_n)Y_n}{n}-\frac{Y_n'Z_n'Z_nY_n}{n}\Big|\Big| \qquad\qquad (B.6.27)\\
&\leq \sup_{\gamma\in\Gamma}\Big|\Big|\frac{Y_n'A_n'(\bar{\gamma}_n)Z_n(\bar{\gamma}_n)Y_n}{n}+\frac{Y_n'Z_n'(\bar{\gamma}_n)A_n(\bar{\gamma}_n)Y_n}{n}\Big|\Big||\hat{\gamma}_n-\gamma_0|=o_p(1)
\end{aligned}$$

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 (B.6.15) 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 (B.6.17), 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 (B.6.19) and (B.6.21) - (B.6.23), 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 (B.6.17) 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 $\Lambda \otimes \Gamma$.

B.6.3 $\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E(\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) \xrightarrow{p} 0$

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

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

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

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

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

Finally,

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

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

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

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

Appendix C

Appendix to Chapter 4

C.1 List of Countries

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

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

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 C.1: List of 91 countries and their isocodes.

Bibliography

[1] Abreu, M., De Groot, H.L.F. and Florax, R.J.G.M. (2005). Space and Growth: A Survey of Empirical Evidence and Methods. *Région et Développement*, 21, 13-44.

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

[3] Amaral, P.V. and Anselin, L. (2013). Finite sample properties of Moran's I test for spatial autocorrelation in tobit models. *Papers in Regional Science*, doi: 10.1111/pirs.12034.

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

[5] Andrews, D.W.K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62(6), 1383-1414.

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

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

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

[9] Anselin, L. (1988b). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, 20, 1-17.

[10] Anselin, L. (2001). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inference*, 97, 113-139.

[11] Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1), 1-26.

[12] Anselin, L., Bera, A.K., Florax, R. and Yoon, M.J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26, 77-104.

[13] 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.

[14] Anselin, L. and Kelejian, H.H. (1997). Testing for Spatial Error Autocorrelation in the Presence of Endogenous Regressions. *International Regional Science Review*, 20 (1 & 2), 153-182.

[15] Anselin, L. and Rey, S.J. (1991). Properties of Tests for Spatial Dependence in Linear Regression Models. *Geographical Analysis*, 23(2), 112-131.

[16] Anselin, L. and Rey, S.J. (1997). Introduction to the Special Issue on Spatial Econometrics. *International Regional Science Review*, 20(1 & 2), 1-7.

[17] Anselin, L., Varga, A. and Acs, Z. (1997). Local Geographic Spillovers between University Research and High Technology Innovations. *Journal of Urban Economics*, 42, 422-448.

[18] Anselin, L., Varga, A. and Acs, Z. (2000). Geographical Spillovers and University Research: A Spatial Econometric Perspective. *Growth and Change*, 31, 501-515.

[19] Arnold, M. and Wied, D. (2010a). Improved GMM Estimation of the Spatial Autoregressive Error Model. *Economics Letters*, 108(1), 65-68.

[20] Arnold, M. and Wied, D. (2010b). Separate estimation of spatial dependence parameters and variance parameters in a spatial model. mimeo.

[21] Arraiz, I., Drukker, D.M., Kelejian, H.H. and Prucha, I.R. (2010). A Spatial Cliff-Ord-type Model with Heteroskedastic Innovations: Small and Large Sample Results. *Journal of Regional Science*, 50(2), 592-614.

[22] Badinger, H. and Egger, P. (2011). Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances. *Papers in Regional Science*, 90(1), 213-235.

[23] Baltagi, B.H., Heun Song, S. and Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117, 123-150.

[24] Baltagi, B.H., Egger, P. and Pfaffermayr, M. (2007). Estimating models of complex FDI: Are there third-country effects?. *Journal of Econometrics*, 140(1), 260-281.

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

[26] Beck, N., Gleditsch, K.S. and Beardsley, K. (2006). Space Is More than Geography: Using Spatial Econometrics in the Study of Political Economy. *International Studies Quarterly*, 50, 27-44.

[27] Bell, K.P. and Bockstael, N.E. (2000). Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data. *The Review of Economics and Statistics*, 82(1), 72-82.

[28] Bera, A. and Yoon, M. (1993). Specification testing with locally misspecified alternatives. *Econometric Theory*, 9, 649-658.

[29] Berndt, E.R. and Savin, N.E. (1977). Conflict among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model. *Econometrica*, 45(5), 1263-1277.

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

[31] Bivand, R. (2010). Computing the Jacobian in Spatial Models: an Applied Survey. *NHH Dept. of Economics Discussion Paper No. 20/2010.*

[32] 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.

[33] Bode, E. and Rey, S.J. (2006). The spatial dimension of economic growth and convergence. *Papers in Regional Science*, 85(2), 171-176.

[34] Bolduc, D., Laferriere, R. and Santarossa, G. (1992). Spatial autoregressive error components in travel flow models. *Regional Science and Urban Economics*, 22, 371-385.

[35] Bramoulle, Y., Djebbari, H. and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150, 41-55.

[36] Brett, C. and Pinkse, J. (1997). Those taxes are all over the map! A test for spatial independence of municipal tax rates in British Columbia. *International Regional Science Review*, 20(1 & 2), 131-151.

[37] Breusch, T.S. and Pagan, A.R. (1980). The Lagrange Multiplier test and its application to model specification in econometrics. *Review of Economic Studies*, 47, 239-254.

[38] Burridge, P. (1980). On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society B*, 42, 107-108.

[39] Can, A. (1990). The measurement of neighborhood dynamics in urban house prices. *Economic Geography*, 66(3), 254-272.

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

[41] Case, A.C. (1991). Spatial patterns in household demand. *Econometrica*, 59(4), 953-965.

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

[43] Cliff, A. and Ord, J.K. (1972). Testing for Spatial Autocorrelation Among Regression Residuals. *Geographical Analysis*, 4, 267-284.

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

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

[46] Coughlin, C. and Segev, E. (2000). Foreign direct investment in China: a spatial econometric study. *The World Economy*, 23(1), 1-23.

[47] Dall'erba, S. and Le Gallo, J. (2008). Regional convergence and the impact of European structural funds over 1989-1999: A spatial econometric analysis. *Papers in Regional Science*, 87(2), 219-244.

[48] Das, D., Kelejian, H.H. and Prucha, I.R. (2003). Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science*, 82, 1-26.

[49] Davies, R.B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64(2), 247-254.

[50] Davies, R.B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1), 33-43.

[51] Debarsy, N. and Ertur, C. (2010). Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics*, 40, 453-470.

[52] Drukker, D.M., Egger, P. and Prucha, I.R. (2011). On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors. mimeo.

[53] Dubin, R.A. (1988). Estimation of Regression Coefficients in the Presence of Spatially Autocorrelated Error Terms. *The Review of Economics and Statistics*, 70(3), 466-474.

[54] Elhorst, J.P. (2003). Specification and Estimation of Spatial Panel Data Models. *International Regional Science Review*, 26(3), 244-268.

[55] Elhorst, P. (2011). Spatial panel models. mimeo

[56] Engle, R.F. (1984). Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics. *Handbook of Econometrics*, volume 2, edited by

Griliches, Z. and Intriligator, M.D. Elsevier Science Publishers BV, 775-826.

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

[58] Fingleton, B. (2001a). Equilibrium and Economic Growth: Spatial Econometric Models and Simulations. *Journal of Regional Science*, 41(1), 117-147.

[59] Fingleton, B. (2001b). Theoretical economic geography and spatial econometrics: dynamic perspectives. *Journal of Economic Geography*, 1(2), 201-225.

[60] Fingleton, B. (2004). Regional Economic Growth and Convergence: Insights from a Spatial Econometric Perspective. *Advances in Spatial Econometrics, Advances in Spatial Science*, edited by Anselin, L., Florax, R.J.G.M. and Rey, S.J. Springer Berlin Heidelberg, 397-432.

[61] Fingleton, B. and Le Gallo, J. (2008a). Finite sample properties of estimators of spatial models with autoregressive, or moving average disturbances and system feedback. *Annals of Economics and Statistics*, 87/88, 39-62.

[62] Fingleton, B. and Le Gallo, J. (2008b). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. *Papers in Regional Science*, 87, 319-339.

[63] Fingleton, B. and McCombie, J.S.L. (1998). Increasing returns and economic growth: some evidence for manufacturing from the European Union regions. *Oxford Economic Papers*, 50(1), 89-105.

[64] Geary, R. (1954). The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5, 115-145.

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

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

[67] Hansen, B.E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2), 413-430.

[68] He, M. (2011). Locally adjusted LM test for spatial autocorrelation in panel data model with fixed effects. mimeo.

[69] Henry, M.S., Schmitt, B. and Piguet, V. (2001). Spatial Econometric Models for Simultaneous Systems: Application to Rural Community Growth in France. *International Regional Science Review*, 24(2), 171-193.

[70] 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.

[71] Hillier, G. and Martellosio, F. (2012). Exact Properties of the Maximum Likelihood Estimator in Spatial Autoregressive Models. mimeo.

[72] Holly, S., Pesaran, M.H. and Yamagata, T. (2011). The spatial and temporal diffusion of house prices in the UK. *Journal of Urban Economics*, 69(1), 2-23.

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

[74] Kapoor, M., Kelejian, H.H. and Prucha, I.R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, 140, 97-130.

[75] 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.

[76] Kelejian, H.H. and Prucha, I.R. (1998). A Generalised Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics*, 17, 99-121.

[77] 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.

[78] 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.

[79] Kelejian, H.H. and Prucha, I.R. (2002). 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics*, 32, 691-707.

[80] Kelejian, H.H. and Prucha, I.R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140, 131-154.

[81] Kelejian, H.H. and Prucha, I.R. (2010). Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances. *Journal of Econometrics*, 157(1), 53-67.

[82] Kelejian, H.H. and Robinson, D.P. (1992). Spatial autocorrelation: A new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics*, 22, 317-331.

[83] Kelejian, H.H. and Robinson, D.P. (1993). A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an

application to a county expenditure model. *Papers in Regional Science*, 72(3), 297-312.

[84] Le Gallo, J. (2004). Space-Time Analysis of GDP Disparities Among European Regions: A Markov Chains Approach. *International Regional Science Review*, 27(2), 138-163.

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

[86] 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.

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

[88] 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/*.

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

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

[91] Lee, L.F. and Liu, X. (2010). Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances. *Econometric Theory*, 26(1), 187-230.

[92] Lee, L.F., Liu, X. and Lin, X. (2010). Specification and estimation of social interaction models with network structures. *Econometrics Journal*, 13, 145-176.

[93] Lee, L.F. and Yu, J. (2010). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, 40, 255-271.

[94] Leenders, R.TH.A.J. (2002). Modeling social influence through network autocorrelation: constructing the weight matrix. *Social Networks*, 24, 21-47.

[95] LeSage, J.P. (1997). Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review*, 20(1 & 2), 113-129.

[96] LeSage, J.P. and Fischer, M.M. (2008). Spatial Growth Regressions: Model Specification, Estimation and Interpretation. *Spatial Economic Analysis*, 3(3), 275-304.

[97] LeSage, J.P. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. CRC Press.

[98] LeSage, J.P. and Parent, O. (2007). Bayesian Model Averaging for Spatial Econometric Models. *Geographical Analysis*, 39, 241-267.

[99] Lin, X. and Lee, L.F. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157, 34-52.

[100] Liu, X., Lee, L.F. and Bollinger, C.R. (2006). Improved Efficient Quasi Maximum Likelihood Estimator of Spatial Autoregressive Models. mimeo.

[101] Lundberg, J. (2006). Using Spatial Econometrics to Analyse Local Growth in Sweden. *Regional Studies*, 40(3), 303-316.

[102] Madariaga, N. and Poncet, S. (2007). FDI in Chinese Cities: Spillovers and Impact on Growth. *The World Economy*, 30(5), 837-862.

[103] 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.

[104] Martellosio, F. (2012). Testing for Spatial Autocorrelation: the Regressors that Make the Power Disappear. *Econometric Reviews*, 31(2), 215-240.

[105] Moran, P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society B*, 10, 243-251.

[106] Moreno, R., Paci, R. and Usai, S. (2003). Spatial spillovers and innovation activity in European regions. *CRENoS Working Paper*, 2003/10.

[107] Mur, J., Herrera, M. and Ruiz, M. (2012). Selecting the Most Adequate Spatial Weighting Matrix: A Study on Criteria. mimeo.

[108] Newey, W.K. and West, K.D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28(3), 777-787.

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

[110] Overmars, K.P., De Koning, G.H.J. and Veldkamp, A. (2003). Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 164, 257-270.

[111] Paelinck, J.H.P. and Klaassen, L.H. (1979). *Spatial Econometrics*. Farnborough: Saxon House.

[112] 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.

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

[114] Pinkse, J., Slade, M.E. and Brett, C. (2002). Spatial price competition: A semiparametric approach. *Econometrica*, 70(3), 1111-1153.

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

[116] Plumper, T. and Neumayer, E. (2010). Model specification in the analysis of spatial dependence. *European Journal of Political Research*, 49, 418-442.

[117] Rey, S.J. and Montouri, B.D. (1999). US Regional Income Convergence: A Spatial Econometric Perspective. *Regional Studies*, 33(2), 143-156.

[118] Saavedra, L.A. (2003). Tests for spatial lag dependence based on method of moments estimation. *Regional Science and Urban Economics*, 33(1), 27-58.

[119] Silvey, S.D. (1959). The Lagrangian multiplier test. *Ann. Math. Statist.*, 30, 389-407.

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

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

[122] Ying, L.G. (2003). Understanding China's recent growth experience: A spatial econometric perspective. *The Annals of Regional Science*, 37, 613-628.