5-2016

On Refined and Robust Inferences for Spatial Econometric Models

Shew Fan LIU
Singapore Management University, sfliu@smu.edu.sg

On Refined and Robust Inferences for Spatial Econometric Models

SHEW FAN LIU

SINGAPORE MANAGEMENT UNIVERSITY

2016

On Refined and Robust Inferences for Spatial Econometric Models

by
Shew Fan Liu

Submitted to School of Economics in partial fulfilment of the
requirements for the Degree of Doctor of Philosophy in Economics

<u>Dissertation Committee:</u>

Zhenlin Yang (Supervisor/Chair)
Professor of Economics and Statistics
Singapore Management University

Yiu Kuen Tse
Professor of Economics
Singapore Management University

Anthony Tay
Associate Professor of Economics
Singapore Management University

Qu Feng
Assistant Professor of Economics
Nanyang Technological University, Singapore

Singapore Management University
2016

Abstract

On Refined and Robust Inferences for Spatial Econometric Models

Shew Fan Liu

Asymptotically refined and heteroskedasticity robust inferences are considered for spatial linear and panel regression models, based on the *quasi maximum likelihood* (QML) or the *adjusted concentrated quasi score* (ACQS) approaches. Refined inferences are achieved through bias correcting the QML estimators, bias correcting the t-ratios for covariate effects, and improving tests for spatial effects; heteroskedasticity-robust inferences are achieved through adjusting the quasi score functions. Several popular spatial linear and panel regression models are considered including the linear regression models with either spatial error dependence (SED), or spatial lag dependence (SLD), or both SED and SLD (SARAR), the linear regression models with higher-order spatial effects, SARAR(p, q), and the fixed-effects panel data models with SED or SLD or both. Asymptotic properties of the new estimators and the new inferential statistics are examined. Extensive Monte Carlo experiments are run, and the results show that the proposed methodologies work really well.

Contents

Acknowledgements

No amount of words are enough to display my gratitude towards my mentor, Professor Zhenlin Yang who is the inspirational force behind this dissertation. It has been a great educational journey and I am honoured to have followed in your footsteps.

I will fail in my duty if I don't acknowledge the love and support of my family - my parents and my husband. Your patience and understanding gave me the courage to persevere and continue even when things seem almost impossible.

Dedication

To Jordan... The one who will always be my biggest contribution to humanity!

CHAPTER 1

Introduction

Spatial dependence is increasingly becoming an integral part of empirical works in economics as a means of modelling the effects of neighbours[1]. As in time series where the concern is to alleviate the estimation problems caused by the lag in time, the analogous case in cross sectional data gives rise to a lag in space. Spatial interaction in general can occur in many forms. For instance peer interaction can cause stratified behaviour in the sample such as herd behaviour in stock markets, innovation spillover effects, localised purchase decisions, etc., while spatial relationships can also occur more naturally due to structural differences in space/cross-section such as geographic proximity, trade agreements, demographic characteristics, etc.[2] Modelling cross relationships over the fabric of space have been prevalent mostly in the empirical literature. Some theoretical

[1] see, e.g., Cliff and Ord (1972, 1973, 1981), Ord (1975), Anselin (1988, 2003), Anselin and Bera (1998), Le Sage and Pace (2009) for some early and comprehensive works

[2] See Case (1991), Pinkse and Slade (1998), Pinkse et al. (2002), Hanushek et al. (2003), Baltagi et al. (2007) to name a few.

results appeared as early as mid 1970's, however, much of the theoretical gaps remained unfilled until the turn of the century. The objective of my dissertation is to contribute in terms of this theoretical effort.

Linear regression models of spatial dependence take the following general functional form:

$$f\left(Y_n, X_n, W_{1n}, \ldots, W_{pn}; \theta, \lambda\right) = \epsilon_n, \tag{1.1}$$

with a dependent variable Y_n conditional on a set of independent variables X_n and spatial weight matrices $W_{1n}, \ldots, W_{pn}$ that capture the relationships among the n spatial units. Parameter vector θ denotes the parameters of the model and $\lambda = \left(\lambda_1, \ldots, \lambda_p\right)'$ denotes the spatial parameters. ϵ_n is an $n \times 1$ vector of model errors. Popular spatial regression models can be written in this form. For example, the spatial autoregressive lagged dependent variable (SLD) model, $Y_n = \lambda W_n Y_n + X_n\beta + \epsilon_n$ can be written in the form, $C_{1n}\left(\lambda\right) Y_n - X_n\beta = \epsilon_n$ where $C_{1n}\left(\lambda\right) = I_n - \lambda W_n$ and I_n is an $n \times n$ identity matrix. The spatial error dependent (SED) model, $Y_n = X_n\beta + u_n$, $u_n = \rho W_n u_n + \epsilon_n$ can be written in the form, $C_{2n}\left(\rho\right)\left[Y_n - X_n\beta\right] = \epsilon_n$ where $C_{2n}\left(\rho\right) = I_n - \rho W_n$. These two models are the building blocks for spatial econometric modelling, and many more general spatial econometric models have been developed based on them.

Of the methods available for spatial model estimation, the maximum likelihood (ML) or quasi-ML (QML) method remains attractive due to its efficiency. As a result of the fast increase in computing power allowing for easier manipulation of large matrices, the initial reluctance for the use of QML estimation as opposed to other easily implementable estimation methods alleviated. As such there had been a growing interest in developing the theoretical aspects behind QML estimation. The present study[3] focuses on the following aspects of QML estimation: (i) asymptotic distribution of the QML estimators of popular spatial

[3] Liu and Yang (2015a,b,c), Yang et al. (2016)

regression models (*ii*) finite-sample bias of the QML estimators (*iii*) refined tests for spatial dependence and covariate effects based on QML estimates, and (*iv*) heteroskedasticity robust QML estimation. Although, some of these aspects are already considered in the literature[4], it is far from complete in the sense most of these works concentrate on particular types of spatial models and hence the issues need to be analysed in relation to other important spatial models that have been ignored so far.

A unique feature of spatial models is that the spatial parameters λ enter the reduced form model and hence the log-likelihood function, in a highly non-linear manner and the spatial dependence maybe strong. As a result, the bias problem in estimating spatial parameters may be quite severe.[5] As such bias correction is particularly important to the applications of this model as it leads potentially to much improved inferences for the regression coefficients.

In studying the finite sample properties of a parameter estimator, say ϑ, $\hat{\vartheta}_n = \arg\{\psi_n(\vartheta) = 0\}$ for a *joint estimating function* (JEF) $\psi_n(\vartheta)$, Rilstone et al. (1996) developed a stochastic expansion from which a bias correction on $\hat{\vartheta}_n$ can be made. In the general spatial econometric model given in (1.1) the vector of parameters contain a set of linear and scale parameters, θ, and a non-linear spatial parameter, λ. Given λ, the constrained estimator $\hat{\theta}(\lambda)$ of θ possesses an explicit expression but the estimation of λ has to be done through numerical optimisation. In such situations, Yang (2015b) argued that it is more effective to work with the *concentrated estimating function* (CEF): $\tilde{\psi}_n(\lambda)$, and to perform a stochastic expansion based on this CEF and hence bias corrections on $\hat{\lambda}_n = \arg\{\tilde{\psi}_n(\lambda) = 0\}$.[6]

In the literature, the SLD model has been extensively studied in terms of the

[4]See Baltagi and Yang (2013), Lee (2004), Lee and Yu (2010), Lin and Lee (2010), Yang (2015b) among others.

[5]Lee (2004) shows the consistency of the QML estimator.

[6]Doing so reduces the dimensionality of the bias correction problem, and also takes into account the additional variability from the estimation of the 'nuisance' parameters, θ.

asymptotic distributions of the QML estimators (Lee, 2004); finite-sample bias corrections on QML estimators (Bao and Ullah, 2007; Yang, 2015b). An interesting phenomenon revealed by Lee (2004) for the SLD model is that the spatial dependence may slow down the rate of convergence of QML estimators of certain model parameters, including the spatial parameter. Subsequent studies also revealed that spatial dependence may cause QML estimators to be biased, and more so with heavier spatial dependence (Baltagi and Yang, 2013; Yang, 2015b). These aspects have not been studied in the context of the SED model. Built upon the works of Lee (2004) and Yang (2015b), this dissertation fills in these gaps. Contrary to the common perceptions, both large and small sample behaviours of the QML estimators for the SED model can be different from those for the SLD (Lee, 2004) model in terms of the rate of convergence and the magnitude of bias. We also derive the second- and third-order biases of the QML estimators of the spatial parameter in the SED model. The key quantities involved in the terms related to the bias of a non-linear estimator are the derivatives of the concentrated log-likelihood function and their expectations. While deriving the analytical solutions of the higher order derivatives may only be a matter of tedious algebraic manipulations, evaluation of their expectations can be very difficult if not impossible. We follow the general method introduced in Yang (2015b) and propose a bootstrap procedure for implementing these bias corrections for the SED model. The method is simple to implement, since no re-estimation of the model parameters is required in every bootstrap iteration. The validity of this procedure when applied to the SED model is established. We argue that once the spatial estimator is bias corrected, the estimators of the other models parameters become nearly unbiased. Monte Carlo results show an excellent performance of the proposed bias correction procedure.

Much effort has been devoted recently to the development of improved inference methods for the spatial econometric models. However, most of the research has been focused on improving inferences for spatial effects in the form of point estimation and testing. Little or no attention has been paid to the development of improved inferences for the covariate effects in the spatial regression models. In practical applications of spatial econometric models, it is central to have a set of reliable inference methods for the covariate effects. As QML estimator of the spatial parameters can be quite biased and hence the standard inferences for spatial effects and covariate effects, based on LM-statistics or $t-$statistics referring to the asymptotic standard normal distribution, can be seriously affected. We adopt the bias correction method of Yang (2015b) to propose methods that 'correct' the standard $t-$statistics for the regression coefficients. Once the biases of non-linear estimators are corrected, the biases of covariate effects and error standard deviations become negligible. We consider in detail three popular spatial regression models: SED model, SLD model, and that with both SLD and SED, also referred to as the SARAR model.[7]

The QML estimators of the spatial panel data models (SPD) are subjected to the same issues on the finite sample bias and finite sample performance of subsequent inferences, but these important issues have not been addressed so far. We focus on the SPD models with fixed effects to provide methods for bias and variance corrections (up to third-order), and then to show how the bias and variance corrections lead to improved $t-$ratios for spatial and covariate effects.[8] While the general stochastic expansions of Yang (2015b) for non-linear estimators are applicable to different models including the SPD models considered in this chapter, the detailed developments of bias corrections, variance corrections and corrections on $t-$ratio vary from one model to another. Furthermore, the trans-

[7] The chapter also extends Yang (2015b) and Liu and Yang (2015a) to linear SARAR model to introduce simple methods for finite-sample bias corrections.

[8] See Lee and Yu (2010) for the asymptotic properties of the fixed effects SPD model.

formation approach to remove the fixed effects (in order to avoid the incidental parameter problem), induces errors that may no longer be independent and identically distributed (iid) even if the original errors are. Thus, the bootstrap method proposed by Yang (2015b) under iid errors, may not be directly applicable. We demonstrate that when the original error distribution is not far from normality, the standard iid bootstrap method can still provide an excellent approximation, due to the fact that the transformed errors are homoskedastic and uncorrelated. When the original errors are extremely non-normal, we show that the wild bootstrap method can improve the approximation. Monte Carlo results reveal that the QML estimators of the spatial parameters can be quite biased, and that a second-order bias correction effectively removes the bias. Furthermore, it shows that inferences for spatial and covariate effects based on the regular $t-$ratios can be misleading, but those based on the proposed $t-$ratios are very reliable.

Although pioneers of spatial econometric literature identified and explored the problem of heteroskedasticity in spatial econometric models, comprehensive treatments of estimation related issues were not considered until recent years.[9] While heteroskedasticity is common in regular cross-section studies, it may be more so for a spatial econometric model due to aggregation, clustering, etc. Hence the assumption of homoskedastic disturbances is likely to be invalid in a spatial context in general. However, much of the present spatial econometrics literature has focused on estimators developed under the assumption that the errors are homoskedastic. Of the available methods, the main focus is on GMM estimation combined with 2SLS. However this estimator may not be the most efficient. As such we explore a QML based robust estimation method which is as easily implementable as a GMM estimator but have the added advantage of being efficient.

In the presence of heteroskedasticity, Lin and Lee (2010) show that the QML

[9]See Lin and Lee (2010), and Baltagi and Yang (2013) among others.

estimator of the spatial autoregressive model with a lagged dependent variable can be inconsistent as a 'necessary' condition for consistency can be violated, and thus propose robust GMM estimators for the model. Inspired by Lin and Lee (2010), we introduce a robust estimator for the SLD model by adjusting the concentrated score function for the spatial parameter to make it robust against unknown heteroskedasticity. For the QML estimator to be consistent under unknown heteroskedasticity, it is necessary that the expected value of the concentrated quasi score function, $\mathrm{E}(\psi_n(\lambda))$ equals to or tends to zero. However, this condition is not necessarily satisfied if the errors are heteroskedastic. Hence we suggest an adjustment to the score function that allows it to reach a probability limit of zero by brute force.[10] Once a heteroskedasticity robust estimator of λ is obtained, the heteroskedasticity robust estimators of the model parameters θ are, $\tilde{\theta}_n = \hat{\theta}_n(\tilde{\lambda}_n)$. The method is very simple and more importantly, it can be easily generalised. We provide formal theories for the consistency and asymptotic normality of the proposed estimator, and the consistency of the robust standard error estimate. We also study the cases under which the regular QML estimator is robust against unknown heteroskedasticity and provide a set of robust inference methods. It is interesting to note that the proposed estimator is computationally as simple as the regular QML estimator, and it also outperforms the latter when it is heteroskedasticity robust.

In order to conduct robust inference on the parameter estimates, an estimate of the standard errors are required which usually involve the estimation of the variance of the adjusted score function. However, the first order variance of score contains the second, third and fourth order moments of $\epsilon_{n,i}$ which vary across i. As such a simple White type estimator is not suitable which makes it infea-

[10]Making the expectation of an estimating function to be zero leads potentially to a finite sample bias corrected estimation. This is in line with Baltagi and Yang (2013) in constructing standardised or heteroskedasticity-robust LM tests with finite sample improvements.

sible to estimate the variance of the score. In this case, we recommend the use of the outer product of the gradients (OPG) of the decomposed numerator of the adjusted score (Baltagi and Yang, 2013). Monte Carlo results show that the proposed ACQS estimator performs superbly. To demonstrate their flexibility and generality, the proposed methods are extended to popular spatial autoregressive models with heteroskedastic innovations including the SARAR(1,1) model, SARAR(p, q) model and the fixed effects spatial panel data model.

The line-up for the rest of the chapters are as follows. Part I consists the chapters that focus on refined inferences for spatial econometric models: Chapter 2 focuses on the asymptotic distribution and finite-sample bias correction of the QML Estimators for the linear SED model. An extension of the methods to a spatial moving average model is also included. Chapter 3 extends Chapter 2 to the linear SARAR model to introduce simple methods for finite-sample bias corrections and for improved t-ratios for covariate and spatial effects. Chapter 4 presents a set of refined inference methods for the fixed-effects spatial panel data models. Part II of this dissertation collects the chapters that focus on robust inferences for spatial models: Chapter 5 moves onto inference methods robust against unknown cross-sectional heteroskedasticity, with a focus on the linear SLD model. The chapter also includes an extension to the SARAR(1,1) model. Chapter 6 introduces a general methodology for heteroskedasticity-robust estimation and inference for all the popular spatial econometric models, with detailed demonstrations given using the linear SARAR(p, q) model, and the fixed-effects spatial panel data models. Chapter 7 concludes the thesis. All proofs and additional details are contained in the Appendices.

Part I

Refined Inferences for Spatial Econometric Models

CHAPTER 2

Asymptotic Distribution and Finite-Sample Bias Correction of QML Estimators for Spatial Error Dependence Model

2.1 Introduction

The conventional way to incorporate spatial autocorrelation in a regression model is to add a spatial lag of the dependent variable or a spatial lag of the error variable into the model, giving rise to a regression model with spatial lag dependence (SLD), or a regression model with spatial error dependence (SED).[1] These two models have over the years become the building blocks for spatial econometric modelling, and many more general spatial econometric models have been developed based on them. [2]

[1]See, among the others, Cliff and Ord (1972, 1973), Ord (1975), Burridge (1980), Cliff and Ord (1981), Anselin (1980, 1988), Anselin and Bera (1998), Anselin (2001).

[2]See, e.g., Anselin (2003), Das et al. (2003), Kelejian and Prucha (1998), and Lee and Liu (2010) for more general spatial regression models; Pinkse (1998) and Fleming (2004) for spatial discrete choices models; and Lee and Yu (2010b) for a survey on spatial panel data models.

Of the methods available for spatial model estimation, the maximum likelihood (ML) or quasi-ML (QML) method remains attractive due to its efficiency. As a result of the fast increase in computing power allowing for easier manipulation of large matrices, the initial reluctance for the use of QML estimation as opposed to other easily implementable estimation methods alleviated.[3] As such there had been a growing interest in developing the theoretical aspects behind QML estimation in recent times which mainly identifies two intriguing issues related the QML estimation of spatial models: asymptotic distribution and finite-sample bias of the ML or QML estimators. Of the two models, the SLD model has been extensively studied in terms of the asymptotic distributions of the ML estimators or QML estimators (Lee, 2004); finite-sample bias corrections on ML estimators or QML estimators (Bao and Ullah, 2007; Bao, 2013; Yang, 2015b). A particularly interesting phenomenon revealed by Lee (2004) for the SLD model is that the spatial dependence may slow down the rate of convergence of QML estimators of certain model parameters, including the spatial parameter. An equally interesting phenomenon revealed by subsequent studies is that spatial dependence may cause QML estimators to be biased, and more so with heavier spatial dependence (Baltagi and Yang, 2013a,b; Yang, 2015b; Liu and Yang, 2015a).

Surprisingly, these issues have not been addressed in terms of the SED model. In particular, the effect of the *degree of spatial dependence* on the convergence rate of the QML estimators has not been formally studied, and methods for correcting finite-sample bias of the QML estimators for the SED model have not been given.[4]

[3]Other estimation methods include GMM (Kelejian and Robinson, 1993; Kelejian and Prucha, 1999; Lee, 2001, 2007; Fingleton, 2008), 2SLS (Kelejian and Prucha, 1998; Lee, 2003), IV estimation (Kelejian and Prucha, 2004), and OLS estimation (Lee, 2002).

[4]Here the *degree of spatial dependence* refers to, e.g., the number of neighbours each spatial unit has, or the connectivity in general. Jin and Lee (2013) studied asymptotic properties of models with both SLD and SED for the purpose of constructing Cox-type tests, but did not study these issues. Further, it is important to know the differences between the SLD model and the SED model in terms of asymptotic and finite sample behaviours, as they may provide a valuable guidance in the specification choice. See also Martellosio (2010) for a related work.

Built upon the works of Lee (2004) and Yang (2015b), this chapter fills in these gaps. Of the two, bias correction is particularly important to the applications of this model as it leads potentially to much improved inferences for the regression coefficients. Contrary to the common perceptions, both large and small sample behaviours of the QML estimators for the SED model can be different from those for the SLD model in terms of the rate of convergence and the magnitude of bias. In summary, the QML estimator of the spatial parameter for the SED model *always* has a convergence rate slower than $\sqrt{n}$ whenever the degree of spatial dependence grows with the increase in sample size n, whereas the QML estimators of regression coefficient and error variance always have $\sqrt{n}$-rate of convergence whether or not the degree of spatial dependence increases with n. In contrast, the QML estimators of all the parameters in the SLD model have $\sqrt{n}$-rate of convergence when the spatially generated regressor is not asymptotically multi-collinear with the original regressors (Lee, 2004, Assumption 8), and a slower than $\sqrt{n}$-rate of convergence occurs in some parameters for non-regular cases where the spatially generated regressor is asymptotically multi-collinear with the original regressors and the degree of spatial dependence grows with the increase of n. Monte Carlo results show that the proposed bias correction procedure works very well for the SED model without compromising on the efficiency of the original QML estimators.

This chapter is organised as follows. Section 2.2 presents results for consistency and asymptotic normality of the QML estimators for the SED model. Section 2.3 presents methods for finite sample bias correction. Section 2.4 extends the study to an alternative SED model where the spatial autoregressive (SAR) error is replaced by a spatial moving average (SMA) error; an undesirable feature of this alternative model specification is revealed. Section 2.5 presents Monte Carlo results and Section 2.6 concludes the chapter.

2.2 Asymptotic Properties of QML Estimators for SED Model

In this section, we examine the asymptotic properties of the QML estimators of the linear regression model with spatial error dependence, giving particular attention to the effect of spatial dependence on the rate of convergence of the QML estimators. We show that the QML estimators of the regression coefficients and the error variance always have the conventional $\sqrt{n}$-rate of convergence, whereas, the QML estimator of the spatial parameter has the conventional $\sqrt{n}$-rate of convergence if the degree of spatial dependence does not grow with the increase in sample size, otherwise it has a slower rate. With an adjustment on the normalisation factor for the score component of the spatial parameter, we establish the joint asymptotic normality for the QML estimators of the model parameters. All proofs are given in Appendix D.

2.2.1 The model and the QML estimation

Consider the following linear regression model with spatial error dependence (SED), where the SED is specified as a spatial autoregressive (SAR) process:

$$Y_n = X_n\beta + u_n, \qquad u_n = \rho W_n u_n + \epsilon_n, \tag{2.1}$$

where Y_n is an $n \times 1$ vector of observations on the dependent variable corresponding to n spatial units, X_n is an $n \times k$ matrix containing the values of k exogenous regressors, W_n is an $n \times n$ spatial weights matrix that summarises the interactions among the spatial units, ϵ_n is an $n \times 1$ vector of independent and identically distributed (iid) disturbances with mean zero and variance σ^2, ρ is the *spatial parameter*, and β denotes the $k \times 1$ vector of regression coefficients.

Let $\theta = (\beta', \sigma^2, \rho)'$ be the vector of model parameters and θ_0 be its true value. Denote $A_n(\rho) = I_n - \rho W_n$ and $A_n = A_n(\rho_0)$ where I_n is an $n \times n$ identity matrix. If A_n^{-1} exists, then Model (2.1) can be written as,

$$Y_n = X_n\beta_0 + A_n^{-1}\epsilon_n, \tag{2.2}$$

leading to $\mathrm{Var}(u_n) = \mathrm{Var}(A_n^{-1}\epsilon_n) = \sigma_0^2(A_n'A_n)^{-1}$.

The linear regression with spatial lag dependence (SLD) model has the form: $Y_n = \rho_0 W_n Y_n + X_n\beta_0 + \epsilon$, which can be rewritten as $Y_n = X_n\beta_0 + \rho_0 G_n X_n \beta_0 + A_n^{-1}\epsilon_n$, where $G_n = W_n A_n^{-1}$. While in both SED and SLD models, the spatial effects generate a non-spherical structure in the disturbance term, the SLD model has an extra *spatially generated regressor*, $G_n X_n \beta_0$ which plays an important role in the identification and estimation of the spatial parameter in the SLD model in a ML estimation framework (Lee, 2004). [5]

While the SLD and SED models have been fundamental to the development of spatial econometric models and methods, an important issue, which is perhaps unique to spatial econometrics models, the effect of the degree of spatial dependence on the asymptotic properties of the QML estimators, in particular, the rate of convergence, was not addressed until Lee (2004) who identified the situations where the rate of convergence can be affected when the spatial dependence increase with the number of observations. However, this issue has not been addressed in the context of SED models. Furthermore, the degree of spatial dependence also has a profound impact on the finite sample performance of parameter estimates.

[5] The first comprehensive treatment of maximum likelihood estimation for the SLD and SED models was given by Ord (1975). More formal results can be found in Anselin (1980). In particular, Anselin (1980) pointed out that the ML estimator of the SED model can be carried out as an application of the general framework of Magnus (1978) for non-spherical errors. See Anselin (1988); and Anselin and Bera (1998) for a detailed survey on the SLD and SED models.

The quasi Gaussian log-likelihood function for the SED model is given by,

$$\ell_n(\theta) = -\tfrac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\rho)| - \tfrac{1}{2\sigma^2}\epsilon_n'(\beta,\rho)\epsilon_n(\beta,\rho), \tag{2.3}$$

where $\epsilon_n(\beta,\rho) = A_n(\rho)(Y_n - X_n\beta)$. Maximizing $\ell_n(\theta)$ gives the ML estimator, $\hat{\theta}_n$ of θ if the errors are indeed Gaussian, otherwise the QML estimator. Given ρ, the log-likelihood function $\ell_n(\theta)$ is partially maximised at,

$$\hat{\beta}_n(\rho) = [X_n'A_n'(\rho)A_n(\rho)X_n]^{-1}X_n'A_n'(\rho)A_n(\rho)Y_n, \text{ and} \tag{2.4}$$

$$\hat{\sigma}_n^2(\rho) = \tfrac{1}{n}Y_n'A_n'(\rho)M_n(\rho)A_n(\rho)Y_n, \tag{2.5}$$

where, $M_n(\rho) = I_n - A_n(\rho)X_n[X_n'A_n'(\rho)A_n(\rho)X_n]^{-1}X_n'A_n'(\rho)$. The concentrated log-likelihood function for ρ upon substituting the constrained QML estimators $\hat{\beta}_n(\rho)$ and $\hat{\sigma}_n^2(\rho)$ into (2.3):

$$\ell_n^c(\rho) = -\tfrac{n}{2}[\log(2\pi) + 1] + \log|A_n(\rho)| - \tfrac{n}{2}\log(\hat{\sigma}_n^2(\rho)). \tag{2.6}$$

Maximising $\ell_n^c(\rho)$ gives the unconstrained QML estimator $\hat{\rho}_n$ of ρ, which in turn gives the unconstrained QML estimators of β and σ^2 as, $\hat{\beta}_n = \hat{\beta}_n(\hat{\rho}_n)$ and $\hat{\sigma}_n^2 = \hat{\sigma}_n^2(\hat{\rho}_n)$.

2.2.2 Consistency and asymptotic normality

The asymptotic properties of the QML estimators of the SED model are built upon the following basic regularity conditions:

Assumption 2.1: *The true ρ_0 is in the interior of the compact parameter set $\mathcal{P}$.*

Assumption 2.2: *$\{\epsilon_{n,i}\}$ are iid with mean 0, variance σ^2, and $\mathrm{E}|\epsilon_{n,i}|^{4+\delta} < \infty, \forall\delta > 0$.*

Assumption 2.3: *X_n has full column rank k, its elements are uniformly bounded constants, and $\lim_{n\to\infty} \frac{1}{n} X_n' A_n'(\rho) A_n(\rho) X_n$ exists and is non-singular for any ρ in a neighbourhood of ρ_0.*

Assumption 2.4: *The elements $\{w_{ij}\}$ of W_n are at most of order h_n^{-1} uniformly for all i and j, where h_n can be bounded or divergent but subject to $\lim_{n\to\infty} \frac{h_n}{n} = 0$; W_n is uniformly bounded in both row and column sums and its diagonal elements are zero.*

Assumption 2.5: *A_n is non-singular and A_n^{-1} is uniformly bounded in both row and column sums. Further, $A_n^{-1}(\rho)$ is uniformly bounded in either row or column sums, uniformly in $\rho \in \mathcal{P}$.*

We allow for the possibility that the degree of spatial dependence, h_n, grows with the sample size n, and the possibility that the error distribution is not normal. These conditions are similar to those in Lee (2004) to ascertain the $\sqrt{n/h_n}$-consistency of the QML estimators of the SLD model. All conditions are very general regularity conditions considered widely in the literature. Assumption 2.1 states that the spatial parameter ρ can only take values in a compact space such that the Jacobian term of the likelihood function, $\log|A_n(\rho)|$, is well defined.[6] The full rank condition of Assumption 2.3 is needed to guarantee that the model does not suffer from multicollinearity. Assumption 2.4 is based on Lee (2004). Assumption 2.5 allows us to write the model in the reduced form (2.2). Uniform boundedness conditions given in Assumptions 2.4 and 2.5 are needed to limit the spatial correlation to a manageable degree. Boundedness on the regressors is not

[6]For this it is necessary that $|I_n - \rho W_n| = \prod_{i=1}^{n}(1-\rho\lambda_i) > 0$, where $\{\lambda_i\}$ are the eigenvalues of W_n. If the eigenvalues of W_n are all real, the parameter space $\mathcal{P}$ can be a closed interval contained in $(\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$, where $\lambda_{\min}$ and $\lambda_{\max}$ are, respectively, the minimum and maximum eigenvalues. If W_n is row-normalised, then $\lambda_{\max} = 1$ and $-1 \leq \lambda_{\min} < 0$ and $\mathcal{P}$ can be a closed interval contained in $(\lambda_{\min}^{-1}, 1)$, where the lower bound can be below -1 (Anselin, 1988). In general, the eigenvalues of W_n may not be all real and in this case Kelejian and Prucha (2010) suggested the interval $(-\tau_n^{-1}, \tau_n^{-1})$, where, $\tau_n = \max_i|\lambda_i|$ is the spectral radius of the weights matrix, and Le Sage and Pace (2009, p. 88-89) suggested interval $(\lambda_s^{-1}, 1)$ where λ_s is the most negative real eigenvalue of W_n as only the real eigenvalues can affect the singularity of $I_n - \lambda W_n$.

restrictive when analysing cross-sectional units, and in case of with stochastic regressors it can be replaced by certain finite moment conditions.

Identification of the model parameters requires that the expected log-likelihood function, $\bar{\ell}_n(\theta) = \mathrm{E}[\ell_n(\theta)]$, has identifiably unique maximisers that converge to θ_0 as $n \to \infty$. (White, 1994, Theorem 3.4; Lee, 2004). The expected log-likelihood function is,

$$\bar{\ell}_n(\theta) = -\tfrac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\rho)| - \tfrac{1}{2\sigma^2}\mathrm{E}\left[\epsilon_n'(\beta,\rho)\epsilon_n(\beta,\rho)\right], \qquad (2.7)$$

which, for a given ρ, is partially maximised at,

$$\beta_n(\rho) = (X_n'A_n'(\rho)A_n(\rho)X_n)^{-1}X_n'A_n'(\rho)A_n(\rho)\mathrm{E}(Y_n) = \beta_0, \text{ and} \qquad (2.8)$$

$$\begin{aligned}\sigma_n^2(\rho) &= \tfrac{1}{n}\mathrm{E}\big\{[Y_n - X_n\beta_n(\rho)]'A_n'(\rho)A_n(\rho)[Y_n - X_n\beta_n(\rho)]\big\}\\ &= \tfrac{1}{n}\mathrm{E}\big\{\mathrm{tr}[\epsilon_n\epsilon_n'A_n'^{-1}A_n'(\rho)A_n(\rho)A_n^{-1}]\big\}\\ &= \tfrac{1}{n}\sigma_0^2\mathrm{tr}[A_n'^{-1}A_n'(\rho)A_n(\rho)A_n^{-1}]. \qquad (2.9)\end{aligned}$$

The resulting concentrated expected log-likelihood function, $\bar{\ell}_n^c(\rho)$ is,

$$\bar{\ell}_n^c(\rho) = \max_{\beta,\sigma^2}\bar{\ell}_n(\theta) = -\tfrac{n}{2}(\log(2\pi)+1) + \log|A_n(\rho)| - \tfrac{n}{2}\log(\sigma_n^2(\rho)). \qquad (2.10)$$

From Assumption 2.3, it is clear that β and σ^2 are identified once ρ is. The latter is guaranteed if $\bar{\ell}_n^c(\rho)$ has an identifiably unique maximiser in $\mathcal{P}$ which converges to ρ_0 as $n \to \infty$, or $\lim_{n\to\infty}\frac{h_n}{n}[\bar{\ell}_n^c(\rho) - \bar{\ell}_n^c(\rho_0)] < 0,\ \forall\rho \neq \rho_0$. The global identification condition for the SED model thus simplifies to a condition on ρ alone.

Assumption 2.6: $\lim_{n\to\infty}\frac{h_n}{n}\left[\log|\sigma_0^2A_n^{-1}A_n'^{-1}| - \log|\sigma_n^2(\rho)A_n^{-1}(\rho)A_n'^{-1}(\rho)|\right] \neq 0, \forall\rho \neq \rho_0$.

This differentiates the SED model from the SLD in the asymptotic behaviours of the QML estimators. The spatially generated regressor $G_nX_n\beta_0$ of the SLD model $Y_n = X_n\beta_0 + \rho_0 G_n X_n \beta_0 + A_n^{-1}\epsilon_n$ can help identifying ρ if it is not asymptotically multi-collinear with the original regressors, giving the conventional $\sqrt{n}$-rate of convergence of $\hat{\rho}_n$ irrespective of whether h_n is bounded or unbounded. When $G_n X_n \beta_0$ is asymptotically collinear with X_n, the convergence rate of $\hat{\rho}_n$ becomes $\sqrt{n/h_n}$. In contrast, $\hat{\rho}_n$ for the SED model always has a $\sqrt{n/h_n}$-rate of convergence. Note that the variance of Y_n of (2.1) is $\sigma_0^2 A_n^{-1} A_n'^{-1}$ and hence the global identification condition given above ensures the uniqueness of the variance matrix. With this global identification condition and the uniform convergence of $\frac{h_n}{n}[\ell_n^c(\rho) - \bar{\ell}_n^c(\rho)]$ to zero in $\mathcal{P}$, the consistency of $\hat{\rho}_n$ follows.

Theorem 2.1 *Under Assumptions 2.1-2.6, the QML estimator $\hat{\rho}_n$ is a consistent estimator of ρ_0.*

Theorem 2.1 and Assumption 2.3 lead immediately to the consistency of $\hat{\beta}_n$ and $\hat{\sigma}_n^2$. However, Theorem 2.1 reveals nothing about the rate of convergence of $\hat{\rho}_n$, and hence the rates of convergence of $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ remain unknown as well. To reveal the exact convergence rates, and at the same time to derive the asymptotic distributions of the QML estimators, consider the score function,

$$S_n(\theta) \equiv \frac{\partial \ell_n(\theta)}{\partial \theta} = \begin{cases} \frac{1}{\sigma^2} X_n' A_n'(\rho) A_n(\rho) u_n(\beta), \\ \frac{1}{2\sigma^4} u_n'(\beta) A_n'(\rho) A_n(\rho) u_n(\beta) - \frac{n}{2\sigma^2}, \\ \frac{1}{\sigma^2} u_n'(\beta) A_n'(\rho) W_n u_n(\beta) - \operatorname{tr}[G_n(\rho)], \end{cases} \tag{2.11}$$

where, $u_n(\beta) = Y_n - X_n\beta$ and $G_n(\rho) = W_n A_n^{-1}(\rho)$. For likelihood-based inferences, the normalised score $\frac{1}{\sqrt{n}} S_n(\theta_0)$ at the true parameter value would be asymptotically normal. Indeed, under Assumptions 2.1-2.5 one can easily show that this is true for β and σ^2 components of $\frac{1}{\sqrt{n}} S_n(\theta_0)$. However, the normalised score for ρ is $O_p(\frac{1}{\sqrt{h_n}})$, see Lemmas A.2 and A.3 in Appendix A. This means that when h_n is divergent, the likelihood function with respect to ρ is too flat so that its normalised score converges to a degenerate distribution. As a result $\hat{\rho}_n$ converges to ρ_0 at a slower rate than the conventional $\sqrt{n}$-rate. A similar phenomenon is observed by Lee (2004) for the spatial parameter as well as the regression coefficients in the SLD model, in the 'non-regular cases' where the spatially generated regressor $G_n X_n \beta_0$, is asymptotically collinear with the regressors.

To account for the effect of spatial dependence on the asymptotic behaviour of the QML estimator $\hat{\rho}_n$ of ρ, and to jointly study the asymptotic distribution of the QML estimator $\hat{\theta}_n$ of θ, we consider the following adjusted score vector:

$$S_n^*(\theta) = K_n S_n(\theta),$$

where, $K_n = \text{diag}(I_k, 1, \sqrt{h_n})$. Hence, $\frac{1}{\sqrt{n}} S_n^*(\theta)$ would have a proper asymptotic behaviour whether h_n is divergent or bounded. Under Assumptions 2.1-2.5, the central limit theorem (CLT) for linear-quadratic forms of Kelejian and Prucha (2001) can be applied to prove the result,

$$\frac{1}{\sqrt{n}} S_n^*(\theta_0) \xrightarrow{D} N(0, \Gamma^*),$$

where, $\Gamma^* = \lim_{n\to\infty} \frac{1}{n}\Gamma_n^*$, $\Gamma_n^* = \text{Var}[S_n^*(\theta_0)] = K_n \Gamma_n K_n'$, $\Gamma_n = \text{Var}[S_n(\theta_0)]$, and

$$\Gamma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' A_n' A_n X_n & \frac{1}{2\sigma_0^3}\gamma X_n' A_n' \iota_n & \frac{1}{\sigma_0}\gamma X_n' A_n' g_n \\ \frac{1}{2\sigma_0^3}\gamma \iota_n' A_n X_n & \frac{n}{4\sigma_0^4}(\kappa+2) & \frac{1}{2\sigma_0^2}(\kappa+2)\text{tr}(G_n) \\ \frac{1}{\sigma_0}\gamma g_n' A_n X_n & \frac{1}{2\sigma_0^2}(\kappa+2)\text{tr}(G_n) & \kappa g_n' g_n + \text{tr}(G_n^s G_n) \end{pmatrix},$$

where, ι_n is an $n \times 1$ vector of ones, $\gamma = \sigma_0^{-3}\text{E}(\epsilon_{n,i}^3)$ is the measure of skewness, $\kappa = \sigma_0^{-4}\text{E}(\epsilon_{n,i}^4) - 3$ is the measure of excess kurtosis, $g_n = \text{diag}(G_n)$, $G_n = G_n(\rho_0)$, and $G_n^s = G_n + G_n'$. The information matrix $\Sigma_n = -\text{E}\left(\frac{\partial^2}{\partial\theta\partial\theta'}\ell_n(\theta_0)\right)$, takes the form:

$$\Sigma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' A_n' A_n X_n & 0 & 0 \\ 0 & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) \\ 0 & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) & \mathrm{tr}(G_n^s G_n) \end{pmatrix},$$

which leads to the adjusted version of the information matrix, $\Sigma_n^* = K_n \Sigma_n K_n'$. One can show that Γ^* exists and its diagonal elements are non-zero and $\Sigma^* = \lim_{n\to\infty} \frac{1}{n}\Sigma_n^*$ exists and is positive definite irrespective of whether h_n is bounded or unbounded. In contrast, if h_n is unbounded,

$$\lim_{n\to\infty} \frac{1}{n}\Gamma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} V_1 & \frac{\gamma}{2\sigma_0^3} V_2 & 0 \\ \frac{\gamma}{2\sigma_0^3} V_2' & \frac{1}{4\sigma_0^4}(\kappa+2) & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ and } \lim_{n\to\infty} \frac{1}{n}\Sigma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} V_1 & 0 & 0 \\ 0 & \frac{1}{2\sigma_0^4} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where, $V_1 = \lim_{n\to\infty} \frac{1}{n} X_n' A_n' A_n X_n$ and $V_2 = \lim_{n\to\infty} \frac{1}{n} X_n' A_n' \iota_n$. Hence, without the adjustment K_n, we cannot derive the asymptotic normality results due to the singularity of the matrices in the asymptotic variance-covariance matrix.

To see that Σ^* is non-singular under a general h_n, consider the determinant of Σ_n^*: $|\Sigma_n^*| = \frac{1}{2\sigma_0^6}\frac{1}{n}|X_n' A_n' A_n X_n| \frac{h_n}{n}[\mathrm{tr}(G_n^s G_n) - \frac{2}{n}\mathrm{tr}^2(G_n)]$. If h_n is bounded then by Assumptions 2.3, 2.4 and 2.5, $|\Sigma_n^*| = O(1)$. Now suppose h_n is unbounded where $\lim_{n\to\infty} h_n = \infty$ such that $\frac{h_n}{n} \to 0$, then $g_{n,ii}$, $\frac{1}{n}\mathrm{tr}(G_n' G_n)$, $\frac{1}{n}\mathrm{tr}(G_n^2)$, and $\frac{1}{n}\mathrm{tr}(G_n)$ are all $O(h_n^{-1})$ and hence by Assumption 2.3, $|\Sigma_n^*| = O(1)$. We have the following theorem for asymptotic normality of QML estimator $\hat{\theta}_n$ of θ_0.

Theorem 2.2 *Under Assumptions 2.1-2.6, we have,*

$$\sqrt{n} K_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma^{*-1}\Gamma^*\Sigma^{*-1}),$$

where, $\Gamma^* = \lim_{n\to\infty} \frac{1}{n}\Gamma_n^*$ *and* $\Sigma^* = \lim_{n\to\infty} \frac{1}{n}\Sigma_n^*$. *If errors* $\{\epsilon_{n,i}\}$ *are normally distributed, then* $\sqrt{n} K_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma^{*-1})$.

Remark 2.1 *For practical applications of Theorem 2.2, note that* h_n*, the degree of spatial dependence and affecting the rate of convergence, is not known in general. However, inference concerning the parameters does not depend on it, since* $\Sigma_n^{*-1}\Gamma_n^*\Sigma_n^{*-1} = (K_n\Sigma_nK_n)^{-1}(K_n\Gamma_nK_n)(K_n\Sigma_nK_n)^{-1} = K_n^{-1}\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}K_n^{-1}$*. Hence,* $AVar(\hat\theta_n - \theta_0) = n^{-1}\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}$.

It is also useful to have the marginal asymptotic distributions of the QML estimators, in particular, the marginal asymptotic distribution of $\hat\rho_n$.

Corollary 2.1 *Under the assumptions of Theorem 2.2, we have,*

$$\sqrt{n}(\hat\beta_n - \beta_0) \ \xrightarrow{D}\ N\big(0,\ \sigma_0^2V_1^{-1}\big),$$
$$\sqrt{n}(\hat\sigma_n^2 - \sigma_0^2) \ \xrightarrow{D}\ N\big[0,\ 2\sigma_0^4T_1 + \kappa\sigma_0^4(T_1 - 2T_2^2T_3)\big],$$
$$\sqrt{\tfrac{n}{h_n}}(\hat\rho_n - \rho_0) \ \xrightarrow{D}\ N\big(0,\ T_4 + \kappa T_5\big);\ \textit{where,}$$

$V_1 = X_n'A_n'A_nX_n$, $T_1 = \lim_{n\to\infty} tr(G_n^sG_n)/tr(C_n^sC_n)$, $T_2 = \lim_{n\to\infty} tr(G_n)/tr(C_n^sC_n)$, $T_3 = \lim_{n\to\infty}\frac{1}{n}\left[tr(G_n^sG_n) - 2g_n'g_n\right]$, $T_4 = \lim_{n\to\infty}\frac{n}{h_n}tr^{-1}(C_n^sC_n)$, $T_5 = \lim_{n\to\infty}\frac{n}{h_n}\frac{g_n'g_n - n^{-1}tr^2(G_n)}{tr^2(C_n^sC_n)}$, $C_n = G_n - \frac{tr(G_n)}{n}I_n$ *and* $C_n^s = C_n' + C_n$.

Corollary 2.1 clearly reveals that only the QML estimator of the spatial parameter has a slower rate of convergence of $\sqrt{n/h_n}$ when h_n is unbounded, which says that the effect of a growing spatial dependence is that the effective sample size for estimating ρ is reduced to n/h_n; $\hat\beta_n$ and $\hat\sigma_n^2$ have the traditional $\sqrt{n}$-convergence rate whether h_n is bounded or unbounded. Intuitively this is correct since unlike in the SLD model where there is a lagged dependent variable W_nY_n, in the SED model, the spatial structure affects only the errors and hypothetically if ρ is known, the model in (2.1) can be simplified to a linear regression model.

We note that due to the block-diagonal structure of Σ_n and the fact that the skewness measure γ appears only in the off-diagonal blocks of Γ_n, the marginal asymptotic distributions do not depend upon γ. For general asymptotic inferences, γ and κ can be consistently estimated by $\hat\gamma_n = \frac{1}{n\hat\sigma_n^3}\sum_{i=1}^n \hat\epsilon_{n,i}^3$ and $\hat\kappa_n =$

$\frac{1}{n\hat{\sigma}_n^4}\sum_{i=1}^n \hat{\epsilon}_{n,i}^4 - 3$, respectively, where $\hat{\epsilon}_{n,i}$ are the QML residuals. Thus, the estimates of Σ_n and Γ_n are obtained by plugging in $\hat{\theta}_n$, $\hat{\gamma}_n$ and $\hat{\kappa}_n$ into Σ_n and Γ_n. These discussions show that the asymptotic inferences for the SED model based on QML estimation are extremely simple. However, an important question remains: how do they perform in finite samples? Take a simple, and a very important special case where the inference concerns the regression coefficients β. While the bias of $\hat{\rho}_n$ does not have much impact on the bias of $\hat{\beta}_n$ and $\hat{\sigma}_n^2$, it does translate into the bias of the variance estimator of $\hat{\beta}_n$ through the term $\hat{V}_n = \frac{1}{n}X_n' A_n'(\hat{\rho}_n)A_n(\hat{\rho}_n)X_n$ (see the end of Section 2.4). This shows the importance of bias correction for the SED model, or perhaps for the more general models with non-spherical errors.

2.3 Finite Sample Bias Correction for the QML Estimators

The problem of estimation bias, arising from the estimation of non-linear parameters has been widely recognised by econometricians[7]. Spatial econometricians too have recognised this issue in estimating spatial econometric models and have successfully tackled this problem for the SLD model (Bao and Ullah, 2007; Bao, 2013; Yang, 2015b). However, no work has been done for the SED model and other spatial models. In a spatial regression context, spatial parameter(s) enter the regression model in a highly non-linear manner and spatial dependence maybe quite strong. As a result, the bias problem in estimating spatial parameter(s) may be quite severe, and hence it is very important to perform bias corrections on spatial estimator(s). Among the various methods for bias corrections, the *stochastic expansion method* of Rilstone et al. (1996) has recently gained more attention.

[7]see, among others, Kiviet, 1995; Hahn and Kuersteiner, 2002; Hahn and Newey, 2004; Bun and Carree, 2005

With the introduction of the bootstrap method by Yang (2015b), its applicability has been greatly expanded (See Efron, 1979, for a general introduction to the bootstrap method).

In this section, we derive the second- and third-order biases of the QML estimator of the spatial parameter in the SED model, based on the technique of stochastic expansion (Rilstone et al., 1996) and bootstrap (Yang, 2015b). As in Yang (2015b), the key quantities involved in the terms related to the bias of a non-linear estimator are the derivatives of the concentrated log-likelihood function and their expectations. While deriving the analytical solutions of the higher-order derivatives may only be a matter of tedious algebraic manipulations, evaluation of their expectations can be very difficult if not impossible. We follow the general method introduced in Yang (2015b) and propose a bootstrap procedure for implementing these bias corrections for the SED model. The validity of this procedure when applied to the SED model is established. Monte Carlo results show an excellent performance of the proposed bias correction procedure. We argue that once the spatial estimator is bias corrected, the estimators of the other models parameters become nearly unbiased. All proofs are given in Appendix D.

2.3.1 The general method for bias correction

In studying the finite sample properties of a parameter estimator, say $\hat{\theta}_n$, defined as $\hat{\theta}_n = \arg\{\psi_n(\theta) = 0\}$ for an estimating function $\psi_n(\theta)$, based on a sample of size n, Rilstone et al. (1996) and Bao and Ullah (2007) developed a stochastic expansion from which a bias correction on $\hat{\theta}_n$ can be made. The vector of parameters θ may contain a set of *linear and scale parameters*, say δ, and a *non-linear parameter*, say ρ, in the sense that given ρ, the constrained estimator $\hat{\delta}_n(\rho)$ of the vector δ possesses an explicit expression and the estimation of ρ has to be done through numerical optimization. In this case, Yang (2015b) argued that it is more effective to work with the *concentrated estimating function* (CEF): $\tilde{\psi}_n(\rho) = \psi_n(\hat{\delta}_n(\rho), \rho)$, and to perform a stochastic expansion on this CEF and hence do the bias correction only on the non-linear estimator defined by, $\hat{\rho}_n = \arg\{\tilde{\psi}_n(\rho) = 0\}$. In doing so, a multi-dimensional problem is reduced to a single-dimensional problem, and the additional variability from the estimation of the 'nuisance' parameters δ is taken into account in bias correcting the estimate of the non-linear parameter ρ.

Let $H_{rn}(\rho) = \frac{d^r}{d\rho^r}\tilde{\psi}_n(\rho), r = 1, 2, 3$. Under some general smoothness conditions on $\tilde{\psi}_n(\rho)$, Yang (2015b) presented a third-order, CEF-based, stochastic expansion for $\hat{\rho}_n$ at the true parameter value ρ_0 as,

$$\hat{\rho}_n - \rho_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(n^{-2}), \tag{2.12}$$

where, $a_{-s/2}$ represents terms of order $O_p(n^{-s/2})$ for $s = 1, 2, 3$, and they are, $a_{-1/2} = \Omega_n\tilde{\psi}_n$; $a_{-1} = \Omega_n H^\circ_{1n} a_{-1/2} + \frac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a^2_{-1/2})$; and $a_{-3/2} = \Omega_n H^\circ_{1n} a_{-1} + \frac{1}{2}\Omega_n H^\circ_{2n}(a^2_{-1/2}) + \Omega_n \mathrm{E}(H_{2n})(a_{-1/2}a_{-1}) + \frac{1}{6}\Omega_n \mathrm{E}(H_{3n})(a^3_{-1/2})$, where, $\tilde{\psi}_n \equiv \tilde{\psi}_n(\rho_0)$, $H_{rn} \equiv H_{rn}(\rho_0), r = 1, 2, 3$, $H^\circ_{rn} = H_{rn} - \mathrm{E}(H_{rn})$ and $\Omega_n = -[\mathrm{E}(H_{1n})]^{-1}$.

The above stochastic expansion leads to a second-order bias, $\mathrm{E}(a_{-1/2} + a_{-1})$, and a third-order bias, $\mathrm{E}(a_{-1/2} + a_{-1} + a_{-3/2})$, which may be used for performing bias corrections on $\hat{\rho}_n$, provided that analytical expressions of the various expected quantities can be derived so that they can be estimated through a plug-in method. Several applications of this plug-in method have appeared in the literature including Bao and Ullah (2007) for the pure spatial autoregressive process, and Bao (2013) for the SLD model. The plug-in method may run into difficulty when the analytical expectations are not available or are difficult/impossible to derive as in the SED model we consider. To overcome this obstacle, Yang (2015b) proposed a simple and yet a very effective bootstrap method to estimate the relevant expected values.

2.3.2 Bias of the QML estimator of the spatial parameter of the SED model

Recall the concentrated log-likelihood function, defined in (2.6). Define the concentrated score function or the CEF for ρ as, $\tilde{\psi}_n(\rho) = \frac{\partial}{\partial\rho}\frac{h_n}{n}\ell_n^c(\rho)$, then,

$$\tilde{\psi}_n(\rho) = -h_n T_{0n}(\rho) + h_n R_{1n}(\rho), \tag{2.13}$$

where, $T_{0n}(\rho) = \frac{1}{n}\mathrm{tr}(G_n(\rho))$ and

$$R_{1n}(\rho) = \frac{Y_n' A_n'(\rho) M_n(\rho) G_n(\rho) M_n(\rho) A_n(\rho) Y_n}{Y_n' A_n'(\rho) M_n(\rho) A_n(\rho) Y_n}, \tag{2.14}$$

leading to $\hat{\rho}_n = \arg\{\tilde{\psi}_n(\rho) = 0\}$. Let $H_{rn}(\rho) = \frac{d^r}{d\rho^r}\tilde{\psi}_n(\rho), r = 1, 2, 3$, then,

$$h_n^{-1} H_{1n}(\rho) = -T_{1n}(\rho) + R_{2n}(\rho) + 2R_{1n}^2(\rho), \tag{2.15}$$

$$h_n^{-1} H_{2n}(\rho) = -2T_{2n}(\rho) + R_{3n}(\rho) + 6R_{1n}(\rho)R_{2n}(\rho) + 8R_{1n}^3(\rho), \tag{2.16}$$

$$\begin{aligned} h_n^{-1} H_{3n}(\rho) = {} & -6T_{3n}(\rho) + R_{4n}(\rho) + 8R_{1n}(\rho)R_{3n}(\rho) + 6R_{2n}^2(\rho) \\ & +48R_{1n}^2(\rho)R_{2n}(\rho) + 48R_{1n}^4(\rho), \end{aligned} \tag{2.17}$$

where, $T_{rn}(\rho) = \frac{1}{n}tr(G_n^{r+1}(\rho)), r = 1, 2, 3$, and

$$R_{jn}(\rho) = \frac{Y_n' A_n'(\rho) M_n(\rho) D_{jn}(\rho) M_n(\rho) A_n(\rho) Y_n}{Y_n' A_n'(\rho) M_n(\rho) A_n(\rho) Y_n}, \ j = 2, 3, 4. \tag{2.18}$$

The full expressions for $D_{jn}(\rho), j = 2, 3, 4$ are given in Appendix C. Clearly, $D_{1n}(\rho) = G_n(\rho)$ in $R_{1n}(\rho)$.

The above expressions show that the key quantities in the third-order stochastic expansion for $\hat{\rho}_n$, are those ratios of quadratic forms $R_{jn}(\rho), j = 1 \ldots, 4$.[8] Now, some specific conditions on R_{jn} are needed to regulate the limiting behaviour of H_{rn} so that the required quantities have finite limits in expectation.

Assumption 2.7: $\mathrm{E}\big(\frac{h_n}{n}\epsilon_n' M_n G_n M_n \epsilon_n\big(\frac{1}{\bar{\sigma}_n^4} - \frac{1}{\sigma_0^4}\big)(\hat{\sigma}_n^2 - \sigma_0^2)\big) = O\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$, *where,* $\bar{\sigma}_n^2$ *lies between* σ_0^2 *and* $\hat{\sigma}_n^2$.

Assumption 2.8:

(i) $h_n^s \mathrm{E}[(R_{1n} - \mathrm{E}R_{1n})^s] = O\big((\frac{h_n}{n})^{\frac{1}{2}}\big), s = 2, 3, 4;$

(ii) $h_n^s \mathrm{E}[(R_{2n} - \mathrm{E}R_{2n})^s] = O\big((\frac{h_n}{n})^{\frac{1}{2}}\big), s = 1, 2;$

(iii) $h_n \mathrm{E}(R_{rn} - \mathrm{E}R_{rn}) = O((\frac{h_n}{n})^{\frac{1}{2}}), r = 3, 4;$

(iv) $h_n^{s+1} \mathrm{E}[(R_{1n} - \mathrm{E}R_{1n})^s (R_{2n} - \mathrm{E}R_{2n})] = O((\frac{h_n}{n})^{\frac{1}{2}}), s = 1, 2,$ and

(v) $h_n^2 \mathrm{E}[(R_{1n} - \mathrm{E}R_{1n})(R_{3n} - \mathrm{E}R_{3n})\big] = O((\frac{h_n}{n})^{\frac{1}{2}}).$

The following Lemma shows the bounded behaviour of the expectations of the quantities in the stochastic expansion.

Lemma 2.1 *Under Assumptions 2.1-2.7, (i)* $h_n R_{in} = O_p(1)$, *(ii)* $\mathrm{E}(h_n R_{in}) = O(1)$, *and (iii)* $h_n R_{in} = \mathrm{E}(h_n R_{in}) + O_p((\frac{h_n}{n})^{\frac{1}{2}}), i = 1, \ldots, 4.$

Given Lemma 2.1 and the regularity conditions, we can prove the following propositions:

Proposition 2.1 *Suppose the SED model specified by* (2.1) *satisfies Assumptions 2.1-2.8. Then, the third-order stochastic expansion given in* (2.12) *holds for the QML estimator* $\hat{\rho}_n$ *of the spatial parameter in the model with* n *replaced by* n/h_n *for the stochastic order:*

$$\hat{\rho}_n - \rho_0 = c_{1n}'\zeta_n + c_{2n}'\zeta_n + c_{3n}'\zeta_n + O_p((\frac{h_n}{n})^2), \tag{2.19}$$

where, $c_{sn}'\zeta_n$ *are of stochastic order* $O((\frac{h_n}{n})^{\frac{s}{2}}), s = 1, 2, 3,$ *with,*

[8] Note that, a function of ρ evaluated at ρ_0 is denoted by dropping the function argument, e.g., $\tilde{\psi}_n = \tilde{\psi}_n(\rho_0), A_n = A_n(\rho_0)$, $G_n = G_n(\rho_0)$, $R_{jn} = R_{jn}(\rho_0)$, $H_{rn} = H_{rn}(\rho_0)$, $T_{rn} = T_{rn}(\rho_0)$.

$\zeta_n = \{\tilde{\psi}_n,\ H_{1n}\tilde{\psi}_n,\ \tilde{\psi}_n^2,\ H_{1n}^2\tilde{\psi}_n,\ H_{2n}\tilde{\psi}_n^2,\ H_{1n}\tilde{\psi}_n^2,\ \psi_n^3\}'$, $c_{1n} = \{\Omega_n, 0'_{6\times 1}\}'$,

$\Omega_n = -\mathrm{E}(H_{1n})^{-1}$, $c_{2n} = \{\Omega_n, \Omega_n^2, \frac{1}{2}\Omega_n^3 E(H_{2n}), 0'_{4\times 1}\}'$, *and*

$c_{3n} = \{\Omega_n, 2\Omega_n^2, \Omega_n^3\mathrm{E}(H_{2n}), \Omega_n^3, \frac{1}{2}\Omega_n^3, \frac{3}{2}\Omega_n^4\mathrm{E}(H_{2n}), \frac{1}{2}\Omega_n^5\mathrm{E}(H_{2n})^2 + \frac{1}{6}\Omega_n^4\mathrm{E}(H_{3n})\}'$.

Remark 2.2 *By letting* $C_{2n} = c_{1n}+c_{2n}$ *and* $C_{3n} = c_{1n}+c_{2n}+c_{3n}$, *the stochastic expansions can be further simplified to* $c'_{1n}\zeta_n$ *(asymptotic),* $C'_{2n}\zeta_n$ *(second-order), and* $C'_{3n}\zeta_n$ *(third order), which are helpful in the bootstrap work introduced later.*

Proposition 2.2 *Under Assumptions 2.1-2.8 and further assuming that a quantity bounded in probability has a finite expectation, a third-order expansion for the bias of* $\hat{\rho}_n$ *is:*

$$Bias(\hat{\rho}_n) = C'_{2n}\mathrm{E}(\zeta_n) + c'_{3n}\mathrm{E}(\zeta_n) + O((\tfrac{h_n}{n})^2), \tag{2.20}$$

and the 2nd and 3rd order bias corrected QML estimators are:

$$\hat{\rho}_n^{bc2} = \hat{\rho}_n - \widehat{C}'_{2n}\widehat{\mathrm{E}}(\zeta_n) \quad \textit{and} \quad \hat{\rho}_n^{bc3} = \hat{\rho}_n - \widehat{C}'_{3n}\widehat{\mathrm{E}}(\zeta_n), \tag{2.21}$$

where, a quantity with a $\widehat{}$ *is the corresponding estimate of that quantity.*

Practical implementation of the bias corrections given in (2.21) depends on the availability of the estimates $\widehat{\mathrm{E}}(\zeta_n)$, and $\widehat{C}_{2n}$ or $\widehat{C}_{3n}$. Note that ζ_n is defined in terms of $\tilde{\psi}_n$ and H_{rn}, and C_{2n} and C_{3n} are defined in terms of $\mathrm{E}(H_{rn}), r = 1, 2, 3$. Given the complicated expressions for $\tilde{\psi}_n$ and H_{rn} defined in (2.13)-(2.17), the conventional method of estimation by deriving the analytical expectations for $\mathrm{E}(\zeta_n)$, and C_{2n} or C_{3n} would be extremely difficult if not impossible. The method of using the sample analogue would not work either due to the fact that $\tilde{\psi}(\hat{\rho}_n) = 0$. These iterate the point raised in Yang (2015b), and hence, the bootstrap method given in same is adopted for the estimation of the quantities in question.

2.3.3 Bootstrap method for implementing the bias correction

From (2.13), and (2.15)-(2.17), we see that $\tilde{\psi}_n$ and H_{rn} are functions of only $R_{jn}, j = 1, \ldots, 4$, i.e., we need to individually estimate the following terms:

$\mathrm{E}(R_{1n}^i), i = 1, \ldots, 5; \quad \mathrm{E}(R_{2n}^j), j = 1, 2; \quad \mathrm{E}(R_{3n}); \quad \mathrm{E}(R_{4n});$

$\mathrm{E}(R_{1n}^i R_{2n}), i = 1, 2, 3; \quad \mathrm{E}(R_{1n} R_{2n}^2); \quad \mathrm{E}(R_{1n}^i R_{3n}), i = 1, 2.$

It is easy to see that,

$$R_{jn} \equiv R_{jn}(e_n, \rho_0) = \frac{e_n' \Lambda_{jn}(\rho_0) e_n}{e_n' M_n(\rho_0) e_n}, \tag{2.22}$$

where $e_n = \sigma_0^{-1} \epsilon_n$, $\Lambda_{jn}(\rho_0) = M_n(\rho_0) D_{jn} M_n(\rho_0)$ with $D_{1n} = G_n$ and $D_{jn}, j = 2, 3$ being defined at the beginning of Appendix C. It follows that all the necessary quantities whose expectations are required can be expressed in terms of e_n and ρ_0. In particular, we can write, $H_{rn} \equiv H_{rn}(e_n, \rho_0)$, and $\zeta_n \equiv \zeta_n(e_n, \rho_0)$. Thus, H_{rn} and ζ_n, and their distributions are *invariant* of β_0 and σ_0^2. The bootstrap procedure for estimating the expectations of the above quantities can be described as follows:

(1) Compute the QML estimators $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\rho}_n)'$ using the original data,

(2) Compute standardised QML residuals, $\hat{e}_n = \hat{\sigma}_n^{-1} A_n(\hat{\rho}_n)(Y_n - X_n \hat{\beta}_n)$.[9] Denote the empirical distribution function (EDF) of centred $\hat{e}_n$ by $\mathcal{F}_n$,

(3) Draw a random sample of size n from $\mathcal{F}_n$, and denote it by $e_{n,b}^*$,

(4) Compute $R_{in}(e_{n,b}^*, \hat{\rho}_n)$, $i = 1, \ldots, 4$, and hence $H_{in}(e_{n,b}^*, \hat{\rho}_n)$, $i = 1, 2, 3$ and $\zeta_n(e_{n,b}^*, \hat{\rho}_n)$,

(5) Repeat (3) and (4) B times, and the bootstrap estimates of $\mathrm{E}(H_{in}), i =$

[9] Whether to bootstrap the standardised QML residuals $\hat{e}_n$ or the original QML residuals $\hat{\epsilon}_n = \hat{\sigma}_n \hat{e}_n$ does not make a difference as R_{jn} are invariant of σ_0. However, use of $\hat{e}_n$ makes the theoretical discussion easier.

1, 2, 3, and $\mathrm{E}(\zeta_n)$ are given by:

$$\hat{\mathrm{E}}(H_{in}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{B} H_{in}(e_{n,b}^*, \hat{\rho}_n), \text{ and } \hat{\mathrm{E}}(\zeta_{in}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{B} \zeta_{in}(e_{n,b}^*, \hat{\rho}_n). \quad (2.23)$$

The proposed bootstrap procedure overcomes the difficulty of analytically evaluating the expectations of very complicated quantities, and is very straightforward since in every bootstrap iteration, no re-estimation of the model parameters is required. The question that remains is its validity, particularly the validity of using $\widehat{C}_{2n}\widehat{\mathrm{E}}(\xi_n)$ in the third-order bias corrections $\widehat{C}_{3n}\widehat{\mathrm{E}}(\xi_n) = \widehat{C}_{2n}\widehat{\mathrm{E}}(\xi_n) + \widehat{c}_{3n}\widehat{\mathrm{E}}(\xi_n)$. We now elaborate using the quantities R_{jn}.

Let $\mathcal{F}_0$ be the CDF of $e_{n,i}$. The EDF $\mathcal{F}_n$ is thus an estimate of $\mathcal{F}_0$. If ρ_0 and $\mathcal{F}_0$ were known, then $\mathrm{E}[R_{jn}(e_n, \rho_0)] \doteq \frac{1}{M}\sum_{m=1}^{M} R_{jn}(e_{n,m}, \rho_0)$, where $e_{n,m}$ is a random sample of size n drawn from $\mathcal{F}_0$ and M is an arbitrarily large number. If ρ is unknown but $\mathcal{F}_0$ is known, $\mathrm{E}[R_{jn}(e_n, \rho_0)]$ can be estimated by $\frac{1}{M}\sum_{m=1}^{M} R_{jn}(e_{n,m}, \hat{\rho}_n)$, giving the Monte Carlo (or parametric bootstrap) estimates of an expectation. In reality, however, both ρ_0 and $\mathcal{F}_0$ are unknown making this Monte Carlo method infeasible. The bootstrap analogue of Model (2.2) takes the form, $Y_{n,b}^* = X_n\hat{\beta}_n + \hat{\sigma}_n A_n^{-1}(\hat{\rho}_n)e_{n,b}^*$, where $(\hat{\beta}_n, \hat{\sigma}_n^2, \hat{\rho}_n)$ are now treated as bootstrap parameters. Based on the generated bootstrap data $(Y_{n,b}^*, W_n, X_n)$ and the bootstrap parameter $\hat{\rho}_n$, one computes R_{jn} defined by (2.14) and (2.18), to give bootstrap analogues of R_{jn}, which are $R_{jn}(e_n^*, \hat{\rho}_n), j = 1, \ldots, 4$. The bootstrap estimates of $\mathrm{E}[R_{jn}(e_n, \rho_0)]$ are thus,

$$\mathrm{E}^*[R_{jn}(e_n^*, \hat{\rho}_n)] \doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B} R_{jn}(e_{n,b}^*, \hat{\rho}_n), \text{ for a large } B,$$

which takes the same form as the Monte Carlo estimate with a known $\mathcal{F}_0$. This gives a heuristic justification on the validity of the bootstrap method.

Formally, denote the second- and third-order bias terms by $b_2(\rho_0, \gamma_0) = C_{2n}'\mathrm{E}(\zeta_n)$

and $b_3(\rho_0, \gamma_0) = c'_{3n}\mathrm{E}(\zeta_n)$, respectively, where $\gamma_0 = \gamma(\mathcal{F}_0)$ denotes the higher (than 2nd) order moments of $\mathcal{F}_0$ that b_2 and b_3 may depend upon. In our QML estimation framework, γ_0 is unknown as $\mathcal{F}_0$ is specified up to only the first two moments. Following the arguments above, the bootstrap estimates of b_2 and b_3 must take the form: $\hat{b}_2 = b_2(\hat{\rho}_n, \hat{\gamma}_n)$ and $\hat{b}_3 = b_3(\hat{\rho}_n, \hat{\gamma}_n)$ where $\hat{\gamma}_n = \gamma(\hat{\mathcal{F}}_n)$. The validity of the bootstrap estimates of bias corrections is thus established.

Proposition 2.3 *Under Assumptions of Proposition 2.1 and further, assuming a quantity bounded in probability has a finite expectation, then,*

$$\mathrm{E}[b_2(\hat{\rho}_n, \hat{\gamma}_n)] = b_2(\rho_0, \gamma_0) + O((\tfrac{h_n}{n})^2), \quad \textit{and} \quad \mathrm{E}[b_3(\hat{\rho}_n, \hat{\gamma}_n)] = b_3(\rho_0, \gamma_0) + o_p((\tfrac{h_n}{n})^2).$$

It follows that $\mathrm{E}(\hat{\rho}_n^{\mathrm{bc2}}) = \rho_0 + O((\frac{h_n}{n})^{\frac{3}{2}})$ *and* $\mathrm{E}(\hat{\rho}_n^{\mathrm{bc3}}) = \rho_0 + O((\frac{h_n}{n})^2)$.

2.4 An Alternative Model Specification

As mentioned in Section 2.2, an alternative to the SED model with an SAR error process is the SED model with a spatial moving average (SMA) error process,

$$Y_n = X_n\beta + u_n, \quad u_n = \epsilon_n - \rho W_n \epsilon_n, \tag{2.24}$$

where, all the quantities are defined in a similar manner as (2.1). The model at the true parameters can be written as $Y_n = X_n\beta_0 + A_n\epsilon_n$, giving, $\mathrm{Var}(u_n) = \sigma_0^2 A_n A_n'$, suggesting a similar non-spherical error structure. The quasi Gaussian log-likelihood function for this model is,

$$\ell_n(\theta) = -\tfrac{n}{2}\log(2\pi\sigma^2) - \log|A_n(\rho)| - \tfrac{1}{2\sigma^2}(Y_n - X_n\beta)' A_n'^{-1}(\rho) A_n^{-1}(\rho)(Y_n - X_n\beta) \tag{2.25}$$

Given ρ, the constrained QML estimators are,

$$\hat{\beta}_n(\rho) = (X_n' A_n'^{-1}(\rho) A_n^{-1}(\rho) X_n)^{-1} X_n' A_n'^{-1}(\rho) A_n^{-1}(\rho) Y_n, \text{ and}$$
$$\hat{\sigma}_n^2(\rho) = \tfrac{1}{n} Y_n' A_n'^{-1}(\rho) M_n(\rho) A_n^{-1}(\rho) Y_n,$$

where, $M_n(\rho) = I_n - A_n^{-1}(\rho) X_n [X_n' A_n'^{-1}(\rho) A_n^{-1}(\rho) X_n]^{-1} X_n' A_n'^{-1}(\rho)$. This results in the following concentrated log-likelihood function by substituting $\hat{\beta}_n(\rho)$ and $\hat{\sigma}_n^2(\rho)$ into (2.25),

$$\ell_n^c(\rho) = -\tfrac{n}{2}[\log(2\pi) + 1] - \log|A_n(\rho)| - \tfrac{n}{2}\log(\hat{\sigma}_n^2(\rho)). \quad (2.26)$$

The unconstrained QML estimator $\hat{\rho}_n$ of ρ maximises $\ell_n^c(\rho)$, and the unconstrained QML estimators of β and σ^2 are given as $\hat{\beta}_n \equiv \hat{\beta}_n(\hat{\rho}_n)$ and $\hat{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\hat{\rho}_n)$, respectively as in Section 2.2.

The QML estimator $\hat{\rho}_n$ of the SMA error model is likely to perform poorer than that of the SAR error model, because the parameter space $\mathcal{P}$ for ρ stays the same, but $\hat{\rho}_n$ now becomes upward biased by comparing (2.26) with (2.6). Thus, when ρ is positive, 0.5 say, $\hat{\rho}_n$ may hit the upper bound of $\mathcal{P}$ when n is small, causing difficulty in estimating ρ.[10] Monte Carlo results given in Section 2.5 confirm this point. See also Martellosio (2010) for related discussions.

Asymptotic Distribution: Consistency and asymptotic normality of $\hat{\theta}_n$ can be proved in a similar manner as in the SED model with SAR errors, under a similar set of regularity conditions. In particular, the Assumption 2.3 has to be adjusted as: $\lim_{n\to\infty} \frac{1}{n} X_n' A'^{-1}(\rho) A^{-1}(\rho) X_n$ *exists and is non-singular uniformly in* ρ *in a neighbourhood of* ρ_0; and replace Assumption 2.6, the identification

[10] A more natural parametrisation for the SMA error model may be $u_n = \epsilon_n + \rho W_n \epsilon_n$, under which $\mathcal{P}$ becomes a closed interval contained in $(-1, -\lambda_{\min}^{-1})$, but the QML estimator $\hat{\rho}_n$ is now downward biased, and hence when ρ_0 is negative and n is small $\hat{\rho}_n$ may hit the lower bound of $\mathcal{P}$, causing the numerical instability of $(I_n + \hat{\rho}_n W_n)^{-1}$.

condition by: *For any* $\rho \neq \rho_0$, $\lim_{n\to\infty} \frac{h_n}{n}\big[\log|\sigma_0^2 A_n' A_n| - \log|\sigma_n^2(\rho) A_n'(\rho) A_n(\rho)|\big] \neq 0$, *where,* $\sigma_n^2(\rho) = \frac{\sigma_0^2}{n}\mathrm{tr}[A_n' A_n'^{-1}(\rho) A_n^{-1}(\rho) A_n]$.

Theorem 2.3 *Under the adjusted Assumptions 2.1-2.6, we have,*

$$\sqrt{n}K_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma^{*-1}\Gamma^*\Sigma^{*-1}),$$

where, $\Gamma^* = \lim_{n\to\infty} \frac{1}{n}\Gamma_n^*$, $\Sigma^* = \lim_{n\to\infty} \frac{1}{n}\Sigma_n^*$, $\Gamma_n^* = K_n \Gamma_n K_n'$, $\Sigma_n^* = K_n \Sigma_n K_n'$,

$$\Gamma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' A_n^{-1'} A_n^{-1} X_n & \frac{1}{2\sigma_0^3}\gamma X_n' A_n^{-1'} \iota_n & \frac{1}{\sigma_0}\gamma X_n' A_n^{-1'} g_n \\ \frac{1}{2\sigma_0^3}\gamma \iota_n' A_n^{-1} X_n & \frac{n}{4\sigma_0^4}(\kappa+2) & \frac{1}{2\sigma_0^2}(\kappa+2)\mathrm{tr}(G_n) \\ \frac{1}{\sigma_0}\gamma g_n' A_n^{-1} X_n & \frac{1}{2\sigma_0^2}(\kappa+2)\mathrm{tr}(G_n) & \kappa g_n' g_n + \mathrm{tr}(G_n^s G_n) \end{pmatrix},$$

$$\Sigma_n = \begin{pmatrix} \frac{1}{\sigma_0^2} X_n' A_n^{-1'} A_n^{-1} X_n & 0 & 0 \\ 0 & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) \\ 0 & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) & \mathrm{tr}(G_n^s G_n) \end{pmatrix},\ \textit{and } G_n = A_n^{-1} W_n.$$

Note that if the errors $\{\epsilon_{n,i}\}$ are normally distributed, then $\sqrt{n}K_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma^{*-1})$. A similar set of results as in Corollary 2.1 can be obtained as well. The proof Theorem 2.3 is omitted as it is very similar to that of Theorem 2.2.

Finite-Sample Bias Correction: To simplify the exposition, we only present the necessary expressions for a second-order bias correction. The derivatives of the *averaged* concentrated log-likelihood function $\frac{h_n}{n}\ell_n^c(\rho)$, are:

$$\begin{aligned} \tilde{\psi}_n(\rho) &= h_n T_{0n}(\rho) - h_n R_{1n}(\rho), \\ h_n^{-1} H_{1n}(\rho) &= T_{1n}(\rho) - R_{2n}(\rho) + 2R_{1n}^2(\rho), \\ h_n^{-1} H_{2n}(\rho) &= 2T_{2n}(\rho) - R_{3n}(\rho) + 6R_{1n}(\rho)R_{2n}(\rho) - 8R_{1n}^3(\rho), \end{aligned}$$

where, $T_{rn}(\rho) = \frac{1}{n}\text{tr}(G_n^{r+1}(\rho)), r = 0, 1, 2,$

$$R_{1n}(\rho) = \frac{Y_n' A_n'^{-1}(\rho) M_n(\rho) G_n(\rho) M_n(\rho) A_n^{-1}(\rho) Y_n}{Y_n' A_n'^{-1}(\rho) M_n(\rho) A_n^{-1}(\rho) Y_n}, \text{ and} \tag{2.27}$$

$$R_{jn}(\rho) = \frac{Y_n' A_n'^{-1}(\rho) M_n(\rho) D_{jn}(\rho) M_n(\rho) A_n^{-1}(\rho) Y_n}{Y_n' A_n'^{-1}(\rho) M_n(\rho) A_n^{-1}(\rho) Y_n}, \ j = 2, 3, \tag{2.28}$$

where, $D_{2n}(\rho)$ and $D_{3n}(\rho)$ are given in Appendix C.

Finally, with the clear definitions of the quantities $\tilde{\psi}_n(\rho)$, $h_n^{-1} H_{1n}(\rho)$ and $h_n^{-1} H_{2n}(\rho)$, the second-order bias correction of the QML estimator $\hat{\rho}_n$ can be carried out using an identical bootstrap procedure as described in Section 2.3. The validity of the bootstrap procedure applied to this model can be proved in a similar manner. While the third-order bias correction can be carried out in the same manner, we found from the Monte Carlo experiments that the second-order bias corrections are more than satisfactory in all the cases considered.

Impact of bias correction: We now offer some details on the impact of bias correcting $\hat{\rho}_n$ on the subsequent inference for β in the form of testing $H_0 : c_0'\beta = 0$. The test statistic based on Corollary 2.1 is $t_n = c_0'\hat{\beta}_n / \sqrt{\hat{\sigma}_n^2 c_0' \hat{V}_n^{-1} c_0 / n}$, where $\hat{V}_n = \frac{1}{n} X_n' A_n'(\hat{\rho}_n) A_n(\hat{\rho}_n) X_n = V_n - (\hat{\rho}_n - \rho_0) X_n'(W_n' A_n + A_n' W_n) X_n / n + (\hat{\rho}_n - \rho_0)^2 X_n' W_n' W_n X_n / n$. As $\hat{\rho}_n$ is downward biased, $\hat{V}_n$ tends to over estimate V_n, and hence $\hat{V}_n^{-1}$ tends to under estimate V_n^{-1}, causing t_n to be more variable and hence size distortions (over rejections). Our Monte Carlo results (unreported for brevity) show that simply replacing $\hat{\rho}_n$ in t_n by $\hat{\rho}_n^{\text{bc2}}$ defined in (2.21) significantly reduces the size distortion. This shows that bias correction has a great potential for improving inferences for the regression coefficients. A formal study on this is interesting, but beyond the scope of this chapter.

2.5 Monte Carlo Experiments

The objective of the Monte Carlo simulations is to investigate the finite sample behaviour of $\hat{\rho}_n$ and the bias corrected $\hat{\rho}_n$, under various spatial layouts, error distributions and the model parameters. The simulations are carried out based on the following data generation processes (DGP):

$$Y_n = \iota_n \beta_0 + X_{1n}\beta_1 + X_{2n}\beta_2 + u_n, \quad u_n = \rho W_n u_n + \epsilon_n,$$

where, ι_n is an $n \times 1$ vector of ones for the intercept term and X_{1n} and X_{2n} are the $n \times 1$ vectors containing the values of two fixed regressors. The parameters of the simulation are initially set to be as: $\beta = (5, 1, 1)'$, $\sigma^2 = 1$, ρ takes values form $\{-0.5, -0.25, 0, 0.25, 0.5\}$ and n take values from $\{50, 100, 200, 500\}$. Each set of results is based on $M = 10,000$ Monte Carlo samples, and $B = 999 + \lfloor n^{0.75} \rfloor$ bootstrap samples within each Monte Carlo sample. The methods for generating X_n, W_n, and the errors are described in Appendix B.

Partial Monte Carlo results are summarised in Tables 2.1-2.4, where in each table, the Monte Carlo means, root mean square errors (rmse) and the standard errors (se) of $\hat{\rho}_n$ and $\hat{\rho}_n^{bc2}$ are reported. The results for $\hat{\rho}_n^{bc3}$ are omitted as $\hat{\rho}_n^{bc2}$ provides satisfactory bias corrections for all the cases and the additional gain of using $\hat{\rho}_n^{bc3}$, although apparent, is quite marginal. Further, the case of queen contiguity (Table 2.2) is replicated by changing the β value to $(0.5, 0.1, 0.1)'$ (Table 2.5), and by changing the σ value to 3 (Table 2.6). We also give some partial results (Tables 2.7 and 2.8) for the SMA error model under the same set of parameters values set out at beginning of this section. It is useful to the note the following general characteristics of the results:

(i) $\hat{\rho}_n$ suffers from severe downward bias for almost all of the ρ values considered. The severity of the bias varies according to variations in (a) the sample size,

(b) the spatial layout, and (c) the distribution of the errors considered.

(ii) $\hat{\rho}_n^{bc2}$ is almost unbiased in all the cases, even at considerably small sample sizes, which ascertains the effectiveness of the proposed bias correction procedure. These corrections can be attained without compromising the efficiency of the original QML estimators.

(iii) The spatial layout has a considerable impact on the finite sample performance of $\hat{\rho}_n$ in terms of the bias, rmse and se. A relatively sparse W_n, as in contiguity schemes, results in lower bias, rmse and se while a relatively dense W_n, as in group interaction scheme, results in the opposite.

(iv) The bias of the original QML estimator seems to worsen as the error distribution deviates from normality. In contrast, $\hat{\rho}_n^{bc2}$ attains a similar level of accuracy in all the cases.

(v) The performance of $\hat{\rho}_n$ is not so sensitive to changes in the values of σ and β in terms of bias and the bias correction works well regardless of the true values set for the parameters.

(vi) The impact of the degree of spatial dependence on the rate of convergence is clearly revealed when comparing the results in Table 2.3 with those in Table 2.4 under the group interaction scheme. When the degree of spatial dependence is stronger as in the case where $k = n^{0.5}$, the rate of convergence is slower than in the case where $k = n^{0.65}$.

As expected, the magnitude of the bias, rmse and se are larger for small sample sizes. When considering the efficiency variations in terms of standard errors it can be seen that the efficiency of the estimators are sensitive to the sample size and the spatial layout. However, the different error distributions does not seem to have a significant effect on standard errors, reiterating the applicability of the proposed bias correction method in terms of robustness.

When the errors follow the SMA process, $u_n = (I_n - W_n)\epsilon_n$, the Monte Carlo results given in Tables 2.7 and 2.8 show that (*i*) the bias becomes positive, (*ii*) the QML estimator $\hat{\rho}_n$ again can be severely biased, and (*iii*) the bias corrected $\hat{\rho}_n$ is almost unbiased. As discussed in Section 2.4, the Monte Carlo results indeed show that when ρ is positive (e.g., 0.5) and n is small (e.g., 50), $\hat{\rho}_n$ can be close to or can hit its upper bound, say 0.9999, causing numerical instability in calculating $A_n^{-1}(\hat{\rho}_n) = (I_n - \hat{\rho}_n W_n)^{-1}$, thus resulting in a poor performance of $\hat{\rho}_n$ and causing difficulty in bootstrapping the bias. This stands in contrast to the SED model with SAR errors where $\hat{\rho}_n$ is downward biased. However, with a larger $n(\geq 100)$, this problem disappears as seen from the results in Tables 2.7 and 2.8. Nevertheless, this does signal to a possible poor performance of the QML estimator for an SMA error model when the sample size is not so large and the true spatial parameter value is positive and big.

Finally, compared to the Monte Carlo results presented in Yang (2015b) for the SLD model, we see that the bias of $\hat{\rho}_n$ is more severe for the SED model, but does not spill over to $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ that much.

Table 2.1

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SAR Errors - Rook Contiguity, REG-1

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.440[.175](.164)	.495[.169](.169)	.445[.166](.157)	.499[.161](.161)	.452[.152](.144)	.503[.147](.147)
	100	.472[.116](.112)	.501[.114](.114)	.471[.112](.108)	.499[.110](.110)	.473[.104](.101)	.500[.102](.102)
	200	.487[.079](.077)	.501[.078](.078)	.486[.077](.075)	.500[.076](.076)	.487[.072](.071)	.500[.071](.071)
	500	.495[.049](.049)	.501[.049](.049)	.495[.049](.048)	.500[.049](.049)	.495[.046](.046)	.500[.046](.046)
.25	50	.202[.192](.186)	.248[.195](.195)	.203[.182](.176)	.248[.184](.184)	.207[.169](.163)	.250[.170](.170)
	100	.228[.130](.128)	.252[.131](.131)	.225[.127](.124)	.248[.127](.127)	.228[.119](.117)	.251[.120](.120)
	200	.239[.091](.090)	.251[.091](.091)	.239[.090](.090)	.250[.090](.090)	.240[.085](.084)	.251[.085](.085)
	500	.246[.057](.057)	.250[.057](.057)	.246[.057](.057)	.251[.058](.058)	.246[.055](.055)	.251[.055](.055)
.00	50	-.032[.192](.189)	.002[.201](.201)	-.035[.184](.181)	-.002[.191](.191)	-.033[.178](.175)	-.002[.184](.184)
	100	-.021[.135](.133)	-.004[.137](.137)	-.018[.131](.130)	.000[.133](.133)	-.019[.124](.123)	-.003[.126](.126)
	200	-.010[.097](.096)	-.001[.098](.098)	-.008[.093](.093)	.001[.094](.094)	-.010[.089](.088)	-.002[.089](.089)
	500	-.005[.060](.060)	-.001[.060](.060)	-.005[.059](.059)	-.001[.059](.059)	-.004[.058](.058)	.001[.058](.058)
-.25	50	-.270[.180](.179)	-.252[.191](.191)	-.273[.171](.170)	-.255[.181](.181)	-.274[.169](.168)	-.257[.178](.178)
	100	-.262[.127](.126)	-.252[.130](.130)	-.261[.124](.123)	-.251[.127](.127)	-.262[.120](.119)	-.252[.123](.123)
	200	-.255[.090](.090)	-.250[.091](.091)	-.255[.088](.088)	-.250[.089](.089)	-.255[.087](.087)	-.250[.088](.088)
	500	-.253[.057](.057)	-.250[.058](.058)	-.252[.057](.057)	-.250[.058](.058)	-.253[.056](.056)	-.250[.057](.057)
-.50	50	-.503[.152](.152)	-.502[.163](.163)	-.503[.144](.144)	-.500[.153](.153)	-.509[.144](.143)	-.507[.153](.153)
	100	-.504[.107](.107)	-.502[.111](.111)	-.503[.104](.104)	-.501[.108](.108)	-.504[.103](.103)	-.502[.106](.106)
	200	-.502[.076](.076)	-.501[.077](.077)	-.502[.074](.074)	-.501[.076](.076)	-.503[.074](.074)	-.502[.075](.075)
	500	-.501[.048](.048)	-.500[.049](.049)	-.501[.047](.047)	-.500[.048](.048)	-.501[.046](.046)	-.501[.047](.047)

Table 2.2

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SAR Errors - Queen Contiguity, REG-1

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.390[.244](.218)	.492[.215](.215)	.395[.232](.206)	.493[.204](.204)	.406[.207](.184)	.501[.181](.181)
	100	.445[.153](.143)	.499[.140](.140)	.449[.145](.135)	.501[.133](.133)	.451[.133](.124)	.501[.122](.122)
	200	.474[.099](.095)	.500[.095](.095)	.474[.098](.095)	.500[.094](.094)	.476[.091](.087)	.500[.087](.087)
	500	.491[.059](.058)	.501[.058](.058)	.490[.059](.058)	.500[.058](.058)	.490[.056](.055)	.500[.055](.055)
.25	50	.144[.270](.248)	.248[.250](.250)	.153[.255](.236)	.254[.238](.238)	.153[.239](.218)	.250[.219](.219)
	100	.196[.179](.171)	.253[.169](.169)	.194[.177](.168)	.249[.166](.166)	.197[.165](.156)	.250[.154](.154)
	200	.221[.121](.117)	.248[.117](.117)	.222[.118](.115)	.249[.114](.114)	.225[.110](.107)	.250[.107](.107)
	500	.240[.073](.073)	.250[.073](.073)	.240[.075](.074)	.250[.074](.074)	.241[.069](.068)	.251[.068](.068)
.00	50	-.101[.294](.276)	-.002[.285](.285)	-.095[.277](.260)	.003[.268](.268)	-.095[.259](.241)	-.001[.247](.247)
	100	-.059[.200](.192)	-.002[.192](.192)	-.059[.197](.188)	-.002[.189](.189)	-.055[.181](.172)	.001[.172](.172)
	200	-.027[.135](.132)	.001[.133](.133)	-.026[.132](.130)	.002[.130](.130)	-.027[.124](.121)	-.002[.121](.121)
	500	-.011[.083](.082)	-.001[.082](.082)	-.011[.082](.081)	.000[.081](.081)	-.010[.079](.079)	.001[.079](.079)
-.25	50	-.339[.299](.285)	-.248[.300](.300)	-.338[.284](.270)	-.249[.283](.283)	-.337[.265](.250)	-.251[.261](.261)
	100	-.308[.211](.203)	-.252[.206](.206)	-.303[.202](.195)	-.248[.198](.198)	-.307[.194](.185)	-.254[.188](.188)
	200	-.277[.142](.140)	-.251[.141](.141)	-.274[.140](.138)	-.249[.139](.139)	-.275[.132](.129)	-.250[.130](.130)
	500	-.262[.089](.089)	-.252[.089](.089)	-.260[.088](.088)	-.250[.088](.088)	-.261[.084](.083)	-.251[.084](.084)
-.50	50	-.576[.291](.281)	-.499[.301](.301)	-.577[.283](.272)	-.502[.290](.290)	-.584[.268](.255)	-.511[.271](.270)
	100	-.548[.208](.203)	-.498[.209](.209)	-.550[.201](.195)	-.501[.201](.201)	-.547[.193](.188)	-.499[.193](.193)
	200	-.524[.144](.142)	-.501[.144](.144)	-.524[.141](.139)	-.501[.141](.141)	-.521[.136](.134)	-.498[.136](.136)
	500	-.511[.090](.089)	-.502[.090](.089)	-.510[.089](.089)	-.501[.089](.089)	-.509[.086](.086)	-.500[.086](.086)

Table 2.3

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SAR Errors - Group Interaction, $k = n^{0.5}$, REG-2

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.277[.403](.335)	.523[.223](.222)	.287[.395](.332)	.524[.223](.222)	.303[.354](.294)	.532[.194](.192)
	100	.375[.233](.197)	.512[.148](.148)	.377[.233](.198)	.511[.149](.149)	.384[.214](.180)	.515[.136](.136)
	200	.424[.160](.141)	.502[.116](.116)	.430[.152](.134)	.506[.111](.111)	.432[.143](.126)	.507[.104](.104)
	500	.454[.106](.096)	.502[.085](.085)	.455[.105](.095)	.502[.085](.085)	.456[.100](.090)	.502[.080](.080)
.25	50	-.082[.548](.437)	.291[.325](.322)	-.078[.541](.431)	.288[.318](.315)	-.061[.507](.401)	.296[.296](.293)
	100	.051[.345](.281)	.268[.220](.219)	.052[.342](.278)	.265[.218](.218)	.068[.309](.249)	.275[.196](.194)
	200	.129[.239](.206)	.259[.171](.171)	.127[.236](.201)	.256[.168](.168)	.131[.220](.184)	.257[.154](.153)
	500	.176[.160](.141)	.254[.126](.126)	.175[.161](.142)	.253[.127](.127)	.179[.153](.135)	.255[.120](.120)
.00	50	-.433[.679](.523)	.040[.419](.417)	-.432[.672](.514)	.034[.412](.411)	-.400[.620](.474)	.055[.378](.375)
	100	-.270[.448](.357)	.018[.288](.288)	-.260[.435](.347)	.020[.280](.280)	-.251[.409](.324)	.025[.263](.261)
	200	-.172[.315](.264)	.009[.223](.223)	-.171[.312](.261)	.008[.221](.221)	-.162[.295](.246)	.012[.209](.209)
	500	-.107[.215](.186)	.002[.167](.167)	-.106[.213](.185)	.002[.166](.166)	-.100[.199](.173)	.006[.156](.155)
-.25	50	-.758[.767](.575)	-.210[.487](.485)	-.746[.753](.567)	-.209[.483](.481)	-.723[.708](.527)	-.195[.448](.445)
	100	-.573[.534](.425)	-.227[.354](.353)	-.574[.530](.420)	-.233[.350](.350)	-.563[.490](.377)	-.228[.314](.313)
	200	-.467[.394](.329)	-.242[.282](.282)	-.466[.382](.315)	-.242[.271](.271)	-.455[.356](.291)	-.236[.250](.250)
	500	-.383[.263](.227)	-.240[.205](.204)	-.381[.263](.228)	-.246[.206](.206)	-.379[.250](.215)	-.245[.194](.194)
-.50	50	-1.057[.828](.614)	-.456[.553](.551)	-1.059[.828](.611)	-.467[.550](.549)	-1.040[.782](.566)	-.454[.505](.503)
	100	-.880[.612](.480)	-.481[.409](.409)	-.875[.598](.465)	-.482[.397](.396)	-.857[.562](.434)	-.472[.369](.368)
	200	-.753[.451](.374)	-.487[.325](.325)	-.751[.445](.369)	-.487[.320](.320)	-.746[.422](.344)	-.487[.299](.299)
	500	-.655[.308](.267)	-.493[.242](.242)	-.659[.311](.267)	-.497[.243](.243)	-.652[.294](.251)	-.492[.228](.228)

Table 2.4

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SAR Errors - Group Interaction, $k = n^{0.65}$, REG-2

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.435[.155](.140)	.504[.119](.119)	.440[.147](.134)	.507[.114](.114)	.441[.133](.119)	.506[.101](.101)
	100	.458[.110](.101)	.502[.091](.091)	.460[.105](.097)	.502[.087](.087)	.462[.094](.086)	.503[.077](.077)
	200	.477[.077](.073)	.503[.069](.068)	.475[.077](.073)	.501[.068](.068)	.478[.069](.065)	.503[.061](.061)
	500	.486[.053](.051)	.501[.050](.050)	.485[.053](.051)	.500[.049](.049)	.487[.050](.048)	.502[.046](.046)
.25	50	.148[.213](.186)	.257[.166](.166)	.151[.205](.179)	.257[.160](.160)	.154[.189](.162)	.257[.144](.144)
	100	.182[.156](.140)	.252[.129](.129)	.183[.151](.135)	.252[.124](.124)	.185[.139](.123)	.252[.112](.112)
	200	.209[.113](.105)	.252[.099](.099)	.211[.109](.102)	.253[.096](.096)	.209[.104](.095)	.250[.090](.090)
	500	.228[.076](.073)	.252[.070](.070)	.227[.077](.073)	.251[.070](.070)	.227[.072](.068)	.251[.066](.066)
.00	50	-.129[.253](.218)	.006[.205](.205)	-.127[.244](.208)	.006[.195](.195)	-.119[.222](.187)	.011[.175](.174)
	100	-.087[.191](.170)	.005[.159](.159)	-.088[.187](.165)	.003[.155](.154)	-.081[.169](.148)	.007[.138](.138)
	200	-.056[.144](.133)	.003[.126](.126)	-.056[.140](.128)	.002[.122](.122)	-.052[.131](.120)	.005[.114](.114)
	500	-.033[.101](.096)	-.001[.093](.093)	-.034[.100](.094)	-.001[.091](.091)	-.030[.093](.088)	.002[.086](.086)
-.25	50	-.395[.273](.231)	-.248[.227](.227)	-.389[.260](.220)	-.244[.216](.216)	-.384[.241](.201)	-.242[.196](.196)
	100	-.351[.218](.193)	-.244[.184](.184)	-.353[.215](.189)	-.247[.180](.180)	-.349[.197](.170)	-.246[.162](.162)
	200	-.319[.170](.156)	-.248[.149](.149)	-.321[.169](.154)	-.251[.147](.147)	-.317[.155](.140)	-.249[.134](.134)
	500	-.290[.122](.115)	-.249[.112](.112)	-.291[.122](.115)	-.251[.112](.112)	-.289[.114](.107)	-.250[.104](.104)
-.50	50	-.647[.276](.234)	-.499[.241](.241)	-.644[.269](.228)	-.499[.236](.236)	-.639[.252](.210)	-.497[.215](.215)
	100	-.616[.241](.212)	-.497[.205](.205)	-.609[.234](.207)	-.492[.200](.200)	-.610[.219](.189)	-.495[.183](.183)
	200	-.580[.193](.176)	-.499[.170](.170)	-.579[.191](.174)	-.499[.168](.168)	-.579[.179](.161)	-.500[.156](.156)
	500	-.547[.141](.133)	-.500[.129](.129)	-.545[.139](.131)	-.498[.128](.128)	-.544[.131](.124)	-.497[.121](.121)

Table 2.5

Replication of Table 2.2 for $\beta = (.5, .1, .1)$

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.395[.242](.218)	.499[.213](.213)	.396[.230](.205)	.497[.200](.200)	.404[.210](.187)	.501[.182](.182)
	100	.446[.150](.140)	.500[.138](.138)	.447[.149](.139)	.499[.137](.137)	.451[.135](.125)	.501[.123](.123)
	200	.474[.100](.096)	.500[.096](.096)	.475[.096](.093)	.500[.092](.092)	.476[.091](.087)	.500[.087](.087)
	500	.490[.059](.058)	.500[.058](.058)	.490[.059](.058)	.500[.058](.058)	.491[.056](.055)	.501[.055](.055)
.25	50	.137[.282](.258)	.246[.258](.258)	.145[.263](.241)	.251[.240](.240)	.152[.246](.225)	.253[.224](.224)
	100	.195[.182](.173)	.252[.172](.172)	.196[.173](.165)	.252[.163](.163)	.195[.162](.152)	.249[.151](.151)
	200	.224[.121](.118)	.250[.118](.118)	.224[.118](.115)	.251[.115](.115)	.226[.111](.108)	.251[.108](.108)
	500	.241[.072](.071)	.251[.071](.071)	.240[.072](.071)	.251[.071](.071)	.241[.070](.070)	.251[.070](.070)
.00	50	-.104[.297](.279)	.004[.286](.286)	-.106[.285](.264)	-.002[.270](.270)	-.098[.269](.250)	.004[.255](.255)
	100	-.059[.201](.192)	-.002[.193](.193)	-.058[.196](.187)	-.001[.188](.188)	-.054[.181](.173)	.002[.173](.173)
	200	-.027[.134](.131)	.001[.132](.132)	-.028[.133](.131)	-.002[.131](.131)	-.027[.124](.121)	-.001[.121](.121)
	500	-.010[.082](.081)	.002[.082](.082)	-.012[.083](.082)	-.001[.082](.082)	-.011[.079](.078)	-.001[.078](.078)
-.25	50	-.352[.305](.288)	-.253[.302](.302)	-.351[.294](.276)	-.254[.289](.289)	-.346[.279](.262)	-.252[.273](.273)
	100	-.302[.208](.202)	-.247[.205](.205)	-.304[.203](.196)	-.249[.199](.199)	-.304[.192](.185)	-.251[.187](.187)
	200	-.275[.142](.140)	-.250[.141](.141)	-.280[.139](.136)	-.255[.137](.137)	-.277[.134](.131)	-.252[.132](.132)
	500	-.261[.090](.089)	-.251[.089](.089)	-.261[.088](.087)	-.251[.088](.088)	-.259[.085](.085)	-.249[.085](.085)
-.50	50	-.591[.300](.286)	-.506[.307](.307)	-.592[.290](.276)	-.508[.294](.294)	-.588[.280](.265)	-.506[.282](.282)
	100	-.549[.207](.201)	-.500[.208](.208)	-.554[.203](.195)	-.506[.201](.201)	-.548[.193](.187)	-.500[.192](.192)
	200	-.524[.144](.142)	-.501[.144](.144)	-.522[.141](.140)	-.499[.142](.142)	-.523[.136](.134)	-.501[.136](.136)
	500	-.509[.091](.090)	-.500[.091](.091)	-.508[.090](.089)	-.499[.090](.090)	-.510[.087](.086)	-.500[.087](.087)

Table 2.6

Replication of Table 2.2 for $\sigma = 3$

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	50	.392[.243](.217)	.499[.210](.210)	.396[.234](.209)	.499[.202](.202)	.404[.212](.189)	.505[.182](.182)
	100	.449[.150](.141)	.501[.139](.139)	.449[.147](.137)	.499[.135](.135)	.452[.134](.125)	.501[.123](.123)
	200	.474[.098](.095)	.500[.094](.094)	.475[.097](.094)	.500[.093](.093)	.474[.091](.087)	.499[.087](.087)
	500	.489[.060](.059)	.499[.059](.059)	.490[.060](.059)	.500[.058](.058)	.490[.056](.055)	.500[.055](.055)
.25	50	.139[.282](.259)	.253[.257](.257)	.136[.271](.246)	.247[.243](.243)	.147[.249](.227)	.255[.224](.223)
	100	.196[.180](.172)	.250[.171](.171)	.195[.174](.165)	.249[.165](.165)	.202[.159](.152)	.253[.151](.151)
	200	.220[.120](.116)	.247[.116](.116)	.225[.119](.116)	.251[.116](.116)	.226[.110](.107)	.251[.107](.107)
	500	.240[.074](.073)	.250[.073](.073)	.240[.072](.071)	.251[.071](.071)	.240[.070](.070)	.250[.070](.070)
.00	50	-.114[.307](.285)	.001[.291](.291)	-.111[.297](.275)	.001[.280](.280)	-.109[.279](.256)	-.001[.259](.259)
	100	-.053[.195](.188)	.003[.189](.189)	-.053[.192](.184)	.001[.185](.185)	-.051[.177](.170)	.002[.171](.171)
	200	-.027[.134](.131)	-.001[.132](.132)	-.028[.132](.129)	-.002[.129](.129)	-.027[.123](.120)	-.002[.121](.121)
	500	-.010[.083](.083)	.001[.083](.083)	-.011[.082](.082)	-.001[.082](.082)	-.011[.079](.078)	-.001[.078](.078)
-.25	50	-.364[.312](.291)	-.258[.306](.305)	-.356[.298](.278)	-.250[.291](.291)	-.355[.286](.266)	-.252[.276](.276)
	100	-.300[.209](.203)	-.248[.207](.207)	-.302[.202](.195)	-.252[.199](.199)	-.297[.187](.181)	-.248[.183](.183)
	200	-.277[.143](.141)	-.252[.142](.142)	-.275[.139](.137)	-.249[.138](.138)	-.274[.134](.132)	-.249[.132](.132)
	500	-.259[.088](.087)	-.249[.087](.087)	-.262[.088](.087)	-.252[.087](.087)	-.260[.085](.085)	-.250[.085](.085)
-.50	50	-.593[.305](.290)	-.501[.312](.312)	-.596[.292](.276)	-.504[.296](.296)	-.599[.281](.263)	-.509[.280](.280)
	100	-.548[.207](.201)	-.503[.208](.208)	-.547[.198](.193)	-.502[.199](.199)	-.543[.192](.187)	-.499[.192](.192)
	200	-.522[.145](.143)	-.499[.145](.145)	-.525[.142](.140)	-.503[.142](.142)	-.522[.136](.134)	-.500[.136](.136)
	500	-.509[.091](.091)	-.500[.091](.091)	-.511[.089](.088)	-.502[.089](.089)	-.510[.086](.086)	-.501[.086](.086)

Table 2.7

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SMA Errors - Queen Contiguity, REG-1

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	100	.554[.154](.145)	.509[.418](.418)	.552[.151](.142)	.509[.318](.318)	.553[.149](.139)	.506[.140](.140)
	200	.527[.101](.097)	.501[.096](.096)	.528[.099](.095)	.502[.095](.095)	.527[.096](.093)	.501[.092](.092)
	500	.510[.059](.058)	.500[.058](.058)	.510[.059](.058)	.500[.058](.058)	.510[.059](.058)	.500[.058](.058)
.25	100	.302[.184](.176)	.256[.178](.178)	.301[.180](.173)	.255[.171](.171)	.292[.171](.166)	.247[.163](.163)
	200	.275[.121](.119)	.251[.117](.117)	.273[.120](.118)	.250[.116](.116)	.274[.115](.112)	.251[.111](.111)
	500	.259[.074](.073)	.250[.073](.073)	.261[.073](.072)	.252[.072](.072)	.260[.071](.071)	.251[.070](.070)
.00	100	.041[.204](.200)	-.001[.196](.196)	.040[.197](.193)	-.002[.188](.188)	.039[.187](.183)	-.001[.179](.179)
	200	.019[.136](.134)	-.002[.132](.132)	.022[.133](.131)	.002[.129](.129)	.021[.129](.127)	.001[.125](.125)
	500	.009[.083](.083)	.001[.083](.083)	.009[.082](.082)	.001[.081](.081)	.008[.081](.080)	.000[.080](.080)
-.25	100	-.214[.217](.214)	-.249[.208](.208)	-.217[.210](.208)	-.251[.202](.202)	-.222[.197](.195)	-.254[.189](.189)
	200	-.234[.145](.144)	-.250[.142](.142)	-.233[.143](.142)	-.249[.140](.140)	-.235[.138](.137)	-.251[.134](.134)
	500	-.245[.089](.089)	-.251[.089](.089)	-.245[.089](.089)	-.251[.089](.089)	-.245[.086](.086)	-.251[.086](.086)
-.50	100	-.472[.218](.216)	-.498[.209](.209)	-.475[.214](.212)	-.500[.205](.205)	-.479[.201](.200)	-.502[.193](.193)
	200	-.489[.149](.149)	-.501[.146](.146)	-.492[.146](.146)	-.503[.143](.143)	-.490[.139](.138)	-.500[.136](.136)
	500	-.495[.092](.092)	-.500[.091](.091)	-.495[.089](.089)	-.500[.089](.089)	-.496[.087](.087)	-.500[.086](.086)

Table 2.8

Empirical Mean[rmse](sd) of Estimators of ρ for SED Model with SMA Errors - Group Interaction, $k = n^{0.5}$, REG-1

		Normal Errors		Mixed Normal Errors		Log-Normal Errors	
ρ	n	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$	$\hat{\rho}_n$	$\hat{\rho}_n^{bc2}$
.50	100	.549[.129](.120)	.508[.128](.127)	.548[.126](.117)	.507[.124](.124)	.548[.121](.111)	.507[.118](.118)
	200	.534[.106](.100)	.503[.104](.104)	.534[.104](.098)	.502[.102](.102)	.533[.099](.094)	.502[.097](.097)
	500	.519[.078](.076)	.501[.078](.078)	.520[.079](.077)	.502[.079](.079)	.519[.077](.074)	.502[.076](.076)
.25	100	.309[.184](.174)	.254[.183](.183)	.310[.179](.169)	.256[.177](.177)	.306[.167](.158)	.253[.165](.165)
	200	.292[.148](.142)	.252[.147](.147)	.292[.147](.141)	.252[.146](.146)	.294[.140](.133)	.254[.138](.138)
	500	.277[.116](.113)	.252[.116](.116)	.276[.116](.113)	.252[.116](.116)	.275[.111](.108)	.251[.111](.111)
.00	100	.071[.234](.223)	.005[.234](.234)	.069[.228](.217)	.004[.227](.227)	.065[.211](.200)	.002[.209](.209)
	200	.051[.197](.190)	.001[.198](.198)	.053[.192](.185)	.004[.192](.192)	.052[.180](.172)	.004[.178](.178)
	500	.032[.152](.149)	-.001[.154](.154)	.032[.150](.146)	.001[.150](.150)	.034[.145](.141)	.003[.145](.145)
-.25	100	-.168[.281](.269)	-.246[.282](.282)	-.174[.269](.258)	-.251[.270](.270)	-.172[.254](.242)	-.246[.253](.253)
	200	-.194[.234](.227)	-.253[.236](.236)	-.187[.233](.225)	-.245[.233](.233)	-.192[.221](.214)	-.249[.222](.222)
	500	-.210[.188](.184)	-.248[.189](.189)	-.211[.188](.184)	-.249[.189](.189)	-.213[.178](.174)	-.251[.179](.179)
-.50	100	-.411[.321](.308)	-.500[.324](.324)	-.408[.315](.302)	-.495[.316](.316)	-.417[.294](.282)	-.503[.296](.296)
	200	-.427[.276](.266)	-.496[.276](.276)	-.427[.272](.262)	-.495[.273](.273)	-.436[.256](.247)	-.502[.257](.257)
	500	-.456[.219](.215)	-.501[.221](.221)	-.453[.223](.218)	-.498[.224](.224)	-.456[.213](.208)	-.501[.214](.214)

2.6 Conclusions

This chapter provide formal results for the asymptotic distribution as well as finite sample bias correction of the QML estimators for the SED model with autoregressive errors of order 1. Comparable results for moving average errors of order 1 has been illustrated as well.

Consistency and the asymptotic normality of the QML estimators has been addressed with a specific attention given to the effect of the degree of spatial dependence on the rate of convergence of the QML estimators of the model parameters. Specifically when the degree spatial dependence, h_n, grows with the sample size n, the QML estimator of the spatial parameter will have a lower rate of convergence (of $\sqrt{n/h_n}$) while the other QML estimators will have a $\sqrt{n}$-rate of convergence irrespective of the behaviour of h_n. Of the finite sample properties of spatial models, a specific attention has been given to the finite sample bias of the QML estimator of the spatial parameter as it enters the model in a highly non-linear manner and thus the estimation of it constitutes the main source of bias. Simulation studies indicate a prominent single direction bias in the estimation of the spatial parameter which in turn affects the subsequent inferences for the other model parameters. The severity of the bias increases as the spatial weights matrix becomes less sparse.

The finite sample results of this chapter demonstrate again that stochastic expansions (Rilstone et al., 1996) coupled with bootstrap (Yang, 2015b) provide a general and effective method for finite sample bias corrections of a non-linear estimator. The suggested theories and methodologies are likely to be appealing to both theorists as well as practitioners alike who are dealing with the SED model or any other regression model that considers a spatial dependence structure in the error process (like SARAR, panel SARAR, spatial dynamic panel data models, etc.).

CHAPTER 3

Improved Inferences for Spatial Regression Models

3.1 Introduction

The maximum likelihood (ML) or quasi-ML (QML) method is popular in the estimation and inference for spatial regression models[1]. However, the ML estimators or quasi-ML (QML) estimators of the spatial parameters can be quite biased [2] and hence the standard inferences for spatial effects and covariate effects, based on LM-statistics or t-statistics referring to the asymptotic standard normal distribution, can be seriously affected. Much effort has been devoted recently to the development of improved inference methods for the spatial econometrics models. However, most of the research has been focused on improving inferences for spatial effects in the form of point estimation [3] and testing [4]. Little or no

[1] Anselin, 1988; Anselin and Bera, 1998; Lee, 2004
[2] Bao and Ullah, 2007; Yang, 2015b; Liu and Yang, 2015a
[3] Bao and Ullah, 2007; Bao, 2013; Liu and Yang, 2015a; Yang, 2015b
[4] Baltagi and Yang, 2013a,b; Robinson and Rossi, 2014a,b; Yang, 2010; Yang, 2015a,b

attention has been paid to the development of improved inferences for the covariate effects in the spatial regression models.

Yang (2015a) proposed a general method for constructing 2nd-order accurate bootstrap LM tests for spatial effects, but the issue of improved inferences for covariate effects was not studied. Yang (2015b) proposed a general method for 3rd-order bias and variance corrections on non-linear estimators which are prone to finite sample bias, and argued that once the biases of non-linear estimators are corrected, the biases of covariate effects and error standard deviations become negligible. He demonstrated the effectiveness of the methods using the linear regression model with spatial lag dependence with results showing that a 2nd-order bias correction is largely sufficient. He further demonstrated that the 2nd-order or 3rd-order corrected t-statistics for spatial effect indeed improve upon the standard t-statistics greatly, but again, no study was carried out to test the performance of the t-statistics for covariate effects, and its improvements.

Evidently, in practical applications of spatial econometrics models, it is central to have a set of reliable inference methods for the covariate effects. In this chapter, we adopt the bias correction method of Yang (2015b) to propose methods that 'correct' the standard t-statistics for the regression coefficients. We demonstrate that by simply replacing the QML estimators of the spatial parameters by their bias corrected versions, the usual t-ratios for the regression coefficients can be greatly improved. We propose further corrections on the standard errors of the 'bias corrected' QML estimators of the regression coefficients, and the resulted t-ratios perform superbly, leading to much more reliable inferences. The proposed methods are simple and can be easily adopted by practitioners. We consider in detail three popular spatial regression models: the linear regression model with spatial error dependence (SED), that with a spatial lag dependence (SLD), and that with both SLD and SED, also referred to as the SARAR model in the literature.[5] Bias-correction on a single spatial estimator has been considered in detail in Yang (2015b) for the SLD model, and in Liu and Yang (2015a) for the SED model. Bias-corrections for the SARAR model have not been formally considered, although briefly discussed in Yang (2015b) under a general outline for bias corrections for a model with a vector of non-linear parameters.

The line-up for the chapter is as follows. Section 3.2 outlines the general method of bias correction on non-linear estimators, and the methods for constructing improved t-statistics for the linear parameters in the model. Sections 3-5 study in detail the improved inference methods for the regression coefficients for, respectively, the SED model, the SLD model, and the SARAR model. Each of Sections 3.3-3.5 is accompanied with a set of Monte Carlo simulation results. Section 3.6 concludes the chapter, and discuss further extensions of the proposed methodology.

3.2 Method of Bias Correction for Non-linear Estimation

From the discussions in the introduction, it is clear that the key for an improved inference for the regression coefficients is to bias correct the QML estimators of the spatial parameters in a spatial regression model. We now outline the method of bias correction on non-linear estimators, not necessarily the QML estimators of the spatial parameters. In studying the finite sample properties of a parameter estimator, say $\hat{\theta}_n$, defined as $\hat{\theta}_n = \arg\{\psi_n(\theta) = 0\}$ for a *joint estimating function* (JEF) $\psi_n(\theta)$, based on a sample of size n, Rilstone et al. (1996) developed a stochastic expansion from which a bias correction on $\hat{\theta}_n$ can be made. The vector of parameters θ may contain a set of *linear and scale parameters*, say α, and

[5] See Anselin and Bera (1998) and Anselin (2001) for excellent reviews on these models.

a few *non-linear parameters*, say δ, in the sense that given δ, the constrained estimator $\tilde{\alpha}_n(\delta)$ of the vector α possesses an explicit expression but the estimation of δ has to be done through numerical optimization. In this case, Yang (2015b) argued that it is more effective to work with the *concentrated estimating function* (CEF): $\tilde{\psi}_n(\delta) = \psi_n(\tilde{\alpha}_n(\delta), \delta)$, and to perform a stochastic expansion based on this CEF and hence bias corrections on the non-linear estimators defined by, $\hat{\delta}_n = \arg\{\tilde{\psi}_n(\delta) = 0\}$, which not only reduces the dimensionality of the bias correction problem (a multi-dimensional problem is reduced to a single-dimensional problem if δ is a scalar parameter), but also takes into account the additional variability from the estimation of the 'nuisance' parameters α.

Let $H_{rn}(\delta) = \nabla^r \tilde{\psi}_n(\delta), r = 1, 2, 3$, be the partial derivatives of $\tilde{\psi}_n(\delta)$, carried out sequentially and element-wise with respect to δ', $\tilde{\psi}_n \equiv \tilde{\psi}_n(\delta_0)$, $H_{rn} \equiv H_{rn}(\delta_0)$, $H_{rn}^{\circ} = H_{rn} - \mathrm{E}(H_{rn}), r = 1, 2, 3$, and $\Omega_n = -[\mathrm{E}(H_{1n})]^{-1}$. Yang (2015b) presents a set of sufficient conditions under which $\hat{\delta}_n$ possesses the following third-order stochastic expansion at δ_0, the true value of δ:

$$\hat{\delta}_n - \delta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(n^{-2}), \tag{3.1}$$

where, $a_{-s/2}$ represents terms of order $O_p(n^{-s/2})$ for $s = 1, 2, 3$:

$$\begin{aligned}
a_{-1/2} &= \Omega_n \tilde{\psi}_n, \\
a_{-1} &= \Omega_n H_{1n}^{\circ} a_{-1/2} + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2} \otimes a_{-1/2}), \\
a_{-3/2} &= \Omega_n H_{1n}^{\circ} a_{-1} + \tfrac{1}{2}\Omega_n H_{2n}^{\circ}(a_{-1/2} \otimes a_{-1/2}) \\
&\quad + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) \\
&\quad + \tfrac{1}{6}\Omega_n \mathrm{E}(H_{3n})(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}),
\end{aligned}$$

with $\otimes$ denoting the Kronecker product.

The key difference between the CEF-based and JEF-based expansions is that

$\mathrm{E}[\tilde{\psi}_n(\delta_0)] \neq 0$ in general, but $\mathrm{E}[\psi_n(\theta_0)] = 0$, which allows a CEF-based bias correction to be derived under a more relaxed condition. Thus, a third-order expansion for the bias of $\hat{\delta}_n$ takes the form:

$$\mathrm{Bias}(\hat{\delta}_n) = b_{-1} + b_{-3/2} + O(n^{-2}), \tag{3.2}$$

where $b_{-1} = \mathrm{E}(a_{-1/2} + a_{-1})$ and $b_{-3/2} = \mathrm{E}(a_{-3/2})$, being respectively the second- and third-order biases of $\hat{\delta}_n$.

If an estimator $\hat{b}_{-1}$ of b_{-1} is available such that $\mathrm{Bias}(\hat{b}_{-1}) = O(n^{-3/2})$, then a second-order bias corrected estimator of δ is, $\delta_n^{\mathrm{bc2}} = \hat{\delta}_n - \hat{b}_{-1}$.

If estimators $\hat{b}_{-1}$ and $\hat{b}_{-3/2}$ of both b_{-1} and $b_{-3/2}$ are available such that $\mathrm{Bias}(\hat{b}_{-1}) = O(n^{-2})$ and $\mathrm{Bias}(\hat{b}_{-3/2}) = O(n^{-2})$, we have a third-order bias corrected estimator of δ as, $\delta_n^{\mathrm{bc3}} = \hat{\delta}_n - \hat{b}_{-1} - \hat{b}_{-3/2}$.

An obvious approach for finding the feasible corrections $\hat{b}_{-1}$ and $\hat{b}_{-3/2}$ is to first find the analytical expressions for b_{-1} and $b_{-3/2}$ and then plugging in $\hat{\theta}_n$ for θ_0. This approach is generally not feasible for two reasons: first, it is often difficult to find these analytical expressions even for known error distributions, and second, even if these expressions are available, it may involve higher-order moments of the errors if they are non-normal, for which estimation may be unstable numerically. To overcome this difficulty, Yang (2015b) proposed a simple and yet very effective bootstrap method to estimate the relevant expected values.

Suppose that the model under consideration takes the form $g(Z_n, \theta_0) = e_n$, and that the key quantities $\tilde{\psi}_n$ and H_{rn} can be expressed as $\tilde{\psi}_n \equiv \tilde{\psi}_n(e_n, \theta_0)$ and $H_{rn} \equiv H_{rn}(e_n, \theta_0), r = 1, 2, 3$. Let $\hat{e}_n = g(Z_n, \hat{\theta}_n)$ be the vector of estimated residuals based on the original data, and $\hat{\mathcal{F}}_n$ be the empirical distribution function (EDF) of $\hat{e}_n$ (centred). When δ is a scalar parameter, the bootstrap estimates of the quantities in the bias terms are:

$$\widehat{\mathrm{E}}(\tilde{\psi}_n^i H_{rn}^j) = \mathrm{E}^*[\tilde{\psi}_n^i(\hat{e}_n^*, \hat{\theta}_n) H_{rn}^j(\hat{e}_n^*, \hat{\theta}_n)], \quad i, j = 0, 1, 2, \ldots, \quad r = 1, 2, 3, \tag{3.3}$$

where E^* denotes the expectation with respect to $\hat{\mathcal{F}}_n$, and $\hat{e}_n^*$ is a vector of n random draws from $\hat{\mathcal{F}}_n$. To make (3.3) practically feasible, the following procedure can be followed.

Bootstrap Algorithm 1 (BA-1):

1. *Compute* $\hat{\theta}_n$ *defined by JEF,* $\hat{e}_n = g(Z_n, \hat{\theta}_n)$*, and EDF* $\hat{\mathcal{F}}_n$ *of the centred* $\hat{e}_n$*;*

2. *Draw a random sample of size* n *from* $\hat{\mathcal{F}}_n$*. Denote the re-sampled vector by* $\hat{e}_{n,b}^*$*,*

3. *Compute* $\tilde{\psi}_n(\hat{e}_{n,b}^*, \hat{\theta}_n)$ *and* $H_{rn}(\hat{e}_{n,b}^*, \hat{\theta}_n), r = 1, 2, 3$*;*

4. *Repeat steps 2.-3. for* B *times, to give approximate bootstrap estimates as,*

$$\mathrm{E}^*[\tilde{\psi}_n^i(\hat{e}_n^*, \hat{\theta}_n) H_{rn}^j(\hat{e}_n^*, \hat{\theta}_n)] \doteq \tfrac{1}{B}\textstyle\sum_{b=1}^B \tilde{\psi}_n^i(\hat{e}_{n,b}^*, \hat{\theta}_n) H_{rn}^j(\hat{e}_{n,b}^*, \hat{\theta}_n),$$

for $i, j = 0, 1, 2, \ldots, \quad r = 1, 2, 3.$

The approximations in the last step can be made arbitrarily accurate by choosing an arbitrarily large B. Yang (2015b) shows that under certain conditions:

$$\begin{aligned}
\mathrm{Bias}(\delta_n^{\mathrm{bc2}}) &= \mathrm{Bias}(\hat{\delta}_n) - \mathrm{E}(\hat{b}_{-1}) = -\mathrm{Bias}(\hat{b}_{-1}) + O(n^{-3/2}) = O(n^{-3/2}), \text{ and} \\
\mathrm{Bias}(\delta_n^{\mathrm{bc3}}) &= \mathrm{Bias}(\hat{\delta}_n) - \mathrm{E}(\hat{b}_{-1}) - \mathrm{E}(\hat{b}_{-3/2}) \\
&= -\mathrm{Bias}(\hat{b}_{-1}) - \mathrm{Bias}(\hat{b}_{-3/2}) + O(n^{-2}) = O(n^{-2}).
\end{aligned}$$

When δ becomes a vector, the non-stochastic and stochastic quantities are mixed in b_{-1} and $b_{-3/2}$. In this case, Yang (2015b) proposed that instead of going through the algebraic procedure to separate the two types of quantities so that the expectations of various quantities can be bootstrapped in one round, the above bootstrap procedure can be revised as follows.

Bootstrap Algorithm 2 (BA-2):

1. *Draw B independent random samples, $\{\hat{e}^*_{n,b},\ b = 1, 2, \ldots, B\}$, from $\hat{\mathcal{F}}_n$,*

2. *Calculate the bootstrap estimates of $E(H_{1n})$ and $E(H_{2n})$,*

$$\widehat{\mathrm{E}}(H_{1n}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{n} H_{1n}(\hat{e}^*_{n,b}, \hat{\theta}_n) \text{ and } \widehat{\mathrm{E}}(H_{2n}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{n} H_{2n}(\hat{e}^*_{n,b}, \hat{\theta}_n)$$

3. *Based on the bootstrap estimates $\hat{\Omega}_n = -\widehat{\mathrm{E}}^{-1}(H_{1n})$ and $\widehat{\mathrm{E}}(H_{2n})$, calculate the bootstrap estimate of, e.g., $\mathrm{E}[H^{\circ}_{2n}(a_{-1/2} \otimes a_{-1/2})]$, as*

$$\tfrac{1}{B}\textstyle\sum_{b=1}^{n} \big\{[H_{2n}(\hat{e}^*_{n,b}, \hat{\theta}_n) - \widehat{\mathrm{E}}(H_{2n})][\hat{\Omega}_n\tilde{\psi}_n(\hat{e}^*_{n,b}, \hat{\theta}_n) \otimes \hat{\Omega}_n\tilde{\psi}_n(\hat{e}^*_{n,b}, \hat{\theta}_n)]\big\}.$$

The other quantities can be handled in a similar manner. This is essentially a two-round bootstrap procedure as it runs the iterations $b = 1, 2, \ldots, B$ two times, based on the same sequence of bootstrap samples. Computationally it is slightly more demanding, but algebraically it is much simpler and thus easier to code. As noted by Yang (2015b), these procedures are time-efficient as the re-estimation of the parameters in the bootstrap process is avoided.

Inferences following bias correction: There are mainly two types of inferences that could benefit from the bias corrections on the non-linear estimators: one is the inference for the non-linear parameters, and the other for the linear parameters. In the framework of linear regressions with spatial dependence, the spatial parameters are the non-linear parameters, and the regression coefficients are the linear parameters. Improved tests for spatial effects have been considered by Baltagi and Yang (2013a,b), Robinson and Rossi (2014a,b), Yang (2010), and Yang (2015a). However, the issue of improved inferences for the regression coefficients has not been considered.

To fix the idea, we focus on the 2nd-order bias corrected $\hat{\delta}_n$, the $\hat{\delta}_n^{\rm bc2}$. Let $\hat{\alpha}_n \equiv \tilde{\alpha}_n(\hat{\delta}_n)$ and $\hat{\alpha}_n^{\rm bc} \equiv \tilde{\alpha}_n(\hat{\delta}_n^{\rm bc2})$, and $\hat{\theta}_n = (\hat{\alpha}_n', \hat{\delta}_n')'$ and $\hat{\theta}_n^{\rm bc} = (\hat{\alpha}_n^{\rm bc\prime}, \hat{\delta}_n^{\rm bc2\prime})'$. Yang (2015b) argued that estimation of the non-linear parameter is the main source of bias and once the non-linear estimator is bias corrected the resulting linear estimators would be nearly unbiased. Let $\Omega_n(\theta_0)$ be the asymptotic variance-covariance (VC) matrix of $\hat{\alpha}_n$. Then, an asymptotic t-statistic for inference for $c_0'\alpha_0$, a linear contrast of α_0, has the familiar form:

$$t_n = (c_0'\hat{\alpha}_n - c_0'\alpha_0)/\sqrt{c_0'\Omega_n(\hat{\theta}_n)c_0}.$$

Simply replacing $\hat{\theta}_n$ by $\hat{\theta}_n^{\rm bc}$, a possibly improved t-statistic results:

$$t_n^{\rm bc} = (c_0'\hat{\alpha}_n^{\rm bc} - c_0'\alpha_0)/\sqrt{c_0'\Omega_n(\hat{\theta}_n^{\rm bc})c_0}.$$

The statistic $t_n^{\rm bc}$ is not fully 2nd-order corrected as it uses the asymptotic variance of $\hat{\alpha}_n$ evaluated at $\hat{\theta}_n^{\rm bc}$. Further, the estimator $\hat{\alpha}_n^{\rm bc}$ is also not fully 2nd-order bias corrected, although it can easily be made so. Let $\hat{\alpha}_n^{\rm bc2}$ be the 2nd-order bias corrected $\hat{\alpha}_n$ or $\hat{\alpha}_n^{\rm bc}$. Let $\Omega_n^{\rm bc2}(\theta_0)$ be the 2nd-order variance of $\hat{\alpha}_n^{\rm bc2}$, and $\hat{\Omega}_n^{\rm bc2}$ be its consistent estimate. A fully 2nd-order corrected t-statistic, using a 2nd-order bias corrected estimator and its 2nd-order variance, is thus:

$$t_n^{\rm bc2} = (c_0'\hat{\alpha}_n^{\rm bc2} - c_0'\alpha_0)/\sqrt{c_0'\hat{\Omega}_n^{\rm bc2}c_0}.$$

Typically, $\Omega_n^{\rm bc2}(\theta_0)$ does not have an explicit expression, but the bootstrap methods described above can be extended to give a consistent estimate of it. See the subsequent sections for details.

3.3 Improved Inferences for the SED Model

In this section, we study the inference methods for the regression coefficients of the SED model. First, in Section 3.3.1, we outline the inferences based on the asymptotic distribution of the QML estimators of the model parameters, then in Section 3.3.2 we outline the method of bias correcting the QML estimator of the spatial parameter, and then in Section 3.3.3 we present the improved inference methods. To assess the finite sample performance of the asymptotic and improved inferences, Monte Carlo results are presented in Section 3.3.4.

3.3.1 Asymptotic inference

Consider the following linear regression model with spatial error dependence (SED), where the SED is specified as a spatial autoregressive (SAR) process:

$$Y_n = X_n\beta + u_n, \quad u_n = \rho W_n u_n + \epsilon_n, \tag{3.4}$$

where Y_n is an $n\times 1$ vector of observations on the dependent variable corresponding to n spatial units, X_n is an $n \times k$ matrix containing the values of k exogenous regressors, W_n is an $n \times n$ spatial weight matrix that summarises the interactions among the spatial units, ϵ_n is an $n \times 1$ vector of independent and identically distributed (iid) disturbances with mean zero and variance σ^2, ρ is the *spatial parameter*, and β denotes the $k \times 1$ vector of regression coefficients. The SED model specific terminology follows that of Chapter 2.

Using the asymptotic Var-Cov given in Corollary 2.1, inference for $c_0'\beta_0$ is carried out based on the following t-ratio:

$$t_{\texttt{SED}} = \frac{c_0'\hat{\beta}_n - c_0'\beta_0}{\sqrt{\hat{\sigma}_n^2 c_0'(X_n'\hat{A}_n'\hat{A}_n X_n)^{-1}c_0}}, \tag{3.5}$$

where c_0 represents a linear contrast of the regression coefficients and $\hat{A}_n = I_n - \hat{\rho}_n W_n$. The t-ratio, $t_{\texttt{SED}}$, is asymptotically $N(0,1)$, and hence inferences concerning β_0 are carried out by referring to the standard normal critical values.

Monte Carlo experiments in Chapter 2 show that $\hat{\rho}_n$ can be seriously downward biased but the bias of $\hat{\rho}_n$ does not spillover much to $\hat{\beta}_n$. This means that the existence of spatial dependence in the regression errors does not affect much the point estimation of the regression coefficients in terms of consistency and finite sample bias. However, it does spill over to the estimate of $\text{Var}(\hat{\beta}_n)$. First, the downward bias of $\hat{\rho}_n$ causes $\hat{\sigma}_n^2$ to be downward biased when n is not large (e.g., 50). Second, from the expression:

$$\begin{aligned} X_n'\hat{B}_n'\hat{B}_nX_n &= X_n'A_n'A_nX_n - (\hat{\rho}_n - \rho_0)X_n'(W_n'A_n + A_n'W_n)X_n \\ &\quad + (\hat{\rho}_n - \rho_0)^2 X_n'W_n'W_nX_n, \end{aligned}$$

we see that the severe bias of $\hat{\rho}_n$ may cause $X_n'\hat{B}_n'\hat{B}_nX_n$ to be severely biased for the estimation of $X_n'A_n'A_nX_n$. For example when $X_n'(W_n'A_n + A_n'W_n)X_n \geq 0$ (in matrix sense),[6] $X_n'\hat{B}_n'\hat{B}_nX_n$ tends to overestimate $X_n'A_n'A_nX_n$, and hence, $\hat{\sigma}_n^2 c_0'(X_n'\hat{B}_n'\hat{B}_nX_n)^{-1}c_0$ tends to underestimate $\text{Var}(c_0'\hat{\beta}_n)$, which makes $t_{\texttt{SED}}$ much more variable than $N(0,1)$ and inferences for β_0 based on $t_{\texttt{SED}}$ defined in (3.5) unreliable. Our Monte Carlo results confirm this point.

3.3.2 Improved inferences for regression coefficients

Following the bias correction results of Chapter 2, by simply replacing $\hat{\rho}$ in $t_{\texttt{SED}}$ defined in (3.5) by $\hat{\rho}_n^{\text{bc2}}$, the second-order bias corrected $\hat{\rho}$, we obtain the following

[6]When W follows the `Group Interaction` scheme, this occurs as long as $(\iota_{n_r}' X_{jr})^2 \geq \frac{n_r - 1 + \rho_0}{(n_r-1)(1-\rho_0)+\rho_0} X_{jr}'X_{jr}$, where n_r is the size of the rth group and X_{jr} contains rth group values of the jth regressor.

potentially improved statistic:

$$t_{\texttt{SED}}^{\text{bc}} = \frac{c_0'\hat{\beta}_n^{\text{bc}} - c_0'\beta_0}{\sqrt{\hat{\sigma}_n^{2,\text{bc}} c_0'(X_n'\hat{B}_n^{\text{bc2}\prime}\hat{B}_n^{\text{bc2}}X_n)^{-1}c_0}}, \tag{3.6}$$

where $\hat{\beta}_n^{\text{bc}} = \tilde{\beta}_n(\hat{\rho}_n^{\text{bc2}})$, $\hat{\sigma}_n^{2,\text{bc}} = \tilde{\sigma}_n^2(\hat{\rho}_n^{\text{bc2}})$, and $\hat{B}_n^{\text{bc2}} = I_n - \hat{\rho}_n^{\text{bc2}}W_n$. Obviously, this statistic is not fully second-order bias corrected. However, Monte Carlo results presented in the next subsection show that it offers a huge improvement over $t_{\texttt{SED}}$. This confirms the point made at the end of Section 3.3.1. However, results also show that when n is not so large, there is still room for further improvement on $t_{\texttt{SED}}^{\text{bc}}$.

Let $F_n(\rho) = [X_n'A_n'(\rho)A_n(\rho)X_n]^{-1}X_n'A_n'(\rho)A_n(\rho)$ such that $\tilde{\beta}_n(\rho) = F_n(\rho)Y_n$ defined in (2.4) and denoting $\tilde{\beta}_n = \tilde{\beta}_n(\rho_0)$, and $\tilde{\beta}_n^{(r)} = \frac{d^r}{d\rho_0^r}\tilde{\beta}_n(\rho_0)$ and $F_n^{(r)} = \frac{d^r}{d\rho_0^r}F_n(\rho_0)$ for $r = 1, 2$, we have the following second-order stochastic expansion for $\hat{\beta}_n = \tilde{\beta}_n(\hat{\rho}_n)$:

$$\begin{aligned} \hat{\beta}_n - \beta_0 &= \tilde{\beta}_n - \beta_0 + \tilde{\beta}_n^{(1)}(\hat{\rho}_n - \rho_0) + \tfrac{1}{2}\tilde{\beta}_n^{(2)}(\hat{\rho}_n - \rho_0)^2 + O_p(n^{-3/2}) \\ &= b_{0n} + \mathrm{E}(\tilde{\beta}_n^{(1)})(a_{-1/2} + a_{-1}) + b_{1n}a_{-1/2} + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}_n^{(2)})a_{-1/2}^2 \\ &\quad + O_p(n^{-3/2}), \end{aligned} \tag{3.7}$$

where $b_{0n} = F_nA_n^{-1}\epsilon_n$, $b_{1n} = F_n^{(1)}A_n^{-1}\epsilon_n$, $\mathrm{E}(\tilde{\beta}_n^{(1)}) = F_n^{(1)}X_n\beta_0$, $\mathrm{E}(\tilde{\beta}_n^{(2)}) = F_n^{(2)}X_n\beta_0$, and $F_n^{(r)}$ are given in Appendix C. This leads immediately to, as a by-product of the bootstrap bias correction for $\hat{\rho}_n$, a fully 2nd-order bias corrected estimator $\hat{\beta}_n^{\text{bc2}}$ of β. Similarly, an expansion as (3.7) can easily be carried out for $\hat{\sigma}_n^2 = \tilde{\sigma}_n^2(\hat{\rho}_n)$, giving a fully 2nd-order bias corrected estimator $\hat{\sigma}_n^{2,\text{bc2}}$ of σ^2.[7] Finally, denoting $g(e_n, \theta_0) \equiv b_{0n} + \mathrm{E}(\tilde{\beta}_n^{(1)})(a_{-1/2} + a_{-1}) + b_{1n}a_{-1/2} + \frac{1}{2}\mathrm{E}(\tilde{\beta}_n^{(2)})a_{-1/2}^2$, the expansion

[7] As $\hat{\beta}_n^{\text{bc}}$ and $\hat{\beta}_n^{\text{bc2}}$ do not differ much, and $\hat{\sigma}_n^{2,\text{bc}}$ and $\hat{\sigma}_n^{2,\text{bc2}}$ also do not differ much, one can simply use $\hat{\beta}_n^{\text{bc}}$ and $\hat{\sigma}_n^{2,\text{bc}}$ in practical applications.

(3.7) leads to a second-order variance expansion:

$$\mathrm{Var}(\hat{\beta}_n) = \mathrm{Var}[g(e_n, \theta_0)] + O(n^{-2}).$$

Further it is easy to see $\mathrm{Var}(\hat{\beta}_n^{\mathrm{bc2}}) = \mathrm{Var}(\hat{\beta}_n) + O(n^{-2})$, and $\mathrm{Var}(\hat{\beta}_n^{\mathrm{bc}}) = \mathrm{Var}(\hat{\beta}_n) + O(n^{-2})$. Obviously, an explicit expression of the above is difficult to obtain, but is not needed as it can be easily estimated by the two-stage bootstrap procedure described below. Recall $a_{-1/2} = \Omega_n \tilde{\psi}_n$, and $a_{-1} = \Omega_n H_{1n}^{\circ} a_{-1/2} + \frac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2}^2) = \Omega_n \tilde{\psi}_n + \Omega_n^2 H_{1n} \tilde{\psi}_n + \frac{1}{2}\Omega_n^3 \mathrm{E}(H_{2n}) \tilde{\psi}_n^2$.

`Stage 1:` Compute $\hat{\theta}_n$ and the QML residuals $\hat{e}_n = \hat{\sigma}_n^{-1} \hat{B}_n (Y_n - X_n \hat{\beta}_n)$. Resample $\hat{e}_n$ to give $\hat{\rho}_n^{\mathrm{bc2}}$, and hence $\hat{\beta}_n^{\mathrm{bc2}}$ and $\hat{\sigma}_n^{2,\mathrm{bc2}}$, using the algorithm `BA-1` given in Section 3.2.

`Stage 2:` Update the QML residuals as $\hat{e}_n^{\mathrm{bc2}} = \hat{\sigma}_n^{\mathrm{bc2},-1} \hat{B}_n^{\mathrm{bc2}} (Y_n - X_n \hat{\beta}_n^{\mathrm{bc2}})$ and compute $g_{n,b}^* \equiv g(\hat{e}_{n,b}^{\mathrm{bc2}*}, \hat{\theta}_n^{\mathrm{bc2}})$ for $b = 1, \ldots, B$, where $\hat{e}_{n,b}^{\mathrm{bc2}*}$ is the bth bootstrap sample drawn from the EDF of $\hat{e}_n^{\mathrm{bc2}}$, and $\hat{\theta}_n^{\mathrm{bc2}} = (\hat{\beta}_n^{\mathrm{bc2}\prime}, \hat{\sigma}_n^{\mathrm{bc2}}, \hat{\rho}_n^{\mathrm{bc2}})'$. The bootstrap estimate of $\mathrm{Var}(\hat{\beta}_n^{\mathrm{bc2}})$, unbiased up to $O(n^{-3/2})$, is thus, $\widehat{\mathrm{Var}}(\hat{\beta}_n^{\mathrm{bc2}}) = \frac{1}{B}\sum_{b=1}^{B} g_{n,b}^* g_{n,b}^{*\prime} - \frac{1}{B}\sum_{b=1}^{B} g_{n,b}^* \frac{1}{B}\sum_{b=1}^{B} g_{n,b}^{*\prime}$.

We have a second-order 'bias corrected' t-statistic as follows:

$$t_{\mathtt{SED}}^{\mathrm{bc2}} = \frac{c_0' \hat{\beta}_n^{\mathrm{bc2}} - c_0' \beta_0}{\sqrt{c_0' \widehat{\mathrm{Var}}(\hat{\beta}_n^{\mathrm{bc2}}) c_0}}. \tag{3.8}$$

3.3.3 Monte Carlo experiments

Finite sample performance of $t_{\mathtt{SED}}$, $t_{\mathtt{SED}}^{\mathrm{bc}}$ and $t_{\mathtt{SED}}^{\mathrm{bc2}}$ is investigated and compared under the following data generating process (DGP):

$$Y_n = \iota_n \beta_0 + X_{1n}\beta_1 + X_{2n}\beta_2 + u_n, \quad u_n = \rho W_n u_n + \epsilon_n,$$

where X_{1n} and X_{2n} are the $n \times 1$ vectors containing the values of two fixed regressors. The parameters of the simulation are initially set to be as: $\beta = (5, 1, 1)'$, $\sigma^2 = 1$, ρ takes values form $\{-0.5, -0.25, 0, 0.25, 0.5\}$ and n take values from $\{50, 100, 200, 500\}$. Each set of Monte Carlo results is based on $M = 10,000$ Monte Carlo samples, and $B = 999 + \lfloor n^{0.75} \rfloor$ bootstrap samples within each Monte Carlo sample. The methods for generating X_n, W_n, and the errors are described in Appendix B.

Table 3.3.1 summarizes some results for $t_{\texttt{SED}}$ and $t^{\text{bc2}}_{\texttt{SED}}$ used for testing $H_0 : \beta_1 = \beta_2$. From the results we see that (i) as n increases, all tests converge in terms of rejection rates, (ii) it is indeed the case that the asymptotic test t_n can be very unreliable in the sense it rejects the true H_0 much too often than it supposes to. The test $t^{\text{bc}}_{\texttt{SED}}$ offers a huge reduction in size distortions, and when $n = 200$ and 500, its rejection rates become very close to their nominal levels. Nevertheless, when $n = 50$ or 100, we see from the tables that there is room for further improvement on $t^{\text{bc}}_{\texttt{SED}}$. The t-statistic $t^{\text{bc2}}_{\texttt{SED}}$ based on the second order corrected variance provides a further improvement on $t^{\text{bc}}_{\texttt{SED}}$ with the rejection rates quite close to the nominal levels even when n is not so large. The results show that the error distribution does not significantly affect the performance of the three tests. The true value of the spatial parameter has little effect on the performance of the two improved tests (except when $n = 50$), but has a significant effect on the asymptotic test: the size distortion gets larger when ρ changes from .5 to $-.5$. Furthermore, the size distortion for the asymptotic test is seen to be quite persistent, which remains to be at least 20% even when $n = 500$. The results (unreported for brevity) show that the tests under a more sparse spatial weight matrix generally have smaller size distortions.

Table 3.3.1 Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in SED Model

Group Interaction, REG2, $\sigma = 1$; Test: $1 = t_{\text{SED}}, 2 = t^{\text{bc}}_{\text{SED}}, 3 = t^{\text{bc2}}_{\text{SED}}$

ρ	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal			Normal Errors			Normal Mixture			Log normal		
		$n = 50$									$n = 200$								
.50	1	.232	.169	.088	.239	.173	.092	.237	.167	.086	.128	.073	.020	.139	.076	.024	.125	.068	.018
	2	.132	.078	.029	.136	.083	.032	.132	.079	.029	.107	.057	.014	.117	.060	.016	.109	.059	.014
	3	.113	.066	.024	.116	.072	.028	.112	.068	.024	.100	.053	.012	.108	.057	.014	.102	.052	.012
.25	1	.252	.185	.102	.254	.188	.104	.255	.186	.100	.153	.092	.030	.138	.080	.026	.141	.084	.028
	2	.133	.083	.038	.132	.083	.037	.136	.082	.035	.114	.060	.018	.116	.063	.017	.109	.058	.015
	3	.117	.076	.034	.120	.075	.033	.121	.073	.031	.108	.056	.016	.109	.057	.014	.101	.053	.013
.00	1	.259	.195	.105	.259	.193	.105	.265	.194	.104	.152	.093	.035	.157	.094	.034	.153	.098	.034
	2	.134	.084	.036	.136	.086	.040	.138	.085	.035	.110	.060	.018	.114	.062	.017	.116	.066	.019
	3	.125	.077	.034	.125	.079	.037	.126	.077	.033	.104	.056	.017	.108	.058	.016	.108	.060	.017
-.25	1	.270	.195	.110	.263	.193	.105	.267	.196	.108	.161	.101	.039	.166	.102	.037	.160	.099	.039
	2	.141	.093	.047	.140	.091	.042	.146	.092	.040	.114	.063	.020	.114	.064	.018	.111	.065	.020
	3	.135	.088	.044	.132	.086	.040	.137	.085	.037	.109	.060	.018	.109	.060	.018	.108	.062	.019
-.50	1	.260	.191	.102	.258	.189	.098	.262	.193	.103	.166	.102	.037	.167	.107	.038	.168	.106	.039
	2	.142	.096	.043	.145	.094	.044	.147	.096	.043	.112	.062	.017	.119	.064	.020	.118	.066	.020
	3	.136	.099	.033	.139	.099	.031	.141	.099	.031	.109	.060	.010	.104	.060	.011	.102	.061	.011
		$n = 100$									$n = 500$								
.50	1	.164	.103	.042	.170	.107	.042	.172	.106	.041	.123	.065	.018	.124	.067	.017	.120	.066	.017
	2	.124	.070	.023	.128	.074	.021	.129	.072	.018	.105	.054	.014	.109	.055	.013	.108	.055	.013
	3	.113	.062	.019	.115	.064	.017	.115	.062	.015	.101	.053	.012	.104	.051	.012	.103	.052	.010
.25	1	.190	.126	.054	.192	.127	.053	.192	.126	.055	.132	.074	.022	.126	.070	.019	.130	.072	.021
	2	.128	.076	.023	.127	.075	.021	.130	.074	.025	.107	.056	.015	.104	.053	.014	.107	.054	.015
	3	.117	.068	.020	.117	.067	.019	.119	.067	.020	.104	.053	.010	.101	.051	.010	.102	.052	.011
.00	1	.200	.133	.058	.197	.128	.056	.204	.133	.058	.132	.077	.024	.136	.077	.024	.134	.075	.024
	2	.124	.070	.024	.123	.072	.023	.126	.073	.025	.105	.057	.015	.107	.056	.014	.107	.056	.015
	3	.116	.064	.021	.114	.066	.021	.119	.070	.023	.103	.050	.011	.105	.051	.010	.103	.051	.010
-.25	1	.201	.132	.060	.204	.137	.059	.199	.129	.057	.135	.077	.023	.135	.076	.021	.133	.076	.021
	2	.124	.072	.027	.123	.071	.024	.117	.068	.023	.104	.056	.014	.104	.053	.013	.105	.056	.013
	3	.116	.067	.026	.115	.066	.022	.109	.063	.022	.102	.051	.011	.102	.050	.012	.102	.054	.013
-.50	1	.198	.137	.058	.195	.130	.057	.203	.133	.058	.137	.077	.023	.134	.077	.024	.133	.077	.024
	2	.118	.068	.024	.117	.069	.026	.120	.071	.025	.105	.058	.014	.100	.054	.015	.101	.055	.013
	3	.110	.065	.021	.111	.060	.020	.115	.068	.021	.104	.051	.010	.100	.051	.011	.101	.051	.010

3.4 Improved Inferences for the SLD Model

This section concerns the improved inference methods for the regression coefficients of the SLD model. Section 3.4.1 outlines the asymptotic results, and Section 3.4.2 the finite sample bias correction results. Section 3.4.3 presents the improved inference methods, and Section 3.4.4 presents Monte Carlo results.

3.4.1 QML estimation and asymptotic inference

The regression model with spatial lag dependence (SLD) takes the form:

$$Y_n = \lambda W_n Y_n + X_n \beta + \epsilon_n, \tag{3.9}$$

Letting $A_n(\lambda) = I_n - \lambda W_n$, the log-likelihood function of $\theta = (\beta', \sigma^2, \lambda)'$ is $\ell_n(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\lambda)| - \frac{1}{2\sigma^2}[A_n(\lambda)Y_n - X_n\beta]'[A_n(\lambda)Y_n - X_n\beta]$.

Given λ, $\ell_n(\theta)$ is maximised at $\tilde{\beta}_n(\lambda) = (X_n'X_n)^{-1}X_n'A_n(\lambda)Y_n$ and $\tilde{\sigma}_n^2(\lambda) = \frac{1}{n}Y_n'A_n'(\lambda)M_nA_n(\lambda)Y_n$, where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. These lead to the concentrated log-likelihood of λ as,

$$\ell_n^c(\lambda) = -\frac{n}{2}[\log(2\pi)+1] - \frac{n}{2}\log\hat{\sigma}_n^2(\lambda) + \log|A_n(\lambda)|. \tag{3.10}$$

Maximizing $\ell_n^c(\lambda)$ gives the unconstrained QML estimator $\hat{\lambda}_n$ of λ. The unconstrained QML estimators of β and σ^2 are thus, $\hat{\beta}_n \equiv \tilde{\beta}_n(\hat{\lambda}_n)$ and $\hat{\sigma}_n^2 \equiv \tilde{\sigma}_n^2(\hat{\lambda}_n)$. Write $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\lambda}_n)'$. Lee (2004) shows that $\hat{\theta}_n$ is asymptotically $N(\theta_0, \Sigma_n^{-1}\Gamma_n\Sigma_n^{-1})$, where

$$\Sigma_n = \begin{pmatrix} \frac{1}{\sigma_0^2}X_n'X_n & 0 & \frac{1}{\sigma_0}X_n'\eta_n \\ 0 & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) \\ \frac{1}{\sigma_0}\eta_n'X_n & \frac{1}{\sigma_0^2}\mathrm{tr}(G_n) & \eta_n'\eta_n + \mathrm{tr}(G_n^sG_n) \end{pmatrix},$$

$$\Gamma_n = \begin{pmatrix} 0 & \frac{1}{2\sigma_0^3}\gamma X_n'\iota_n & \frac{1}{\sigma_0}\gamma X_n' g_n \\ \frac{1}{2\sigma_0^3}\gamma\iota_n' X_n & \frac{n}{4\sigma_0^4}\kappa & \frac{1}{2\sigma_0^2}\gamma\iota_n'\eta_n + \frac{1}{2\sigma_0^2}\kappa\mathrm{tr}(G_n) \\ \frac{1}{\sigma_0}\gamma g_n' X_n & \frac{1}{2\sigma_0^2}\gamma\iota_n'\eta_n + \frac{1}{2\sigma_0^2}\kappa\mathrm{tr}(G_n) & \kappa g_n' g_n + 2\gamma g_n'\eta_n \end{pmatrix} + \Sigma_n,$$

ι_n, γ and κ are measures of skewness and excess kurtosis of $\epsilon_{n,i}$, respectively, $g_n = \mathrm{diag}(G_n)$, $G_n = G_n(\rho_0) = W_n A_n^{-1}(\lambda_0)$, $G_n^s = G_n + G_n'$ and $\eta_n = \sigma_0^{-1} G_n X_n \beta_0$.

Letting V_{n1} be the sub-matrix of $\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}$ corresponding to β and $\hat{V}_{n1}$ be its estimate, an asymptotic t-statistic for inferences for $c_0'\beta_0$ is thus,

$$t_{\texttt{SLD}} = \frac{c_0'\hat{\beta}_n - c_0'\beta_0}{\sqrt{c_0'\hat{V}_{n1}c_0}}, \tag{3.11}$$

which is asymptotically $N(0,1)$. Finite sample properties of $t_{\texttt{SLD}}$ is of interest.

As $\tilde{\beta}_n(\hat{\lambda}) = \beta_0 + (\lambda - \hat{\lambda})(X_n'X_n)^{-1}X_n'G_nX_n\beta_0 + o_p(1)$, any estimation bias of $\hat{\lambda}$ is quickly passed down to the QML estimator of $\beta_0$and thus the t-statistic computed using $\tilde{\beta}_n(\hat{\lambda})$ and the variance estimate $\hat{V}_{n1}$ can be unreliable.

3.4.2 Bias corrections

As an illustration to his general bias correction method, Yang (2015b) studied the SLD model in detail. Letting $\tilde{\psi}_n(\lambda) = \frac{\partial}{\partial\lambda}\ell_n^c(\lambda)$, where $\ell_n^c(\lambda)$ is given in (3.10), we have, $\tilde{\psi}_n(\lambda) = -h_n T_{0n}(\lambda) + h_n R_{1n}(\lambda)$, and

$$H_{1n}(\lambda) = -T_{1n}(\lambda) - R_{2n}(\lambda) + 2R_{1n}^2(\lambda), \tag{3.12}$$

$$H_{2n}(\lambda) = -2T_{2n}(\lambda) - 6R_{1n}(\lambda)R_{2n}(\lambda) + 8R_{1n}^3(\lambda), \tag{3.13}$$

$$H_{3n}(\lambda) = -6T_{3n}(\lambda) + 6R_{2n}^2(\lambda) - 48R_{1n}^2(\lambda)R_{2n}(\lambda) + 48R_{1n}^4(\lambda), \tag{3.14}$$

where $T_{rn}(\lambda) = n^{-1}\mathrm{tr}(G_n^{r+1}(\lambda)), r = 0, 1, 2, 3,\ G_n(\lambda) = W_n A_n^{-1}(\lambda)$,

$$R_{1n}(\lambda) = \frac{Y_n'A_n'(\lambda)M_nW_nY_n}{Y_n'A_n'(\lambda)M_nA_n(\lambda)Y_n} \quad \text{and} \quad R_{2n}(\lambda) = \frac{Y_n'W_n'M_nW_nY_n}{Y_n'A_n'(\lambda)M_nA_n(\lambda)Y_n}. \tag{3.15}$$

Bootstrap estimates of biases: The two key ratios can be written as:

$$R_{1n}(e_n, \theta_0) = \frac{e_n' M_n G_n e_n + e_n' M_n \eta_n}{e_n' M_n e_n},$$
$$R_{2n}(e_n, \theta_0) = \frac{e_n' G_n' M_n G_n e_n + 2e_n' G_n' M_n \eta_n + \eta_n' M_n \eta_n}{e_n' M_n e_n},$$

where $e_n = \sigma_0^{-1}\epsilon_n$. Hence, $\tilde{\psi}_n = \tilde{\psi}_n(e_n, \theta_0)$ and $H_{rn} = H_{rn}(e_n, \theta_0)$ $r = 1, 2, 3$ are expressed in terms of e_n and θ_0. So, the bias corrections are carried out using an estimate of $R_{1n}(e_n, \theta_0)$ and $R_{2n}(e_n, \theta_0)$. See Yang (2015b) for details. Let $\hat{\lambda}_n^{\mathrm{bc2}}$ be the second-order bias corrected $\hat{\lambda}_n$, and let $\hat{\beta}_n^{\mathrm{bc}} = \tilde{\beta}(\hat{\lambda}_n^{\mathrm{bc2}})$ and $\hat{\sigma}_n^{2,\mathrm{bc}} = \tilde{\sigma}_n^2(\hat{\lambda}_n^{\mathrm{bc2}})$.

3.4.3 Improved inferences for regression coefficients

Replacing $\hat{\lambda}_n$ by $\hat{\lambda}_n^{\mathrm{bc2}}$ in the definition of $t_{\mathtt{SLD}}$, we obtain:

$$t_{\mathtt{SLD}}^{\mathrm{bc}} = \frac{c_0' \hat{\beta}_n^{\mathrm{bc}} - c_0' \beta_0}{\sqrt{c_0' \hat{V}_{n1}^{\mathrm{bc}} c_0}}, \tag{3.16}$$

where $\hat{V}_{n1}^{\mathrm{bc}}$ is V_{n1} evaluated at $\hat{\lambda}_n^{\mathrm{bc2}}$, $\hat{\beta}_n^{\mathrm{bc}}$, $\hat{\sigma}_n^{2,\mathrm{bc}}$, $\hat{\gamma}_n^{\mathrm{bc}}$, and $\hat{\kappa}_n^{\mathrm{bc}}$.

Let $\mathbb{X}_n = X_n(X_n' X_n)^{-1}$. Now, to further improve $t_{\mathtt{SLD}}^{\mathrm{bc}}$, note that

$$\begin{aligned} \hat{\beta}_n - \beta_0 &= \tilde{\beta}_n - \beta_0 - (\hat{\lambda}_n - \lambda_0)\mathbb{X}_n' G_n X_n \beta_0 - (\hat{\lambda}_n - \lambda_0)\mathbb{X}_n' G_n \epsilon_n \\ &= \mathbb{X}_n' \big[\epsilon_n - (a_{-1/2} + a_{-1}) G_n X_n \beta_0 - a_{-1/2} G_n \epsilon_n\big] + O_p(n^{-3/2}). \end{aligned} \tag{3.17}$$

This leads immediately to a 2nd-order bias corrected estimator $\hat{\beta}_n^{\mathrm{bc2}}$ of β, and a second-order expansion for $\mathrm{Var}(\hat{\beta}_n)$ as,

$$\mathrm{Var}(\hat{\beta}_n) = \mathbb{X}_n' \mathrm{Var}\big[\epsilon_n - (a_{-1/2} + a_{-1}) G_n X_n \beta_0 - a_{-1/2} G_n \epsilon_n\big] \mathbb{X}_n + O(n^{-2}).$$

We have, $\mathrm{Var}(\hat{\beta}_n^{\mathrm{bc2}}) = \mathrm{Var}(\hat{\beta}_n) + O(n^{-2})$. An expansion can be carried out for

$\hat{\sigma}_n^2$ in terms of $\hat{\lambda}_n$, leading to a 2nd-order bias corrected estimator $\sigma_n^{2,\text{bc2}}$ of σ^2.[8] A two-stage bootstrap procedure can be followed to give a consistent estimate of $V = X_n' \text{Var}\big[\epsilon_n - (a_{-1/2} + a_{-1})G_n X_n \beta_0 - a_{-1/2} G_n \epsilon_n\big] X_n$: first, run the algorithm `BA-1` to give 2nd-order bias corrected estimators $\hat{\lambda}_n^{\text{bc2}}$, $\hat{\beta}_n^{\text{bc2}}$ and $\sigma_n^{2,\text{bc2}}$; then update the residuals and run the algorithm `BA-1` again using the updated residuals to give a sequence of bootstrap values for V, and hence the bootstrap estimate $\widehat{\text{Var}}(\hat{\beta}_n^{\text{bc2}})$ of $\text{Var}(\hat{\beta}_n)$. The resulted 2nd-order bias corrected t-statistic is:

$$t_{\texttt{SLD}}^{\text{bc2}} = \frac{c_0' \hat{\beta}_n^{\text{bc2}} - c_0' \beta_0}{\sqrt{c_0' \widehat{\text{Var}}(\hat{\beta}_n^{\text{bc2}}) c_0}}. \tag{3.18}$$

3.4.4 Monte Carlo experiments

Finite sample performance of $t_{\texttt{SLD}}$, $t_{\texttt{SLD}}^{\text{bc}}$ and $t_{\texttt{SLD}}^{\text{bc2}}$ is investigated under the DGP:

$$Y_n = \lambda W_n Y_n + \iota_n \beta_0 + X_{1n} \beta_1 + X_{2n} \beta_2 + \epsilon_n,$$

where all the quantities are generated as in those for the SED model. Parameters for the Monte Carlo simulation are also set to be the same before.

Table 3.4.1 summarises some empirical sizes of the tests $t_{\texttt{SLD}}$, $t_{\texttt{SLD}}^{\text{bc}}$ and $t_{\texttt{SLD}}^{\text{bc2}}$ when used for testing $H_0 : \beta_1 = \beta_2$ under the `Group Interaction` scheme. From the results we see that (i) as n increases, all tests converge in terms of sizes, (ii) it is indeed the case that the asymptotic test $t_{\texttt{SLD}}$ can be very unreliable in the sense that it rejects the true H_0 much too often than it supposes to. The test $t_{\texttt{SLD}}^{\text{bc2}}$ offers a huge reduction in size distortions, with the empirical sizes getting close to their nominal levels faster than in the SED case. Nevertheless, when $n = 50$, the results show that $t_{\texttt{SLD}}^{\text{bc}}$ needs further improvements, and indeed the test $t_{\texttt{SLD}}^{\text{bc2}}$ based on the second-order corrected variance offers the desired improvements.

[8] Again, $\hat{\beta}_n^{\text{bc}}$ and $\hat{\beta}_n^{\text{bc2}}$, and $\hat{\sigma}_n^{2,\text{bc}}$ and $\hat{\sigma}_n^{2,\text{bc2}}$ do not differ much. Hence in practical applications, one can use the simpler versions $\hat{\beta}_n^{\text{bc}}$ and $\hat{\sigma}_n^{2,\text{bc}}$.

Table 3.4.1 Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in SLD Model

`Group Interaction, REG2,` $\sigma = 1$; `Test:` $1 = t_{\mathtt{SLD}}, 2 = t^{\mathtt{bc}}_{\mathtt{SLD}}, 3 = t^{\mathtt{bc2}}_{\mathtt{SLD}}$

ρ	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal			Normal Errors			Normal Mixture			Log normal		
		$n = 50$									$n = 200$								
.50	1	.161	.095	.028	.162	.095	.029	.174	.120	.058	.109	.057	.013	.113	.059	.014	.125	.068	.019
	2	.113	.062	.016	.117	.068	.017	.142	.098	.049	.102	.051	.012	.104	.052	.011	.114	.061	.016
	3	.095	.045	.010	.100	.045	.010	.100	.054	.014	.093	.048	.011	.098	.048	.011	.100	.050	.010
.25	1	.156	.095	.028	.160	.095	.026	.171	.107	.041	.115	.060	.013	.111	.056	.012	.123	.063	.014
	2	.113	.062	.014	.117	.065	.015	.139	.085	.033	.106	.055	.012	.103	.050	.011	.112	.054	.012
	3	.095	.044	.009	.093	.045	.010	.098	.050	.012	.100	.049	.010	.097	.046	.010	.100	.047	.010
.00	1	.158	.097	.030	.157	.090	.028	.139	.073	.021	.115	.056	.012	.115	.061	.014	.114	.061	.015
	2	.115	.065	.018	.116	.062	.017	.110	.060	.015	.106	.051	.010	.104	.053	.011	.105	.054	.012
	3	.100	.048	.012	.093	.049	.012	.099	.053	.015	.099	.046	.009	.099	.048	.010	.098	.049	.010
-.25	1	.163	.096	.033	.161	.099	.032	.122	.067	.019	.112	.057	.015	.112	.058	.012	.111	.058	.011
	2	.117	.068	.020	.124	.069	.020	.100	.049	.012	.105	.054	.013	.105	.053	.010	.108	.055	.011
	3	.095	.052	.014	.102	.053	.015	.100	.050	.010	.100	.049	.012	.100	.050	.010	.103	.052	.009
-.50	1	.167	.100	.033	.161	.099	.034	.113	.062	.017	.119	.065	.016	.108	.056	.012	.105	.057	.014
	2	.124	.069	.020	.126	.074	.022	.094	.046	.012	.108	.058	.013	.107	.056	.013	.099	.052	.011
	3	.099	.055	.016	.106	.051	.015	.103	.050	.011	.099	.051	.011	.099	.049	.011	.100	.051	.011
		$n = 100$									$n = 500$								
.50	1	.131	.070	.018	.127	.067	.017	.133	.077	.027	.106	.053	.011	.111	.057	.014	.107	.056	.012
	2	.105	.055	.013	.103	.051	.011	.117	.066	.023	.101	.048	.009	.104	.051	.012	.102	.052	.011
	3	.098	.054	.010	.099	.049	.018	.093	.048	.010	.099	.050	.010	.098	.048	.011	.097	.048	.010
.25	1	.127	.068	.019	.130	.073	.019	.145	.087	.024	.110	.060	.012	.109	.057	.013	.109	.054	.011
	2	.103	.052	.014	.109	.056	.014	.120	.069	.018	.103	.055	.011	.100	.051	.010	.104	.050	.010
	3	.096	.049	.010	.093	.050	.010	.095	.046	.009	.098	.050	.010	.099	.050	.009	.099	.049	.010
.00	1	.133	.070	.019	.130	.071	.017	.128	.073	.018	.107	.055	.012	.109	.058	.012	.108	.055	.013
	2	.105	.054	.013	.109	.057	.012	.111	.059	.013	.101	.051	.011	.102	.053	.010	.101	.051	.012
	3	.100	.050	.010	.099	.050	.009	.099	.052	.011	.099	.050	.010	.097	.049	.009	.100	.049	.011
-.25	1	.133	.071	.020	.134	.077	.021	.130	.065	.013	.103	.052	.014	.107	.054	.012	.105	.054	.012
	2	.109	.054	.014	.112	.060	.015	.110	.051	.009	.097	.048	.012	.100	.049	.011	.099	.050	.010
	3	.100	.046	.011	.099	.047	.010	.103	.049	.010	.099	.050	.011	.099	.049	.010	.099	.050	.010
-.50	1	.128	.071	.017	.132	.074	.018	.113	.057	.012	.105	.056	.013	.108	.056	.011	.107	.054	.011
	2	.106	.057	.013	.112	.060	.012	.094	.044	.008	.098	.052	.011	.102	.050	.009	.100	.050	.009
	3	.099	.050	.010	.099	.052	.011	.100	.049	.010	.099	.050	.010	.099	.049	.010	.100	.049	.009

3.5 Improved Inferences for the SARAR Model

In this section, finite sample bias of the SARAR model and improved inference methods for the regression coefficients is given. Neither issue has been formally considered due to its complexity, and hence the results in this section constitute important contributions, in particular considering the fact that the SARAR model is more versatile and practically more useful than either the SLD or SED models. Section 3.5.1 outlines the QML estimation and the asymptotic inference method. Section 3.5.2 presents detailed results for bias correcting the QML estimators of the spatial parameters. Section 3.5.3 presents improved inference methods for regression coefficients. Section 3.5.4 presents Monte Carlo results.

3.5.1 QML estimation and asymptotic inference

Combining the SED and SLD models considered above, we have the spatial autoregressive model with autoregressive errors, also known as the SARAR model:

$$Y_n = \lambda W_{1n} Y_n + X_n \beta + u_n, \quad u_n = \rho W_{2n} u_n + \epsilon_n. \tag{3.19}$$

Let $\delta = (\lambda, \rho)'$. Gaussian log-likelihood function of $\theta = (\beta', \sigma^2, \lambda, \rho)'$ is $\ell_n(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log|A_{1n}(\lambda)| + \log|A_{2n}(\rho)| - \frac{1}{2\sigma^2}\epsilon_n'(\beta,\delta)\epsilon_n(\beta,\delta)$, where $\epsilon_n(\beta,\delta) = Y_n(\delta) - X_n(\rho)\beta$, $A_{1n}(\lambda) = I_n - \lambda W_{1n}$, $A_{2n}(\rho) = I_n - \rho W_{2n}$, $X_n(\rho) = A_{2n}(\rho)X_n$ and $Y_n(\delta) = A_{2n}(\rho)A_{1n}(\lambda)Y_n$. The constrained QML estimators of β and σ^2, are $\tilde{\beta}_n(\delta) = [X_n'(\rho)X_n(\rho)]^{-1}X_n'(\rho)Y_n(\delta)$ and $\tilde{\sigma}_n^2(\delta) = \frac{1}{n}Y_n'(\delta)M_n(\rho)Y_n(\delta)$, where $M_n(\rho) = I_n - A_{2n}(\rho)X_n[X_n'A_{2n}'(\rho)A_{2n}(\rho)X_n]^{-1}X_n'A_{2n}'(\rho)$. Then, the concentrated Gaussian log-likelihood function for δ is,

$$\ell_n^c(\delta) = -\frac{n}{2}[\ln(2\pi)+1] - \frac{n}{2}\ln(\hat{\sigma}_n^2(\delta)) + \ln|A_{1n}(\lambda)| + \ln|A_{2n}(\rho)|. \tag{3.20}$$

Maximizing (3.20) gives the QML estimator $\hat{\delta}_n$ of δ, and thus the QML estimators of β and σ^2 as $\hat{\beta}_n \equiv \tilde{\beta}_n(\hat{\delta}_n)$ and $\hat{\sigma}_n^2 \equiv \tilde{\sigma}_n^2(\hat{\delta}_n)$. Write $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\delta}_n')'$. The concentrated score function upon dividing by n is,

$$\tilde{\psi}_n(\delta) = \begin{cases} -\frac{1}{n}\mathrm{tr}(G_{1n}(\lambda)) + \frac{Y_n'(\delta)M_n(\rho)\bar{B}_n(\delta)Y_n(\delta)}{Y_n'(\delta)M_n(\rho)Y_n(\delta)}, \\ -\frac{1}{n}\mathrm{tr}(G_{2n}(\rho)) + \frac{Y_n'(\delta)M_n(\rho)G_{2n}(\rho)M_n(\rho)Y_n(\delta)}{Y_n'(\lambda)M_n(\rho)Y_n(\delta)}, \end{cases} \tag{3.21}$$

where $G_{1n}(\lambda) = W_{1n}A_{1n}^{-1}(\lambda)$, $G_{2n}(\rho) = W_{2n}A_{2n}^{-1}(\rho)$ and $\bar{B}_n(\delta) = A_{2n}(\rho)G_{1n}(\lambda)A_{2n}^{-1}(\rho)$.

Jin and Lee (2013) shows that under some regularity conditions, $\hat{\theta}_n$ is asymptotically normal with mean θ_0 and asymptotic VC matrix $\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}$, where

$$\Sigma_n = \begin{pmatrix} \frac{1}{\sigma_0^2}X_n'A_{2n}'A_{2n}X_n & 0 & \frac{1}{\sigma_0}X_n'A_{2n}'\mu_n & 0 \\ 0 & \frac{n}{2\sigma_0^4} & \frac{1}{\sigma_0^2}\mathrm{tr}(\bar{B}_n) & \frac{1}{\sigma_0^2}\mathrm{tr}(G_{2n}) \\ \frac{1}{\sigma_0}\mu_n'A_{2n}X_n & \frac{1}{\sigma_0^2}\mathrm{tr}(\bar{B}_n) & \mu_n'\mu_n + \mathrm{tr}(\bar{B}_n^s\bar{B}_n) & \mathrm{tr}(G_{2n}^s\bar{B}_n) \\ 0 & \frac{1}{\sigma_0^2}\mathrm{tr}(G_{2n}) & \mathrm{tr}(G_{2n}^s\bar{B}_n) & \mathrm{tr}(G_{2n}^sG_{2n}) \end{pmatrix},$$

$$\Gamma_n = \begin{pmatrix} 0 & \frac{\gamma}{2\sigma_0^3}X_n'A_{2n}'\iota_n & \frac{\gamma}{\sigma_0}X_n'A_{2n}'\bar{b}_n & \frac{\gamma}{\sigma_0}X_n'A_{2n}'g_{2n} \\ \sim & \frac{n\kappa}{4\sigma_0^4} & \frac{\kappa}{2\sigma_0^2}\mathrm{tr}(\bar{B}_n) + \frac{\gamma}{2\sigma_0^2}\iota_n'\mu_n & \frac{\kappa}{2\sigma_0^2}\mathrm{tr}(G_{2n}) \\ \sim & \sim & \kappa\bar{b}_n'\bar{b}_n + 2\gamma\bar{b}_n'\mu_n & \kappa g_{2n}'\bar{b}_n + \gamma g_{2n}'\mu_n \\ \sim & \sim & \sim & \kappa g_{2n}'g_{2n} \end{pmatrix} + \Sigma_n,$$

ι_n, γ and κ are defined in the earlier sections, $\mu_n = \sigma_0^{-1}A_{2n}G_{1n}X_n\beta_0$, $\bar{b}_n = \mathrm{diag}(\bar{B}_n)$, $g_{2n} = \mathrm{diag}(G_{2n})$, $\bar{B}_n^s = \bar{B}_n + \bar{B}_n'$ and $G_{2n}^s = G_{2n} + G_{2n}'$.

Letting V_{n1} be the sub-matrix of $\Sigma_n^{-1}\Gamma_n\Sigma_n^{-1}$ corresponding to β and $\hat{V}_{n1}$ be its estimate, an asymptotic t-statistic for inferences for $c_0'\beta_0$ is thus,

$$t_{\texttt{SARAR}} = \frac{c_0'\hat{\beta}_n - c_0'\beta_0}{\sqrt{c_0'\hat{V}_{n1}c_0}} \ \sim\ N(0,1). \tag{3.22}$$

3.5.2 Bias corrections

Given that QML estimators of the spatial parameters in the SED and SLD models can both be seriously biased, there is a good reason to believe that they will remain so when the spatial effects are combined. Hence bias corrections for the QML estimators of the SARAR model would be useful. To conduct bias correction, we need the higher-order partial derivatives of $\tilde{\psi}_n(\delta)$, $H_{rn}(\delta) = \nabla^r \tilde{\psi}_n(\delta), r = 1, 2, 3$, where the partial derivatives are obtained sequentially and element-wise with respect to δ'. Define, $T_{rn} = \text{tr}(G_{1n}^r(\lambda))$ and $K_{rn} = \text{tr}(G_{2n}^r(\rho))$, $r = 0, 1, 2, 3$. Also define the following quantities,

$$R_{1n}(\delta) = \frac{Y_n'(\delta)M_n(\rho)\bar{B}_n(\delta)Y_n(\delta)}{Y_n'(\delta)M_n(\rho)Y_n(\delta)},$$

$$R_{2n}(\delta) = \frac{Y_n'(\delta)\bar{B}_n'(\delta)M_n(\rho)\bar{B}_n(\delta)Y_n(\delta)}{Y_n'(\delta)M_n(\rho)Y_n(\delta)},$$

$$S_{rn}(\delta) = \frac{Y_n'(\delta)M_n(\rho)D_{rn}(\rho)M_n(\rho)Y_n(\delta)}{Y_n'(\delta)M_n(\rho)Y_n(\delta)}, \ r = 1, 2, 3, 4,$$

$$Q_{rn}^{\dagger}(\delta) = \frac{Y_n'(\delta)M_n(\rho)D_{rn}(\rho)M_n(\rho)\bar{B}_n(\delta)Y_n(\delta)}{Y_n'(\delta)M_n(\rho)Y_n(\delta)}, \ r = 1, 2, 3,$$

$$Q_{rn}^{\ddagger}(\delta) = \frac{Y_n'(\delta)\bar{B}_n'(\delta)M_n(\rho)D_{rn}(\rho)M_n(\rho)\bar{B}_n(\delta)Y_n(\delta)}{Y_n'(\delta)M_n(\rho)Y_n(\delta)}, \ r = 1, 2,$$

where $D_{1n}(\rho) = G_{2n}(\rho)$, and $D_{rn}(\rho),\ r = 2, 3, 4$, are given in Appendix C. These quantities have the following properties,

$$\tfrac{d}{d\lambda}R_{1n}(\delta) = 2R_{1n}^2(\delta) - R_{2n}(\delta), \qquad \tfrac{d}{d\lambda}R_{2n}(\delta) = 2R_{1n}(\delta)R_{2n}(\delta),$$

$$\tfrac{d}{d\lambda}S_{rn}(\delta) = 2R_{1n}(\delta)S_{rn}(\delta) - 2Q_{rn}^{\dagger}(\delta), \qquad \tfrac{d}{d\lambda}Q_{rn}^{\dagger}(\delta) = 2R_{1n}(\delta)Q_{rn}^{\dagger}(\delta) - Q_{rn}^{\ddagger}(\delta),$$

$$\tfrac{d}{d\lambda}Q_{rn}^{\ddagger}(\delta) = 2R_{1n}(\delta)Q_{rn}^{\ddagger}(\delta), \qquad \tfrac{d}{d\rho}R_{1n}(\delta) = 2R_{1n}(\delta)S_{1n}(\delta) - 2Q_{1n}^{\dagger}(\delta),$$

$$\tfrac{d}{d\rho}R_{2n}(\delta) = 2R_{2n}(\delta)S_{1n}(\delta) - 2Q_{1n}^{\ddagger}(\delta), \qquad \tfrac{d}{d\rho}S_{rn}(\delta) = 2S_{1n}(\delta)S_{rn}(\delta) + S_{r+1,n}(\delta),$$

$$\tfrac{d}{d\rho}Q_{rn}^{\dagger}(\delta) = 2S_{1n}(\delta)Q_{rn}^{\dagger}(\delta) + Q_{r+1,n}^{\dagger}(\delta), \qquad \tfrac{d}{d\rho}Q_{rn}^{\ddagger}(\delta) = 2S_{1n}(\delta)Q_{rn}^{\ddagger}(\delta) + Q_{r+1,n}^{\ddagger}(\delta).$$

Write $\tilde{\psi}_n(\delta) = (\tilde{\psi}_{1n}(\delta),\ \tilde{\psi}_{2n}(\delta))'$, where $\tilde{\psi}_{1n}(\delta) = -T_{0n}(\lambda) + R_{1n}(\delta)$ and $\tilde{\psi}_{2n}(\delta) = -K_{0n}(\rho) + S_{1n}(\delta)$. Denote the partial derivatives of $\tilde{\psi}_{rn}(\delta)$ by adding superscripts λ and/or ρ sequentially, e.g., $\tilde{\psi}_{1n}^{\lambda\lambda}(\delta) = \frac{\partial^2}{\partial\lambda^2}\tilde{\psi}_{1n}(\delta)$, and $\tilde{\psi}_{2n}^{\lambda\rho\lambda}(\delta) = \frac{\partial^3}{\partial\lambda\partial\rho\partial\lambda}\tilde{\psi}_{2n}(\delta)$.

Thus, $H_{1n}(\delta)$ has 1st row $\{\tilde{\psi}_{1n}^{\lambda}(\delta), \tilde{\psi}_{1n}^{\rho}(\delta)\}$ and 2nd row $\{\tilde{\psi}_{2n}^{\lambda}(\delta), \tilde{\psi}_{2n}^{\rho}(\delta)\}$:

$$H_{1n}(\delta) = \begin{pmatrix} -T_{1n}(\lambda) - R_{2n}(\delta) + 2R_{1n}^2(\delta), & -2Q_{1n}^{\dagger}(\delta) + 2R_{1n}(\delta)S_{1n}(\delta) \\ -2Q_{1n}^{\dagger}(\delta) + 2R_{1n}(\delta)S_{1n}(\delta), & -K_{1n}(\rho) + S_{2n}(\delta) + 2S_{1n}^2(\delta) \end{pmatrix}.$$

$H_{2n}(\delta)$ has rows $\{\tilde{\psi}_{1n}^{\lambda\lambda}(\delta), \tilde{\psi}_{1n}^{\lambda\rho}(\delta), \tilde{\psi}_{1n}^{\rho\lambda}(\delta), \tilde{\psi}_{1n}^{\rho\rho}(\delta)\}$ and $\{\tilde{\psi}_{2n}^{\lambda\lambda}(\delta), \tilde{\psi}_{2n}^{\lambda\rho}(\delta), \tilde{\psi}_{2n}^{\rho\lambda}(\delta), \tilde{\psi}_{2n}^{\rho\rho}(\delta)\}$, where

$\tilde{\psi}_{1n}^{\lambda\lambda}(\delta) = -2T_{2n}(\lambda) - 6R_{1n}(\delta)R_{2n}(\delta) + 8R_{1n}^3(\delta)$,

$\tilde{\psi}_{1n}^{\lambda\rho}(\delta) = 2Q_{1n}^{\ddagger}(\delta) - 8R_{1n}(\delta)Q_{1n}^{\dagger}(\delta) - 2R_{2n}(\delta)S_{1n}(\delta) + 8R_{1n}^2(\delta)S_{1n}(\delta)$,

$\tilde{\psi}_{1n}^{\rho\rho}(\delta) = -2Q_{2n}^{\dagger}(\delta) - 8S_{1n}(\delta)Q_{1n}^{\dagger}(\delta) + 2R_{1n}(\delta)S_{2n}(\delta) + 8R_{1n}(\delta)S_{1n}^2(\delta)$,

$\tilde{\psi}_{2n}^{\rho\rho}(\delta) = -2K_{2n}(\rho) + S_{3n}(\delta) + 6S_{1n}(\delta)S_{2n}(\delta) + 8S_{1n}^3(\delta)$

$\tilde{\psi}_{1n}^{\lambda\rho}(\delta) = \tilde{\psi}_{1n}^{\rho\lambda}(\delta) = \tilde{\psi}_{2n}^{\lambda\lambda}(\delta)$ and $\tilde{\psi}_{1n}^{\rho\rho}(\delta) = \tilde{\psi}_{2n}^{\lambda\rho}(\delta) = \tilde{\psi}_{2n}^{\rho\lambda}(\delta)$

$H_{3n}(\delta)$ is obtained by taking partial derivatives w.r.t. δ' for every element of $H_{2n}(\delta)$. It has elements:

$$\tilde{\psi}_{1n}^{\lambda\lambda\lambda}(\delta) = -6T_{3n}(\lambda) + 6R_{2n}^2(\delta) - 48R_{1n}^2(\delta)R_{2n}(\delta) + 48R_{1n}^4(\delta),$$

$$\begin{aligned}\tilde{\psi}_{1n}^{\lambda\lambda\rho}(\delta) = {} & 12R_{2n}(\delta)Q_{1n}^{\dagger}(\delta) + 12R_{1n}(\delta)Q_{1n}^{\ddagger}(\delta) - 24R_{1n}(\delta)R_{2n}(\delta)S_{1n}(\delta) \\ & -48R_{1n}^2(\delta)Q_{1n}^{\dagger}(\delta) + 48R_{1n}^3(\delta)S_{1n}(\delta),\end{aligned}$$

$$\begin{aligned}\tilde{\psi}_{1n}^{\lambda\rho\rho}(\delta) = {} & 2Q_{2n}^{\ddagger}(\delta) + 16Q_{1n}^{\dagger 2}(\delta) + 8S_{1n}(\delta)Q_{1n}^{\ddagger}(\delta) - 8R_{1n}(\delta)Q_{2n}^{\dagger}(\delta) \\ & -64R_{1n}(\delta)S_{1n}(\delta)Q_{1n}^{\dagger}(\delta) - 2R_{2n}(\delta)S_{2n}(\delta) - 8R_{2n}(\delta)S_{1n}^2(\delta) \\ & +8R_{1n}^2(\delta)S_{2n}(\delta) + 48R_{1n}^2(\delta)S_{1n}^2(\delta),\end{aligned}$$

$$\begin{aligned}\tilde{\psi}_{1n}^{\rho\rho\rho}(\delta) = {} & -2Q_{3n}^{\dagger}(\delta) - 12S_{2n}(\delta)Q_{1n}^{\dagger}(\delta) - 12S_{1n}(\delta)Q_{2n}^{\dagger}(\delta) - 48S_{1n}^2(\delta)Q_{1n}^{\dagger}(\delta) \\ & +24R_{1n}(\delta)S_{1n}(\delta)S_{2n}(\delta) + 2R_{1n}(\delta)S_{3n}(\delta) + 48R_{1n}(\delta)S_{1n}^3(\delta),\end{aligned}$$

$$\begin{aligned}\tilde{\psi}_{2n}^{\rho\rho\rho}(\delta) = {} & -6K_{3n}(\rho) + S_{4n}(\delta) + 6S_{2n}^2(\delta) + 8S_{1n}(\delta)S_{3n}(\delta) \\ & +48S_{1n}^2(\delta)S_{2n}(\delta) + 48S_{1n}^4(\delta),\end{aligned}$$

$\tilde{\psi}_{1n}^{\lambda\lambda\rho}(\delta) = \tilde{\psi}_{1n}^{\lambda\rho\lambda}(\delta) = \tilde{\psi}_{1n}^{\rho\lambda\lambda}(\delta) = \tilde{\psi}_{2n}^{\lambda\lambda\lambda}(\delta)$, $\tilde{\psi}_{1n}^{\rho\rho\rho}(\delta) = \tilde{\psi}_{2n}^{\lambda\rho\rho}(\delta) = \tilde{\psi}_{2n}^{\rho\lambda\rho}(\delta) = \tilde{\psi}_{2n}^{\rho\rho\lambda}(\delta)$ and $\tilde{\psi}_{1n}^{\lambda\rho\rho}(\delta) = \tilde{\psi}_{1n}^{\rho\lambda\rho}(\delta) = \tilde{\psi}_{2n}^{\lambda\lambda\rho}(\delta) = \tilde{\psi}_{1n}^{\rho\rho\lambda}(\delta) = \tilde{\psi}_{2n}^{\lambda\rho\lambda}(\delta) = \tilde{\psi}_{2n}^{\rho\lambda\lambda}(\delta)$.

Bootstrap estimates of biases: The R-, S- and Q-ratios at $\delta = \delta_0$ defined above can all be written as functions of θ_0 and $e_n = \sigma_0^{-1}\epsilon_n$, given X_n and $W_{rN}, r =$

1, 2 and using the relations $M_nA_{2n}X_n = 0$ and $W_{1n}Y_n = G_{1n}\left(X_n\beta_0 + A_{2n}^{-1}\epsilon_n\right)$:

$$\begin{aligned}
R_{1n}(\theta_0, e_n) &= \frac{e_n' M_n(\mu_n + \bar{B}_n e_n)}{e_n' M_n e_n}, \\
R_{2n}(\theta_0, e_n) &= \frac{(\mu_n + \bar{B}_n e_n)' M_n(\mu_n + \bar{B}_n e_n)}{e_n' M_n e_n}, \\
S_{rn}(\theta_0, e_n) &= \frac{e_n' M_n D_{rn} M_n e_n}{e_n' M_n e_n}, \ r = 1, 2, 3, 4, \\
Q_{rn}^{\dagger}(\theta_0, e_n) &= \frac{e_n' M_n D_{rn} M_n(\mu_n + \bar{B}_n e_n)}{e_n' M_n e_n}, \ r = 1, 2, 3, \\
Q_{rn}^{\ddagger}(\theta_0, e_n) &= \frac{(\mu_n + \bar{B}_n e_n)' M_n D_{rn} M_n(\mu_n + \bar{B}_n e_n)}{e_n' M_n e_n}, \ r = 1, 2,
\end{aligned}$$

where $\bar{B}_n = \bar{B}_n(\delta_0)$. As a result, we have $\tilde{\psi}_n = \tilde{\psi}_n(\theta_0, e_n)$ and $H_{rn} = H_{rn}(\theta_0, e_n)$, $r = 1, 2, 3$. The bias terms, b_{-1} and $b_{-3/2}$, can be estimated using the `Bootstrap Algorithm 2` (BA-2), described in Section 3.2.

Let $\hat{\delta}_n^{\text{bc2}} = (\hat{\lambda}_n^{\text{bc2}}, \hat{\rho}_n^{\text{bc2}})'$ be the 2nd-order bias corrected version of $\hat{\delta}_n$. Let $\hat{\beta}_n^{\text{bc}} = \tilde{\beta}(\hat{\delta}_n^{\text{bc2}})$ and $\hat{\sigma}_n^{2,\text{bc}} = \tilde{\sigma}_n^2(\hat{\delta}_n^{\text{bc2}})$. As expected, which can also be inferred from the results given in Section 3.5.4, the QML estimators can be severely biased and a 2nd-order bias correction effectively eliminates the bias. To conserve space, we do not report the Monte Carlo results for the finite sample biases of the QML estimators and the bias corrected QML estimators of the SARAR model.

3.5.3 Improved inferences for regression coefficients

Replacing $\hat{\delta}_n$ by $\hat{\delta}_n^{\text{bc2}}$ in the definition of $t_{\texttt{SARAR}}$, we obtain a statistic which is expected to have a better finite sample performance:

$$t_{\texttt{SARAR}}^{\text{bc}} = \frac{c_0'\hat{\beta}_n^{\text{bc}} - c_0'\beta_0}{\sqrt{c_0'\hat{V}_{n1}^{\text{bc}}c_0}}, \tag{3.23}$$

where $\hat{V}_{n1}^{\text{bc}}$ is V_{n1} evaluated at $\hat{\delta}_n^{\text{bc2}}$, $\tilde{\beta}_n^{\text{bc}}$, $\hat{\sigma}_n^{2,\text{bc}}$, $\hat{\gamma}_n^{\text{bc}}$, and $\hat{\kappa}_n^{\text{bc}}$. The last two are the estimates of γ and κ, the skewness and excess kurtosis of $\epsilon_{n,i}$ involved in Γ_n.

Given $\tilde{\beta}_n = \tilde{\beta}_n(\delta_0)$, let $\tilde{\beta}_n^{(r)}$ be the rth derivative with respect to δ_0', $r = 1, 2$. Also define $F_n(\rho) = [X_n' A_{2n}'(\rho) A_{2n}(\rho) X_n]^{-1} X_n' A_{2n}'(\rho) A_{2n}(\rho)$ where we have $\tilde{\beta}_n(\delta) = F_n(\rho) A_{1n}(\lambda) Y_n$ and $F_n^{(r)} = F_n^{(r)}(\rho_0)$ is the rth derivative with respect to δ_0', $r = 1, 2$. Assuming that $\mathrm{E}(\tilde{\beta}_n^{(r)})$ exists and that $\tilde{\beta}_n^{(r)} - \mathrm{E}(\tilde{\beta}_n^{(r)}) = O_p(n^{-1/2})$, $r = 1, 2$, by a Taylor expansion, we have,

$$
\begin{aligned}
\tilde{\beta}_n(\hat{\delta}_n) - \beta_0 &= \tilde{\beta}_n - \beta_0 + \tilde{\beta}_n^{(1)}(\hat{\delta}_n - \delta_0) + \tfrac{1}{2}\tilde{\beta}_n^{(2)}[(\hat{\delta}_n - \delta_0) \otimes (\hat{\delta}_n - \delta_0)] \\
&\quad + O_p(n^{-3/2}), \\
&= b_{0n} + \mathrm{E}(\tilde{\beta}_n^{(1)})(a_{-1/2} + a_{-1}) + b_{1n} a_{-1/2} \\
&\quad + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}_n^{(2)})(a_{-1/2} \otimes a_{-1/2}) + O_p(n^{-3/2}),
\end{aligned}
$$

where $b_{0n} = F_n A_{2n}^{-1} \epsilon_n$, $b_{1n} = (-F_n G_{1n} A_{2n}^{-1} \epsilon_n,\ F_n^{(1)} A_{2n}^{-1} \epsilon_n)$,
$\mathrm{E}(\tilde{\beta}_n^{(1)}) = (-F_n G_{1n} X_n \beta_0,\ F_n^{(1)} X_n \beta_0)$ and
$\mathrm{E}(\tilde{\beta}_n^{(2)}) = (0_{k\times 1},\ -F_n^{(1)} G_{1n} X_n \beta_0,\ -F_n^{(1)} G_{1n} X_n \beta_0,\ F_n^{(2)} X_n \beta_0)$.
The expressions for $F_n^{(1)}$ and $F_n^{(2)}$ are given in Appendix C. This leads to a second order expansion for $\mathrm{Var}(\hat{\beta}_n)$ or $\mathrm{Var}(\hat{\beta}_n^{bc2})$: $\mathrm{Var}(\hat{\beta}_n^{\mathrm{bc2}}) = \mathrm{Var}[b_{0n} + \mathrm{E}(\tilde{\beta}_n^{(1)})(a_{-1/2} + a_{-1}) + b_{1n} a_{-1/2} + \frac{1}{2}\mathrm{E}(\tilde{\beta}_n^{(2)})(a_{-1/2} \otimes a_{-1/2})] + O_p(n^{-2})$, where $a_{-1/2} = \Omega_n \tilde{\psi}_n$ and $a_{-1} = \Omega_n \tilde{\psi}_n + \Omega_n(\tilde{\psi}_n' \otimes H_{1n})\mathrm{vec}(\Omega_n) + \frac{1}{2}\Omega_n \mathrm{E}(H_{2n})(\Omega_n \otimes \Omega_n)(\tilde{\psi}_n \otimes \tilde{\psi}_n)$, (see Yang, 2015b). One can easily obtain the 2nd-order bias corrected estimators $\hat{\beta}_n^{\mathrm{bc2}}$ and $\hat{\sigma}_n^{2,\mathrm{bc2}}$, but again Monte Carlo results (not reported for brevity) show that they do not differ much from the corresponding 'plug-in' estimators. A similar two stage bootstrap procedure as given in Section 3.3, but based on the algorithm `BA-2` presented in Section 3.2, can be applied to obtain an estimate of this variance term, $\widehat{\mathrm{Var}}(\hat{\beta}_n^{\mathrm{bc2}})$. We have a second order bias corrected t-statistic as follows:

$$
t_{\mathtt{SARAR}}^{\mathrm{bc2}} = \frac{c_0' \hat{\beta}_n^{\mathrm{bc2}} - c_0' \beta_0}{\sqrt{c_0' \widehat{\mathrm{Var}}(\hat{\beta}_n^{\mathrm{bc2}}) c_0}}. \tag{3.24}
$$

3.5.4 Monte Carlo experiments

The methods for bias correction and for improved inferences introduced above for the SARAR model are investigated for their finite sample performance under the following DGP:

$$Y_n = \lambda W_{1n} Y_n + \iota_n \beta_0 + X_{1n}\beta_1 + X_{2n}\beta_2 + u_n, \quad u_n = \rho W_{2n} u_n + \epsilon_n,$$

where all the quantities are generated in a similar manner to those for the SED model. The two spatial weight matrices are taken to be the same. The parameters are set to be the same as before, where λ and ρ both take values from $\{-0.5, -0.25, 0, 0.25, 0.5\}$.

We focus on the finite sample performance of the three tests $t_{\texttt{SARAR}}$, $t^{\text{bc}}_{\texttt{SARAR}}$ and $t^{\text{bc2}}_{\texttt{SARAR}}$. The results for the finite sample bias of the QML estimators are available from the authors upon request. Tables 5.1-5.3 report empirical sizes of $t_{\texttt{SARAR}}$, $t^{\text{bc}}_{\texttt{SARAR}}$ and $t^{\text{bc2}}_{\texttt{SARAR}}$ when used for testing $H_0 : \beta_1 = \beta_2$, under the `Group Interaction` spatial layouts described in Appendix B. Similar conclusions are drawn from the Monte Carlo results for the SARAR model as those for the two sub models considered in the earlier sections: (i) as n increases, all tests converge in terms of sizes, (ii) the asymptotic test $t_{\texttt{SARAR}}$ remains unreliable in the sense it rejects the true H_0 much too often than it supposes to, (iii) the test $t^{\text{bc}}_{\texttt{SARAR}}$ offers immediate reduction in size distortions, and (iv) $t^{\text{bc2}}_{\texttt{SARAR}}$ generally offers further improvements. Furthermore, like the asymptotic test for the SED model, $t_{\texttt{SARAR}}$ can have a size distortion that is very persistent, having values that are at least 24% even when $n = 500$. The results (unreported to conserve space) under `Rook` and `Queen Contiguity` show similar patterns, but the differences are of a lesser degree due to the weaker spatial dependence (less number of neighbours) under these two spatial layouts.

Table 3.5.1 Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in SARAR Model

Group Interaction, REG2, $\sigma = 1, \lambda = 0.5$; Test: $1 = t_{\texttt{SARAR}}, 2 = t^{\texttt{bc}}_{\texttt{SARAR}}, 3 = t^{\texttt{bc2}}_{\texttt{SARAR}}$

ρ	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal			Normal Errors			Normal Mixture			Log normal		
		$n = 50$									$n = 200$								
.50	1	.197	.115	.040	.201	.122	.044	.197	.122	.040	.141	.078	.022	.140	.082	.028	.131	.078	.021
	2	.120	.068	.020	.123	.073	.023	.146	.084	.028	.113	.056	.014	.117	.061	.017	.116	.061	.015
	3	.115	.062	.017	.119	.068	.023	.128	.074	.024	.105	.050	.011	.107	.056	.016	.106	.052	.013
.25	1	.191	.109	.031	.180	.110	.031	.183	.109	.035	.147	.085	.025	.152	.089	.028	.150	.083	.025
	2	.118	.067	.020	.116	.069	.022	.120	.066	.021	.108	.056	.012	.112	.061	.012	.111	.058	.012
	3	.109	.061	.016	.103	.058	.019	.103	.055	.017	.100	.050	.009	.102	.054	.011	.101	.051	.010
.00	1	.191	.110	.031	.177	.099	.028	.191	.114	.037	.150	.089	.026	.137	.075	.016	.138	.084	.020
	2	.111	.054	.015	.100	.054	.016	.117	.065	.021	.104	.055	.012	.116	.061	.014	.124	.066	.017
	3	.098	.047	.012	.095	.046	.013	.100	.055	.018	.097	.050	.010	.102	.051	.010	.105	.052	.012
-.25	1	.173	.100	.025	.170	.096	.027	.184	.108	.033	.158	.093	.030	.131	.074	.018	.120	.062	.014
	2	.094	.048	.011	.098	.049	.016	.111	.059	.020	.108	.054	.013	.123	.068	.019	.118	.066	.017
	3	.108	.048	.009	.108	.051	.013	.090	.047	.016	.099	.049	.012	.102	.055	.010	.095	.054	.010
-.50	1	.182	.104	.030	.162	.085	.023	.177	.100	.034	.127	.072	.020	.120	.061	.013	.119	.063	.013
	2	.097	.049	.013	.085	.043	.010	.102	.059	.019	.115	.066	.017	.122	.063	.015	.135	.074	.019
	3	.100	.048	.011	.091	.052	.009	.092	.046	.014	.105	.060	.013	.095	.046	.009	.091	.052	.009
		$n = 100$									$n = 500$								
.50	1	.169	.099	.027	.163	.097	.029	.171	.103	.031	.124	.068	.018	.126	.070	.017	.124	.073	.018
	2	.115	.058	.014	.122	.064	.017	.115	.059	.016	.102	.053	.013	.107	.053	.011	.106	.056	.011
	3	.101	.049	.013	.107	.057	.015	.105	.055	.013	.098	.049	.012	.100	.049	.010	.100	.050	.010
.25	1	.165	.094	.029	.172	.101	.028	.163	.095	.030	.130	.073	.023	.134	.073	.020	.130	.074	.018
	2	.106	.054	.012	.116	.056	.011	.111	.056	.013	.105	.056	.015	.106	.057	.014	.101	.053	.012
	3	.095	.045	.010	.101	.047	.009	.098	.049	.011	.099	.052	.014	.100	.053	.013	.099	.049	.010
.00	1	.177	.103	.031	.176	.098	.032	.165	.100	.035	.138	.075	.021	.135	.072	.019	.133	.076	.020
	2	.105	.054	.012	.102	.052	.013	.105	.056	.014	.106	.055	.013	.099	.052	.009	.106	.054	.012
	3	.093	.046	.011	.099	.048	.011	.095	.051	.013	.103	.053	.011	.099	.049	.009	.101	.053	.010
-.25	1	.170	.100	.027	.164	.095	.029	.170	.098	.029	.131	.074	.020	.135	.077	.022	.132	.075	.022
	2	.096	.047	.010	.097	.048	.010	.102	.048	.011	.101	.055	.013	.102	.053	.012	.098	.053	.011
	3	.095	.054	.011	.099	.050	.009	.099	.052	.010	.096	.051	.012	.099	.051	.011	.100	.051	.010
-.50	1	.158	.091	.026	.151	.086	.022	.145	.087	.024	.128	.071	.018	.144	.076	.022	.129	.072	.019
	2	.090	.046	.010	.091	.044	.010	.091	.048	.011	.094	.046	.011	.107	.054	.014	.093	.050	.011
	3	.090	.054	.009	.099	.047	.009	.098	.052	.009	.092	.045	.011	.103	.051	.013	.099	.050	.010

Table 3.5.2 Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in SARAR Model

`Group Interaction, REG2,` $\sigma = 1, \lambda = 0.0$; `Test:` $1 = t_{\texttt{SARAR}}, 2 = t^{\texttt{bc}}_{\texttt{SARAR}}, 3 = t^{\texttt{bc2}}_{\texttt{SARAR}}$

ρ	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal			Normal Errors			Normal Mixture			Log-normal		
		$n = 50$									$n = 200$								
.50	1	.186	.107	.027	.196	.127	.050	.188	.114	.037	.134	.079	.020	.132	.066	.019	.135	.074	.018
	2	.130	.073	.020	.123	.076	.027	.132	.079	.027	.111	.058	.014	.107	.055	.013	.116	.061	.013
	3	.128	.068	.017	.113	.071	.025	.118	.069	.023	.101	.052	.012	.098	.049	.010	.104	.053	.010
.25	1	.199	.125	.047	.187	.113	.039	.204	.130	.048	.140	.083	.023	.143	.086	.029	.153	.087	.025
	2	.115	.064	.021	.126	.071	.018	.111	.065	.020	.117	.066	.017	.107	.060	.016	.108	.058	.015
	3	.110	.061	.020	.112	.061	.015	.108	.062	.019	.107	.060	.014	.098	.053	.014	.096	.050	.013
.00	1	.184	.110	.034	.184	.107	.033	.203	.126	.043	.157	.093	.029	.155	.094	.027	.153	.089	.027
	2	.110	.061	.017	.114	.062	.020	.127	.074	.022	.106	.058	.015	.112	.060	.014	.110	.056	.014
	3	.097	.054	.015	.095	.054	.017	.106	.059	.017	.099	.053	.013	.103	.054	.013	.100	.052	.012
-.25	1	.192	.114	.039	.189	.109	.036	.194	.122	.039	.127	.072	.016	.136	.072	.019	.127	.068	.016
	2	.110	.059	.018	.112	.063	.017	.117	.067	.021	.107	.061	.014	.129	.069	.019	.128	.072	.019
	3	.095	.050	.015	.095	.051	.013	.099	.055	.017	.099	.053	.012	.111	.054	.015	.092	.049	.012
-.50	1	.194	.114	.038	.177	.100	.030	.183	.115	.033	.156	.095	.028	.123	.067	.014	.150	.090	.030
	2	.105	.058	.018	.102	.052	.014	.112	.062	.016	.106	.053	.012	.127	.071	.020	.105	.056	.014
	3	.098	.049	.014	.098	.052	.011	.099	.047	.012	.098	.049	.011	.105	.054	.012	.096	.050	.013
		$n = 100$									$n = 500$								
.50	1	.172	.105	.030	.168	.096	.032	.173	.099	.030	.129	.074	.021	.129	.068	.016	.125	.072	.017
	2	.122	.067	.017	.122	.065	.017	.110	.054	.014	.109	.055	.013	.107	.053	.011	.107	.057	.010
	3	.107	.060	.014	.109	.056	.014	.100	.049	.012	.102	.051	.012	.099	.050	.009	.100	.052	.010
.25	1	.175	.102	.030	.171	.101	.031	.171	.110	.036	.136	.077	.018	.136	.077	.022	.128	.075	.019
	2	.113	.057	.013	.108	.055	.013	.115	.064	.016	.106	.053	.011	.103	.057	.014	.102	.053	.012
	3	.098	.049	.011	.096	.049	.010	.105	.056	.014	.100	.047	.010	.099	.052	.012	.097	.050	.011
.00	1	.173	.103	.030	.175	.103	.034	.180	.107	.031	.137	.079	.021	.134	.081	.022	.126	.077	.021
	2	.098	.051	.014	.105	.056	.013	.110	.054	.013	.111	.053	.012	.105	.055	.015	.103	.057	.012
	3	.097	.052	.013	.094	.050	.011	.098	.047	.011	.105	.050	.012	.099	.051	.014	.100	.053	.012
-.25	1	.180	.109	.032	.159	.094	.028	.165	.099	.030	.136	.077	.023	.142	.082	.021	.136	.071	.019
	2	.104	.052	.011	.094	.046	.012	.101	.049	.011	.101	.053	.011	.109	.055	.011	.099	.049	.010
	3	.093	.046	.010	.091	.054	.010	.099	.055	.011	.098	.050	.010	.105	.053	.010	.100	.050	.010
-.50	1	.172	.106	.029	.159	.093	.026	.158	.096	.025	.146	.076	.020	.134	.078	.017	.138	.076	.021
	2	.101	.048	.011	.090	.045	.009	.093	.048	.009	.103	.050	.010	.103	.053	.009	.101	.051	.011
	3	.096	.054	.010	.098	.049	.009	.096	.054	.010	.100	.047	.010	.100	.050	.009	.100	.049	.011

Table 3.5.3 Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in SARAR Model

`Group Interaction, REG2,` $\sigma = 1, \lambda = -.25$; `Test:` $1 = t_{\mathtt{SARAR}}, 2 = t^{\mathtt{bc}}_{\mathtt{SARAR}}, 3 = t^{\mathtt{bc2}}_{\mathtt{SARAR}}$

ρ	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal			Normal Errors			Normal Mixture			Log-normal		
		$n = 50$									$n = 200$								
.50	1	.196	.119	.045	.203	.126	.047	.188	.115	.045	.129	.073	.020	.144	.083	.022	.133	.072	.016
	2	.121	.070	.020	.122	.076	.022	.138	.085	.030	.112	.063	.014	.116	.060	.012	.111	.059	.013
	3	.114	.066	.017	.117	.072	.020	.122	.074	.022	.104	.055	.013	.106	.054	.012	.093	.046	.009
.25	1	.198	.123	.042	.205	.128	.043	.205	.130	.054	.143	.081	.025	.150	.085	.022	.151	.084	.025
	2	.108	.059	.018	.109	.057	.020	.112	.066	.025	.122	.062	.018	.122	.065	.015	.112	.055	.014
	3	.103	.056	.015	.104	.055	.017	.109	.064	.022	.110	.056	.016	.106	.057	.011	.100	.051	.013
.00	1	.192	.115	.037	.180	.109	.038	.199	.127	.051	.144	.086	.023	.129	.075	.016	.156	.091	.028
	2	.115	.065	.017	.118	.065	.017	.106	.062	.020	.123	.066	.017	.114	.065	.015	.113	.056	.013
	3	.103	.058	.014	.101	.056	.015	.104	.059	.020	.110	.059	.014	.100	.053	.011	.103	.050	.012
-.25	1	.196	.114	.032	.186	.108	.038	.194	.115	.042	.136	.075	.023	.125	.065	.018	.153	.090	.026
	2	.107	.052	.016	.109	.060	.019	.114	.069	.022	.123	.068	.018	.120	.062	.017	.106	.056	.011
	3	.099	.050	.013	.098	.051	.015	.098	.057	.017	.112	.060	.015	.101	.048	.012	.097	.052	.011
-.50	1	.188	.113	.040	.188	.111	.037	.186	.116	.040	.120	.064	.016	.114	.055	.011	.150	.091	.025
	2	.111	.061	.018	.089	.049	.014	.098	.055	.015	.117	.063	.015	.126	.063	.015	.105	.051	.012
	3	.095	.051	.015	.093	.055	.013	.099	.051	.015	.106	.055	.012	.099	.045	.009	.097	.049	.011
		$n = 100$									$n = 500$								
.50	1	.175	.100	.029	.171	.099	.030	.167	.098	.033	.132	.069	.016	.131	.070	.017	.133	.072	.021
	2	.116	.059	.016	.126	.067	.018	.113	.061	.014	.110	.058	.012	.108	.055	.010	.109	.056	.012
	3	.100	.051	.013	.111	.057	.015	.104	.055	.014	.104	.053	.011	.100	.048	.010	.102	.052	.011
.25	1	.179	.102	.032	.172	.103	.034	.170	.099	.031	.132	.079	.023	.125	.074	.019	.138	.076	.020
	2	.114	.059	.014	.111	.063	.015	.108	.057	.014	.109	.060	.014	.106	.055	.012	.107	.052	.012
	3	.102	.052	.012	.099	.053	.013	.098	.052	.011	.104	.056	.013	.100	.051	.009	.103	.051	.011
.00	1	.176	.106	.030	.178	.103	.032	.158	.093	.029	.135	.077	.025	.129	.077	.020	.128	.071	.019
	2	.099	.054	.012	.108	.055	.012	.096	.044	.011	.105	.056	.015	.099	.049	.012	.100	.050	.011
	3	.099	.055	.010	.096	.048	.010	.099	.045	.011	.101	.053	.013	.100	.049	.011	.099	.050	.011
-.25	1	.177	.102	.031	.165	.098	.031	.162	.097	.029	.139	.082	.026	.139	.079	.022	.130	.077	.020
	2	.096	.048	.009	.101	.050	.013	.101	.053	.013	.106	.059	.014	.104	.053	.011	.099	.050	.012
	3	.099	.052	.010	.099	.050	.013	.100	.050	.011	.101	.056	.014	.099	.050	.010	.100	.050	.012
-.50	1	.169	.102	.029	.160	.100	.032	.159	.100	.035	.143	.085	.023	.140	.084	.024	.126	.074	.023
	2	.098	.047	.010	.095	.049	.012	.099	.053	.012	.107	.054	.012	.111	.059	.014	.098	.053	.013
	3	.096	.051	.009	.096	.049	.011	.099	.052	.012	.105	.053	.011	.108	.055	.012	.099	.052	.012

3.6 Conclusions

This chapter considers inference problems for the regression coefficients β in linear regression models with spatial dependence, where the estimation of the spatial parameters may incur severe bias. It is shown that while the existence of spatial dependence does not have a big impact on the point estimation of the regression coefficients in terms of consistency and bias, it can have a huge impact on the usual t-statistics for β. We propose simple ways to correct the t-statistics, and the resulted 2nd-order corrected t-statistics perform superbly. Considering the effectiveness and the simplicity of the proposed methods, they are recommended for practical applications.

Central to the proposed inference methods for regression coefficients in this chapter is the general bias correction methods for non-linear estimators proposed in Yang (2015b). The proposed methods have a great potential to be extended to more advanced models such as higher-order SARAR models, spatial panel data models, dynamic panel data models, non-linear spatial regression models and non-linear spatial panel data models. They are equally applicable to non-spatial models as well. Among these, the extension to a higher-order SARAR incurs only some extra algebra.

The classical approach to the problem considered in this chapter is to directly bootstrap the original t-statistic to give asymptotically refined approximations to the finite sample critical values, taking advantage of the underlining statistic being asymptotically pivotal. However, bootstrapping a Wald-type or a likelihood ratio statistic requires the re-estimation of all parameters in every bootstrap iteration, and thus is computationally much more demanding compared to our approach, in particular when the model contains more non-linear parameters. Nevertheless, it would be interesting as a future research to compare the two approaches.[9]

[9] We thank a referee for raising this issue.

CHAPTER 4

Bias Correction and Refined Inferences for Fixed Effects Spatial Panel Data Models

4.1 Introduction

Panel data models with spatial and social interactions have received a belated but recently increasing attention by econometricians, since Anselin (1988).[1] Spatial panel data (SPD) models are differentiated by whether they are static or dynamic and whether they contain random effects or fixed effects. Popular methods of model estimation and inferences are quasi maximum likelihood (QML) and generalised method of moments (GMM).[2]

It has been recognised through the studies of spatial regression models that

[1]See, among others, Baltagi et al. (2003, 2013), Kapoor et al. (2007), Yu et al. (2008, 2012), Yu and Lee (2010), Lee and Yu (2010a,b), Baltagi and Yang (2013a,b), and Su and Yang (2015b).

[2]See Lee and Yu (2010a, 2015) and Anselin et al. (2008) for general accounts on issues related to SPD model specifications, parameter estimation, etc.

QML estimators of the spatial parameter(s), though efficient, can be quite biased [3], and more so with a denser spatial weight matrix (Chapters 2 and 3). As a result the subsequent model inferences (based on t-ratios) can be seriously affected.[4]

Evidently, the QML estimators of the SPD models are subjected to the same issues on the finite sample bias and finite sample performance of subsequent inferences, but these important issues have not been addressed.[5] Given the popularity of the SPD models among the applied researchers, it is highly desirable to have a set of simple and reliable methods for parameter estimation and model inference. In this chapter, we focus on the SPD models with fixed effects to provide methods for bias and variance corrections (up to third-order) by extending the methods of Yang (2015b),[6] and then to show how the bias and variance corrections lead to improved t-ratios for spatial and covariate effects. Lee and Yu (2010b) investigate the asymptotic properties for the QML estimation of this model based on *direct* and *transformation* approaches. The latter approach is more attractive as it provides consistent estimators for all the common parameters, which is crucial in the developments of the methods for finite sample bias-corrections and refined inferences.[7]

[3]Lee, 2004; Bao and Ullah, 2007; Bao, 2013; Yang, 2015b

[4]Methods of bias-correcting the QML estimators of the spatial parameter(s) have been given for the spatial lag (SL) model (Bao and Ullah, 2007; Bao, 2013; Yang, 2015b), the spatial error (SE) model (Chapter 2), and the spatial lag and error (SLE) model (Chapter 3). The improved t-ratios for the SL effect is given in Yang (2015b), and improved t-ratios for the covariate effects are given in Chapter 3 for the SL, SE and SLE models, respectively.

[5]The importance of bias correction for models with non-linear parameters is seen from the large literature on the regular dynamic panels (see, e.g., Nickell (1981), Kiviet (1995), Hahn and Kuersteiner (2002), Hahn and Newey (2004), Bun and Carree (2005), Hahn and Moon (2006), and Arellano and Hahn (2005).

[6]The fixed effects model has the advantage of robustness because fixed effects are allowed to depend on included regressors. It also provides a unified model framework for different random effects models considered in, e.g., Anselin (1988), Kapoor et al. (2007) and Baltagi et al. (2013). However, fixed effects model encounters incidental parameter problem (Neyman and Scott, 1948; Lancaster, 2000).

[7]Lee and Yu (2010b) observe that when conducting a direct estimation using the likelihood function where all the common parameters and the fixed effects are estimated together, the estimate of the variance parameter is inconsistent when T is finite while n is large. With data transformations to eliminate the fixed effects, the incidental parameter problem is avoided, and the ratio of n and T does not affect the asymptotic properties of estimates as the data are

We note that while the general stochastic expansions of Yang (2015b) for non-linear estimators are applicable to different models including the SPD models considered in this chapter, the detailed developments of bias corrections, variance corrections and corrections on t-ratio vary from one model to another. Furthermore, the transformation approach induces errors that may no longer be independent and identically distributed (iid) even if the original errors are. Thus, the bootstrap method proposed by Yang (2015b) under iid errors, may not be directly applicable. We demonstrate in this chapter that when the original error distribution is not far from normality, the standard iid bootstrap method can still provide an excellent approximation, due to the fact that the transformed errors are homoskedastic and uncorrelated. When the original errors are extremely non-normal, we show that the wild bootstrap method can improve the approximation. Monte Carlo results reveal that the QML estimators of the spatial parameters can be quite biased, in particular for the models with spatial error dependence, and that a second-order bias correction effectively removes the bias. Furthermore, Monte Carlo results show that inferences for spatial and covariate effects based on the regular t-ratios can be misleading, but these based on the proposed t-ratios are very reliable. We emphasize that while corrections on bias and variance of a point estimator are important, it is more important to correct the t-ratios so that practical applications of the models and methods are more reliable. The methods presented in this chapter show a plausible way to do so. They are simple and yet quite general as the spatial regression models are embedded as special cases.

The rest of the chapter is organised as follows. Section 4.2 introduces the spatial panel data model allowing both spatial lag and spatial error, and both time-specific effects and individual-specific effects, and its QML estimation based on the transformed likelihood function. Section 4.3 presents a third-order stochas-

pooled. The QML estimators so derived are shown to be consistent, and, except for the variance estimate, are identical to those from the direct approach.

tic expansion for the QML estimators of the spatial parameters, a third-order expansion for the bias, and a third-order expansion for the variance of the QML estimators of the spatial parameters. Section 4.3 also addresses issues on the bias of QML estimators of other model parameters, and on the inferences following bias and variance corrections. Section 4.4 introduces the bootstrap methods for estimating various quantities in the expansions, and presents theories for the validity of these methods. Section 4.5 presents Monte Carlo results. Section 4.6 discusses and concludes the chapter.

4.2 The Model and QML Estimation

For the spatial panel data (SPD) model with fixed effects (FE), we can investigate the case with both spatial lag and spatial error, where n is large and T could be finite or large. We include both individual effects and time effects to have a robust specification. The FE-SPD model under consideration is

$$Y_{nt} = \lambda_0 W_{1n} Y_{nt} + X_{nt}\beta_0 + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}, \quad U_{nt} = \rho_0 W_{2n} U_{nt} + V_{nt}, \tag{4.1}$$

for $t = 1, 2, \ldots, T$, where, for a given t, $Y_{nt} = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is an $n \times 1$ vector of observations on the response variable, X_{nt} is an $n \times k$ matrix containing the values of k non-stochastic, individually and time varying regressors, $V_{nt} = (v_{1t}, v_{2t}, \ldots, v_{nt})'$ is an $n \times 1$ vector of errors where $\{v_{it}\}$ are independent and identically distributed (iid) for all i and t with mean 0 and variance σ_0^2, $\mathbf{c}_{n0}$ is an $n \times 1$ vector of fixed individual effects, and α_{t0} is the fixed time effect with l_n being an $n \times 1$ vector of ones. W_{1n} and W_{2n} are given $n \times n$ spatial weights matrices where W_{1n} generates the 'direct' spatial effects among the spatial units in their response values Y_{nt}, and W_{2n} generates cross-sectional dependence among the disturbances U_{nt}. In practice, W_{1n} and W_{2n} may be the same.

In Lee and Yu (2010b), QML estimation of (4.1) is considered by using either a direct approach or a transformation approach. The direct approach is to estimate the regression parameters jointly with the individual and time effects, which yields a bias of order $O(T^{-1})$ due to the estimation of individual effects and a bias of order $O(n^{-1})$ due to the estimation of time effects. The transformation approach eliminates the individual and time effects and then implements the estimation, which yields consistent estimates of the common parameters when either n or T is large. In the current chapter, we will follow the transformation approach so that it is free from the incidental parameter problem.

To eliminate individual effects, define $J_T = (I_T - \frac{1}{T} l_T l_T')$ and let $[F_{T,T-1}, \frac{1}{\sqrt{T}} l_T]$ be the orthonormal eigenvector matrix of J_T, where $F_{T,T-1}$ is the $T \times (T-1)$ sub-matrix corresponding to the eigenvalues of one, I_T is a $T \times T$ identity matrix and l_T is a $T \times 1$ vector of ones.[8] To eliminate the time effects, let J_n and $F_{n,n-1}$ be similarly defined, and let W_{1n} and W_{2n} be row normalised.[9] For any $n \times T$ matrix $[Z_{n1}, \cdots, Z_{nT}]$, define the $(n-1) \times (T-1)$ transformed matrix as

$$[Z_{n1}^*, \ldots, Z_{n,T-1}^*] = F_{n,n-1}'[Z_{n1}, \ldots, Z_{nT}]F_{T,T-1}. \tag{4.2}$$

This leads to, for $t = 1, \ldots, T-1$, Y_{nt}^*, U_{nt}^*, V_{nt}^*, and $X_{nt,j}^*$ for the jth regressor. As in Lee and Yu (2010), let $X_{nt}^* = [X_{nt,1}^*, X_{nt,2}^*, \ldots, X_{nt,k}^*]$, and $W_{hn}^* = F_{n,n-1}' W_{hn} F_{n,n-1}, h = 1, 2$. The transformed model we will work on thus takes the form:

$$Y_{nt}^* = \lambda_0 W_{1n}^* Y_{nt}^* + X_{nt}^* \beta_0 + U_{nt}^*, \quad U_{nt}^* = \rho_0 W_{2n}^* U_{nt}^* + V_{nt}^*, \quad t = 1, \ldots, T-1. \tag{4.3}$$

[8] As discussed in Lee and Yu (2010b, Footnote 12), the first difference and Helmert transformation have often been used to eliminate the individual effects. A special selection of $F_{T,T-1}$ gives rise to the Helmert transformation where $\{V_{nt}\}$ are transformed to $(\frac{T-t}{T-t+1})^{1/2}[V_{nt} - \frac{1}{T-t}(V_{n,t+1} + \cdots + V_{nT})]$, which is of particular interest for dynamic panel data models.

[9] When W_{jn} are not row normalised, the linear SARAR presentation of (4.4) for the spatial panel model will no longer hold. In that case, a likelihood formulation would not be feasible.

After the transformations, the effective sample size becomes $N = (n-1)(T-1)$. Stacking the vectors and matrices, i.e., letting $\mathbf{Y}_N = (Y_{n1}^{*\prime}, \ldots, Y_{n,T-1}^{*\prime})'$, $\mathbf{U}_N = (U_{n1}^{*\prime}, \ldots, U_{n,T-1}^{*\prime})'$, $\mathbf{V}_N = (V_{n1}^{*\prime}, \ldots, V_{n,T-1}^{*\prime})'$, $\mathbf{X}_N = (X_{n1}^{*\prime}, \ldots, X_{n,T-1}^{*\prime})'$, and denoting $\mathbf{W}_{hN} = I_{T-1} \otimes W_{hn}^*, h = 1, 2$, we have the following compact expression for the transformed model:

$$\mathbf{Y}_N = \lambda_0 \mathbf{W}_{1N} \mathbf{Y}_N + \mathbf{X}_N \beta_0 + \mathbf{U}_N, \quad \mathbf{U}_N = \rho_0 \mathbf{W}_{2N} \mathbf{U}_N + \mathbf{V}_N, \tag{4.4}$$

which is in form identical to the spatial autoregressive model with autoregressive errors (SARAR), showing that the QML estimation of the two-way fixed effects panel SARAR model is similar to that of the linear SARAR model. The key difference is that the elements of $\mathbf{V}_N$ may not be iid though they are uncorrelated and homoskedastic as shown below. This may have a certain impact on the bootstrap method (see next section for details).

It is easy to show that the transformed errors $\{v_{it}^*\}$ are uncorrelated for all i and t by using the identity $(V_{n1}^{*\prime}, \ldots, V_{n,T-1}^{*\prime})' = (F_{T,T-1}' \otimes F_{n,n-1}')(V_{n1}', \ldots, V_{nT}')'$,

$$E(V_{n1}^{*\prime}, \ldots, V_{n,T-1}^{*\prime})'(V_{n1}^{*\prime}, \cdots, V_{n,T-1}^{*\prime}) = \sigma_0^2 (F_{T,T-1}' \otimes F_{n,n-1}')(F_{T,T-1} \otimes F_{n,n-1}) = \sigma_0^2 I_N.$$

Hence, $\{v_{it}^*\}$ are iid $N(0, \sigma_0^2)$ if the original errors $\{v_{it}\}$ are iid $N(0, \sigma_0^2)$. It follows that the (quasi) Gaussian log likelihood function for (4.3) is,

$$\ell_N(\theta) = -\tfrac{N}{2} \ln(2\pi\sigma^2) + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \tfrac{1}{2\sigma^2} \mathbf{V}_N'(\zeta) \mathbf{V}_N(\zeta), \tag{4.5}$$

where $\zeta = (\beta', \lambda, \rho)'$, $\theta = (\beta', \sigma^2, \lambda, \rho)'$, $\mathbf{A}_N(\lambda) = I_N - \lambda \mathbf{W}_{1N}$, $\mathbf{B}_N(\rho) = I_N - \rho \mathbf{W}_{2N}$, and $\mathbf{V}_N(\zeta) = \mathbf{B}_N(\rho)[\mathbf{A}_N(\lambda)\mathbf{Y}_N - \mathbf{X}_N \beta]$.

Now, letting $\mathbf{Y}_N(\lambda) = \mathbf{A}_N(\lambda)\mathbf{Y}_N$ and $\mathbf{X}_N(\rho) = \mathbf{B}_N(\rho)\mathbf{X}_N$, the constrained QML estimators of β and σ^2, given λ and ρ, can be expressed in the following simple form:

$$\tilde{\beta}_N(\lambda,\rho) = [\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)\mathbf{B}_N(\rho)\mathbf{Y}_N(\lambda), \quad (4.6)$$

$$\tilde{\sigma}^2_N(\lambda,\rho) = N^{-1}\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda), \quad (4.7)$$

where $\mathbf{M}_N(\rho) = \mathbf{B}'_N(\rho)\{I_N - \mathbf{X}_N(\rho)[\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)\}\mathbf{B}_N(\rho)$. Substituting $\tilde{\beta}_N(\lambda,\rho)$ and $\tilde{\sigma}^2_N(\lambda,\rho)$ back into (4.5) gives the concentrated log likelihood function of (λ,ρ):

$$\ell^c_N(\lambda,\rho) = -\tfrac{N}{2}(\ln(2\pi)+1) + \ln|\mathbf{A}_N(\lambda)| + \ln|\mathbf{B}_N(\rho)| - \tfrac{N}{2}\ln\tilde{\sigma}^2_N(\lambda,\rho). \quad (4.8)$$

Maximizing $\ell^c_N(\lambda,\rho)$ in (4.8) gives the unconstrained QML estimators $\hat{\lambda}_N$ and $\hat{\rho}_N$ of λ and ρ, and substituting $(\hat{\lambda}_N,\hat{\rho}_N)$ back into (4.6) and (4.7) gives the unconstrained QML estimators of β and σ^2 as $\hat{\beta}_N \equiv \tilde{\beta}_N(\hat{\lambda}_N,\hat{\rho}_N)$ and $\hat{\sigma}^2_N \equiv \tilde{\sigma}^2_N(\hat{\lambda}_N,\hat{\rho}_N)$.[10] Write $\hat{\theta}_N = (\hat{\beta}'_N, \hat{\lambda}_N, \hat{\rho}_N, \hat{\sigma}^2_N)'$. Lee and Yu (2010b) show that $\hat{\theta}_N$ is $\sqrt{N}$-consistent and asymptotically normal under some mild conditions. These conditions and the asymptotic variance of $\hat{\theta}_N$ are given in Appendix E to facilitate the subsequent developments for the higher-order results. It follows that the QML estimators for any of the sub-models discussed below will be $\sqrt{N}$-consistent and asymptotically normal as well, where N can be $(n-1)(T-1)$, $n(T-1)$, $(n-1)T$, or nT.

The linear SARAR representation (4.4) has greatly facilitated the QML estimation of the general FE-SPD model. It is also very helpful for the subsequent developments in bias and variance corrections. Obviously, it contains as special cases the spatial regression models. Based on this representation, the results de-

[10] Numerical maximization of $\ell^c_N(\lambda,\rho)$ can be computationally demanding if N is large due to the need of repeated calculations of the two determinants. Following simplifications help alleviate such a burden: $|\mathbf{A}_N(\lambda)| = |I_{n-1} - \lambda W^*_{1n}|^{T-1} = \left(\frac{1}{1-\lambda}|I_n - \lambda W_{1n}|\right)^{T-1} = \left(\frac{1}{1-\lambda}\prod_{i=1}^n(1-\lambda\omega_{1i})\right)^{T-1}$, where ω_{1i} are the eigenvalues of W_{1n}, the middle equation from Lee and Yu (2010), and the last equation is from Griffith (1988). Similarly the determinant of $|\mathbf{B}_N(\rho)|$ is calculated.

veloped for this general model can easily be reduced to suit simpler models. For example, setting ρ or λ to zero in (4.4) gives an FE-SPD model with only the SL effect or an FE-SPD model with only the SE effect; dropping either α_{t0} or $\mathbf{c}_{n0}$ in (4.1) (or dropping either $F_{n,n-1}$ or $F_{T,T-1}$ in (4.2)) leads to a sub-model with only the individual-specific effects or a sub-model with only the time-specific effects; and finally, dropping both $\mathbf{c}_{n0}$ and α_{t0} in (4.1) leads to the SARAR regression model. On the other hand, the spatial panel model considered in this chapter can also be extended to include more spatial lag terms in both the response and the disturbance, in particular the former.[11] Software can be developed to facilitate the end users of the methodologies developed in this chapter.

4.3 Third-Order Bias and MSE for FE-SPD Model

4.3.1 Third-order stochastic expansions for non-linear estimators

Yang (2015b) presents a general method for up to third-order bias and variance corrections on a set of non-linear estimators based on stochastic expansions and bootstrap. The stochastic expansions provide tractable approximations to the bias and variance of the non-linear estimators and the bootstrap make these expansions practically implementable. The method is demonstrated, through a linear SAR model, to be very effective in correcting the bias and improving inferences. It was emphasised in Yang (2015b) that, in estimating a model with both linear and non-linear parameters, the main source of bias and the main difficulty in correcting the bias are associated with the estimation of the non-linear parameters, and hence one should focus on the concentrated estimation equations. By doing so,

[11] See Lee and Yu (2015, 2016) for more discussions and for the related issue on parameter identification.

the dimensionality of the problem can be greatly reduced, and more importantly, the additional variations from the estimation of the linear and scale parameters are captured in correcting the non-linear estimators, thus making the bias and variance corrections more effective. The method is summarised as follows.

Let δ be the vector of non-linear parameters of a model, and $\hat{\delta}_N$ defined as $\hat{\delta}_N = \arg\{\tilde{\psi}_N(\delta) = 0\}$, be its $\sqrt{N}$-consistent estimator, with $\tilde{\psi}_N(\delta)$ is the concentrated estimating function (CEF) and $\tilde{\psi}_N(\delta) = 0$ the concentrated estimating equation (CEE). Let $H_{rN}(\delta) = \nabla^r \tilde{\psi}_N(\delta), r = 1, 2, 3$, where the partial derivatives are carried out sequentially and element-wise, with respect to δ'. Let $\tilde{\psi}_N \equiv \tilde{\psi}_N(\delta_0)$, $H_{rN} \equiv H_{rN}(\delta_0)$ and $H^{\circ}_{rN} = H_{rN} - \mathrm{E}(H_{rN}), r = 1, 2, 3$. Note that hereafter the expectation operator 'E' corresponds to the true model parameters θ_0. Define $\Omega_N = -[\mathrm{E}(H_{1N})]^{-1}$. Yang (2015b), extending Rilstone et al. (1996) and Bao and Ullah (2007), gives a set of sufficient conditions for a third-order stochastic expansion of $\hat{\delta}_N = \arg\{\tilde{\psi}_N(\delta) = 0\}$, based on a general CEF $\tilde{\psi}_N(\delta)$, which are restated here to facilitate the development of higher-order results for the FE-SPD model:

Assumption G1. *$\hat{\delta}_N$ solves $\tilde{\psi}_N(\delta) = 0$ and $\hat{\delta}_N - \delta_0 = O_p(N^{-1/2})$.*

Assumption G2. *$\tilde{\psi}_N(\delta)$ is differentiable up to the rth order for δ in a neighbourhood of δ_0, $\mathrm{E}(H_{rN}) = O(1)$, and $H^{\circ}_{rN} = O_p(N^{-1/2})$, $r = 1, 2, 3$.*

Assumption G3. *$[\mathrm{E}(H_{1N})]^{-1} = O(1)$, and $H^{-1}_{1N} = O_p(1)$.*

Assumption G4. *$\|H_{rN}(\delta) - H_{rN}(\delta_0)\| \leq \|\delta - \delta_0\| \, U_N$ for δ in a neighbourhood of δ_0, $r = 1, 2, 3$, and $\mathrm{E}\,|U_N| \leq c < \infty$ for some constant c.*

Under these conditions, a third-order stochastic expansion for $\hat{\delta}_N$ is:

$$\hat{\delta}_N - \delta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(N^{-2}), \tag{4.9}$$

where $a_{-s/2}$ represents a term of order $O_p(N^{-s/2})$ for $s = 1, 2, 3$, having the expressions

$$\begin{aligned}
a_{-1/2} &= \Omega_N \tilde{\psi}_N, \\
a_{-1} &= \Omega_N H^{\circ}_{1N} a_{-1/2} + \tfrac{1}{2}\Omega_N \mathrm{E}(H_{2N})(a_{-1/2} \otimes a_{-1/2}), \\
a_{-3/2} &= \Omega_N H^{\circ}_{1N} a_{-1} + \tfrac{1}{2}\Omega_N H^{\circ}_{2N}(a_{-1/2} \otimes a_{-1/2}) \\
&\quad + \tfrac{1}{2}\Omega_N \mathrm{E}(H_{2N})(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) \\
&\quad + \tfrac{1}{6}\Omega_N \mathrm{E}(H_{3N})(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}),
\end{aligned}$$

where $\otimes$ denotes the Kronecker product. In moving from the stochastic expansion given in (4.9) to third-order expansions for the bias, MSE and variance of $\hat{\delta}_N$, it is assumed that $\mathrm{E}(\tilde{\psi}_N) = O(N^{-1})$ and that *a quantity bounded in probability has a finite expectation.* The latter is a simplifying assumption to ensure that the remainders are of the stated order. A third-order expansion for the bias of $\hat{\delta}_N$ is

$$\mathrm{Bias}(\hat{\delta}_N) = b_{-1} + b_{-3/2} + O(N^{-2}), \tag{4.10}$$

where $b_{-1} = \mathrm{E}(a_{-1/2} + a_{-1})$ and $b_{-3/2} = \mathrm{E}(a_{-3/2})$, being the second- and third-order biases. Similarly, a third-order expansion for the mean squared error (MSE) of $\hat{\delta}_N$ is

$$\mathrm{MSE}(\hat{\delta}_N) = m_{-1} + m_{-3/2} + m_{-2} + O(N^{-5/2}), \tag{4.11}$$

where $m_{-1} = \mathrm{E}(a_{-1/2}a'_{-1/2})$, $m_{-3/2} = \mathrm{E}(a_{-1/2}a'_{-1} + a_{-1}a'_{-1/2})$, $m_{-2} = \mathrm{E}(a_{-1}a'_{-1} + a_{-1/2}a'_{-3/2} + a_{-3/2}a'_{-1/2})$, and the third-order expansion for the variance of $\hat{\delta}_N$ is

$$\mathrm{Var}(\hat{\delta}_N) = v_{-1} + v_{-3/2} + v_{-2} + O(N^{-5/2}), \tag{4.12}$$

where $v_{-1} = \mathrm{Var}(a_{-1/2})$, $v_{-3/2} = \mathrm{Cov}(a_{-1/2}, a_{-1}) + \mathrm{Cov}(a_{-1}, a_{-1/2})$, and $v_{-2} = \mathrm{Cov}(a_{-1/2},\ a_{-3/2}) + \mathrm{Cov}(a_{-3/2}, a_{-1/2}) + \mathrm{Var}(a_{-1} + a_{-3/2})$; or simply $v_{-1} = m_{-1}$, $v_{-3/2} = m_{-3/2}$, and $v_{-2} = m_{-2} - b^2_{-1}$.

Therefore, we can improve the statistical inference in finite samples by correcting the bias and standard deviation of estimates. From (4.10), we can use

$$\delta_N^{\text{bc2}} = \hat{\delta}_N - b_{-1} \;\text{ or }\; \delta_N^{\text{bc3}} = \hat{\delta}_N - b_{-1} - b_{-3/2},$$

to yield an estimator unbiased up to order $O(N^{-1})$ or $O(N^{-3/2})$. With estimated b_{-1} and $b_{-3/2}$, feasible δ_N^{bc2} and δ_N^{bc3} can be constructed.

Similar procedures can be applied to increase the precision of variance estimate. By (4.12), if $\hat{b}_{-1} - b_{-1} = O_p(N^{-3/2})$ and $\hat{b}_{-3/2} - b_{-3/2} = O_p(N^{-2})$, we have

$$\text{Var}(\delta_N^{\text{bc3}}) = v_{-1} + v_{-3/2} + v_{-2} - 2\text{ACov}(\hat{\delta}_N, \hat{b}_{-1}) + O(N^{-5/2}), \tag{4.13}$$

and $\text{Var}(\delta_N^{\text{bc2}}) = \text{Var}(\delta_N^{\text{bc3}}) + O(N^{-5/2})$, where ACov denotes asymptotic covariance.

4.3.2 Third-order bias and variance for spatial estimators

The general expansions summarised in Section 4.3.1 are applicable to different models including the FE-SPD model we consider,, but the detailed developments for the corrections on bias, variance, and t-ratio vary from one model to another. Furthermore, the transformation approach induces errors that are no longer iid, rendering the bootstrap method of Yang (2015b) for estimating the correction terms not directly applicable. In this subsection, we first derive all the quantities required for the third-order expansions for the FE-SPD model, and then discuss conditions under which the results (4.9)-(4.13) hold under the FE-SPD model instead of going through the detailed proofs of them. As seen from Section 4.2, the set of non-linear parameters in the FE-SPD model are $\delta = (\lambda, \rho)'$. The CEF leading to the QML estimator $\hat{\delta}_N = (\hat{\lambda}_N, \hat{\rho}_N)$ is $\tilde{\psi}_N(\delta) = \frac{1}{N}\frac{\partial}{\partial\delta}\ell_N^c(\delta)$, which is shown to have the form:

$$\tilde{\psi}_N(\delta) = \begin{cases} -T_{0N}(\lambda) + \dfrac{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, \\ -K_{0N}(\rho) - \dfrac{\mathbf{Y}'_N(\lambda)\mathbf{M}^{(1)}_N(\rho)\mathbf{Y}_N(\lambda)}{2\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, \end{cases} \tag{4.14}$$

where $T_{0N}(\lambda) = \frac{1}{N}\text{tr}(\mathbf{W}_{1N}\mathbf{A}_N^{-1}(\lambda))$, $K_{0N}(\rho) = \frac{1}{N}\text{tr}(\mathbf{W}_{2N}\mathbf{B}_N^{-1}(\rho))$, and $\mathbf{M}^{(1)}_N(\rho) = \frac{d}{d\rho}\mathbf{M}_N(\rho)$.[12] To derive the rth order derivative, $H_{rN}(\delta)$, of $\tilde{\psi}_N(\delta)$ w.r.t. δ', $r = 1, 2, 3$, for up to third-order bias correction, define $T_{rN}(\lambda) = \frac{1}{N}\text{tr}[(\mathbf{W}_{1N}\mathbf{A}_N^{-1}(\lambda))^{r+1}]$, and $K_{rN}(\rho) = \frac{1}{N}\text{tr}[(\mathbf{W}_{2N}\mathbf{B}_N^{-1}(\rho))^{r+1}], r = 0, 1, 2, 3$. Let $\mathbf{M}^{(k)}_N(\rho)$ be the kth derivative of $\mathbf{M}_N(\rho)$ w.r.t. ρ, $k = 1, 2, 3, 4$. Define

$$R_{1N}(\delta) = \frac{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, \quad R_{2N}(\delta) = \frac{\mathbf{Y}'_N\mathbf{W}'_{1N}\mathbf{M}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)};$$
$$Q^{\dagger}_{kN}(\delta) = \frac{\mathbf{Y}'_N(\lambda)\mathbf{M}^{(k)}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)}, \quad Q^{\ddagger}_{kN}(\delta) = \frac{\mathbf{Y}'_N\mathbf{W}'_{1N}\mathbf{M}^{(k)}_N(\rho)\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)};$$
$$S_{kN}(\delta) = \frac{\mathbf{Y}'_N(\lambda)\mathbf{M}^{(k)}_N(\rho)\mathbf{Y}_N(\lambda)}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\lambda)},$$

which have the following properties:

$$\frac{\partial R_{1N}(\delta)}{\partial\lambda} = 2R^2_{1N}(\delta) - R_{2N}(\delta), \qquad \frac{\partial R_{2N}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)R_{2N}(\delta),$$
$$\frac{\partial Q^{\dagger}_{kN}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)Q^{\dagger}_{kN}(\delta) - Q^{\ddagger}_{kN}(\delta), \quad \frac{\partial Q^{\ddagger}_{kN}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)Q^{\ddagger}_{kN}(\delta),$$
$$\frac{\partial S_{kN}(\delta)}{\partial\lambda} = 2R_{1N}(\delta)S_{kN}(\delta) - 2Q^{\dagger}_{kN}(\delta);$$

$$\frac{\partial R_{1N}(\delta)}{\partial\rho} = Q^{\dagger}_{1N}(\delta) - R_{1N}(\delta)S_{1N}(\delta), \qquad \frac{\partial R_{2N}(\delta)}{\partial\rho} = Q^{\ddagger}_{1N}(\delta) - R_{2N}(\delta)S_{1N}(\delta),$$
$$\frac{\partial Q^{\dagger}_{kN}(\delta)}{\partial\rho} = Q^{\dagger}_{k+1,N}(\delta) - Q^{\dagger}_{kN}(\delta)S_{1N}(\delta), \quad \frac{\partial Q^{\ddagger}_{kN}(\delta)}{\partial\rho} = Q^{\ddagger}_{k+1,N}(\delta) - Q^{\ddagger}_{kN}(\delta)S_{1N}(\delta),$$
$$\frac{\partial S_{kN}(\delta)}{\partial\rho} = S_{k+1,N}(\delta) - S_{kN}(\delta)S_{1N}(\delta).$$

[12] Lee and Yu (2010b) provide a useful identity: $(I_{n-1} - \lambda W^*_{hn})^{-1} = F'_{n,n-1}(I_{n-1} - \lambda W_{hn})^{-1}F_{n,n-1}$. Based on this, the inverses of $\mathbf{A}_N(\lambda)$ and $\mathbf{B}_N(\lambda)$ can easily be calculated as they are block-diagonal. The conditions for the $\sqrt{N}$-consistency of $\hat{\delta}_N$ are given in Lee and Yu (2010b), and in Appendix E.

Write $\tilde{\psi}_N(\delta) = (\tilde{\psi}_{1N}(\delta), \tilde{\psi}_{2N}(\delta))'$ with $\tilde{\psi}_{1N}(\delta) = -T_{0N}(\lambda) + R_{1N}(\delta)$ and $\tilde{\psi}_{2N}(\delta) = -K_{0N}(\rho) - S_{1N}(\delta)$. Denote the partial derivatives of $\tilde{\psi}_{jN}(\delta)$ by adding superscripts λ and/or ρ sequentially, e.g., $\tilde{\psi}_{2N}^{\lambda\rho\lambda}(\delta) = \frac{\partial^3}{\partial\lambda\partial\rho\partial\lambda}\tilde{\psi}_{2N}(\delta)$. Thus, $H_{1N}(\delta)$ has 1st row $\{\tilde{\psi}_{1N}^{\lambda}(\delta), \tilde{\psi}_{1N}^{\rho}(\delta)\}$ and 2nd row $\{\tilde{\psi}_{2N}^{\lambda}(\delta), \tilde{\psi}_{2N}^{\rho}(\delta)\}$, which gives

$$H_{1N}(\delta) = \begin{pmatrix} -T_{1N}(\lambda) - R_{2N}(\delta) + 2R_{1N}^2(\delta), & Q_{1N}^{\dagger}(\delta) - R_{1N}(\delta)S_{1N}(\delta) \\ Q_{1N}^{\dagger}(\delta) - R_{1N}(\delta)S_{1N}(\delta), & -K_{1N}(\rho) - \frac{1}{2}S_{2N}(\delta) + \frac{1}{2}S_{1N}^2(\delta) \end{pmatrix}.$$

$H_{2N}(\delta)$ has rows:

$\{\tilde{\psi}_{1N}^{\lambda\lambda}(\delta), \tilde{\psi}_{1N}^{\lambda\rho}(\delta), \tilde{\psi}_{1N}^{\rho\lambda}(\delta), \tilde{\psi}_{1N}^{\rho\rho}(\delta)\}$ and $\{\tilde{\psi}_{2N}^{\lambda\lambda}(\delta), \tilde{\psi}_{2N}^{\lambda\rho}(\delta), \tilde{\psi}_{2N}^{\rho\lambda}(\delta), \tilde{\psi}_{2N}^{\rho\rho}(\delta)\}$, where

$\tilde{\psi}_{1N}^{\lambda\lambda}(\delta) = -2T_{2N}(\lambda) - 6R_{1N}(\delta)R_{2N}(\delta) + 8R_{1N}^3(\delta)$,

$\tilde{\psi}_{1N}^{\lambda\rho}(\delta) = -Q_{1N}^{\ddagger}(\delta) + 4R_{1N}(\delta)Q_{1N}^{\dagger}(\delta) + R_{2N}(\delta)S_{1N}(\delta) - 4R_{1N}^2(\delta)S_{1N}(\delta)$,

$\tilde{\psi}_{1N}^{\rho\rho}(\delta) = Q_{2N}^{\dagger}(\delta) - 2Q_{1N}^{\dagger}(\delta)S_{1N}(\delta) + 2R_{1N}(\delta)S_{1N}^2(\delta) - R_{1N}(\delta)S_{2N}(\delta)$,

$\tilde{\psi}_{2N}^{\rho\rho}(\delta) = -2K_{2N}(\rho) - \frac{1}{2}S_{3N}(\delta) + \frac{3}{2}S_{1N}(\delta)S_{2N}(\delta) - S_{1N}^3(\delta)$,

$\tilde{\psi}_{2N}^{\lambda\lambda}(\delta) = \tilde{\psi}_{1N}^{\rho\lambda}(\delta) = \tilde{\psi}_{1N}^{\lambda\rho}(\delta)$, and $\tilde{\psi}_{2N}^{\lambda\rho}(\delta) = \tilde{\psi}_{2N}^{\rho\lambda}(\delta) = \tilde{\psi}_{1N}^{\rho\rho}(\delta)$.

$H_{3N}(\delta)$ is obtained by differentiating elements of $H_{2N}(\delta)$ w.r.t. δ':

$\tilde{\psi}_{1N}^{\lambda\lambda\lambda}(\delta) = -6T_{3N}(\lambda) + 6R_{2N}^2(\delta) - 48R_{1N}^2(\delta)R_{2N}(\delta) + 48R_{1N}^4(\delta)$,

$\tilde{\psi}_{1N}^{\lambda\lambda\rho}(\delta) = -6Q_{1N}^{\dagger}(\delta)R_{2N}(\delta) + 12R_{1N}(\delta)R_{2N}(\delta)S_{1N}(\delta) - 6R_{1N}(\delta)Q_{1N}^{\ddagger}(\delta)$,

$\quad +24R_{1N}^2(\delta)[Q_{1N}^{\dagger}(\delta) - R_{1N}(\delta)S_{1N}(\delta)]$,

$\tilde{\psi}_{1N}^{\lambda\rho\lambda}(\delta) = 2Q_{1N}^{\ddagger}(\delta)R_{1N}(\delta) + 12R_{1N}(\delta)R_{2N}(\delta)S_{1N}(\delta) - 6R_{1N}(\delta)Q_{1N}^{\dagger}(\delta)$

$\quad +8R_{1N}^2(\delta)Q_{1N}^{\dagger}(\delta) - 20R_{1N}^3(\delta)S_{1N}(\delta)$,

$\tilde{\psi}_{1N}^{\lambda\rho\rho}(\delta) = -Q_{2N}^{\ddagger}(\rho) + 2Q_{1N}^{\ddagger}(\rho)S_{1N}(\delta) - 2R_{2N}(\delta)S_{1N}^2(\delta) - 16R_{1N}(\delta)S_{1N}(\delta)Q_{1N}^{\dagger}(\delta)$

$\quad +R_{2N}(\delta)S_{2N}(\delta) + 4R_{1N}(\delta)Q_{2N}^{\dagger}(\delta) + 12R_{1N}^2(\delta)S_{1N}^2(\delta) - 4R_{1N}^2(\delta)S_{2N}(\delta) + 4Q_{1N}^{\dagger 2}(\delta)$,

$\tilde{\psi}_{1N}^{\rho\rho\lambda}(\delta) = -Q_{2N}^{\ddagger}(\delta) + 4Q_{2N}^{\dagger}(\delta)R_{1N}(\delta) + 2Q_{1N}^{\ddagger}(\delta)S_{1N}(\delta) - 16R_{1N}(\delta)Q_{1N}^{\dagger}(\delta)S_{1N}(\delta)$

$\quad -R_{2N}(\delta)S_{2N}(\delta) + 12R_{1N}^2(\delta)S_{1N}^2(\delta) - 2R_{2N}(\delta)S_{1N}^2(\delta) - 4S_{1N}^2(\delta)S_{2N}(\delta) + 4Q_{1N}^{\dagger 2}(\delta)$,

$\tilde{\psi}_{1N}^{\rho\rho\rho}(\delta) = Q_{3N}^{\dagger}(\delta) - 3Q_{2N}^{\dagger}(\delta)S_{1N}(\delta) + 6Q_{1N}^{\dagger}(\delta)S_{1N}^2(\delta) - 3Q_{1N}^{\dagger}(\delta)S_{2N}(\delta)$

$\quad -6R_{1N}(\delta)S_{1N}^3(\delta) + 6R_{1N}(\delta)S_{1N}(\delta)S_{2N}(\delta) - R_{1N}(\delta)S_{3N}(\delta)$,

$\tilde{\psi}_{2N}^{\rho\rho\lambda}(\delta) = Q_{3N}^{\dagger}(\delta) - R_{1N}(\delta)S_{3N}(\delta) - 3Q_{1N}^{\dagger}(\delta)S_{2N}(\delta) + 6R_{1N}(\delta)S_{1N}(\delta)S_{2N}(\delta)$

$-3S_{1N}(\delta)Q^{\dagger}_{2N}(\delta) + 6S^2_{1N}(\delta)Q^{\dagger}_{1N}(\delta) - 6R_{1N}(\delta)S^3_{1N}(\delta)$,

$\tilde{\psi}^{\rho\rho\rho}_{2N}(\delta) = -6K_{3N}(\rho) - \frac{1}{2}S_{4N}(\delta) + 2S_{1N}(\delta)S_{3N}(\delta) + \frac{3}{2}S^2_{2N}(\delta) - 6S_{2N}(\delta)S^2_{1N}(\delta)$

$+3S^4_{1N}(\delta)$.

$\tilde{\psi}^{\rho\lambda\lambda}_{1N}(\delta) = \tilde{\psi}^{\lambda\rho\lambda}_{1N}(\delta) = \tilde{\psi}^{\lambda\lambda\lambda}_{2N}(\delta)$, $\tilde{\psi}^{\rho\lambda\rho}_{1N}(\delta) = \tilde{\psi}^{\lambda\rho\rho}_{1N}(\delta) = \tilde{\psi}^{\lambda\lambda\rho}_{2N}(\delta)$, $\tilde{\psi}^{\rho\rho\lambda}_{1N}(\delta) = \tilde{\psi}^{\lambda\rho\lambda}_{2N}(\delta) = \tilde{\psi}^{\rho\lambda\lambda}_{2N}(\delta)$, and $\tilde{\psi}^{\rho\rho\rho}_{1N}(\delta) = \tilde{\psi}^{\lambda\rho\rho}_{2N}(\delta) = \tilde{\psi}^{\rho\lambda\rho}_{2N}(\delta)$.

Expressions of $\mathbf{M}^{(k)}_N(\rho)$, ρ, $k = 1, 2, 3, 4$, are lengthy and are given in Appendix E.

For general results (4.9)-(4.13) for the CEF $\tilde{\psi}_N(\delta)$ of the FE-SPD model, it is sufficient that $\tilde{\psi}_N(\delta)$ satisfies Assumptions G1-G4. First the $\sqrt{N}$-consistency of $\hat{\delta}_N$ in Assumption G1 is given in Theorem E.1 in Appendix E. The differentiability of $\tilde{\psi}_N(\delta)$ in Assumption G2 is obvious. From Section 4.4.1 we see that the $R-$, S- and Q-quantities at δ_0 are all ratios of quadratic forms in $\mathbf{V}_N$, having the same denominator $\mathbf{V}'_N\mathbf{M}^{\diamond}_N\mathbf{V}_N$ where $\mathbf{M}^{\diamond}_N = I_N - \mathbf{X}_N(\rho_0)[\mathbf{X}'_N(\rho_0)\mathbf{X}_N(\rho_0)]^{-1}\mathbf{X}'_N(\rho_0)$. It can be shown that $\frac{1}{N}\mathbf{V}'_N\mathbf{M}^{\diamond}_N\mathbf{V}_N = \sigma^2_0 + o_p(1)$. Hence, with Assumptions E1-E8 in Appendix E, for the H-quantities to have proper stochastic behaviour, it would require the existence of the 6th moment of v_{it} for a second-order bias correction, and the existence of the 10th moment of v_{it} for a third-order bias correction. Variance corrections have stronger moment requirements. However, these moment requirements are no more than those under a joint estimating equation with analytical approach. The condition $\mathrm{E}(\tilde{\psi}_N) = O(N^{-1})$ is required so that b_{-1} is $O(N^{-1})$. This condition is not restrictive given $\sqrt{N}(\hat{\delta}_N - \delta_0)$ converges to a centred bivariate normal distribution as $N \to \infty$, (Lee and Yu, 2010b), implies that $\mathrm{E}(\tilde{\psi}_N) = o(N^{-1/2})$. The other conditions are likely to hold. With these and Assumptions E1-E8 in Appendix E, the results (4.9)-(4.13) are likely to hold. Hence, we do not present detailed proofs of the results (4.9)-(4.13) for the FE-SPD model, but rather focus on the validity of the bootstrap methods for the practical implementation of the bias and variance corrections.

4.3.3 Reduced models

Letting either $\rho = 0$ or $\lambda = 0$ leads to two sub-models, the FE-SPD model with SL dependence and the FE-SPD model with SE dependence. Bias and variance corrections become much simpler in these cases, in particular the former.

FE-SPD model with SL dependence. The necessary terms for up to a third-order bias and variance correction for the FE-SPD model with SL dependence are:

$$
\begin{aligned}
R_{1N}(\lambda) &= \tfrac{\mathbf{Y}'_N(\lambda)\mathbf{M}^0_N\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}^0_N\mathbf{Y}_N(\lambda)}, \quad R_{2N}(\lambda) = \tfrac{\mathbf{Y}'_N\mathbf{W}'_{1N}\mathbf{M}^0_N\mathbf{W}_{1N}\mathbf{Y}_N}{\mathbf{Y}'_N(\lambda)\mathbf{M}^0_N\mathbf{Y}_N(\lambda)}, \\
\tilde{\psi}_N(\lambda) &= -T_{0N}(\lambda) + R_{1N}(\lambda), \\
H_{1N}(\lambda) &= -T_{1N}(\lambda) - R_{2N}(\lambda) + 2R^2_{1N}(\lambda), \\
H_{2N}(\lambda) &= -2T_{2N}(\lambda) - 6R_{1N}(\lambda)R_{2N}(\lambda) + 8R^3_{1N}(\lambda), \\
H_{3N}(\lambda) &= -6T_{3N}(\lambda) + 6R^2_{2N}(\lambda) - 48R^2_{1N}(\lambda)R_{2N}(\lambda) + 48R^4_{1N}(\lambda),
\end{aligned}
$$

where $\mathbf{M}^0_N \equiv \mathbf{M}_N(0) = I_N - \mathbf{X}_N(\mathbf{X}'_N\mathbf{X}_N)^{-1}\mathbf{X}'_N$. These results contain, as a special case, the results for linear SAR model considered in Yang (2015b), showing the usefulness of the linear SARAR representation (4.4) for the FE-SPD model.

FE-SPD model with SE dependence. The necessary terms for up to third-order bias and variances correction for the FE-SPD model with only SE dependence are:

$$
\begin{aligned}
S_{kN}(\rho) &= \tfrac{\mathbf{Y}'_N\mathbf{M}^{(k)}_N(\rho)\mathbf{Y}_N}{\mathbf{Y}'_N\mathbf{M}_N(\rho)\mathbf{Y}_N}, \ k = 1, 2, 3, 4, \\
\tilde{\psi}_N(\rho) &= -K_{0N}(\rho) - \tfrac{1}{2}S_{1N}(\rho), \\
H_{1N}(\rho) &= -K_{1N}(\rho) - \tfrac{1}{2}S_{2N}(\rho) + \tfrac{1}{2}S^2_{1N}(\rho), \\
H_{2N}(\rho) &= -2K_{2N}(\rho) - \tfrac{1}{2}S_{3N}(\rho) + \tfrac{3}{2}S_{1N}(\rho)S_{2N}(\rho) - S^3_{1N}(\rho), \\
H_{3N}(\rho) &= -6K_{3N}(\rho) - \tfrac{1}{2}S_{4N}(\delta) + 2S_{1N}(\delta)S_{3N}(\delta) + \tfrac{3}{2}S^2_{2N}(\delta) \\
&\quad -6S_{2N}(\delta)S^2_{1N}(\delta) + 3S^4_{1N}(\delta).
\end{aligned}
$$

These results contain, as a special case, the results for the linear SED model considered in Chapter 2. Again, these results show the usefulness of the linear SARAR representation for the fixed effects spatial panel data model given in (4.4).

Simplifications to a one-way fixed effects model are easily done. If the model contains only individual-specific effects, $t = 1, \ldots, T-1$ and $N = n(T-1)$, and when model contains only the time-specific effects, $t = 1, \ldots, T$ and $N = (n-1)T$.

4.3.4 Bias correction for non-spatial estimators

Note that $\hat{\beta}_N = \tilde{\beta}_N(\hat{\delta}_N)$ and $\hat{\sigma}^2_N = \tilde{\sigma}^2_N(\hat{\delta}_N)$, where $\tilde{\beta}_N(\delta)$ and $\tilde{\sigma}^2_N(\delta)$ are the constrained QML estimators of β and σ^2 defined in (4.6) and (4.7). As $\tilde{\beta}_N(\delta_0)$ is an unbiased estimator of β, and $\frac{N}{N-k}\tilde{\sigma}^2_N(\delta_0)$ is an unbiased estimator of σ^2, it is natural to expect that, with a bias-corrected QML estimator $\hat{\delta}^{\rm bc}_N$ of δ, $\hat{\beta}^{\rm bc}_N = \tilde{\beta}_N(\hat{\delta}^{\rm bc}_N)$ and $\hat{\sigma}^{2,\rm bc}_N = \frac{N}{N-k}\tilde{\sigma}^2_N(\hat{\delta}^{\rm bc}_N)$ would be much less biased than the original QML estimators. Thus, with a bias-corrected non-linear estimator, the QML estimators of the linear and scale parameters may be automatically bias-corrected, making the overall bias correction much easier. This is another point stressed by Yang (2015b) in supporting the arguments that one should use CEE to perform bias correction on non-linear parameters. We now present some results to support this point. First, $\hat{\beta}_N \equiv \tilde{\beta}_N(\hat{\delta}_N) = \mathbf{F}_N(\hat{\rho}_N)\mathbf{Y}_N(\hat{\lambda}_N)$, where $\mathbf{F}_N(\rho) = [\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)\mathbf{B}_N(\rho)$, by (4.6). Let $\tilde{\beta}^{(k)}_N(\delta)$ be the kth derivative of $\tilde{\beta}_N(\delta)$ w.r.t. δ', and $\mathbf{F}^{(k)}_N(\rho)$ the kth derivative of $\mathbf{F}_N(\rho)$ w.r.t. ρ. Assume $\mathrm{E}(\tilde{\beta}^{(k)}_N)$ exists and $\tilde{\beta}^{(k)}_N - \mathrm{E}(\tilde{\beta}^{(k)}_N) = O_p(N^{-1/2}), k = 1, 2$. By a Taylor series expansion, we obtain,[13]

$$\begin{aligned}\tilde{\beta}_N(\hat{\delta}_N) &= \tilde{\beta}_N + \tilde{\beta}^{(1)}_N(\hat{\delta}_N - \delta_0) + \tfrac{1}{2}\tilde{\beta}^{(2)}_N[(\hat{\delta}_N - \delta_0)\otimes(\hat{\delta}_N - \delta_0)] + O_p(N^{-\frac{3}{2}}), \quad (4.15)\\ &= \tilde{\beta}_N + \mathrm{E}(\tilde{\beta}^{(1)}_N)(\hat{\delta}_N - \delta_0) + b_N a_{-\frac{1}{2}} + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}^{(2)}_N)(a_{-\frac{1}{2}}\otimes a_{-\frac{1}{2}}) + O_p(N^{-\frac{3}{2}}),\end{aligned}$$

[13]A notational convention is followed: $\tilde{\beta}_N \equiv \tilde{\beta}_N(\delta_0)$, $\tilde{\beta}^{(k)}_N \equiv \tilde{\beta}^{(k)}_N(\delta_0)$, $\mathbf{F}_N \equiv \mathbf{F}_N(\rho_0)$, $\mathbf{A}_N = \mathbf{A}_N(\lambda_0)$, $\mathbf{B}_N = \mathbf{B}_N(\rho_0)$, etc.

where $\mathrm{E}(\tilde{\beta}_N^{(1)}) = [-\mathbf{F}_N\mathbf{G}_N\mathbf{X}_N\beta_0, \mathbf{F}_N^{(1)}\mathbf{X}_N\beta_0]$, $b_N = [-\mathbf{F}_N\mathbf{G}_N\mathbf{B}_N^{-1}\mathbf{V}_N,\ \mathbf{F}_N^{(1)}\mathbf{B}_N^{-1}\mathbf{V}_N]$, $\mathbf{G}_N = \mathbf{W}_{1N}\mathbf{A}_N^{-1}$, and $\mathrm{E}(\tilde{\beta}_N^{(2)}) = [0_{k\times 1},\ -\mathbf{F}_N^{(1)}\mathbf{G}_N\mathbf{X}_N\beta_0, -\mathbf{F}_N^{(1)}\mathbf{G}_N\mathbf{X}_N\beta_0, \mathbf{F}_N^{(2)}\mathbf{X}_N\beta_0]$. Recall $a_{-1/2} = \Omega_N\tilde{\psi}_N$. It is easy to see that the expansion (4.15) holds when $\hat{\delta}_N$ is replaced by $\hat{\delta}_N^{\mathrm{bc2}}$. Thus,

$$\begin{aligned}
\mathrm{Bias}(\hat{\beta}_N) &= \mathrm{E}(\tilde{\beta}_N^{(1)})\mathrm{Bias}(\hat{\delta}_N) + \mathrm{E}(b_N a_{-\frac{1}{2}}) + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}_N^{(2)})\mathrm{E}(a_{-\frac{1}{2}} \otimes a_{-\frac{1}{2}}) + O(N^{-\frac{3}{2}}), \\
\mathrm{Bias}(\hat{\beta}_N^{\mathrm{bc2}}) &= \mathrm{E}(b_N a_{-\frac{1}{2}}) + \tfrac{1}{2}\mathrm{E}(\tilde{\beta}_N^{(2)})\mathrm{E}(a_{-\frac{1}{2}} \otimes a_{-\frac{1}{2}}) + O(N^{-\frac{3}{2}}). \qquad (4.16)
\end{aligned}$$

The key term $\mathrm{E}(\tilde{\beta}_N^{(1)})\mathrm{Bias}(\hat{\delta}_N)$ of order $O(N^{-1})$ in the bias of $\tilde{\beta}_N(\hat{\delta}_N)$ is absorbed into the error term when $\hat{\delta}_N$ is replaced by $\hat{\delta}_N^{\mathrm{bc2}}$ in defining the estimator for β_0. Thus, it can be expected that the resulting bias reduction can be big, and the estimator $\hat{\beta}_N^{\mathrm{bc2}} = \tilde{\beta}_N(\hat{\delta}_N^{\mathrm{bc2}})$ is essentially second-order bias-corrected, if $\mathrm{E}(b_N a_{-1/2}) + \frac{1}{2}\mathrm{E}(\tilde{\beta}_N^{(2)})\mathrm{E}(a_{-1/2} \otimes a_{-1/2})$ is 'small'. In general, using (4.16), $\hat{\beta}_N^{\mathrm{bc2}}$ can easily be further bias-corrected to be 'truly' second-order unbiased. However, our Monte Carlo results given in Section 4.5 suggest that this may not be necessary. Finally, $\mathbf{F}_N^{(k)}(\rho), k = 1, 2$, can be easily derived.

Now, from (4.7), $\hat{\sigma}_N^2 = \tilde{\sigma}_N^2(\hat{\delta}_N) = \frac{1}{N}\mathbf{Y}_N'(\hat{\lambda}_N)\mathbf{M}_N(\hat{\rho}_N)\mathbf{Y}_N(\hat{\lambda}_N) \equiv \frac{1}{N}Q_N(\hat{\delta}_N)$. Let $Q_N^{(k)}(\delta)$ be the kth partial derivative of $Q_N(\delta)$ w.r.t. δ', and similarly $Q_N^{(k)} \equiv Q_N^{(k)}(\delta_0)$. Assume $\frac{1}{N}\mathrm{E}(Q_N^{(k)}) = O(1)$ and $\frac{1}{N}[Q_N^{(k)} - \mathrm{E}(Q_N^{(k)})] = O_p(N^{-1/2})$ for $k = 1, 2$. A Taylor series expansion gives,

$$\begin{aligned}
\tilde{\sigma}_N^2(\hat{\delta}_N) &= \tilde{\sigma}_N^2 + \tfrac{1}{N}Q_N^{(1)}(\hat{\delta}_N - \delta_0) + \tfrac{1}{2N}Q_N^{(2)}[(\hat{\delta}_N - \delta_0) \otimes (\hat{\delta}_N - \delta_0)] + O_p(N^{-\frac{3}{2}}), \\
&= \tilde{\sigma}_N^2 + \tfrac{1}{N}\mathrm{E}(Q_N^{(1)})(\hat{\delta}_N - \delta_0) + q_N a_{-\frac{1}{2}} + \tfrac{1}{2N}\mathrm{E}(Q_N^{(2)})(a_{-\frac{1}{2}} \otimes a_{-\frac{1}{2}}) \\
&\quad + O_p(N^{-\frac{3}{2}}) \qquad (4.17)
\end{aligned}$$

where the expressions for q_N and $\mathrm{E}(Q_N^{(k)}), k = 1, 2$, are given in Appendix E.

The expansion (4.17) holds when $\hat{\delta}_N$ is replaced by $\hat{\delta}_N^{\text{bc2}}$. It follows that

$$\begin{aligned}
\text{Bias}[\tfrac{N}{N-k}\tilde{\sigma}_N^2(\hat{\delta}_N)] &= \tfrac{1}{N-k}\text{E}(Q_N^{(1)})\text{Bias}(\hat{\delta}_N) + \tfrac{N}{N-k}\text{E}(q_N a_{-\frac{1}{2}}) \\
&\quad + \tfrac{1}{2(N-k)}\text{E}(Q_N^{(2)})\text{E}(a_{-\frac{1}{2}} \otimes a_{-\frac{1}{2}}) + O(N^{-\frac{3}{2}}), \\
\text{Bias}[\tfrac{N}{N-k}\tilde{\sigma}_N^2(\hat{\delta}_N^{\text{bc2}})] &= \tfrac{N}{N-k}\text{E}(q_N a_{-\frac{1}{2}}) + \tfrac{1}{2(N-k)}\text{E}(Q_N^{(2)})\text{E}(a_{-\frac{1}{2}} \otimes a_{-\frac{1}{2}}) \quad (4.18) \\
&\quad + O(N^{-\frac{3}{2}}).
\end{aligned}$$

Again, the key bias term $\frac{1}{N-k}\text{E}(Q_N^{(1)})\text{Bias}(\hat{\delta}_N)$ is removed when $\hat{\delta}_N$ is replaced by $\hat{\delta}_N^{\text{bc2}}$ in defining the estimator for σ_0^2, and our Monte Carlo results in Section 4.5 show that $\frac{N}{N-k}\tilde{\sigma}_N^2(\hat{\delta}_N^{\text{bc2}})$ is nearly unbiased for σ_0^2. In any case, one can always use (4.18) to carry out further bias correction on $\frac{N}{N-k}\tilde{\sigma}_N^2(\hat{\delta}_N^{\text{bc2}})$.

4.3.5 Inferences following bias and variance corrections

It would be interesting to further investigate the impacts of bias and variance corrections for spatial estimators on the statistical inferences concerning the model parameters. The latter issue is of a great practical relevance, as being able to assess the covariate effects in a reliable manner may be the most desirable feature in any econometric modelling activity. Unfortunately, this issue has not been addressed for the spatial panel data regression models.

One of the most interesting type of inferences for a spatial model would be the testing for the existence of spatial effects. With the availability of QML estimators $\hat{\delta}_N$ and its asymptotic variance $\Omega_N\text{E}(\tilde{\psi}_N\tilde{\psi}'_N)\Omega_N$, one can easily carry out a Wald test. However, given the fact that $\hat{\delta}_N$ can be quite biased, it is questionable that this asymptotic test would be reliable when N is not large. With the bias and variance correction results presented in Section 4.3, one can easily construct various 'bias-corrected' Wald tests. For testing $H_0 : \lambda = \rho = 0$, i.e., the joint non-existence of both types of spatial effects, we have,

$$\mathcal{W}_{N,jk}^{\text{SARAR}} = (\hat{\delta}_N^{\text{bc}j})'\text{Var}_k^{-1}(\hat{\delta}_N^{\text{bc}j})\hat{\delta}_N^{\text{bc}j}, \tag{4.19}$$

where $\hat{\delta}_N^{\text{bc}j}$ is the jth-order bias-corrected $\hat{\delta}_N$ and $\text{Var}_k(\hat{\delta}_N^{\text{bc}j})$ is the kth-order corrected variance of $\hat{\delta}_N^{\text{bc}j}$. When $j = k = 1$, $\hat{\delta}_N^{\text{bc1}} = \hat{\delta}_N$, $\text{Var}_1(\hat{\delta}_N^{\text{bc1}}) = \Omega_N \text{E}(\tilde{\psi}_N \tilde{\psi}_N')\Omega_N$, and the test is an asymptotic Wald test. The details on estimating $\text{Var}_k(\hat{\delta}_N^{\text{bc}j})$, in particular, $\text{Var}_3(\hat{\delta}_N^{\text{bc3}})$, are given at the end of Section 4.4.

Similarly, for testing the non-existence of one type of spatial effects, allowing the existence of the other type of spatial effects, i.e., $H_0 : \lambda = 0$, allowing ρ, or $H_0 : \rho = 0$ allowing λ, we have, respectively,

$$\mathcal{W}_{N,jk}^{\text{SAR}} = \hat{\lambda}_N^{\text{bc}j} / \sqrt{\text{Var}_{11,k}(\hat{\delta}_N^{\text{bc}j})} \quad \text{or} \quad \mathcal{W}_{N,jk}^{\text{SED}} = \hat{\rho}_N^{\text{bc}j} / \sqrt{\text{Var}_{22,k}(\hat{\delta}_N^{\text{bc}j})}, \tag{4.20}$$

where $\text{Var}_{ii,k}(\hat{\delta}_N^{\text{bc}j})$ denotes the i-th diagonal element of $\text{Var}_k(\hat{\delta}_N^{\text{bc}j})$. We can construct improved tests for testing the non-existence of spatial effect in the two reduced models, i.e., testing $H_0 : \lambda = 0$, given $\rho = 0$, or $H_0 : \rho = 0$, given $\lambda = 0$:

$$\mathcal{T}_{N,jk}^{\text{SAR}} = \hat{\lambda}_N^{\text{bc}j} / \sqrt{\text{Var}_k(\hat{\lambda}_N^{\text{bc}j})} \quad \text{or} \quad \mathcal{T}_{N,jk}^{\text{SED}} = \hat{\rho}_N^{\text{bc}j} / \sqrt{\text{Var}_k(\hat{\rho}_N^{\text{bc}j})}, \tag{4.21}$$

where $\text{Var}_k(\hat{\lambda}_N^{\text{bc}j})$ and $\text{Var}_k(\hat{\rho}_N^{\text{bc}j})$ are the k-order corrected variances of the jth-order bias-corrected estimators based on the corresponding reduced models described in Section 4.3.3.

Another important type of inference concerns the covariate effects, i.e., the testing or confidence interval construction for $c'\beta_0$, a linear combination of the regression parameters. For an improved inference, we need the bias-corrected variance estimator for $\hat{\beta}_N^{\text{bc2}}$. By (4.15) with $\hat{\delta}_N$ being replaced by $\hat{\delta}_N^{\text{bc2}}$, we have, $\text{Var}(\hat{\beta}_N^{\text{bc2}}) = \text{Var}\big[\tilde{\beta}_N + \text{E}(\tilde{\beta}_N^{(1)})(a_{-1/2} + a_{-1}) + b_N a_{-1/2} + \frac{1}{2}\text{E}(\tilde{\beta}_N^{(2)})(a_{-1/2} \otimes a_{-1/2})\big] +$

$O_p(N^{-2})$. This variance can be easily estimated based on the bootstrap method described at the end of Section 4.4. For testing $H_0 : c'\beta_0 = 0$, the following two statistics may be used:

$$\mathcal{T}_{N,11} = c'\hat{\beta}_N/\sqrt{c'\widehat{\text{AVar}}(\hat{\beta}_N)c}, \text{ and } \mathcal{T}_{N,22} = c'\hat{\beta}_N^{\text{bc2}}/\sqrt{c'\widehat{\text{Var}}(\hat{\beta}_N^{\text{bc2}})c}, \tag{4.22}$$

where $\widehat{\text{AVar}}(\hat{\beta}_N)$ is the estimate of the asymptotic variance of $\hat{\beta}_N$ and $\widehat{\text{Var}}(\hat{\beta}_N^{\text{bc2}})$ is the bootstrap estimate of $\text{Var}(\hat{\beta}_N^{\text{bc2}})$ (see end of Section 4.4). These results can be simplified to suit the simpler models. Monte Carlo results show that inferences based on $\mathcal{T}_{N,22}$ are much more reliable than inferences based on $\mathcal{T}_{N,11}$.

4.4 Bootstrap for Feasible Bias and Variance Corrections

For practical purpose, we need to evaluate the expectations of $a_{-s/2}$ for $s = 1, 2, 3$, and the expectations of their cross products. Thus, we need to compute expectations of all the R-, S-, and Q-ratios of quadratic forms defined below (4.14), expectations of their powers, and expectations of cross products of powers, which seem impossible analytically. The use of a joint estimating equation (JEE) as in Bao and Ullah (2007) and Bao (2013) may offer a possibility. However, even for a second-order bias correction of a simple SAR model (Bao, 2013), the formulae are seen to be very complicated already. Further, the analytical approach runs into another problem with variance corrections and higher-order bias corrections – it may involve higher than fourth moments of the errors of which estimation may not be stable numerically. In the current chapter, we follow Yang (2015b) to use the CEE, $\tilde{\psi}_N(\delta) = 0$, which not only reduces the dimensionality but also captures additional bias a nd variability from the estimation of linear and scale parameters, making the bias correction more effective. We then use bootstrap to estimate these expectations involved in the bias and variance corrections, which overcomes the difficulty in analytically evaluating the expectations and avoids the direct estimation of higher-order moments of the errors.

4.4.1 The bootstrap method

We follow Yang (2015b) and propose a bootstrap procedure for the FE-SPD model with SARAR effects. Note $\mathbf{Y}_N(\lambda_0) = \mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}(\rho_0)\mathbf{V}_N$, $\mathbf{W}_{1N}\mathbf{Y}_N = \mathbf{G}_N[\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}(\rho_0)\mathbf{V}_N]$, where $\mathbf{G}_N \equiv \mathbf{G}_N(\lambda_0) = \mathbf{W}_{1N}\mathbf{A}^{-1}(\lambda_0)$, and $\mathbf{M}_N(\rho)\mathbf{X}_N = 0$. The R-ratios, S-ratios and Q-ratios at $\delta = \delta_0$ defined below (4.14) can all be written as functions of $\zeta_0 = (\beta_0', \delta_0')'$ and $\mathbf{V}_N$, given $\mathbf{X}_N$ and $\mathbf{W}_{jN}, j = 1, 2$:

$$R_{1N}(\zeta_0, \mathbf{V}_N) = \frac{\mathbf{V}_N'\mathbf{B}_N'^{-1}\mathbf{M}_N\mathbf{G}_N(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.23}$$

$$R_{2N}(\zeta_0, \mathbf{V}_N) = \frac{(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)'\mathbf{G}_N'\mathbf{M}_N\mathbf{G}_N(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.24}$$

$$Q_{kN}^{\dagger}(\zeta_0, \mathbf{V}_N) = \frac{(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)'\mathbf{M}_N^{(k)}\mathbf{G}_N(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.25}$$

$$Q_{kN}^{\ddagger}(\zeta_0, \mathbf{V}_N) = \frac{(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)'\mathbf{G}_N'\mathbf{M}_N^{(k)}\mathbf{G}_N(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.26}$$

$$S_{kN}(\zeta_0, \mathbf{V}_N) = \frac{(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)'\mathbf{M}_N^{(k)}(\mathbf{X}_N\beta_0 + \mathbf{B}_N^{-1}\mathbf{V}_N)}{\mathbf{V}_N'\mathbf{M}_N^{\diamond}\mathbf{V}_N}, \tag{4.27}$$

where $\mathbf{M}_N^{\diamond} = I_N - \mathbf{X}_N(\rho_0)[\mathbf{X}_N'(\rho_0)\mathbf{X}_N(\rho_0)]^{-1}\mathbf{X}_N'(\rho_0)$ and $\mathbf{M}_N^{(k)} \equiv \mathbf{M}_N^{(k)}(\rho_0)$. It follows that $\tilde{\psi}_N = \tilde{\psi}_N(\zeta_0, \mathbf{V}_N)$ and $H_{rN} = H_{rN}(\zeta_0, \mathbf{V}_N), r = 1, 2, 3$. Now, define the QML estimate of the error vector $\mathbf{V}_N$ in the FE-SPD model (4.4):

$$\hat{\mathbf{V}}_N = \mathbf{B}_N(\hat{\rho}_N)[\mathbf{A}(\hat{\lambda}_N)\mathbf{Y}_N - \mathbf{X}_N\hat{\beta}_N]. \tag{4.28}$$

Let $\hat{\mathbf{V}}_N^{\star}$ be a bootstrap sample based on $\hat{\mathbf{V}}_N$. The bootstrap analogues of various quantities are $\tilde{\psi}_N^{\star} \equiv \tilde{\psi}_N(\hat{\zeta}_N, \mathbf{V}_N^{\star})$ and $H_{rN}^{\star} \equiv H_{rN}(\hat{\zeta}_N, \mathbf{V}_N^{\star})$, $r = 1, 2, 3$. Thus, the bootstrap estimates of the quantities in bias and variance corrections are,

$$\begin{aligned}\widehat{\mathrm{E}}(\tilde{\psi}_N \otimes H_{rN}) &= \mathrm{E}^{\star}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star}) \otimes H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star})\big], \text{ and} \\ \widehat{\mathrm{E}}(\tilde{\psi}_N \otimes \tilde{\psi}_N \otimes \tilde{\psi}_N) &= \mathrm{E}^{\star}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_N^{\star})\big],\end{aligned}$$

where $\mathrm{E}^{\star}$ denotes the expectation with respect to the bootstrap distribution. The bootstrap estimates of other quantities are defined in the same manner.[14] To make these bootstrap expectations practically feasible, we first follow Yang (2015b) and propose the following *iid bootstrap* procedure:

Algorithm 4.1 *(iid Bootstrap)*

1. *Compute* $\hat{\zeta}_N$ *and* $\hat{\mathbf{V}}_N$*, and centre* $\hat{\mathbf{V}}_N$*.*
2. *Draw a bootstrap sample* $\hat{\mathbf{V}}_{N,b}^{\star}$*, i.e., make* N *random draws from the elements of centred* $\hat{\mathbf{V}}_N$*.*
3. *Compute* $\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_{N,b}^{\star})$ *and* $H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}_{N,b}^{\star}), r = 1, 2, 3$*.*
4. *Repeat steps 2-3 for* B *times to give approximate bootstrap estimates as*

$$\begin{aligned}\widehat{\mathrm{E}}(\tilde{\psi}_N \otimes H_{rN}) &\doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_{N,b}^{\star}) \otimes H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}_{N,b}^{\star})\big], \text{ and} \\ \widehat{\mathrm{E}}(\tilde{\psi}_N \otimes \tilde{\psi}_N \otimes \tilde{\psi}_N) &\doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\big[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_{N,b}^{\star}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_{N,b}^{\star}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}_{N,b}^{\star})\big].\end{aligned}$$

The approximation in the last step of Algorithm (4.1) can be made arbitrarily accurate by choosing an arbitrarily large B, and that the scale parameter σ_0^2 and its QML estimator $\hat{\sigma}_N^2$ do not play a role in the bootstrap process as they

[14]To facilitate the bootstrapping, the $a_{-s/2}$ in (4.9) can be re-expressed so that the random quantities are put together, using the well-known properties of Kronecker product: $(A \otimes B)(C \otimes D) = AC \otimes BD$ and $\mathrm{vec}(ACB) = (B' \otimes A)\mathrm{vec}(C)$, where 'vec' vectorizes a matrix by stacking its columns. For example, $H_{1N}\Omega_N\tilde{\psi}_N = (\tilde{\psi}_N' \otimes H_{1N})\mathrm{vec}(\Omega_N)$, and $a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2} = (\Omega_N \otimes \Omega_N \otimes \Omega_N)(\tilde{\psi}_N \otimes \tilde{\psi}_N \otimes \tilde{\psi}_N)$. Alternatively, one can follow the 'two-step' procedure given in Yang (2015b, Sec. 4).

are hidden in either $\mathbf{V}_N$ or $\hat{\mathbf{V}}_N$. The iid bootstrap procedure requires that the underlining error vector $\mathbf{V}_N$ contains iid elements, which may not be true if the original errors are not normal. However, the fact that the elements of $\mathbf{V}_N$ are uncorrelated and homoskedastic suggests that applying the iid bootstrap may give a good approximation although it may not be strictly valid. Nevertheless, when the original errors are non-normal, the following *wild bootstrap* or *perturbation* procedure can be used.

Algorithm 4.2 *(Wild Bootstrap)*

1. *Compute* $\hat{\zeta}_N$ *and* $\hat{\mathbf{V}}_N$*, and centre* $\hat{\mathbf{V}}_N$.
2. *Compute* $\hat{\mathbf{V}}^{\star}_{N,b} = \hat{\mathbf{V}}_N \odot \boldsymbol{\varepsilon}_b$*, where* $\odot$ *denotes the Hadamard product, and* $\boldsymbol{\varepsilon}_b$ *is an* N*-vector of iid draws from a distribution of mean zero and all higher moments 1, and is independent of* $\hat{\mathbf{V}}_N$.[15]
3. *Compute* $\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})$ *and* $H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}), r = 1, 2, 3$.
4. *Repeat steps 2-3 for* B *times to give approximate bootstrap estimates as*

$$
\begin{aligned}
\widehat{\mathrm{E}}(\tilde{\psi}_N \otimes H_{rN}) &\doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\left[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes H_{rN}(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})\right], \text{ and} \\
\widehat{\mathrm{E}}(\tilde{\psi}_N \otimes \tilde{\psi}_N \otimes \tilde{\psi}_N) &\doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B}\left[\tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b}) \otimes \tilde{\psi}_N(\hat{\zeta}_N, \hat{\mathbf{V}}^{\star}_{N,b})\right].
\end{aligned}
$$

Note that the common applications of the wild bootstrap method are to handle the problem of unknown heteroskedasticity, which clearly is not the main purpose of this chapter. In our model, the (transformed) errors are homoskedastic in the usual sense, i.e., variances are constant. Also, the errors are uncorrelated. However, the transformed errors are, strictly speaking, heteroskedastic in the sense

[15] We are unaware of the existence of such a distribution. However, the two-point distribution suggested by Mammen (1993): $\varepsilon_{b,i} = -(\sqrt{5}-1)/2$ or $(\sqrt{5}+1)/2$ with probability $(\sqrt{5}+1)/(2\sqrt{5})$ or $(\sqrt{5}-1)/(2\sqrt{5})$, has mean zero, and second and third moments 1. Another two-point distribution: $\varepsilon_{b,i} = -1$ or 1 with equal probability, has all the odd moments zero and even moments 1. See Liu (1988) and Davidson and Flachaire (2008) for more details on wild bootstrap.

that their third and higher order moments may not be constant. The wild bootstrap here aims to capture these non-constant higher-order moments. Also, there may be higher-order dependence, which the wild bootstrap is not able to capture. We see in the next section that this can be ignored.

4.4.2 Validity of the bootstrap method

In discussing the validity of the bootstrap method, we concentrate on bias corrections. The fact that the elements of the transformed errors $\mathbf{V}_N = \{v^*_{it}\}$ are uncorrelated and homoskedastic (up to second moment) across i and t, and its observed counterpart $\hat{\mathbf{V}}_N$ is consistent provide the theoretical base for the proposed iid bootstrap method. However, these may not be sufficient for the iid bootstrap method to be strictly valid, as our estimation requires matching of the higher-order bootstrap moments with those of v^*_{it}. There are important special cases under which the classical iid bootstrap method is strictly valid.

First, we note that the original errors $\{v_{it}\}$ are iid normal, the transformed errors $\{v^*_{it}\}$ are again iid normal. Further, Lemma 4.1 shows that if the original errors $\{v_{it}\}$ are iid with mean zero, variance σ_0^2, and cumulants $k_r = 0, r = 3, 4, \ldots$, then the transformed errors $\{v^*_{it}\}$ will also have mean zero, variance σ_0^2, and rth cumulant being zero for $r = 3, 4, \ldots$. Furthermore, the rth order joint cumulants of the transformed errors are also zero. The iid bootstrap procedure essentially falls into the general framework of Yang (2015b) and hence its validity is fully established. We have the following proposition.

Proposition 4.1 *Suppose the conditions leading to the third-order bias expansion* (4.10) *are satisfied by the FE-SPD model. Assume further that the* r*th cumulant* k_r *of* $\{v_{it}\}$ *is 0,* $r = 3, \ldots, 10$*. Then the iid bootstrap method stated in Algorithm 4.1 is valid, i.e., Bias*$(\hat{\delta}_N^{bc2}) = O(N^{-3/2})$ *and Bias*$(\hat{\delta}_N^{bc3}) = O(N^{-2})$*.*

Second, for the important sub-model with individual effects only and small T, the transformed errors, $[V_{n1}^*, \ldots, V_{n,T-1}^*] = [V_{n1}, \ldots, V_{n,T}]F_{T,T-1}$ are iid across i, i.e., the rows of the matrix $[V_{n1}^*, \ldots, V_{n,T-1}^*]$ are iid whether the original errors are normal or non-normal, where $N = n(T-1)$. As T is small and fixed, the asymptotics depend only on n. The bootstrap thus proceeds by randomly drawing the rows of the QML estimate of $[V_{n1}^*, \ldots, V_{n,T-1}^*]$. We have the following proposition.

Proposition 4.2 *Suppose the conditions leading to the third-order bias expansion* (4.10) *are satisfied by the FE-SPD model with only individual effects. Assume further that the* r*th cumulant* k_r *of* $\{v_{it}\}$ *exists,* $r = 3, \ldots, 10$*, and* T *is fixed. Then the bootstrap method making iid draws from the rows of the QML estimates of* $[V_{n1}^*, \ldots, V_{n,T-1}^*]$ *is valid, i.e., Bias*$(\hat{\delta}_N^{bc2}) = O(N^{-3/2})$ *and Bias*$(\hat{\delta}_N^{bc3}) = O(N^{-2})$.

For the general FE-SPD model with two-way fixed effects, T being small or large, and the original errors being iid but not necessarily normal, the classical iid bootstrap may not be strictly valid, because the transformed errors (on which the iid bootstrap depend) are not guaranteed to be iid, although they are uncorrelated with mean zero and constant variance σ_0^2. In particular, the transformed errors may not be independent, and their higher-order moments (3rd-order and higher) may not be constant. On the other hand, making random draws from the empirical distribution function (EDF) of the centred $\hat{\mathbf{V}}_N$ gives bootstrap samples that are of iid elements. Thus, the classical iid bootstrap does not fully *mimic* or *recreate* the random structure of $\mathbf{V}_N$, rendering its strict validity questionable. The following proposition says that the wild bootstrap described in Algorithm 4.2 is valid.

Proposition 4.3 *Suppose the conditions leading to the third-order bias expansion* (4.10) *are satisfied by the FE-SPD model. Assume further that the* r*th cumulant* k_r *of* $\{v_{it}\}$ *exists for* $r = 3, \ldots, 10$*. Then the wild bootstrap method stated in Algorithm 4.2 is valid for the general FE-SPD model, provided that the* *joint cumulants of the transformed errors* $\{v_{it}^*\}$ *up to* r*th order,* $r = 3, \ldots, 10$*, are negligible.*

Proof: We now present a collective discussion/proof of the Propositions 4.1-4.3. Very importantly, we want to 'show' that the classical iid bootstrap method can give a very good approximation in cases it is not strictly valid, i.e., the 'missing parts' can be ignored numerically.

Let $\mathbb{V}_{nT} = (V_{n1}', \ldots, V_{nT}')'$ be the vector of original errors in Model (4.1), which contains iid elements of mean zero, variance σ_0^2, cumulative distribution function (CDF) $\mathcal{F}$, and cumulants $k_r, r = 3, 4, \ldots, 10$. Let $\mathbb{F}_{nT,N} = F_{T,T-1} \otimes F_{n,n-1}$ be the $nT \times N$ transformation matrix. We have

$$\mathbf{V}_N = \mathbb{F}_{nT,N}' \mathbb{V}_{nT}. \tag{4.29}$$

For convenience, denote the elements of $\mathbf{V}_N$ by $\mathbf{v}_i$, and the ith column of $\mathbb{F}_{nT,N}$ by $\mathbf{f}_i, i = 1, \ldots, N$. Let $\kappa_r(\cdot)$ denote the rth cumulant of a random variable, and $\kappa(\cdot, \ldots, \cdot)$ the joint cumulants of random variables. Let $\odot$ denote the Hadamard product. A vector raised to rth power is operated element-wise. From the definition of the bias terms $b_{-s/2}, s = 2, 3$, we see that $b_{-s/2} \equiv b_{-s/2}(\zeta_0, \boldsymbol{\kappa}_N)$ where $\boldsymbol{\kappa}_N$ contains the cumulants or joint cumulants of $\{\mathbf{v}_i\}$. From (4.23)-(4.28), it is clear that the bootstrap estimates of $b_{-s/2}$ are such that $\hat{b}_{-s/2} \equiv b_{-s/2}(\hat{\zeta}_N, \hat{\boldsymbol{\kappa}}_N^\star)$ where $\hat{\boldsymbol{\kappa}}_N^\star$ contains the cumulants of $\{\mathbf{v}_i^\star\}$ w.r.t. the bootstrap distribution. With the $\sqrt{N}$-consistency of $\hat{\theta}_N$, how the set $\hat{\boldsymbol{\kappa}}_N^\star$ match the set $\boldsymbol{\kappa}_N$, becomes central to the validity of the bootstrap method. Following lemmas reveal their relationship.

Lemma 4.1 *If the elements of* $\mathbb{V}_{nT}$ *are iid with mean zero, variance* σ_0^2*, CDF* $\mathcal{F}$*, and higher-order cumulants* $k_r, r = 3, 4, \ldots$*, then,*

(*a*) $\kappa_1(\mathbf{v}_i) = 0,\ \kappa_2(\mathbf{v}_i) = \sigma_0^2,$ *and* $\kappa_r(\mathbf{v}_i) = k_r\ a_{r,i},\ r \geq 3,\ i = 1, \ldots, N,$

(*b*) $\kappa(\mathbf{v}_i, \mathbf{v}_j) = 0$ *for* $i \neq j,$ *and* $\kappa(\mathbf{v}_{i_1}, \ldots, \mathbf{v}_{i_r}) = k_r\ a_{i_1,\ldots,i_r},\ r \geq 3,$ *where*

$a_{r,i} = l'_{nT}\mathbf{f}_i^r$, $a_{i_1,\ldots,i_r} = l'_{nT}(\mathbf{f}_{i_1} \odot \cdots \odot \mathbf{f}_{i_r})$, *and* $\{i_1, \ldots, i_r\}$ *are not all the same.*

Lemma 4.1 shows clearly that the higher-order cumulants or joint cumulants of $\{\mathbf{v}_i\}$ are proportional to the higher-order cumulants k_r of the original errors $\{v_{it}\}$. This suggests that when $k_r = 0, r = 3, \ldots, 10$, $\{\mathbf{v}_i\}$ are essentially iid and hence the conclusion of Proposition 4.1 holds in light of the results of Yang (2015b) for the iid bootstrap. Similarly, the conclusion of Proposition 4.2 also holds. When $k_r \neq 0$ for some or all $r = 3, \ldots, 10$, $\{\mathbf{v}'_i\}$ are no longer iid. First, $a_{r,i}$ are constant across i only when $r = 1$ and 2, i.e., $a_{1,i} = 0$ and $a_{2,i} = 1$. Thus, $\kappa_r(\mathbf{v}_i), r \geq 3$, are not constant across i unless $k_r = 0$. Second, $\mathbf{v}'_is$ are not independent as $a_{i_1,\ldots,i_r} \neq 0$ for $r \geq 3$. The latter may cause more problem as it is known that the iid bootstrap is unable to capture dependence. However, noting that the proportionality constants $a_{i_1,\ldots,i_r}$ are all pure numbers, being the sum of element-wise products of the orthonormal vectors $\{\mathbf{f}_i\}$, intuitively they should be small, and the larger the r, the smaller the $a_{i_1,\ldots,i_r}$.[16] These suggest that the higher-order dependence among $\{\mathbf{v}_i\}$ can largely be ignored. The question left is how well the two sets of cumulants match.

Lemma 4.2 *Let* $\mathbf{v}^\star$ *be a random draw from* $\{\mathbf{v}_i, i = 1, \ldots, N\}$. *Then, under the conditions of Lemma 4.1, we have* $\kappa_1^\star(\mathbf{v}^\star) = 0, \kappa_2^\star(\mathbf{v}^\star) = \sigma_0^2 + O_p(N^{-1/2})$ *and* $\kappa_r^\star(\mathbf{v}^\star) = k_r\bar{a}_r + O_p(N^{-1/2}), r \geq 3$, *where* $\bar{a}_r = \frac{1}{N}\sum_{i=1}^N a_{r,i}$, *and* $\kappa_r^\star(\cdot)$ *denotes rth cumulant w.r.t. the EDF* $\mathcal{G}_N$ *of* $\{\mathbf{v}_i, i = 1, \ldots, N\}$.

Lemma 4.2 shows that the *iid bootstrap* is able to capture, to a certain degree, the higher-order moments of $\mathbf{v}_i$ ($\bar{a}_r$ versus $a_{r,i}$), but is unable to capture the

[16] We are unable to further characterise these quantities. However, as they are pure numbers depending on n and T through $F_{T,T-1}$ and $F_{n,n-1}$, it should be indicative to present some of their values. With the eigenvector-based transformations defined above (4.2) and calculated using Matlab `eig` function, we have, for $n = 100$ and $T = 3$, $a_{1,2,3} = -5.6e^{-5}$, $a_{1,2,3,4} = 3.4e^{-5}$, and $a_{1,2,3,4,5} = -3.7e^{-7}$; and for $n = 200$, the same set of numbers become $2.3e^{-5}$, $-3.8e^{-6}$ and $1.3e^{-8}$. With Helmert transformations (see Footnote 5), these numbers become much smaller ($< 1.0e^{-19}$).

higher-order dependence. As argued below Lemma 4.1, the latter does not have a significant effect as such dependence is weak and negligible since both $\{a_{r,i}\}$ and their variability are not big and get smaller as r increases,[17] the results of Lemmas 4.1-4.3 strongly suggest that the simple iid bootstrap method may be able to give a good approximation in situations where the original errors are not far from normal.

Lemma 4.3 *Suppose Assumptions A1-A8 and the conditions of Lemma 4.1 hold. Let* $\hat{\mathbf{v}}^\star$ *be a random draw from the EDF* $\hat{\mathcal{G}}_N$ *of* $\{\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_N\}$*, and* $\mathbf{v}^\star$ *a random draw from the EDF* $\mathcal{G}_N$ *of* $\{\mathbf{v}_1 \ldots, \mathbf{v}_N\}$*. Then,*

$$\kappa_r^\star(\hat{\mathbf{v}}^\star) = \kappa_r^\star(\mathbf{v}^\star) + O_p(N^{-1/2}),\ \ or\ \kappa_r(\hat{\mathcal{G}}_N) = \kappa_r(\mathcal{G}_N) + O_p(N^{-1/2}),\ r \geq 3,$$

where $\kappa_r^\star(\hat{\mathbf{v}}^\star)$ *is the* r*th cumulant of* $\hat{\mathbf{v}}^\star$ *w.r.t.* $\hat{\mathcal{G}}_N$*, and* $\kappa_r^\star(\mathbf{v}^\star)$ *is the* r*th cumulant of* $\mathbf{v}^\star$ *w.r.t.* $\mathcal{G}_N$*.*

In case of severe non-normality of the original errors so that the transformed errors are far from being iid, it may be more important to be able to match the even moments, in particular the kurtosis, than the odd moments as $a_{r,i}$ is typically small on average with moderate variability when r is odd, see Footnote 16. This point is also reflected by the fact that the variance of the joint score function (given in Theorem A.1) is free from the third cumulant of the original error. In this spirit, the simple two-point distribution with equal probability described in Footnote 10 may provide satisfactory results.

Lemma 4.4 *Suppose Assumptions A1-A8 and the conditions of Lemma 4.1 hold. Let* $\hat{\mathbf{v}}_i^\star = \hat{\mathbf{v}}_i\varepsilon^\star$*, where* $\varepsilon^\star$ *is independent of* $\hat{\mathbf{v}}_i$*, having a distribution with*

[17] Again, we are unable to further characterise these pure constants. To have some concrete idea, we have calculated the mean and standard deviation of $\{a_{r,i}\}$ for $n = 100, T = 3$ and $r = 3, 4, 5, 6$: $(-.0020, .0827)$, $(.1245, .0679)$, $(-.0010, .0425)$, $(.0308, .0299)$. When $n = 500$, the same set of values becomes: $(.0008, .0751)$, $(.1141, .0714)$, $(.0010, .0360)$, $(.0263, .0281)$. With Helmert transformations, these numbers become slightly bigger.

mean 0 and rth moment 1, $r \geq 2$. Then,

$$\mathrm{E}^{\star}(\hat{\mathbf{v}}_i^{\star}) = 0, \ \textit{and} \ \mathrm{E}^{\star}[(\hat{\mathbf{v}}_i^{\star})^r] = \hat{\mathbf{v}}_i^r, \ r \geq 2,$$

where $\mathrm{E}^{\star}$ corresponds to the distribution of $\varepsilon^{\star}$.

Lemma 4.3 shows that moving from the model errors to their observed counterparts introduces errors of smaller order and hence can be ignored asymptotically. With the results of Lemma 4.4, the validity of the wild bootstrap follows. The proofs of Lemmas 4.1-4.4 are given in Appendix E.

Variance corrections. A final note is given to the variance correction. The bootstrap estimate of a bias term or a variance term typically has a bias of order $O(N^{-1})$ multiplied by the order of that term, i.e., $\mathrm{Bias}(\hat{b}_{-1}) = O(N^{-2})$, $\mathrm{Bias}(\hat{v}_{-1}) = O(N^{-2})$, $\mathrm{Bias}(\hat{v}_{-3/2}) = O(N^{-5/2})$, etc. This is sufficient for achieving a third-order bias correction, but not for a third-order variance correction. Thus, to achieve a third-order variance correction (up to $O(N^{-2})$), a further correction on the bootstrap estimate $\hat{v}_{-1}$ of v_{-1} is desirable. Yang (2015b) proposed a method based on the first-order variance term obtained from the joint estimating function. To avoid algebraic complications, we adopt a simple approximation method: replacing $\hat{v}_{-1}$ evaluated at the original QML estimator $\hat{\theta}_N$, by $\hat{v}_{-1}^{\mathrm{bc}}$ evaluated at the second-order bias-corrected QML estimator $\hat{\theta}_N^{\mathrm{bc2}}$. Monte Carlo results show that this approximation works well.

To have a third-order variance correction for $\hat{\delta}_N^{\mathrm{bc3}}$, we also need to estimate $\mathrm{ACov}(\hat{\delta}_N, \hat{b}_{-1})$ in (4.13). We write $\mathrm{ACov}(\hat{\delta}_N, \hat{b}_{-1}) = \mathrm{ACov}(\hat{\delta}_N, \hat{\zeta}_N)\mathrm{E}(b'_{-1,\zeta_0})$, where b_{-1,ζ_0} is the partial derivative of b_{-1} w.r.t. ζ'_0, and $\mathrm{ACov}(\hat{\delta}_N, \hat{\zeta}_N)$ is the sub-matrix of $\mathrm{E}\big(\frac{\partial}{\partial\theta'_0}\psi_N(\theta_0)\big)^{-1}\mathrm{Var}\big(\psi_N(\theta_0)\big)\mathrm{E}\big(\frac{\partial}{\partial\theta'_0}\psi_N(\theta_0)\big)^{-1}$, where $\psi_N(\theta) = \frac{\partial}{\partial\theta'}\ell_N(\theta)$. The detailed expressions of $\psi_N(\theta) = \frac{\partial}{\partial\theta'}\ell_N(\theta)$, $\mathrm{Var}\big(\psi_N(\theta_0)\big)$, and $\mathrm{E}\big(\frac{\partial}{\partial\theta'_0}\psi_N(\theta_0)\big)$ are given in Theorem E.1 in Appendix E. We estimate $\mathrm{E}(b_{-1,\zeta_0})$ by $\hat{b}_{-1,\hat{\zeta}_N}$, the numerical derivatives. $\mathrm{E}(\frac{\partial}{\partial\theta_0'}\psi_N(\theta_0))$ can simply be estimated by the plug-in method as it involves only the parameter-vector θ_0. $\mathrm{Var}(\frac{\partial}{\partial\theta_0}\ell_N(\theta_0))$ involves k_4, the fourth cumulant of the original errors, besides the parameter-vector θ_0. The results of Lemmas 4.1-4.3 suggest that k_4 can be consistently estimated by $\hat{k}_4 = \bar{a}_4^{-1}\kappa_4(\hat{\mathbf{V}}_N)$, where $\kappa_4(\hat{\mathbf{V}}_N)$ is the fourth sample cumulant of the QML residuals $\hat{\mathbf{V}}_N$, and $\bar{a}_4$ is given in Lemma 4.2.

Finally, to estimate $\widehat{\mathrm{Var}}(\hat{\beta}_N^{\mathrm{bc2}})$ in (4.22): we need to (i) calculate the estimates of all non-stochastic quantities with analytical expressions by plugging in $\hat{\delta}_N^{\mathrm{bc2}}$ and $\hat{\beta}_N^{\mathrm{bc2}}$ for δ_0 and β_0, (ii) calculate new QML residuals based on $\hat{\delta}_N^{\mathrm{bc2}}$ and $\hat{\beta}_N^{\mathrm{bc2}}$, and (iii) bootstrap the new residuals to give bootstrap estimates of the other quantities in $\mathrm{Var}(\hat{\beta}_N^{\mathrm{bc2}})$, including Ω_N and $\mathrm{E}(H_{2N})$, and hence the final estimate $\widehat{\mathrm{Var}}(\hat{\beta}_N^{\mathrm{bc2}})$ of $\mathrm{Var}(\hat{\beta}_N^{\mathrm{bc2}})$. The estimates of Ω_N and $\mathrm{E}(H_{2N})$ from the early stage bootstrap based on the original QML estimators $\hat{\delta}_N$ and $\hat{\beta}_N$ can be used.

4.5 Monte Carlo Study

We present Monte Carlo results to show (i) the finite sample performance of the QML estimator $\hat{\delta}_N$ and the bias-corrected QML estimators $\hat{\delta}_N^{\mathrm{bc2}}$ and $\hat{\delta}_N^{\mathrm{bc3}}$, (ii) the impact of bias corrections for $\hat{\delta}_N$ on the estimations for β and σ^2, and (iii) the impact of bias and variance correction on the inferences for spatial or regression coefficients. The simulations are carried out based on the following data generation process (DGP): for $t = 1, \ldots, T$

$$Y_{nt} = \lambda_0 W_{1n} Y_{nt} + X_{1nt}\beta_{10} + X_{2nt}\beta_{20} + \mathbf{c}_{n0} + \alpha_{t0} l_n + U_{nt}, \quad U_{nt} = \rho_0 W_{2n} U_{nt} + V_{nt}.$$

For all the Monte Carlo experiments, $\beta_0 = (\beta_{10}, \beta_{20})'$ is set to $(1,1)'$, $\sigma_0^2 = 1$, λ_0 and ρ_0 take values from $\{-0.5, -0.25, 0, 0.25, 0.5\}$, $n = \{25, 50, 100, 200, 500\}$, and $T = \{3, 10\}$. Each set of Monte Carlo results is based on $M = 5000$ Monte

Carlo samples, and $B = 999$ bootstrap samples within each Monte Carlo sample. The $F_{T,T-1}$ and $F_{n,n-1}$ defined above (4.2) are used and calculated using Matlab `eig` function. The weight matrices, the regressors, and the idiosyncratic errors are generated as follows. The manner in which weight matrices, regressors, and error distribution are generated are presented in Appendix E.

The estimators of spatial parameters. The finite sample performance of the QML estimators and bias-corrected QML estimators of the spatial parameters is investigated. Monte Carlo results are summarised in Tables 4.1a, 4.1b, 4.2, 4.3a and 3b, where Tables 4.1a-4.1b correspond to the model with $\rho = 0$, i.e., the spatial lag dependence model; Table 4.2 the model with $\lambda = 0$, i.e., the spatial error dependence model; and Tables 4.3a-4.3b the general model. All the reported results correspond to the iid bootstrap method given in Algorithm 4.1. The results (unreported for brevity) using the wild bootstrap method described in Algorithm 4.2 show that the wild bootstrap gives almost identical results as the iid bootstrap, consistent with remarks below Lemma 4.2.

From Tables 4.1a and 4.1b, we see that regular QML estimators of the spatial parameters can be very biased, depending on the spatial layouts, the true values of the parameters, and the way that the regressors are generated. First, when the number of cross sectional units increases from 50 to 500, the magnitude of the bias becomes small. The bias is apparent for $n = 50$ and negligible for $n = 500$, which implies that bias correction is especially needed for the data with a small sample size. Also, when the spatial weights matrix becomes denser (from the queen matrix to the group interaction matrix), the bias of regular QML estimators becomes larger. When the true value of spatial effect parameter becomes larger in absolute value, the bias becomes larger. Either reducing the magnitude of the regression parameters β or increasing the value of the error standard deviation increases the bias of the QML estimator of the spatial parameter. The magnitude of the bias is also influenced by the way that the regressors are generated. The DGPs with normal errors and log-normal errors give a smaller bias than the DGP with normal mixture errors. For the bias correction, we see that our bias correction procedure works very well, independent of the spatial layouts, model parameters, and the way the regressors being generated. We see that even for the small sample case of $n = 50$, the bias correction procedure produces nearly unbiased estimates. By comparing $\hat{\lambda}_n^{\mathrm{bc2}}$ and $\hat{\lambda}_n^{\mathrm{bc3}}$, we see that in most of the situations considered, a second-order bias correction has essentially removed the bias of the QML estimators and the third-order bias correction might not be needed.

The results in Table 4.2 show that the patterns observed from the spatial lag model for the regular QML estimators and bias corrections generally hold for the spatial error model. A noticeable difference is that the regular QML estimator of the spatial error parameter can be much more biased and the bias can be much more persistent than the QML estimator of the spatial lag parameter in the spatial lag model. Therefore, the bias correction procedures developed in the current chapter works even more effectively for the spatial error model. Furthermore, unlike the case of spatial lag model, the magnitude of β and σ does not affect the performance of $\hat{\rho}_N$ much.

From Tables 4.3a and 4.3b where the third-order bias correction results are omitted for brevity, we see that the general patterns we observed for the two special models hold for the general model as well. However, we observe that the QML estimator of the spatial error parameter can be much more biased than the QML estimator of the spatial lag parameter, in particular when the regressors are generated in a non-iid manner. The bias of the QML estimator of the spatial error parameter can be very persistent and even when $n = 500$, there can still exist very noticeable bias.

The results show that in general the QML estimators of the spatial panel data models need to be bias-corrected even when sample size is not small, and that the proposed bias correction method is very effective in removing the bias. As far as the bias correction is concerned, a simple iid bootstrap may well serve the purpose. The method can easily be applied and thus is recommended to the practitioners.

The estimators of non-spatial parameters. The finite sample properties of $\hat{\beta}_N$ and $\hat{\sigma}^2_N$, and their bias-corrected versions $\hat{\beta}^{\text{bc}}_N$ and $\hat{\sigma}^{2,\text{bc}}_N$ defined in Section 4.3.4 are investigated. Monte Carlo results reveal some interesting phenomena. The biases of the non-spatial estimators $\hat{\beta}_N$ and $\hat{\sigma}^2_N$ depend very much on whether $\hat{\lambda}_N$ is biased, not much on whether $\hat{\rho}_N$ is biased. In general the biases of $\hat{\beta}_N$ and $\hat{\sigma}^2_N$ are not problems of serious concern (at most 6-7% for the experiments considered). Consistent with the discussions in Section 4.3.4, $\hat{\beta}^{\text{bc}}_N$ is nearly unbiased in general. When the error distribution is skewed, $\hat{\sigma}^{2,\text{bc}}_N$ may still encounter a bias of less than 5% when $n = 50$ and $T = 3$, and in this case the method given in Section 4.3.4 can be applied for further bias correction. Partial results are summarised in Table 4.4.

Inferences following bias and variance corrections. To demonstrate the potential gains from bias and variance corrections, we present Monte Carlo results concerning the finite sample performance of various tests for spatial effects, and the tests concerning the regression coefficients, presented in Section 4.3.5. Partial results are summarised in Tables 4.5a-4.5c, and 4.6. More comprehensive results are available from the authors upon request.

Table 4.5a presents the empirical sizes of, respectively, the joint tests for the lack of both SLD and SED effects given in (4.19), and the one-directional tests for the lack of SLD effect allowing the presence of SED effect or the lack of SED effect allowing the presence of SLD effect, given in (4.20). The results show that the third-order bias and variance corrections on the spatial estimators lead to tests that can have a much better finite sample performance over the tests based on the original estimates and asymptotic variances. The tests based on second-order corrections offer improvements over the asymptotic ones but may not be satisfactory. All the reported results are based on the wild bootstrap with the perturbation distribution being the simple two-point (1 and -1) distribution with equal probability. Consistent with the results of Section 4.2, in case of severe non-normality such as the log-normal errors, the wild bootstrap perform better than the iid bootstrap; in case of normal errors, the iid bootstrap performs slightly better than the wild bootstrap and both show excellent performance of the third-order corrected Wald tests. Due to its robustness, the wild bootstrap may be a better choice in the case of testing for spatial effects. Tables 4.5b and 4.5c present the empirical sizes of the tests given in (4.21) for the two simpler models, from which the same conclusions are drawn.

Table 4.6 presents partial results for the empirical sizes of the tests for the equality of the two regression slopes given in (4.22), based on iid bootstrap. The results show that the tests with merely second-order bias and variance corrections significantly outperforms the standard tests with the original estimate and asymptotic variance. With smaller values of the slope parameters, the size distortion for the standard tests becomes more persistent. The results (unreported for brevity) shows that when the spatial dependence becomes weaker the performance of the asymptotic test improves, but is still outperformed by the proposed bias-corrected test.

Table 4.1a. Empirical Mean[rmse](sd) of Estimators of λ, 2FE-SPD Model with SLD, $T = 3, \beta = (\mathbf{1}, \mathbf{1})', \sigma = 1$

λ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$
	(a) Queen Contiguity, REG1			(b) Group Interaction, REG2		
	Normal Error, n=50					
.50	.484[.120](.119)	.502[.120](.120)	.502[.120](.120)	.469[.095](.089)	.497[.088](.088)	.499[.088](.088)
.25	.234[.142](.141)	.248[.143](.143)	.250[.143](.143)	.210[.130](.124)	.250[.123](.123)	.251[.123](.123)
.00	-.010[.158](.158)	.001[.161](.161)	.002[.161](.161)	-.049[.167](.159)	-.001[.160](.160)	.001[.160](.160)
-.25	-.258[.161](.161)	-.251[.164](.164)	-.250[.165](.165)	-.303[.189](.182)	-.250[.184](.184)	-.248[.184](.184)
-.50	-.504[.163](.163)	-.503[.166](.166)	-.502[.167](.167)	-.565[.214](.204)	-.509[.208](.208)	-.507[.208](.208)
	Normal Mixture, n=50					
.50	.483[.119](.117)	.500[.118](.118)	.501[.118](.118)	.470[.091](.086)	.498[.084](.084)	.499[.084](.084)
.25	.238[.139](.139)	.253[.141](.141)	.254[.141](.141)	.209[.128](.121)	.248[.120](.120)	.249[.120](.120)
.00	-.013[.155](.154)	-.002[.157](.157)	-.001[.157](.157)	-.048[.160](.152)	-.001[.153](.153)	.001[.153](.153)
-.25	-.257[.158](.158)	-.251[.161](.161)	-.250[.162](.162)	-.301[.188](.181)	-.248[.182](.182)	-.247[.183](.183)
-.50	-.504[.163](.163)	-.503[.166](.166)	-.503[.167](.167)	-.556[.206](.199)	-.500[.203](.203)	-.498[.203](.203)
	Log-normal Error, n=50					
.50	.485[.111](.110)	.501[.111](.111)	.502[.111](.111)	.470[.090](.085)	.497[.083](.083)	.498[.083](.083)
.25	.239[.133](.133)	.253[.134](.134)	.254[.134](.134)	.212[.122](.116)	.249[.115](.115)	.251[.115](.115)
.00	-.010[.146](.146)	.001[.149](.149)	.002[.149](.149)	-.045[.154](.147)	.000[.147](.147)	.002[.147](.147)
-.25	-.255[.151](.151)	-.249[.154](.154)	-.248[.154](.154)	-.302[.178](.171)	-.251[.173](.173)	-.250[.173](.173)
-.50	-.498[.152](.152)	-.499[.155](.155)	-.499[.156](.156)	-.556[.204](.196)	-.503[.200](.200)	-.501[.200](.200)
	Normal Error, n=100					
.50	.493[.079](.078)	.502[.078](.078)	.502[.078](.078)	.482[.067](.065)	.500[.064](.064)	.501[.064](.064)
.25	.243[.095](.095)	.251[.095](.095)	.252[.095](.095)	.222[.096](.092)	.248[.092](.092)	.248[.092](.092)
.00	-.007[.110](.109)	.000[.110](.110)	.000[.110](.110)	-.031[.123](.119)	.000[.120](.120)	.001[.120](.120)
-.25	-.255[.114](.114)	-.250[.115](.115)	-.250[.115](.115)	-.289[.146](.141)	-.254[.143](.143)	-.253[.143](.143)
-.50	-.503[.117](.117)	-.501[.118](.118)	-.501[.118](.118)	-.538[.162](.158)	-.503[.162](.162)	-.503[.162](.162)
	Normal Mixture, n=100					
.50	.490[.078](.078)	.499[.078](.078)	.500[.078](.078)	.482[.067](.065)	.500[.065](.065)	.500[.065](.065)
.25	.241[.095](.095)	.249[.095](.095)	.250[.095](.095)	.224[.095](.091)	.250[.091](.091)	.250[.091](.091)
.00	-.006[.106](.106)	.001[.107](.107)	.002[.107](.107)	-.034[.122](.117)	-.002[.118](.118)	-.002[.118](.118)
-.25	-.255[.112](.112)	-.250[.113](.113)	-.250[.113](.113)	-.286[.144](.140)	-.251[.142](.142)	-.250[.142](.142)
-.50	-.502[.117](.117)	-.499[.119](.119)	-.499[.119](.119)	-.535[.160](.156)	-.500[.159](.159)	-.500[.159](.159)
	Log-normal Error, n=100					
.50	.492[.075](.075)	.501[.075](.075)	.501[.075](.075)	.482[.065](.062)	.500[.062](.062)	.500[.062](.062)
.25	.242[.091](.091)	.250[.091](.091)	.250[.091](.091)	.225[.093](.090)	.250[.090](.090)	.250[.090](.090)
.00	-.006[.102](.102)	.001[.103](.103)	.001[.103](.103)	-.029[.116](.113)	.001[.113](.113)	.002[.113](.113)
-.25	-.255[.110](.110)	-.250[.111](.111)	-.250[.111](.111)	-.283[.138](.134)	-.249[.136](.136)	-.248[.136](.136)
-.50	-.503[.112](.112)	-.500[.113](.113)	-.500[.113](.113)	-.526[.157](.154)	-.492[.159](.159)	-.495[.159](.159)
	Normal Error, n=500					
.50	.498[.033](.033)	.500[.033](.033)	.500[.033](.033)	.495[.034](.033)	.500[.033](.033)	.500[.033](.033)
.25	.249[.040](.040)	.251[.041](.041)	.251[.041](.041)	.242[.050](.049)	.249[.049](.049)	.249[.049](.049)
.00	-.001[.047](.047)	.000[.047](.047)	.000[.047](.047)	-.009[.065](.064)	.000[.065](.065)	.000[.065](.065)
-.25	-.252[.050](.050)	-.251[.050](.050)	-.251[.050](.050)	-.260[.080](.079)	-.249[.079](.079)	-.249[.079](.079)
-.50	-.501[.050](.050)	-.501[.050](.050)	-.501[.050](.050)	-.514[.096](.095)	-.501[.095](.095)	-.501[.095](.095)
	Normal Mixture, n=500					
.50	.498[.033](.033)	.500[.033](.033)	.500[.033](.033)	.495[.034](.033)	.500[.033](.033)	.500[.033](.033)
.25	.249[.040](.040)	.250[.040](.040)	.250[.040](.040)	.242[.050](.049)	.249[.049](.049)	.249[.049](.049)
.00	-.002[.045](.045)	-.001[.045](.045)	-.001[.045](.045)	-.007[.066](.066)	.002[.066](.066)	.002[.066](.066)
-.25	-.251[.048](.048)	-.250[.048](.048)	-.250[.048](.048)	-.261[.081](.081)	-.250[.081](.081)	-.250[.081](.081)
-.50	-.501[.050](.050)	-.500[.050](.050)	-.500[.050](.050)	-.514[.095](.094)	-.501[.094](.094)	-.501[.094](.094)
	Log-normal Error, n=500					
.50	.498[.032](.032)	.500[.032](.032)	.500[.032](.032)	.496[.034](.034)	.501[.034](.034)	.501[034](.034)
.25	.248[.040](.040)	.250[.040](.040)	.250[.040](.040)	.243[.050](.049)	.250[.049](.049)	.250[.049](.049)
.00	-.003[.046](.046)	-.001[.046](.046)	-.001[.046](.046)	-.009[.065](.064)	.000[.064](.064)	.000[.064](.064)
-.25	-.250[.048](.048)	-.249[.048](.048)	-.249[.048](.048)	-.259[.080](.080)	-.248[.080](.080)	-.248[.080](.080)
-.50	-.501[.049](.049)	-.501[.049](.049)	-.501[.049](.049)	-.514[.095](.094)	-.501[.095](.095)	-.501[.095](.095)

Table 4.1b. Empirical Mean[rmse](sd) of Estimators of λ, 2FE-SPD Model with SLD, $T = 3, \beta = (.\mathbf{5}, .\mathbf{5})', \sigma = 1$

λ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\lambda}_N^{\text{bc3}}$
	(a) Queen Contiguity, REG1			(b) Group Interaction, REG2		
	Normal Error, n=50					
.50	.477[.133](.132)	.500[.133](.133)	.500[.132](.132)	.449[.122](.111)	.498[.105](.105)	.500[.105](.105)
.25	.231[.157](.156)	.251[.159](.159)	.252[.158](.158)	.179[.171](.156)	.248[.150](.150)	.250[.150](.150)
.00	-.015[.176](.175)	.000[.180](.180)	.002[.180](.180)	-.086[.214](.196)	-.002[.191](.191)	.001[.191](.191)
-.25	-.261[.180](.180)	-.252[.185](.185)	-.251[.185](.185)	-.348[.247](.227)	-.252[.224](.224)	-.249[.224](.224)
-.50	-.505[.185](.184)	-.502[.190](.190)	-.501[.190](.190)	-.609[.283](.262)	-.504[.261](.261)	-.502[.262](.262)
	Normal Mixture, n=50					
.50	.478[.133](.132)	.501[.133](.133)	.500[.132](.132)	.449[.120](.109)	.498[.103](.103)	.500[.103](.103)
.25	.229[.158](.157)	.248[.159](.159)	.249[.159](.159)	.180[.168](.153)	.248[.147](.147)	.250[.147](.147)
.00	-.017[.174](.173)	-.002[.177](.177)	.000[.177](.177)	-.088[.212](.193)	-.003[.188](.188)	.000[.188](.188)
-.25	-.260[.176](.176)	-.251[.181](.181)	-.250[.181](.181)	-.346[.247](.227)	-.250[.224](.224)	-.247[.225](.225)
-.50	-.502[.181](.181)	-.499[.186](.186)	-.499[.186](.186)	-.608[.281](.260)	-.503[.260](.260)	-.500[.260](.260)
	Log-normal Error, n=50					
.50	.480[.123](.122)	.502[.123](.123)	.502[.122](.122)	.454[.112](.102)	.502[.097](.097)	.504[.097](.097)
.25	.229[.148](.147)	.249[.150](.149)	.250[.149](.149)	.184[.157](.143)	.251[.138](.138)	.254[.138](.138)
.00	-.013[.162](.161)	.002[.165](.165)	.003[.165](.165)	-.079[.193](.176)	.003[.172](.172)	.006[.172](.172)
-.25	-.258[.168](.167)	-.248[.172](.172)	-.247[.172](.172)	-.341[.225](.206)	-.247[.203](.203)	-.244[.203](.203)
-.50	-.504[.173](.172)	-.501[.177](.177)	-.501[.178](.178)	-.598[.258](.239)	-.495[.239](.239)	-.493[.240](.240)
	Normal Error, n=100					
.50	.490[.090](.090)	.502[.090](.090)	.502[.089](.089)	.469[.087](.081)	.499[.079](.079)	.500[.079](.079)
.25	.242[.108](.108)	.253[.109](.109)	.253[.109](.109)	.205[.127](.119)	.248[.117](.117)	.248[.117](.117)
.00	-.003[.122](.122)	.006[.123](.123)	.006[.123](.123)	-.058[.166](.155)	-.004[.153](.153)	-.003[.153](.153)
-.25	-.256[.130](.129)	-.250[.131](.131)	-.249[.131](.131)	-.313[.192](.181)	-.249[.179](.179)	-.249[.179](.179)
-.50	-.505[.131](.131)	-.503[.133](.133)	-.503[.133](.133)	-.578[.223](.209)	-.506[.209](.208)	-.506[.209](.209)
	Normal Mixture, n=100					
.50	.491[.088](.088)	.502[.088](.088)	.502[.088](.088)	.470[.087](.082)	.500[.080](.080)	.500[.079](.079)
.25	.241[.105](.105)	.252[.106](.106)	.252[.106](.106)	.207[.124](.116)	.249[.113](.113)	.250[.113](.113)
.00	-.010[.120](.120)	-.002[.121](.121)	-.001[.121](.121)	-.056[.160](.150)	-.001[.148](.148)	-.001[.148](.148)
-.25	-.254[.129](.129)	-.248[.131](.131)	-.247[.131](.131)	-.314[.195](.184)	-.251[.182](.182)	-.250[.182](.182)
-.50	-.503[.130](.130)	-.500[.131](.131)	-.500[.132](.132)	-.567[.217](.207)	-.496[.206](.206)	-.495[.206](.206)
	Log-normal Error, n=100					
.50	.490[.084](.084)	.502[.084](.084)	.502[.084](.084)	.470[.084](.079)	.500[.077](.077)	.500[.077](.077)
.25	.235[.102](.101)	.246[.102](.102)	.246[.102](.102)	.208[.120](.113)	.250[.110](.110)	.251[.110](.110)
.00	-.005[.116](.116)	.004[.117](.117)	.004[.117](.117)	-.050[.151](.143)	.003[.141](.141)	.004[.141](.141)
-.25	-.258[.121](.121)	-.252[.123](.123)	-.252[.123](.123)	-.316[.185](.172)	-.253[.171](.171)	-.253[.171](.171)
-.50	-.502[.125](.125)	-.499[.126](.126)	-.499[.126](.126)	-.565[.208](.197)	-.495[.197](.197)	-.495[.197](.197)
	Normal Error, n=500					
.50	.498[.039](.039)	.500[.039](.039)	.500[.039](.039)	.490[.050](.049)	.501[.048](.048)	.501[.048](.048)
.25	.247[.048](.048)	.250[.048](.048)	.250[.048](.048)	.234[.073](.071)	.250[.071](.071)	.250[.071](.071)
.00	-.001[.055](.055)	.001[.055](.055)	.001[.055](.055)	-.021[.097](.094)	.000[.094](.094)	.000[.094](.094)
-.25	-.251[.058](.058)	-.250[.058](.058)	-.250[.058](.058)	-.275[.117](.114)	-.249[.113](.113)	-.249[.113](.113)
-.50	-.500[.060](.060)	-.499[.061](.061)	-.499[.061](.061)	-.530[.139](.136)	-.500[.135](.135)	-.500[.135](.135)
	Normal Mixture, n=500					
.50	.499[.039](.039)	.501[.039](.039)	.501[.039](.039)	.490[.048](.047)	.501[.047](.047)	.501[.047](.047)
.25	.247[.048](.048)	.249[.048](.048)	.249[.048](.048)	.233[.074](.072)	.249[.071](.071)	.249[.071](.071)
.00	.000[.054](.054)	.002[.055](.055)	.002[.055](.055)	-.020[.095](.093)	.002[.092](.092)	.002[.092](.092)
-.25	-.250[.059](.059)	-.249[.059](.059)	-.249[.059](.059)	-.279[.119](.116)	-.253[.115](.115)	-.253[.115](.115)
-.50	-.501[.059](.059)	-.500[.060](.060)	-.500[.060](.060)	-.529[.137](.134)	-.499[.133](.133)	-.499[.133](.133)
	Log-normal Error, n=500					
.50	.497[.037](.037)	.500[.037](.037)	.500[.037](.037)	.491[.047](.046)	.502[.046](.046)	.502[.046](.046)
.25	.248[.048](.048)	.250[.048](.048)	.250[.048](.048)	.234[.072](.070)	.251[.069](.069)	.251[.069](.069)
.00	-.002[.053](.053)	.000[.053](.053)	.000[.053](.053)	-.020[.094](.092)	.001[.091](.091)	.001[.091](.091)
-.25	-.252[.057](.057)	-.251[.058](.058)	-.251[.058](.058)	-.277[.116](.112)	-.250[.112](.112)	-.251[.112](.112)
-.50	-.499[.059](.059)	-.499[.059](.059)	-.499[.059](.059)	-.530[.139](.136)	-.498[.135](.135)	-.499[.135](.135)

Table 4.2. Empirical Mean[rmse](sd) of Estimators of λ - 2FE-SPD Model with SED, $T = 3, \beta = (1,1)', \sigma = 1$

λ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\mathrm{bc2}}$	$\hat{\lambda}_N^{\mathrm{bc3}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\mathrm{bc2}}$	$\hat{\lambda}_N^{\mathrm{bc3}}$
	(a) Queen Contiguity, REG1			(b) Group Interaction, REG2		
	Normal Error, n=50					
.50	.481[.144](.142)	.500[.143](.143)	.500[.142](.142)	.457[.139](.132)	.503[.116](.116)	.503[.115](.115)
.25	.233[.171](.170)	.252[.171](.171)	.254[.171](.171)	.177[.202](.188)	.258[.167](.167)	.260[.167](.166)
.00	-.018[.190](.189)	-.001[.190](.190)	.001[.191](.190)	-.115[.266](.240)	-.004[.221](.221)	-.001[.220](.220)
-.25	-.271[.202](.201)	-.255[.203](.203)	-.254[.204](.204)	-.382[.299](.268)	-.250[.256](.256)	-.249[.256](.256)
-.50	-.516[.203](.202)	-.503[.205](.205)	-.502[.206](.206)	-.637[.321](.290)	-.496[.287](.287)	-.497[.288](.288)
	Normal Mixture, n=50					
.50	.480[.139](.138)	.500[.138](.138)	.500[.137](.137)	.458[.137](.130)	.504[.114](.114)	.504[.113](.113)
.25	.233[.166](.165)	.252[.166](.166)	.251[.166](.166)	.168[.210](.194)	.251[.172](.172)	.250[.171](.171)
.00	-.016[.186](.185)	.002[.186](.186)	.003[.186](.186)	-.108[.258](.234)	.004[.214](.214)	.003[.214](.214)
-.25	-.267[.195](.194)	-.252[.196](.196)	-.250[.197](.197)	-.381[.293](.262)	-.248[.251](.251)	-.249[.251](.251)
-.50	-.511[.198](.197)	-.498[.200](.200)	-.498[.201](.201)	-.636[.313](.282)	-.493[.280](.280)	-.495[.281](.281)
	Log-normal Error, n=50					
.50	.483[.135](.133)	.504[.134](.134)	.503[.133](.133)	.454[.136](.128)	.502[.112](.112)	.502[.111](.111)
.25	.237[.160](.159)	.256[.161](.160)	.255[.160](.160)	.174[.196](.181)	.257[.160](.160)	.256[.160](.160)
.00	-.012[.179](.179)	.006[.180](.180)	.005[.180](.180)	-.105[.242](.218)	.009[.199](.199)	.002[.199](.199)
-.25	-.264[.186](.186)	-.248[.188](.188)	-.249[.188](.188)	-.368[.273](.247)	-.233[.235](.235)	-.239[.236](.235)
-.50	-.512[.191](.191)	-.499[.194](.194)	-.499[.194](.194)	-.632[.305](.275)	-.489[.272](.272)	-.489[.274](.273)
	Normal Error, n=100					
.50	.490[.096](.095)	.500[.095](.095)	.500[.095](.095)	.467[.107](.102)	.501[.093](.093)	.501[.093](.093)
.25	.241[.119](.119)	.251[.119](.119)	.251[.118](.118)	.196[.152](.142)	.252[.132](.132)	.251[.132](.132)
.00	-.011[.132](.132)	-.001[.132](.132)	.000[.132](.132)	-.074[.192](.177)	-.002[.171](.171)	-.002[.171](.171)
-.25	-.259[.141](.140)	-.249[.141](.141)	-.249[.141](.141)	-.333[.215](.199)	-.255[.199](.199)	-.255[.199](.199)
-.50	-.510[.142](.142)	-.501[.143](.143)	-.501[.143](.143)	-.574[.220](.207)	-.500[.215](.215)	-.500[.215](.215)
	Normal Mixture, n=100					
.50	.489[.095](.094)	.500[.094](.094)	.500[.094](.094)	.465[.104](.098)	.500[.090](.090)	.500[.090](.090)
.25	.240[.118](.117)	.250[.117](.117)	.250[.117](.117)	.196[.149](.139)	.253[.130](.130)	.253[.130](.130)
.00	-.010[.130](.130)	.001[.130](.130)	.001[.130](.130)	-.073[.189](.174)	.000[.168](.168)	.000[.168](.168)
-.25	-.260[.138](.138)	-.250[.138](.138)	-.249[.138](.138)	-.327[.211](.196)	-.249[.197](.197)	-.249[.197](.197)
-.50	-.510[.138](.138)	-.501[.139](.139)	-.501[.139](.139)	-.569[.220](.209)	-.495[.219](.219)	-.495[.219](.219)
	Log-normal Error, n=100					
.50	.494[.088](.088)	.505[.088](.088)	.505[.088](.088)	.465[.107](.101)	.501[.092](.092)	.500[.092](.092)
.25	.240[.110](.110)	.251[.110](.110)	.251[.110](.110)	.198[.145](.135)	.256[.126](.126)	.256[.126](.125)
.00	-.006[.126](.126)	.004[.127](.126)	.003[.127](.126)	-.064[.174](.162)	.010[.156](.156)	.010[.156](.156)
-.25	-.259[.136](.136)	-.250[.136](.136)	-.249[.136](.136)	-.320[.200](.188)	-.239[.189](.188)	-.239[.189](.189)
-.50	-.508[.135](.135)	-.500[.136](.136)	-.500[.136](.136)	-.561[.214](.205)	-.485[.215](.215)	-.486[.215](.215)
	Normal Error, n=500					
.50	.497[.041](.041)	.499[.041](.041)	.499[.041](.041)	.487[.060](.059)	.500[.057](.057)	.500[.057](.057)
.25	.249[.051](.051)	.251[.051](.051)	.251[.051](.051)	.226[.090](.087)	.249[.083](.083)	.249[.083](.083)
.00	-.003[.058](.058)	-.001[.058](.058)	-.001[.058](.058)	-.033[.121](.116)	.000[.112](.112)	.000[.112](.112)
-.25	-.252[.062](.061)	-.250[.062](.062)	-.250[.062](.062)	-.292[.148](.142)	-.249[.137](.137)	-.249[.137](.137)
-.50	-.500[.063](.063)	-.499[.063](.063)	-.499[.063](.063)	-.549[.170](.162)	-.499[.158](.158)	-.499[.158](.158)
	Normal Mixture, n=500					
.50	.498[.040](.040)	.500[.040](.040)	.500[.040](.040)	.485[.060](.058)	.499[.056](.056)	.499[.056](.056)
.25	.247[.051](.051)	.250[.051](.051)	.250[.051](.051)	.226[.091](.088)	.250[.084](.084)	.249[.084](.084)
.00	-.001[.058](.058)	.001[.058](.058)	.001[.058](.058)	-.035[.120](.114)	-.001[.110](.110)	-.002[.110](.110)
-.25	-.252[.062](.062)	-.250[.062](.062)	-.250[.062](.062)	-.291[.146](.140)	-.249[.136](.136)	-.249[.136](.136)
-.50	-.504[.063](.063)	-.502[.063](.063)	-.502[.063](.063)	-.551[.173](.165)	-.500[.161](.161)	-.500[.161](.161)
	Log-normal Error, n=500					
.50	.498[.040](.040)	.500[.040](.040)	.500[.040](.040)	.485[.062](.060)	.500[.058](.058)	.499[.058](.058)
.25	.249[.050](.050)	.251[.050](.050)	.251[.050](.050)	.227[.088](.085)	.252[.081](.081)	.252[.081](.081)
.00	-.003[.057](.056)	-.001[.056](.056)	-.001[.056](.056)	-.030[.112](.108)	.006[.104](.104)	.005[.104](.104)
-.25	-.251[.060](.060)	-.249[.060](.060)	-.249[.060](.060)	-.290[.141](.135)	-.245[.131](.130)	-.246[.130](.130)
-.50	-.503[.062](.062)	-.501[.062](.062)	-.501[.062](.062)	-.545[.168](.162)	-.492[.158](.157)	-.493[.158](.157)

Table 4.3a. Empirical Mean[rmse](sd) of Estimators of λ and ρ, 2FE-SPD Model with SARAR, $T=3, \beta=(1,1)', \sigma=1$, `Queen Contiguity, REG-1`

λ	ρ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\mathrm{bc2}}$	$\hat{\rho}_N$	$\hat{\rho}_N^{\mathrm{bc2}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\mathrm{bc2}}$	$\hat{\rho}_N$	$\hat{\rho}_N^{\mathrm{bc2}}$
		(a) Normal Error, $n=50$				(b) Log-normal Error, $n=50$			
.50	.50	.484[.116](.115)	.500[.116](.116)	.483[.143](.142)	.500[.143](.143)	.486[.105](.104)	.502[.105](.105)	.484[.131](.130)	.502[.131](.131)
	.25	.484[.119](.117)	.501[.118](.118)	.226[.176](.174)	.242[.175](.175)	.485[.114](.113)	.501[.113](.113)	.233[.162](.161)	.250[.161](.161)
	.00	.483[.118](.116)	.500[.117](.117)	-.019[.192](.191)	-.002[.192](.192)	.486[.110](.109)	.503[.110](.110)	-.015[.177](.176)	.002[.177](.177)
	-.25	.482[.124](.122)	.500[.123](.123)	-.267[.202](.202)	-.251[.203](.203)	.487[.112](.111)	.503[.112](.112)	-.265[.193](.193)	-.249[.193](.193)
	-.50	.484[.125](.123)	.500[.124](.124)	-.513[.208](.208)	-.498[.209](.209)	.489[.111](.110)	.505[.111](.111)	-.514[.195](.194)	-.499[.196](.196)
-.50	.50	-.502[.158](.158)	-.500[.161](.161)	.486[.144](.143)	.504[.144](.144)	-.502[.145](.145)	-.500[.148](.148)	.486[.132](.131)	.504[.132](.131)
	.25	-.506[.165](.165)	-.504[.168](.168)	.232[.174](.173)	.249[.174](.174)	-.505[.152](.151)	-.503[.155](.154)	.233[.161](.160)	.250[.160](.160)
	.00	-.501[.163](.163)	-.499[.167](.167)	-.006[.187](.187)	.010[.187](.187)	-.499[.159](.159)	-.497[.162](.162)	-.018[.180](.179)	-.001[.180](.180)
	-.25	-.500[.164](.164)	-.498[.168](.168)	-.262[.209](.209)	-.246[.210](.210)	-.501[.152](.152)	-.499[.155](.155)	-.263[.197](.197)	-.246[.197](.197)
	-.50	-.506[.169](.169)	-.505[.172](.172)	-.518[.207](.206)	-.503[.208](.208)	-.498[.157](.157)	-.497[.160](.160)	-.513[.194](.194)	-.498[.195](.195)
		(c) Normal Error, $n=100$				(d) Log-normal Error, $n=100$			
.50	.50	.494[.078](.077)	.502[.078](.078)	.490[.096](.096)	.499[.096](.096)	.490[.078](.078)	.499[.078](.078)	.493[.090](.090)	.502[.090](.090)
	.25	.490[.080](.080)	.499[.080](.080)	.244[.117](.116)	.253[.117](.117)	.491[.081](.080)	.500[.080](.080)	.243[.111](.111)	.252[.111](.111)
	.00	.493[.083](.083)	.502[.083](.083)	-.011[.132](.131)	-.002[.131](.131)	.494[.079](.079)	.503[.079](.079)	-.009[.126](.126)	.001[.126](.126)
	-.25	.491[.084](.083)	.500[.083](.083)	-.258[.142](.142)	-.249[.142](.142)	.490[.077](.077)	.499[.077](.077)	-.264[.138](.137)	-.254[.138](.137)
	-.50	.490[.079](.078)	.499[.078](.078)	-.509[.142](.141)	-.499[.142](.142)	.493[.077](.077)	.501[.077](.077)	-.509[.137](.137)	-.499[.137](.137)
-.50	.50	-.494[.118](.118)	-.493[.119](.119)	.492[.094](.094)	.501[.094](.094)	-.503[.106](.106)	-.503[.107](.107)	.491[.089](.088)	.500[.088](.088)
	.25	-.501[.119](.119)	-.500[.121](.121)	.242[.117](.117)	.251[.117](.117)	-.502[.112](.112)	-.501[.113](.113)	.240[.111](.111)	.249[.111](.111)
	.00	-.496[.115](.115)	-.495[.117](.117)	-.008[.133](.133)	.001[.133](.133)	-.498[.114](.114)	-.498[.115](.115)	-.007[.129](.129)	.003[.128](.128)
	-.25	-.505[.118](.118)	-.504[.120](.120)	-.258[.143](.143)	-.248[.143](.143)	-.497[.112](.112)	-.496[.113](.113)	-.257[.136](.136)	-.248[.136](.136)
	-.50	-.501[.118](.118)	-.500[.120](.120)	-.504[.148](.148)	-.495[.149](.149)	-.505[.109](.109)	-.504[.110](.110)	-.507[.137](.137)	-.498[.138](.137)
		(e) Normal Error, $n=500$				(f) Log-normal Error, $n=500$			
.50	.50	.497[.033](.033)	.499[.033](.033)	.499[.041](.041)	.501[.041](.041)	.499[.030](.030)	.501[.030](.030)	.497[.040](.040)	.499[.040](.040)
	.25	.497[.033](.033)	.499[.033](.033)	.247[.052](.052)	.249[.052](.052)	.499[.032](.032)	.501[.032](.032)	.249[.050](.050)	.250[.050](.050)
	.00	.499[.033](.033)	.501[.033](.033)	.001[.057](.057)	.003[.058](.057)	.498[.033](.033)	.500[.033](.033)	-.001[.057](.057)	.001[.057](.057)
	-.25	.498[.033](.032)	.499[.033](.033)	-.254[.062](.062)	-.252[.062](.062)	.498[.033](.033)	.500[.033](.033)	-.250[.061](.061)	-.248[.061](.061)
	-.50	.498[.032](.032)	.500[.032](.032)	-.503[.062](.062)	-.501[.062](.062)	.497[.032](.032)	.499[.032](.032)	-.501[.062](.062)	-.499[.062](.062)
-.50	.50	-.502[.049](.049)	-.501[.049](.049)	.498[.041](.041)	.500[.041](.041)	-.499[.049](.049)	-.499[.049](.049)	.498[.040](.040)	.500[.040](.040)
	.25	-.503[.051](.051)	-.502[.051](.051)	.249[.051](.051)	.250[.051](.051)	-.500[.051](.051)	-.499[.051](.051)	.248[.050](.050)	.250[.050](.050)
	.00	-.501[.050](.050)	-.501[.050](.050)	-.001[.060](.060)	.001[.060](.060)	-.501[.051](.051)	-.500[.052](.052)	-.002[.058](.058)	.000[.058](.058)
	-.25	-.502[.051](.050)	-.502[.051](.051)	-.253[.061](.061)	-.251[.061](.061)	-.499[.051](.051)	-.498[.051](.051)	-.252[.062](.062)	-.250[.062](.062)
	-.50	-.500[.049](.049)	-.499[.049](.049)	-.501[.063](.063)	-.499[.064](.064)	-.500[.048](.048)	-.500[.049](.049)	-.503[.062](.062)	-.502[.062](.062)

Table 4.3b. Empirical Mean[rmse](sd) of Estimators of λ and ρ, 2FE-SPD Model with SARAR, $T = 3, \beta = (1,1)', \sigma = 1$, `Group Interaction, REG-2`

λ	ρ	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\rho}_N$	$\hat{\rho}_N^{\text{bc2}}$	$\hat{\lambda}_N$	$\hat{\lambda}_N^{\text{bc2}}$	$\hat{\rho}_N$	$\hat{\rho}_N^{\text{bc2}}$
		(a) Normal Error, $n = 50$				(b) Log-normal Error, $n = 50$			
.50	.50	.484[.095](.094)	.499[.092](.092)	.453[.156](.149)	.500[.129](.129)	.484[.089](.088)	.500[.087](.087)	.456[.146](.140)	.505[.121](.121)
	.25	.480[.103](.101)	.497[.099](.099)	.162[.238](.221)	.248[.194](.194)	.484[.096](.095)	.501[.093](.093)	.161[.237](.220)	.251[.193](.193)
	.00	.481[.104](.102)	.498[.100](.100)	-.120[.298](.272)	.001[.243](.243)	.486[.097](.096)	.501[.093](.093)	-.120[.301](.276)	.005[.247](.247)
	-.25	.481[.104](.102)	.496[.100](.100)	-.408[.362](.326)	-.257[.299](.299)	.488[.097](.096)	.502[.094](.094)	-.407[.365](.330)	-.252[.306](.306)
	-.50	.484[.099](.098)	.498[.096](.096)	-.685[.400](.354)	-.512[.335](.334)	.491[.095](.095)	.504[.093](.093)	-.682[.413](.370)	-.506[.354](.354)
-.50	.50	-.527[.218](.216)	-.499[.218](.218)	.453[.158](.150)	.501[.130](.130)	-.522[.214](.213)	-.494[.215](.215)	.458[.147](.141)	.507[.123](.122)
	.25	-.534[.237](.235)	-.505[.237](.236)	.164[.235](.219)	.251[.191](.191)	-.524[.226](.225)	-.495[.227](.227)	.171[.220](.205)	.259[.179](.179)
	.00	-.532[.239](.237)	-.504[.239](.239)	-.117[.301](.277)	.004[.249](.249)	-.528[.239](.237)	-.501[.239](.239)	-.114[.293](.270)	.010[.242](.242)
	-.25	-.530[.237](.235)	-.504[.237](.237)	-.407[.357](.320)	-.257[.295](.295)	-.519[.240](.240)	-.494[.241](.241)	-.396[.349](.317)	-.243[.293](.293)
	-.50	-.524[.233](.232)	-.500[.233](.233)	-.689[.403](.355)	-.518[.337](.336)	-.528[.251](.250)	-.505[.252](.252)	-.661[.399](.364)	-.489[.345](.345)
		(c) Normal Error, $n = 250$				(d) Log-normal Error, $n = 250$			
.50	.50	.497[.044](.044)	.501[.044](.044)	.477[.082](.079)	.500[.074](.074)	.497[.043](.043)	.500[.042](.042)	.477[.081](.078)	.500[.073](.073)
	.25	.497[.043](.043)	.500[.043](.043)	.209[.124](.117)	.250[.110](.110)	.497[.042](.042)	.500[.042](.042)	.209[.119](.112)	.250[.105](.105)
	.00	.497[.041](.040)	.499[.040](.040)	-.056[.161](.151)	.001[.142](.142)	.498[.040](.040)	.500[.039](.039)	-.056[.158](.148)	.002[.138](.138)
	-.25	.498[.038](.038)	.500[.038](.038)	-.327[.204](.189)	-.253[.178](.178)	.498[.038](.038)	.500[.038](.038)	-.322[.194](.180)	-.247[.169](.169)
	-.50	.499[.035](.035)	.500[.035](.035)	-.590[.232](.214)	-.501[.203](.203)	.500[.035](.035)	.501[.035](.035)	-.588[.229](.211)	-.497[.200](.200)
-.50	.50	-.508[.123](.122)	-.498[.122](.122)	.476[.082](.078)	.499[.073](.073)	-.509[.122](.121)	-.498[.121](.121)	.476[.081](.078)	.500[.073](.073)
	.25	-.510[.118](.118)	-.502[.118](.118)	.213[.121](.115)	.253[.108](.108)	-.504[.118](.118)	-.496[.118](.118)	.210[.120](.113)	.251[.106](.106)
	.00	-.507[.116](.116)	-.500[.116](.116)	-.063[.167](.155)	-.005[.146](.146)	-.509[.113](.113)	-.502[.113](.113)	-.058[.161](.150)	.000[.140](.140)
	-.25	-.502[.105](.105)	-.497[.105](.105)	-.326[.201](.186)	-.252[.175](.175)	-.507[.105](.105)	-.502[.105](.105)	-.320[.192](.179)	-.245[.169](.169)
	-.50	-.506[.099](.099)	-.502[.099](.099)	-.592[.235](.216)	-.503[.204](.204)	-.503[.100](.100)	-.499[.100](.100)	-.589[.234](.217)	-.498[.205](.205)
		(e) Normal Error, $n = 500$				(f) Log-normal Error, $n = 500$			
.50	.50	.498[.030](.030)	.500[.030](.030)	.484[.065](.063)	.500[.060](.060)	.498[.030](.030)	.500[.030](.030)	.484[.065](.063)	.501[.060](.060)
	.25	.499[.029](.029)	.500[.029](.029)	.220[.098](.093)	.248[.089](.089)	.498[.029](.029)	.500[.029](.029)	.223[.096](.092)	.252[.087](.087)
	.00	.500[.027](.027)	.501[.027](.027)	-.040[.128](.122)	.001[.116](.116)	.500[.027](.027)	.501[.027](.027)	-.044[.128](.120)	-.001[.114](.114)
	-.25	.500[.025](.025)	.501[.025](.025)	-.303[.160](.151)	-.249[.144](.144)	.500[.025](.025)	.501[.025](.025)	-.305[.158](.148)	-.249[.141](.141)
	-.50	.499[.023](.023)	.500[.023](.023)	-.562[.187](.176)	-.496[.168](.168)	.499[.022](.022)	.500[.022](.022)	-.565[.192](.180)	-.497[.172](.172)
-.50	.50	-.505[.087](.087)	-.500[.087](.087)	.485[.065](.063)	.500[.060](.060)	-.505[.085](.085)	-.499[.085](.085)	.484[.064](.062)	.501[.059](.059)
	.25	-.507[.082](.082)	-.503[.082](.082)	.220[.098](.094)	.248[.089](.089)	-.504[.081](.081)	-.500[.081](.081)	.223[.096](.092)	.252[.088](.088)
	.00	-.503[.075](.075)	-.500[.075](.075)	-.041[.131](.124)	.000[.118](.118)	-.502[.075](.075)	-.499[.075](.075)	-.044[.127](.119)	-.001[.113](.113)
	-.25	-.504[.070](.070)	-.502[.070](.070)	-.303[.161](.152)	-.249[.145](.145)	-.501[.071](.071)	-.499[.071](.071)	-.303[.159](.150)	-.248[.143](.143)
	-.50	-.501[.065](.065)	-.499[.065](.065)	-.569[.192](.179)	-.503[.171](.171)	-.502[.065](.065)	-.500[.065](.065)	-.562[.187](.176)	-.494[.168](.168)

4.6 Conclusion

We have introduced a general method for finite sample bias and variance corrections of the QML estimators of the two-way fixed effects spatial panel data models where the spatial interactions can be in the form of either spatial lag or spatial error, or both, and the panels can be either short or long. We have demonstrated that bias and variance corrections lead to refined inferences for the spatial effects as well as covariate effects. The proposed methods are seen to be very easy to implement, and very effective. If only bias-correction is of concern, a second-order correction using iid bootstrap suffices. For improved inferences for the spatial parameters, a third-order variance correction seems necessary and a wild bootstrap method seems to perform better. However, for improved inferences concerning the regression coefficients (the covariate effects), the second-order bias and variance corrections seem sufficient, and the resulting inferences can be much more reliable than those based on the standard asymptotic methods. The latter observation is perhaps the most important one in this study as being able to assess the covariate effects in a reliable manner may be the most desirable feature of the econometric modelling activities. All the methods proposed in the current chapter can easily be built into the standard statistical software to facilitate the practical applications. Further extensions of the proposed methods are desirable and possible such as the FE-SPD models of higher-order spatial effects, but are beyond the scope of the chapter. Nevertheless, the results presented in this chapter reinforce that the general methodology of bias and variance corrections of Yang (2015b), based on stochastic expansion and bootstrap, is indeed a promising approach in handling the bias issues, and in providing refined inference methods.

Table 4.4. Empirical Means of the Non-Spatial Estimators, 2FE-SPD Model with SLD
`Group Interaction`, `REG2`, $T = 3$

λ	$\hat\beta_{1N}$	$\hat\beta_{2N}$	$\hat\sigma^2_N$	$\hat\beta^{\text{bc}}_{1N}$	$\hat\beta^{\text{bc}}_{2N}$	$\hat\sigma^{2,\text{bc}}_N$	$\hat\beta_{1N}$	$\hat\beta_{2N}$	$\hat\sigma^2_N$	$\hat\beta^{\text{bc}}_{1N}$	$\hat\beta^{\text{bc}}_{2N}$	$\hat\sigma^{2,\text{bc}}_N$
	(a) $\beta = (1,1)', \sigma = 1$						(b) $\beta = (.5,.5)', \sigma = 1$					
	Normal Error, n=50											
.50	1.041	1.035	0.984	0.996	0.998	0.992	0.533	0.530	0.985	0.496	0.499	0.991
.25	1.039	1.030	0.982	0.997	0.995	0.992	0.532	0.524	0.981	0.498	0.496	0.991
.00	1.035	1.023	0.980	0.997	0.992	0.992	0.529	0.519	0.978	0.498	0.494	0.991
-.25	1.032	1.023	0.978	0.997	0.995	0.992	0.524	0.519	0.975	0.496	0.496	0.992
-.50	1.030	1.019	0.974	0.999	0.994	0.989	0.527	0.514	0.970	0.501	0.494	0.990
	Normal Mixture, n=50											
.50	1.040	1.031	0.975	0.996	0.994	0.982	0.532	0.520	0.981	0.495	0.490	0.988
.25	1.041	1.030	0.973	1.000	0.996	0.982	0.531	0.523	0.973	0.497	0.495	0.983
.00	1.038	1.030	0.973	1.001	0.998	0.984	0.526	0.518	0.973	0.495	0.493	0.986
-.25	1.035	1.025	0.966	1.001	0.997	0.980	0.524	0.515	0.963	0.496	0.492	0.979
-.50	1.028	1.023	0.969	0.997	0.997	0.985	0.521	0.520	0.962	0.496	0.500	0.981
	Log-normal Error, n=50											
.50	1.036	1.031	0.944	0.994	0.995	0.951	0.529	0.523	0.946	0.493	0.493	0.952
.25	1.036	1.032	0.947	0.996	0.999	0.957	0.529	0.521	0.946	0.496	0.494	0.956
.00	1.028	1.020	0.936	0.992	0.990	0.947	0.525	0.519	0.944	0.495	0.494	0.957
-.25	1.029	1.019	0.942	0.996	0.992	0.955	0.522	0.517	0.943	0.494	0.494	0.959
-.50	1.026	1.017	0.940	0.996	0.993	0.956	0.518	0.514	0.926	0.494	0.494	0.945
	Normal Error, n=100											
.50	1.028	1.023	0.993	1.000	0.999	0.996	0.526	0.521	0.993	0.501	0.499	0.996
.25	1.027	1.019	0.991	1.000	0.996	0.995	0.524	0.517	0.990	0.500	0.496	0.995
.00	1.023	1.020	0.990	0.998	0.999	0.996	0.524	0.516	0.991	0.501	0.496	0.997
-.25	1.020	1.020	0.989	0.996	1.000	0.995	0.521	0.514	0.988	0.499	0.496	0.995
-.50	1.024	1.018	0.988	1.002	0.999	0.995	0.520	0.514	0.986	0.500	0.497	0.994
	Normal Mixture, n=100											
.50	1.026	1.022	0.990	0.998	0.998	0.993	0.523	0.518	0.988	0.497	0.497	0.991
.25	1.024	1.019	0.987	0.998	0.996	0.992	0.525	0.519	0.986	0.501	0.498	0.990
.00	1.022	1.018	0.985	0.997	0.996	0.990	0.522	0.515	0.985	0.499	0.496	0.991
-.25	1.023	1.018	0.987	1.000	0.998	0.994	0.523	0.517	0.983	0.501	0.499	0.991
-.50	1.022	1.019	0.982	1.000	1.001	0.989	0.518	0.515	0.983	0.498	0.498	0.992
	Log-normal Error, n=100											
.50	1.024	1.021	0.973	0.997	0.998	0.977	0.524	0.518	0.969	0.499	0.497	0.972
.25	1.025	1.023	0.964	1.000	1.002	0.968	0.522	0.516	0.966	0.498	0.496	0.971
.00	1.023	1.015	0.963	0.999	0.995	0.969	0.520	0.514	0.962	0.497	0.495	0.968
-.25	1.022	1.016	0.970	0.999	0.997	0.977	0.520	0.516	0.964	0.499	0.498	0.972
-.50	1.021	1.012	0.960	1.000	0.995	0.966	0.516	0.514	0.958	0.497	0.498	0.967
	Normal Error, n=250											
.50	1.011	1.010	0.997	0.999	0.998	0.999	0.512	0.512	0.997	0.499	0.499	0.998
.25	1.010	1.009	0.996	0.998	0.997	0.998	0.512	0.512	0.996	0.500	0.500	0.998
.00	1.009	1.009	0.996	0.998	0.997	0.998	0.509	0.509	0.996	0.497	0.497	0.998
-.25	1.009	1.010	0.996	0.997	0.998	0.999	0.508	0.511	0.995	0.497	0.500	0.998
-.50	1.009	1.010	0.995	0.998	0.999	0.998	0.511	0.510	0.994	0.500	0.499	0.997
	Normal Mixture, n=250											
.50	1.014	1.013	0.997	1.002	1.000	0.998	0.513	0.509	0.996	0.500	0.497	0.997
.25	1.012	1.010	0.993	1.000	0.998	0.995	0.512	0.511	0.995	0.500	0.498	0.996
.00	1.010	1.011	0.995	0.998	0.999	0.997	0.510	0.512	0.993	0.498	0.500	0.996
-.25	1.012	1.011	0.996	1.001	1.000	0.998	0.510	0.510	0.997	0.498	0.498	1.000
-.50	1.009	1.008	0.994	0.998	0.997	0.996	0.510	0.509	0.993	0.499	0.498	0.996
	Log-normal Error, n=250											
.50	1.011	1.010	0.986	0.999	0.998	0.987	0.511	0.511	0.982	0.498	0.498	0.983
.25	1.012	1.013	0.985	1.000	1.001	0.987	0.513	0.513	0.986	0.501	0.501	0.988
.00	1.010	1.009	0.983	0.998	0.998	0.985	0.511	0.511	0.984	0.499	0.499	0.987
-.25	1.010	1.009	0.982	0.999	0.997	0.985	0.512	0.510	0.984	0.500	0.498	0.987
-.50	1.007	1.007	0.985	0.996	0.997	0.987	0.509	0.508	0.983	0.498	0.497	0.986

Table 4.5a. Empirical Sizes: Two-Sided Tests of Spatial Dependence in SARAR Model
Group Interaction, REG2, $T = 3, \beta = (1,1)', \sigma = 1$

n	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal Errors		
					$H_0 : \lambda = \rho = 0$					
50	$\mathcal{W}_{11}$	.1974	.1288	.0546	.1918	.1232	.0450	.1616	.1062	.0456
	$\mathcal{W}_{22}$	.1896	.1196	.0516	.1846	.1222	.0470	.1584	.1008	.0408
	$\mathcal{W}_{33}$	.1520	.0906	.0388	.1428	.0874	.0302	.1318	.0778	.0300
100	$\mathcal{W}_{11}$	.1732	.1048	.0348	.1652	.0964	.0384	.1416	.0860	.0286
	$\mathcal{W}_{22}$	.1754	.1116	.0366	.1684	.1070	.0388	.1416	.0858	.0284
	$\mathcal{W}_{33}$	.1290	.0764	.0224	.1228	.0734	.0266	.1192	.0676	.0208
250	$\mathcal{W}_{11}$	.1406	.0808	.0208	.1364	.0736	.0198	.1104	.0620	.0162
	$\mathcal{W}_{22}$	.1390	.0788	.0234	.1350	.0758	.0206	.1170	.0712	.0196
	$\mathcal{W}_{33}$	.1148	.0618	.0174	.1102	.0576	.0154	.1026	.0564	.0170
500	$\mathcal{W}_{11}$	.1334	.0740	.0176	.1168	.0682	.0142	.1128	.0630	.0136
	$\mathcal{W}_{22}$	.1358	.0752	.0178	.1270	.0674	.0176	.1338	.0730	.0196
	$\mathcal{W}_{33}$	.1088	.0548	.0128	.1000	.0528	.0118	.1096	.0552	.0118
					$H_0 : \lambda = 0$, (true $\rho = 0$)					
50	$\mathcal{W}_{11}$	.1660	.1024	.0392	.1436	.0920	.0320	.1450	.0920	.0360
	$\mathcal{W}_{22}$	.1622	.1044	.0382	.1578	.0968	.0378	.1590	.0970	.0410
	$\mathcal{W}_{33}$	.1354	.0842	.0294	.1260	.0758	.0246	.1284	.0798	.0286
100	$\mathcal{W}_{11}$	.1362	.0798	.0256	.1352	.0812	.0268	.1302	.0734	.0230
	$\mathcal{W}_{22}$	.1532	.0908	.0282	.1494	.0906	.0294	.1332	.0758	.0230
	$\mathcal{W}_{33}$	.1174	.0668	.0212	.1162	.0686	.0202	.1186	.0670	.0178
250	$\mathcal{W}_{11}$	.1232	.0732	.0174	.1228	.0690	.0158	.1134	.0576	.0154
	$\mathcal{W}_{22}$	.1266	.0726	.0170	.1238	.0682	.0160	.1174	.0616	.0154
	$\mathcal{W}_{33}$	.1126	.0630	.0132	.1100	.0594	.0118	.1052	.0542	.0126
500	$\mathcal{W}_{11}$	.1108	.0578	.0142	.1094	.0556	.0116	.1116	.0616	.0138
	$\mathcal{W}_{22}$	.1198	.0588	.0148	.1120	.0576	.0128	.1198	.0662	.0160
	$\mathcal{W}_{33}$	.1050	.0530	.0122	.1030	.0524	.0098	.1070	.0572	.0130
					$H_0 : \rho = 0$ (true $\lambda = 0$)					
50	$\mathcal{W}_{11}$	.1730	.1054	.0392	.1714	.1070	.0382	.1498	.0902	.0328
	$\mathcal{W}_{22}$	.1366	.0850	.0326	.1418	.0822	.0312	.1202	.0692	.0192
	$\mathcal{W}_{33}$	.1268	.0794	.0280	.1214	.0710	.0262	.1056	.0598	.0170
100	$\mathcal{W}_{11}$	.1604	.0980	.0268	.1478	.0856	.0250	.1292	.0710	.0198
	$\mathcal{W}_{22}$	.1302	.0758	.0252	.1274	.0732	.0260	.1142	.0672	.0220
	$\mathcal{W}_{33}$	.1124	.0630	.0198	.1056	.0612	.0196	.0952	.0568	.0164
250	$\mathcal{W}_{11}$	.1358	.0742	.0192	.1304	.0724	.0192	.1030	.0506	.0122
	$\mathcal{W}_{22}$	.1216	.0694	.0166	.1226	.0670	.0176	.1036	.0552	.0168
	$\mathcal{W}_{33}$	.1074	.0570	.0132	.1054	.0556	.0126	.0880	.0456	.0132
500	$\mathcal{W}_{11}$	.1306	.0704	.0158	.1126	.0600	.0140	.0976	.0514	.0124
	$\mathcal{W}_{22}$	.1208	.0682	.0170	.1110	.0590	.0150	.1154	.0616	.0146
	$\mathcal{W}_{33}$	.1030	.0528	.0114	.0928	.0466	.0106	.0966	.0478	.0116

Note: $\mathcal{W}_{jj}$ are defined in (4.19) for joint tests and (4.20) for one-directional tests.

Table 4.5b. Empirical Sizes: Two-Sided Tests of $H_0 : \lambda = 0$ in SLD Model
Group Interaction, REG2, $T = 3, \beta = (1,1)', \sigma = 1$. $\mathcal{T}_{jj}$ are defined in (4.21)

n	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal Errors		
50	$\mathcal{T}_{11}$	.1422	.0850	.0232	.1254	.0676	.0190	.1068	.0552	.0140
	$\mathcal{T}_{22}$	.1348	.0808	.0212	.1154	.0586	.0162	.1042	.0586	.0134
	$\mathcal{T}_{33}$	.1120	.0616	.0146	.0992	.0472	.0126	.0918	.0484	.0102
100	$\mathcal{T}_{11}$	.1224	.0622	.0174	.1186	.0660	.0136	.1070	.0590	.0116
	$\mathcal{T}_{22}$	.1142	.0604	.0128	.1214	.0654	.0158	.1108	.0600	.0130
	$\mathcal{T}_{33}$	.1004	.0478	.0102	.1046	.0518	.0118	.0958	.0502	.0084
250	$\mathcal{T}_{11}$	.1148	.0584	.0176	.1042	.0540	.0112	.1006	.0512	.0142
	$\mathcal{T}_{22}$	.1130	.0622	.0172	.1128	.0604	.0128	.1140	.0572	.0150
	$\mathcal{T}_{33}$	.1006	.0526	.0130	.0946	.0506	.0086	.0996	.0466	.0124
500	$\mathcal{T}_{11}$	.1126	.0560	.0106	.1082	.0528	.0122	.0970	.0472	.0082
	$\mathcal{T}_{22}$	.1154	.0646	.0140	.1066	.0564	.0118	.1064	.0554	.0106
	$\mathcal{T}_{33}$	.1010	.0554	.0110	.0972	.0484	.0104	.0960	.0474	.0080

Table 4.5c. Empirical Sizes: Two-Sided Tests of $H_0 : \rho = 0$ in SED Model
`Group Interaction, REG2,` $T = 3, \beta = (1, 1)', \sigma = 1.$ $\mathcal{T}_{jj}$ `are defined in` (4.21)

n	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal Errors		
50	$\mathcal{T}_{11}$	.1572	.0920	.0282	.1492	.0846	.0236	.1282	.0666	.0164
	$\mathcal{T}_{22}$	.1386	.0758	.0234	.1242	.0734	.0220	.1030	.0572	.0152
	$\mathcal{T}_{33}$	.1146	.0620	.0172	.1152	.0640	.0176	.0928	.0518	.0142
100	$\mathcal{T}_{11}$	.1420	.0798	.0224	.1324	.0738	.0142	.1170	.0598	.0126
	$\mathcal{T}_{22}$	.1274	.0736	.0202	.1248	.0700	.0160	.1010	.0550	.0140
	$\mathcal{T}_{33}$	.1116	.0594	.0154	.1054	.0540	.0112	.0840	.0444	.0116
250	$\mathcal{T}_{11}$	.1224	.0630	.0140	.1128	.0568	.0114	.1028	.0544	.0124
	$\mathcal{T}_{22}$	.1190	.0656	.0172	.1096	.0560	.0142	.1056	.0566	.0166
	$\mathcal{T}_{33}$	.1006	.0518	.0124	.0882	.0450	.0114	.0880	.0466	.0114
500	$\mathcal{T}_{11}$	.1124	.0578	.0120	.1126	.0526	.0098	.1004	.0518	.0116
	$\mathcal{T}_{22}$	.1136	.0624	.0142	.1202	.0604	.0148	.1164	.0610	.0178
	$\mathcal{T}_{33}$	.0952	.0492	.0098	.1004	.0482	.0108	.0982	.0476	.0126

Table 4.6. Empirical Sizes: Two-Sided Tests of $H_0 : \beta_1 = \beta_2$ in SARAR Model
`Group Interaction, REG2,` $T = 3, \sigma = 1, \lambda = \rho = 0$

n	Test	10%	5%	1%	10%	5%	1%	10%	5%	1%
		Normal Errors			Normal Mixture			Log-normal Errors		
50	$\mathcal{T}_{11}$	.1608	.1020	.0386	.1630	.1046	.0386	.1604	.0978	.0344
	$\mathcal{T}_{22}$	.1154	.0650	.0214	.1190	.0678	.0206	.1138	.0614	.0204
100	$\mathcal{T}_{11}$	.1334	.0744	.0228	.1344	.0794	.0218	.1334	.0782	.0218
	$\mathcal{T}_{22}$	.1012	.0546	.0138	.1042	.0536	.0126	.1032	.0534	.0120
250	$\mathcal{T}_{11}$	.1240	.0642	.0166	.1210	.0680	.0204	.1196	.0670	.0184
	$\mathcal{T}_{22}$	.1066	.0524	.0120	.1060	.0564	.0152	.1018	.0580	.0114
500	$\mathcal{T}_{11}$	.1092	.0548	.0116	.1100	.0564	.0140	.1154	.0616	.0200
	$\mathcal{T}_{22}$	.0958	.0472	.0092	.0978	.0472	.0100	.1022	.0536	.0146
50	$\mathcal{T}_{11}$	.1624	.1004	.0376	.1624	.1024	.0390	.1610	.0992	.0376
	$\mathcal{T}_{22}$	.1136	.0654	.0196	.1204	.0666	.0208	.1136	.0640	.0216
100	$\mathcal{T}_{11}$	.1282	.0742	.0196	.1394	.0810	.0208	.1420	.0808	.0250
	$\mathcal{T}_{22}$	.0968	.0496	.0114	.1068	.0540	.0090	.1060	.0564	.0118
250	$\mathcal{T}_{11}$	.1254	.0688	.0190	.1224	.0642	.0140	.1146	.0622	.0180
	$\mathcal{T}_{22}$	.1050	.0568	.0142	.1024	.0480	.0094	.0990	.0526	.0132
500	$\mathcal{T}_{11}$	.1240	.0626	.0152	.1130	.0594	.0130	.1220	.0650	.0160
	$\mathcal{T}_{22}$	.1102	.0502	.0124	.0978	.0482	.0096	.1084	.0552	.0122

Note: $\beta = (1, 1)'$ for upper panel, and $(.5, .5)'$ for lower panel. $\mathcal{T}_{jj}$ are defined in (4.22).

Part II

Robust Inferences for Spatial Econometric Models

CHAPTER 5

Adjusted QML Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity and Non-normality

5.1 Introduction

While heteroskedasticity is common in regular cross-section studies, it may be more so for a spatial econometrics model due to aggregation, clustering, etc. Anselin (1988) identifies that heteroskedasticity can broadly occur due to "idiosyncrasies in model specification and affect the statistical validity of the estimated model". This may be due to the misspecification of the model that feeds to the disturbance term or may occur more naturally in the presence of peer interactions. Heteroskedasticity may also occur if the model deals with a mix of aggregate and non aggregate data, the aggregation may cause errors to be heteroskedastic.[1] As such, the assumption of homoskedastic disturbances is likely to be invalid in a

[1]See, e.g., Glaeser et al. (1996), Le Sage and Pace (2009), Lin and Lee (2010), Kelejian and Prucha (2010), for more discussions.

spatial context in general. However, much of the present spatial econometrics literature has focused on estimators developed under the assumption that the errors are homoskedastic.

Although Anselin raised the issue of heteroskedasticity in spatial models as early as in 1988, and made an attempt to provide tests of spatial effects robust to unknown heteroskedasticity, comprehensive treatments of estimation related issues were not considered until recent years.[2] We introduce a robust estimator for the SAR model by adjusting the concentrated quasi score function for the spatial parameter. It turns out that the method is simple and can be easily generalised to suit more general models.[3] For heteroskedasticity robust inferences, we propose an outer-product-of-gradient (OPG) method for estimating the asymptotic variance of estimators. We provide formal theories for the consistency and asymptotic normality of the proposed estimator, and the consistency of the robust standard error estimate. Extensive Monte Carlo results show that the proposed estimator generally outperforms its GMM counterparts in terms of efficiency and sensitivity to the magnitude of model parameters in particular the regression coefficients. The Monte Carlo results also show that the proposed robust standard error estimate performs well. We also study the cases under which the regular QML estimator is robust against unknown heteroskedasticity and provide a set of robust inference methods. It is interesting to note that the prosed estimator is computationally as simple as the regular QML estimator, and it also outperforms the regular QML estimator when the latter is heteroskedasticity robust.

[2]e.g., Kelejian and Prucha (2007, 2010), Le Sage (1997), Lin and Lee (2010), Arraiz et al. (2010), Badinger and Egger (2011), Jin and Lee (2012), Baltagi and Yang (2013b), and Doğan and Taşpinar (2014). Lin and Lee (2010) formally illustrate that the traditional quasi maximum likelihood (QML) and generalised method of moments (GMM) estimators are inconsistent in general when the SAR model suffers from heteroskedasticity, and provide heteroskedasticity robust GMM estimators by adjusting the usual quadratic moment conditions.

[3]The efficiency of an ML estimator may be the driving force for exploiting a likelihood-based estimator for achieving robustness against various model misspecifications such as heteroskedasticity and non-normality.

To demonstrate their flexibility and generality, the proposed methods are then extended to the popular spatial autoregressive model with spatial autoregressive disturbances (SARAR(1,1)) with heteroskedastic innovations. Kelejian and Prucha (2010) formally treat this model with a three-step estimation procedure. Monte Carlo results show that the ACQS estimator performs better in finite sample than the three-step estimator.

The rest of the chapter is organised as follows. Section 5.2 examines the cases where the regular QML estimator of the SAR model is consistent under unknown heteroskedasticity, and provides methods for robust inferences. Section 5.3 introduces the ACQS estimator that is generally robust against unknown heteroskedasticity, and presents methods for robust inferences. Section 5.4 presents the Monte Carlo results for the SAR model. Section 5.5 extends the proposed methods to the popular SARAR(1,1) model and discusses further possible extensions. Section 5.6 concludes the chapter. All technical details are given in Appendix F.

5.2 QML Estimation of SAR Models

In this section[4], we first outline the QML estimation of the SAR model under the assumptions that the errors are independent and identically distributed (iid). Then, we examine the properties of the QML estimator of the SAR model when the errors are independent but not identically distributed (inid). We provide conditions under which the regular QML estimator is robust against heteroskedasticity of unknown form, derive its asymptotic distribution, and provide heteroskedasticity robust estimator of its asymptotic variance.

[4] Some general notation will be followed in this chapter: $|\cdot|$ and $\text{tr}(\cdot)$ denote, respectively, the determinant and trace of a square matrix; A' denotes the transpose of a matrix A; $\text{diag}(\cdot)$ denotes the diagonal matrix formed by a vector or the diagonal elements of a square matrix; $\text{diagv}(\cdot)$ denotes the column vector formed by the diagonal elements of a square matrix; and a vector raised to a certain power is operated element-wise.

5.2.1 The model and the QML estimation

Consider the spatial autoregressive or SAR model of the form:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + \epsilon_n, \tag{5.1}$$

where X_n is an $n \times k$ matrix of regressors, W_n is a known $n \times n$ spatial weights matrix, ϵ_n is an $n \times 1$ vector of disturbances of independent and identically distributed (iid) elements with mean zero and variance σ^2, β is a $k \times 1$ vector of regression coefficients and λ is the spatial parameter. The Gaussian log-likelihood of $\theta = (\beta', \sigma^2, \lambda)'$ is,

$$\ell_n(\theta) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln(\sigma^2) + \ln|A_n(\lambda)| - \tfrac{1}{2\sigma^2}\epsilon_n'(\beta,\lambda)\epsilon_n(\beta,\lambda), \tag{5.2}$$

where $A_n(\lambda) = I_n - \lambda W_n$, I_n is an $n \times n$ identity matrix, and $\epsilon_n(\beta,\lambda) = A_n(\lambda)Y_n - X_n\beta$. Given λ, $\ell_n(\theta)$ is maximised at $\hat{\beta}_n(\lambda) = (X_n'X_n)^{-1}X_n'A_n(\lambda)Y_n$ and $\hat{\sigma}_n^2(\lambda) = \frac{1}{n}Y_n'A_n'(\lambda)M_nA_n(\lambda)Y_n$, where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. Substituting $\hat{\beta}_n(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ in $\ell_n(\theta)$, we get the concentrated Gaussian log-likelihood function for λ as,

$$\ell_n^c(\lambda) = -\tfrac{n}{2}[\ln(2\pi)+1] - \tfrac{n}{2}\ln(\hat{\sigma}_n^2(\lambda)) + \ln|A_n(\lambda)|. \tag{5.3}$$

Maximizing $\ell_n^c(\lambda)$ gives the unconstrained QML estimator $\hat{\lambda}_n$ of λ, and thus the QML estimators of β and σ^2 as $\hat{\beta}_n \equiv \hat{\beta}(\hat{\lambda}_n)$ and $\hat{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\hat{\lambda}_n)$. Denote $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\lambda}_n)'$, the QML estimator of θ.

Under regularity conditions, Lee (2004) establishes the consistency and asymptotic normality of the QML estimator $\hat{\theta}_n$. In particular, $\hat{\lambda}_n$ and $\hat{\beta}_n$ may have a slower than $\sqrt{n}$-rate of convergence if the degree of spatial dependence (or the number of neighbours each spatial unit has) grows with the sample size. The QML estimator and its asymptotic distribution developed by Lee are robust for non-normality of the errors. However, some important issues remain: (i) conditions under which the regular QML estimator $\hat{\theta}_n$ remains consistent when errors are heteroskedastic, (ii) methods to adjust the QML estimator $\hat{\theta}_n$ so that it becomes consistent under unknown heteroskedasticity, and (iii) methods of estimating the variance of the (adjusted) QML estimator robust for unknown heteroskedasticity.

5.2.2 Robustness of QML estimator under unknown heteroskedasticity

It is accepted that the regular QML estimator of the usual linear regression model without spatial dependence, developed under homoskedastic errors, is still consistent when the errors are in fact heteroskedastic. However, for correct inferences the standard error of the estimator has to be adjusted to account for this unknown heteroskedasticity (White, 1980). Suppose now we have a linear regression model with spatial dependence as given in (6.21) with disturbances that are inid with means zero and variances $\sigma^2 h_{n,i},\ i = 1, \ldots, n$, where $h_{n,i} > 0$ and $\frac{1}{n}\sum_{i=1}^{n} h_{n,i} = 1$.[5] Consider the score function derived from (5.2),

$$\psi_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial \theta} = \begin{cases} \frac{1}{\sigma^2} X_n' \epsilon_n(\beta, \lambda), \\ \frac{1}{2\sigma^4}[\epsilon_n'(\beta, \lambda)\epsilon_n(\beta, \lambda) - n\sigma^2], \\ \frac{1}{\sigma^2} Y_n' W_n' \epsilon_n(\beta, \lambda) - \mathrm{tr}[G_n(\lambda)], \end{cases} \tag{5.4}$$

where $G_n(\lambda) = W_n A^{-1}(\lambda)$. It is well known that for an extremum estimator, such as the QML estimator $\hat{\theta}_n$ we consider, to be consistent, a necessary condition is that $\text{plim}_{n\to\infty} \frac{1}{n}\psi_n(\theta_0) = 0$ at the true parameter θ_0 (Amemiya, 1985). This is

[5] Note that σ^2 is the average of $\mathrm{Var}(\epsilon_{n,i})$. Under homoskedasticity, $h_{n,i} = 1, \forall i$. For generality, we allow $h_{n,i}$ to depend on n, for each i. This parametrisation, a non-parametric version of Breusch and Pagan (1979), is useful as it allows the estimation of the average scale parameter. See Section 5.3 for more details.

always the case for the β and σ^2 components of $\psi_n(\theta_0)$ whether or not the errors are homoskedastic. However, it may not be the case for the λ component of $\psi_n(\theta_0)$. Let $h_n = (h_{n,1}, \ldots, h_{n,n})'$, $g_n = (g_{n,1}, \ldots, g_{n,n})' = \text{diagv}(G_n)$, $\bar{g}_n = \frac{1}{n}\sum_{i=1}^n g_{n,i}$, $H_n = \text{diag}(h_n)$. Let $\text{Cov}(g_n, h_n)$ denote the sample covariance between the two vectors g_n and h_n. We have, similarly to Lin and Lee (2010),

$$\begin{aligned} \frac{1}{n}\frac{\partial}{\partial\lambda}\ell_n(\theta_0) &= \frac{1}{n}\text{tr}(H_n G_n - G_n) + o_p(1) \\ &= \frac{1}{n}\sum_{i=1}^n (h_{n,i} - 1)(g_{n,i} - \bar{g}_n) + o_p(1) \\ &= \text{Cov}(g_n, h_n) + o_p(1). \end{aligned} \tag{5.5}$$

Therefore, for $\hat{\theta}_n$ to be consistent, it is necessary that as $n \to \infty$, $\text{Cov}(g_n, h_n) \to 0$; in other words, when $\lim_{n\to\infty} \text{Cov}(g_n, h_n) \neq 0$, $\hat{\theta}_n$ cannot be consistent.

Lin and Lee (2010) noted that this condition is satisfied if almost all the diagonal elements of the matrix G_n are equal. In fact, by Cauchy-Schwartz inequality, this condition is satisfied if $\text{Var}(g_n) \to 0$, which boils down to $\text{Var}(k_n) \to 0$, where k_n is the vector of number of neighbours for each unit.[6] Furthermore, if heteroskedasticity occurs due to reasons unrelated to the number of neighbours, for example, due to the nature of the exogenous regressors X_n, then the required condition will still be satisfied. These discussions suggest that the regular QML estimator of the SAR model derived under homoskedasticity can still be consistent when in fact the errors are heteroskedastic. following regularity conditions.[7]

Assumption 5.1: *The true parameter λ_0 is in the interior of a compact parameter set Λ.*[8]

[6]This is because (i) $G_n = W_n + \lambda W_n^2 + \lambda^2 W_n^3 + \ldots$, if $|\lambda| < 1$ and $w_{n,ij} < 1$, and (ii) the diagonal elements of $W_n^r, r \geq 2$ inversely relate to k_n, see Anselin (2003). In fact, when W_n is row-normalised and symmetric, $\text{diag}(W_n^2) = \{k_{n,i}^{-1}\}$. $\text{Var}(k_n) = o(1)$ can be seen to be true for many popular spatial layouts such as Rook, Queen, group interactions such that variation in groups sizes becomes small when n gets large, etc., see Yang (2010).

[7]A quantity defined at the true parameter is represented with a suppressed variable notation, e.g., $A_n \equiv A_n(\lambda_0)$ and $G_n \equiv G_n(\lambda_0)$.

[8]For QML-type estimation, the parameter space Λ must be such that $A_n(\lambda)$ is non-singular $\forall\lambda \in \Lambda$. If the eigenvalues of W_n are all real, then $\Lambda = (w_{\min}^{-1}, w_{\max}^{-1})$ where $w_{\min}$ and $w_{\max}$

Assumption 5.2: *$\epsilon_n \sim (0, \sigma_0^2 H_n)$, where $H_n = \text{diag}(h_{n,1}, \ldots, h_{n,n})$, such that $\frac{1}{n}\sum_{i=1}^n h_{n,i} = 1$ and $h_{n,i} > 0, \forall i$ and $E|\epsilon_{n,i}|^{4+\delta} < c$ for some $\delta > 0$ and constant c for all n and i.*

Assumption 5.3: *The elements of the $n \times k$ regressor matrix X_n are uniformly bounded for all n, X_n has the full rank k, and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is non-singular.*

Assumption 5.4: *The spatial weights matrix W_n is uniformly bounded in absolute value in both row and column sums and its diagonal elements are zero.*

Assumption 5.5: *The matrix A_n is non-singular and A_n^{-1} is uniformly bounded in absolute value in both row and column sums. Further, $A_n^{-1}(\lambda)$ is uniformly bounded in either row or column sums, uniformly in $\lambda \in \Lambda$.*

Assumption 5.6: *The limit $\lim_{n\to\infty} \frac{1}{n}(G_n X_n \beta_0)' M_n (G_n X_n \beta_0) = k$, where either $k > 0$, or $k = 0$ but $\lim_{n\to\infty} \frac{1}{n} \ln |\sigma_0^2 A_n^{-1} A_n'^{-1}| - \frac{1}{n} \ln |\sigma_n^2(\lambda) A_n^{-1}(\lambda) A_n'^{-1}(\lambda)| \neq 0$, whenever $\lambda \neq \lambda_0$, where $\sigma_n^2(\lambda) = \frac{1}{n}\sigma_0^2 \text{tr}(H_n A_n'^{-1} A_n'(\lambda) A_n(\lambda) A_n^{-1})$.*

Assumptions 5.2 and 5.3 are similar to those from Lin and Lee (2010). Assumption 5.2 implies that $\{h_{n,i}\}$ as well as the third and fourth moments of $\epsilon_{n,i}$ are uniformly bounded for all n and i. Assumptions 5.2 and 5.3 imply that $\lim_{n\to\infty} \frac{1}{n} X_n' H_n X_n$ exists and is non-singular. Assumptions 5.4 and 5.5 are standard for the SAR model. The uniform boundedness conditions limit the spatial dependence to a manageable level (Kelejian and Prucha, 1999). Assumption 5.6 is the heteroskedastic version of the identification condition introduced by Lee (2004) for the homoskedastic SAR model.

For the log-likelihood and score functions given in (6.23) and (5.4), let $\mathbb{I}_n =$

are, respectively, the smallest and the largest eigenvalues of W_n; if, W_n is row normalised, then $w_{\max} = 1$ and $w_{\min}^{-1} < -1$, and $\Lambda = (w_{\min}^{-1}, 1)$ (Anselin, 1988). In general, the eigenvalues of W_n may not be all real as W_n can be asymmetric. Le Sage and Pace (2009, p. 88-89) argue that only the purely real eigenvalues can affect the singularity of $A_n(\lambda)$. Consequently, for W_n with complex eigenvalues, the interval of λ that guarantees non-singular $A_n(\lambda)$ is $(w_s^{-1}, 1)$ where w_s is the most negative real eigenvalue of W_n. Kelejian and Prucha (2010) suggest Λ be $(-\tau_n^{-1}, \tau_n^{-1})$ where τ_n is the spectral radius of W_n, or $(-1, 1)$ after normalization.

$-\frac{1}{n}\mathrm{E}\big[\frac{\partial^2}{\partial\theta\partial\theta'}\ell_n(\theta_0)\big]$ and $\Sigma_n = \frac{1}{n}\mathrm{E}\big[\frac{\partial}{\partial\theta}\ell_n(\theta_0)\frac{\partial}{\partial\theta'}\ell_n(\theta_0)\big]$, with their exact expressions deferred to the next subsection. We have the following results (recall $g_n = \mathrm{diagv}(G_n)$ and let $q_n = \mathrm{diagv}(G_n' G_n)$).

Theorem 5.1 *Given $Cov(g_n, h_n) = o(1)$ and $Cov(q_n, h_n) = o(1)$, and under Assumptions 5.1-5.6, we have as $n \to \infty$, $\hat{\theta}_n \xrightarrow{p} \theta_0$; under Assumptions 5.1-5.6 and $Cov(g_n, h_n) = o(n^{-1/2})$, we have as $n \to \infty$,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \mathbb{I}^{-1}\Sigma\ \mathbb{I}^{-1}), \tag{5.6}$$

where $\mathbb{I} = \lim_{n\to\infty}\mathbb{I}_n$ and $\Sigma = \lim_{n\to\infty}\Sigma_n$ both assumed to exist and $\mathbb{I}$ is non-singular.

5.2.3 Robust standard errors of the QML estimators

Asymptotically valid inference for θ based on the QML estimators $\hat{\theta}_n$ requires a consistent estimator of the asymptotic variance given in Theorem 5.1.[9] Under unknown heteroskedasticity designated by H_n, we have:

$$\mathbb{I}_n = \begin{pmatrix} \frac{1}{\sigma_0^2 n}X_n'X_n & 0 & \frac{1}{\sigma_0^2 n}X_n'\eta_n \\ \sim & \frac{1}{2\sigma_0^4} & \frac{1}{\sigma_0^2 n}\mathrm{tr}(H_nG_n) \\ \sim & \sim & \frac{1}{\sigma_0^2 n}\eta_n'\eta_n + \frac{1}{n}\mathrm{tr}(H_nG_n'G_n + G_n^2) \end{pmatrix},$$

where $\eta_n = G_nX_n\beta_0$. Thus a consistent estimator of $\mathbb{I}_n$ can still be obtained by 'plugging' $\hat{\theta}_n$ for θ_0, $G_n(\hat{\theta}_n)$ for G_n and $\hat{H}_n = \frac{1}{\hat{\sigma}_n^2}\mathrm{diag}(\hat{\epsilon}_{n,1}^2, \ldots, \hat{\epsilon}_{n,n}^2)$ for H_n, in line with the idea of White (1980), where $\{\hat{\epsilon}_{n,i}\}$ are the QML residuals. However, this approach fails in estimating the variance of the score, Σ_n, as its σ_0^2-element:

$$\Sigma_{n,\sigma^2\sigma^2} = \frac{1}{4n\sigma_0^4}\sum_{i=1}^n(\kappa_{n,i} + 2h_{n,i}^2),$$

[9] This is simple under homoskedasticity as the sample analogues of $\mathbb{I}_n$ and Σ_n can be used as consistent estimators.

cannot be consistently estimated unless the excess kurtosis measures $\{\kappa_{n,i}\}$ are all zero or $\{\epsilon_{n,i}\}$ are normally distributed. This means that robust inference methods for σ_0^2 is not available. Obviously, σ^2 is typically not the main parameter that inferences concern, although the consistency of its QML estimator (shown in Theorem 5.1) is crucial. Thus, to circumvent this problem, we focus on λ and β as those are the main parameters that inferences concern. First, based on the concentrated score function for λ, we obtain the robust variance of $\hat{\lambda}_n$, and then based on the relationship between $\hat{\beta}_n$ and $\hat{\lambda}_n$ we obtain the robust variance of $\hat{\beta}_n$. Detailed developments in this regard are presented in next section.

5.3 Robust Estimation

As argued in Lin and Lee (2010) and further discussed in Section 5.2, the necessary condition for the consistency of the QML estimator, $\lim_{n\to\infty} \text{Cov}(g_n, h_n) = 0$, can be violated when h_n is proportional to the number of neighbours k_n for each spatial unit and $\lim_{n\to\infty} \text{Var}(k_n) \neq 0$.[10] To solve this problem, Lin and Lee (2010) propose robust GMM and optimal robust GMM estimators for the SAR model. In this chapter, we introduce an adjusted concentrated quasi score (ACQS) estimator for the SAR model by adjusting the concentrated score function for the spatial parameter to make it robust against unknown heteroskedasticity. The method is very simple and more importantly it can be easily generalised to suit more general models (see Section 5.5). Furthermore, the method of adjustment takes into account the estimation of the β and σ^2 parameters, thus can be expected to have a good finite sample performance. Indeed, the Monte Carlo results presented in Section 5.4 show an excellent finite sample performance of the proposed esti-

[10]For example, when W_n corresponds to group interactions (circular world spatial layout can be a special case), and the group sizes are generated from a fixed discrete distribution, we have $\lim_{n\to\infty} \text{Var}(k_n) \neq 0$. In fact, in many empirical situations, the spatial weight matrix is constructed from economic or geographic distance, and hence does not satisfy the condition $\text{Cov}(g_n, h_n) = o(1)$.

mator. For robust inferences concerning the spatial or regression parameters, we introduce OPG estimators of the variances of the ACQS estimators.

5.3.1 The method

Given the problems associated with the λ-element of $\psi_n(\theta_0)$ in (6.27), in asymptotically attaining the limit desired to ensure consistency of the related extremum estimator under heteroskedasticity, one can look at an adjustment to the score function that allows it to reach a probability limit of zero. This method is in line with Lin and Lee (2010)'s treatment to the quadratic moments of the form $\mathrm{E}(\epsilon_n' P_n \epsilon_n) = 0$, where $\mathrm{tr}(P_n) = 0$ is replaced with $\mathrm{diag}(P_n) = 0$ to give a consistent GMM estimator under unknown heteroskedasticity. Following this, if we adjust the last component of $\psi_n(\theta_0)$ as, $\sigma_0^{-2}[Y_n' W_n' \epsilon_n - \epsilon_n' \mathrm{diag}(G_n)\epsilon_n]$, we see $\mathrm{plim}\frac{1}{n\sigma_0^2}[Y_n' W_n' \epsilon_n - \epsilon_n' \mathrm{diag}(G_n)\epsilon_n] = 0$, in light of (5.5). This adjustment is asymptotically valid in the sense that it give estimators consistent under unknown heteroskedasticity. However, the finite sample performance of the estimators is not guaranteed as the variations from the estimation of β and σ^2 are unaccounted for.

Now consider the average concentrated score function derived by concentrating out β and σ^2, i.e., replacing β and σ^2 by $\hat{\beta}_n(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ in the last component of (5.4), or taking the derivative of (5.3), and then dividing the resulting concentrated score function by n,

$$\tilde{\psi}_n(\lambda) = \frac{Y_n' A_n'(\lambda) M_n [G_n(\lambda) - \frac{1}{n}\mathrm{tr}(G_n(\lambda)) I_n] A_n(\lambda) Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n}. \tag{5.7}$$

$\tilde{\psi}_n(\lambda)$ captures the variability coming from estimating β and σ^2. Under QML estimation framework, the QML estimator of λ is equivalently defined as $\hat{\lambda}_n = \arg\{\tilde{\psi}_n(\lambda) = 0\}$. Solving $\tilde{\psi}_n(\lambda) = 0$ is equivalent to solving $Y_n' A_n'(\lambda) M_n [G_n(\lambda) - \frac{1}{n}\mathrm{tr}(G_n(\lambda)) I_n] A_n(\lambda) Y_n = 0$, and for $\hat{\lambda}_n$ to remain consistent under unknown heteroskedasticity, it is necessary that $\frac{1}{n}\mathrm{E}[Y_n' A_n' M_n(G_n - \frac{1}{n}\mathrm{tr}(G_n)I_n)A_n Y_n]$ equals or tends to zero, see van der Vaart (1998, ch. 5). This is not true if there exists unknown heteroskedasticity and the conditions stated in Theorem 5.1 are violated.

Our idea is to adjust the numerator of (5.7) so that its expectation at λ_0 is zero even under unknown heteroskedasticity.[11] Since $\mathrm{E}(Y_n' A_n' M_n G_n A_n Y_n) = \sigma_0^2 \mathrm{tr}(H_n M_n G_n) = \sigma_0^2 \mathrm{tr}(H_n \mathrm{diag}(M_n G_n))$, this suggests that one should replace $\frac{1}{n}\mathrm{tr}(G_n)I_n$ by $\mathrm{diag}(M_n G_n)$. However, this leads to $\mathrm{E}(Y_n' A_n' M_n \mathrm{diag}(M_n G_n) A_n Y_n) = \sigma_0^2 \mathrm{tr}(H_n M_n \mathrm{diag}(M_n G_n)) \neq \mathrm{E}(Y_n' A_n' M_n G_n A_n Y_n)$. Thus, to cancel the effect of the additional M_n, one should instead replace $\frac{1}{n}\mathrm{tr}(G_n)I_n$ by $\mathrm{diag}(M_n)^{-1}\mathrm{diag}(M_n G_n)$. Hence, $\tilde{\psi}_n(\lambda)$ is adjusted by replacing $G_n(\lambda) - \frac{1}{n}\mathrm{tr}(G_n(\lambda))I_n$ by, $G_n^{\circ}(\lambda) = G_n(\lambda) - \mathrm{diag}(M_n)^{-1}\mathrm{diag}(M_n G_n(\lambda))$. This gives an adjusted concentrated score function,

$$\tilde{\psi}_n^*(\lambda) = \frac{Y_n' A_n'(\lambda) M_n G_n^{\circ}(\lambda) A_n(\lambda) Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n}, \tag{5.8}$$

and hence a ACQS estimator of λ_0 as,

$$\tilde{\lambda}_n = \arg\{\tilde{\psi}_n^*(\lambda) = 0\}. \tag{5.9}$$

Once a heteroskedasticity robust estimator of λ is obtained, the heteroskedasticity robust estimators (or the ACQS estimators) of β and σ^2 are, respectively, $\tilde{\beta}_n = \hat{\beta}_n(\tilde{\lambda}_n)$ and $\tilde{\sigma}_n^2 = \hat{\sigma}_n^2(\tilde{\lambda}_n)$ as the estimating functions (first two components of $\psi_n(\theta)$) leading to $\hat{\beta}_n(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ defined below (5.2) are robust to unknown heteroskedasticity. More discussions on this will follow.

Jin and Lee (2012) proposed a heteroskedasticity robust *root estimator* of λ by solving the quadratic equation: $Y_n' A_n'(\lambda) M_n P_n A_n(\lambda) Y_n = 0$, where P_n is an $n \times n$

[11] Making the expectation of an estimating function to be zero leads potentially to a finite sample bias corrected estimation. This is in line with Baltagi and Yang (2013a,b) in constructing standardised or heteroskedasticity-robust LM tests with finite sample improvements. See also Kelejian and Prucha (2001, 2010) and Lin and Lee (2010) for some useful methods in handling the linear-quadratic forms of heteroskedastic random vectors.

matrix such that $M_n P_n$ has a zero diagonal. As there are two roots and only one is consistent, they gave criteria to choose the consistent root. When P_n matrix is parameter dependent, they suggested using some initial consistent estimates to come up with an estimate, say $\widehat{P}_n$, of P_n, and then solve $Y_n' A_n'(\lambda) M_n \widehat{P}_n A_n(\lambda) Y_n = 0$. Clearly, $G_n^{\circ}(\lambda)$ defined above is a choice for P_n although an initial estimate of λ, say $\hat{\lambda}_n^0$, is needed to obtain $\widehat{P}_n = G_n^{\circ}(\hat{\lambda}_n^0)$. Jin and Lee also suggest this. This approach is attractive as the root estimator has a closed-form expression and thus can handle a super large data. However, it can be ambiguous in practice in choosing a consistent root as the selection criterion is parameter dependent. Furthermore, our Monte Carlo simulation shows that $Y_n' A_n'(\lambda) M_n \widehat{P}_n A_n(\lambda) Y_n = 0$ tends to give non-real roots when $|\lambda|$ is not small, say ≥ 0.5, in particular when λ is negative, and when n is not very large. In contrast, this problem does not occur to the ACQS estimator $\tilde{\lambda}_n$. Thus, the ACQS estimator $\tilde{\lambda}_n$ complements Jin and Lee's (2012) root estimator. More discussions follow.

Remark 5.1 *It turns out that the ACQS estimators of the SAR model are computationally as simple as the original QML estimators, but the former are generally consistent under unknown heteroskedasticity while preserving the nature of being robust against non-normality.*

Remark 5.2 *The proposed methods can be easily extended to more advanced models (spatial or non-spatial) as demonstrated in Section 5.5 and Chapter 6. However, it is not clear to us how to extend the GMM estimators of Lin and Lee (2010) to a more general model, and the root estimator of Jin and Lee (2012) may run into difficulty for a more general model as when there are two (or more) quadratic functions of two (or more) unknowns, it is difficult to choose the consistent roots.*

Remark 5.3 *The correction $G_n^{\circ}(\lambda)$ as opposed to the intuitively appealing cor-*

rection $G_n(\lambda) - \mathrm{diag}(G_n(\lambda))$ has better finite sample performance since the adjustment is made directly on the concentrated score function which contains the variability accruing from the estimation of β and σ^2.

5.3.2 Asymptotic distribution of the ACQS estimators

To ensure that the adjusted estimation function given in (5.8) uniquely identifies λ_0, the Assumption 5.6 needs to be adjusted as follows. Let $\Omega_n(\lambda) = A'_n(\lambda)[G_n(\lambda) - \mathrm{diag}(G_n(\lambda))]A_n(\lambda)$.

Assumption 5.6*: $\lim_{n\to\infty} \frac{1}{n}[\beta_0' X_n' A_n'^{-1}\Omega_n(\lambda)A_n^{-1}X_n\beta_0 + \sigma_0^2 \mathrm{tr}(H_n A_n'^{-1}\Omega_n(\lambda)A_n^{-1})] \neq 0, \forall \lambda \neq \lambda_0$.

The central limit theorem (CLT) for linear quadratic forms of Kelejian and Prucha (2001) allows for heteroskedasticity and can be used to prove the asymptotic normality of the ACQS estimator. First, the normalised and adjusted concentrated score function has the following representation at λ_0,

$$\sqrt{n}\tilde{\psi}_n^* \equiv \sqrt{n}\tilde{\psi}_n^*(\lambda_0) = \tfrac{1}{\sqrt{n}\sigma_0^2}\big(\epsilon_n' B_n \epsilon_n + c_n'\epsilon_n\big) + o_p(1), \tag{5.10}$$

where $B_n = M_n G_n^\circ$ and $c_n = M_n G_n^\circ X_n \beta_0$. As $\hat{\sigma}_n^2(\lambda_0) = \frac{1}{n}\epsilon_n' M_n \epsilon_n = \frac{1}{n}\mathrm{E}(\epsilon_n' M_n \epsilon_n) + o_p(1) = \frac{\sigma_0^2}{n}\mathrm{tr}(H_n M_n) + o_p(1) = \sigma_0^2 + o_p(1)$, it follows that $\hat{\sigma}_n^{-2}(\lambda_0) = \sigma_0^{-2} + o_p(1)$.

Let $\tau_n(\cdot)$ denote the first-order standard deviation and $\tau_n^2(\cdot)$ the first-order variance of a normalised quantity, e.g., $\tau_n^2(\tilde{\psi}_n^*)$ is the first-order term of $\mathrm{Var}(\sqrt{n}\tilde{\psi}_n^*)$, and $\tau_n^2(\tilde{\lambda}_n)$ is the first-order term of $\mathrm{Var}(\sqrt{n}\tilde{\lambda}_n)$. By Lemma A.3, we have,

$$\begin{aligned}\tau_n^2(\tilde{\psi}_n^*) &= \frac{1}{n}\sum_{i=1}^n (b_{n,ii}^2 h_{n,i}^2 \kappa_{n,i} + \frac{2}{\sigma_0^4} b_{n,ii} c_{n,i} s_{n,i}) \\ &\quad + \tfrac{1}{n}\mathrm{tr}[H_n B_n (H_n B_n + H_n B_n')] + \tfrac{1}{n\sigma_0^2} c_n' H_n c_n,\end{aligned} \tag{5.11}$$

where $b_{n,ii}$ are the diagonal elements of B_n, $s_{n,i} = \mathrm{E}(\epsilon_{n,i}^3)$, and $\kappa_{n,i}$ is the excess kurtosis of $\epsilon_{n,i}$ which together with H_n are defined in Section 5.2.3. Now by the CLT for linear-quadratic forms of Kelejian and Prucha (2001), we have,

$$\frac{\sqrt{n}\tilde{\psi}_n^*}{\tau_n(\tilde{\psi}_n^*)} \xrightarrow{D} N(0,1). \tag{5.12}$$

This result leads to the following theorem.

Theorem 5.2 *Under Assumptions 5.1-5.5 and 5.6*, the ACQS estimator $\tilde{\lambda}_n$ is consistent and asymptotically normal, i.e., as $n \to \infty$, $\tilde{\lambda}_n \xrightarrow{p} \lambda_0$, and*

$$\sqrt{n}(\tilde{\lambda}_n - \lambda_0) \xrightarrow{D} N\big(0,\ \lim_{n\to\infty} \tau_n^2(\tilde{\lambda}_n)\big),$$

where $\tau_n^2(\tilde{\lambda}_n) = \Phi_n^{-2}\tau_n^2(\tilde{\psi}_n^*)$, $\Phi_n = \frac{1}{n}\mathrm{tr}[H_n(G_n^\circ G_n + G_n^{\circ\prime} G_n - \dot{G}_n^\circ)] + \frac{1}{n\sigma_0^2}c_n'\eta_n$, and $\dot{G}_n^\circ = \frac{d}{d\lambda}G_n^\circ = G_n^2 - \mathrm{diag}(M_n)^{-1}\mathrm{diag}(M_n G_n^2)$

. Now consider the ACQS estimators $\tilde{\beta}_n$ and $\tilde{\sigma}_n^2$ of β_0 and σ_0^2 defined below (5.9). Using the relation $A_n(\tilde{\lambda}_n) = A_n - (\tilde{\lambda}_n - \lambda_0)W_n$, we can write,

$$\tilde{\beta}_n = \hat{\beta}_n(\lambda_0) - (\tilde{\lambda}_n - \lambda_0)(X_n'X_n)^{-1}X_n'G_nA_nY_n, \text{ and} \tag{5.13}$$

$$\tilde{\sigma}_n^2 = \hat{\sigma}_n^2(\lambda_0) - 2(\tilde{\lambda}_n - \lambda_0)\tfrac{1}{n}Y_n'W_n'M_nA_nY_n + (\tilde{\lambda}_n - \lambda_0)^2\tfrac{1}{n}Y_n'W_n'M_nW_nY_n \tag{5.14}$$

The asymptotic properties of $\tilde{\beta}_n$ and $\tilde{\sigma}_n^2$ are summarised in the following theorem.

Theorem 5.3 *Under Assumptions 5.1-5.5 and 5.6*, the ACQS estimators $\tilde{\beta}_n$ and $\tilde{\sigma}_n^2$ are consistent, i.e., as $n \to \infty$, $\tilde{\beta}_n \xrightarrow{p} \beta_0$ and $\tilde{\sigma}_n^2 \xrightarrow{p} \sigma_0^2$, and further $\tilde{\beta}_n$ is asymptotically normal, i.e.,*

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) \xrightarrow{D} N\big[0, \lim_{n\to\infty}(X_n'X_n)^{-1}X_n'\mathbb{A}_nX_n(X_n'X_n)^{-1}\big],$$

where $\mathbb{A}_n = n\sigma_0^2H_n + \tau_n^2(\tilde{\lambda}_n)\eta_n\eta_n' - 2\Phi_n^{-1}(\sigma_0^{-2}\mathrm{diag}(B_n)s_n + H_nc_n)\eta_n'$, $\eta_n = G_nX_n\beta_0$, *and* $s_n = \mathrm{E}(\epsilon_n^3)$.[12]

Clearly, the applicability of the results of Theorems 5.2 and 5.3 for making inferences for λ or β depends on the availability of a consistent estimator of $\tau_n^2(\tilde{\psi}_n^*)$. The plug-in method based on (5.11) does not work due to the involvement of higher-order moments $s_{n,i}$ and $\kappa_{n,i}$.

5.3.3 Robust standard errors of the ACQS estimators

Following the discussions in Section 5.2.3, we focus on λ and β for robust inferences. In order to carry out inference for model parameters using the ACQS procedure, we need a consistent estimate of $\tau_n^2(\tilde{\lambda}_n)$. Given this, consistent estimates of $\tau_n^2(\tilde{\beta}_n) = (X_n'X_n)^{-1}X_n'\mathbb{A}_nX_n(X_n'X_n)^{-1}$ immediately follow. The first-order variance of the adjusted score as given in (5.11) contains second, third and fourth moments of ϵ_i which vary across i, and hence a simple White-type estimator (White, 1980) may not be suitable, which in turn makes $\tau_n^2(\tilde{\lambda}_n)$ infeasible. Hence, we follow the idea of Baltagi and Yang (2013b) to decompose the numerator of the adjusted score into a sum of uncorrelated terms, and then use the outer product of gradients (OPG) method to estimate the variance of the score function which in turn leads to a consistent estimate of $\tau^2(\tilde{\lambda}_n)$. Let the numerator of (5.10) be,

$$Q_n(\epsilon_n) = \epsilon_n' B_n \epsilon_n + c_n' \epsilon_n. \tag{5.15}$$

Clearly, Q_n is not a sum of uncorrelated components, but can be made to be so by the technique of Baltagi and Yang (2013b). Decompose the non-stochastic matrix

[12]Similarly, $\sqrt{n}(\tilde{\sigma}_n^2 - \sigma_0^2) \xrightarrow{D} N\big(0, \lim_{n\to\infty} \tau_n^2(\tilde{\sigma}_n^2)\big)$, where the first-order variance of $\sqrt{n}\tilde{\sigma}_n^2$, $\tau_n^2(\tilde{\sigma}_n^2) = \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(\epsilon_{n,i}^2) + \frac{4}{n^2}\sigma_0^4\tau_n^2(\tilde{\lambda}_n)\mathrm{tr}^2(H_nG_n) + \frac{4}{n^2}\mathrm{Cov}(\epsilon_n'\epsilon, \epsilon_n'B_n\epsilon_n + c_n'\epsilon_n)\mathrm{tr}(H_nG_n)\Phi_n^{-1} = O(1)$, suggesting that $\tilde{\sigma}_n^2$ is root-n consistent. However, similar to the regular QML estimator, this result cannot be used for inference for σ_0^2 as the key element in the variance formula $\frac{1}{n}\sum_{i=1}^n \mathrm{Var}(\epsilon_{n,i}^2) = \frac{\sigma_0^4}{n}\sum_{i=1}^n(\kappa_{n,i} + 2h_{n,i}^2)$ cannot be consistently estimated.

B_n as,

$$B_n = B_n^u + B_n^l + B_n^d, \tag{5.16}$$

where B_n^u, B_n^l and B_n^d are, respectively, the upper triangular, lower triangular and diagonal matrices of B_n. Let $\zeta_n = (B_n^{u\prime} + B_n^l)\epsilon_n$. Then, $Q_n(\epsilon_n)$ can be written as,

$$Q_n(\epsilon_n) = \textstyle\sum_{i=1}^n \epsilon_{n,i}(\zeta_{n,i} + b_{n,ii}\epsilon_{n,i} + c_{n,i}), \tag{5.17}$$

where $\epsilon_{n,i}$, $\zeta_{n,i}$ and $c_{n,i}$ are, respectively, the elements of ϵ_n, ζ_n and c_n. Equation (5.17) expresses $Q_n(\epsilon_n)$ as a sum of n uncorrelated terms $\{\epsilon_{n,i}(\zeta_{n,i}+b_{n,ii}\epsilon_{n,i}+c_{n,i})\}$, and hence its OPG gives a consistent estimate of the variance of $Q_n(\epsilon_n)$, which in turn leads to a consistent estimate of $\tau_n^2(\tilde{\psi}_n^*)$, the first-order variance of $\sqrt{n}\psi_n^*$, as:

$$\tilde{\tau}_n^2(\tilde{\psi}_n^*) = \tfrac{1}{n\tilde{\sigma}_n^4}\textstyle\sum_{i=1}^n \left(\tilde{\epsilon}_{n,i}(\tilde{\zeta}_{n,i} + \tilde{b}_{n,ii}\tilde{\epsilon}_{n,i} + \tilde{c}_{n,i})\right)^2, \tag{5.18}$$

where $\tilde{\epsilon}_{n,i}$ are the residuals computed from the ACQS estimators $\tilde{\theta}_n = (\tilde{\beta}_n', \tilde{\sigma}_n^2, \tilde{\lambda}_n)'$.

Let $\tilde{H}_n = \frac{1}{\tilde{\sigma}_n^2}\text{diag}(\tilde{\epsilon}_{1n}^2, \ldots, \tilde{\epsilon}_{nn}^2)$. Let $\tilde{\Phi}_n$ be Φ_n evaluated at $\tilde{\theta}_n$ and $\tilde{H}_n$, $\tilde{\eta}_n = \tilde{G}_n X_n \tilde{\beta}_n$, and $\tilde{G}_n = G_n(\tilde{\lambda}_n)$. Define the estimators of $\tau_n^2(\tilde{\lambda}_n)$ and $\tau_n^2(\tilde{\beta}_n)$ as,

$$\tilde{\tau}_n^2(\tilde{\lambda}_n) = \tilde{\Phi}_n^{-2}\tilde{\tau}_n^2(\tilde{\psi}_n^*), \text{ and} \tag{5.19}$$

$$\tilde{\tau}_n^2(\tilde{\beta}_n) = (X_n'X_n)^{-1}X_n'\tilde{\mathbb{A}}_n X_n(X_n'X_n)^{-1}, \tag{5.20}$$

where $\tilde{\mathbb{A}}_n = n\tilde{\sigma}_n^2\tilde{H}_n + \tilde{\tau}_n^2(\tilde{\lambda}_n)\tilde{\eta}_n\tilde{\eta}_n' - 2\tilde{\Phi}_n^{-1}(\tilde{\sigma}_n^{-2}\tilde{B}_n^d\tilde{s}_n + \tilde{H}_n\tilde{c}_n)\tilde{\eta}_n'$ and $\tilde{s}_n = \tilde{\epsilon}_n^3$. Note Φ_n can be estimated by $-\frac{d}{d\lambda_0}\tilde{\psi}_n^*|_{\lambda_0=\tilde{\lambda}_n}$ as Φ_n is the 1st-order term of $-\text{E}(\frac{d}{d\lambda_0}\tilde{\psi}_n^*)$.

Theorem 5.4 *If Assumptions 5.1-5.5 and 5.6* hold, then we have as* $n \to \infty$, $\tilde{\tau}_n^2(\tilde{\lambda}_n) - \tau_n^2(\tilde{\lambda}_n) \xrightarrow{p} 0$; *and* $\tilde{\tau}_n^2(\tilde{\beta}_n) - \tau_n^2(\tilde{\beta}_n) \xrightarrow{p} 0$.

Finally, when the conditions of Theorem 5.1 are satisfied so the QML estimators are consistent, the robust variances of $\hat{\lambda}_n$ and $\hat{\beta}_n$ can be obtained from the results of Theorems 5.2-5.4. Starting with the concentrated score $\tilde{\psi}_n$ given in (5.7), one obtains $\tau^2(\hat{\lambda}_n)$ by simply replacing G_n° by $G_n - \frac{1}{n}\text{tr}(G_n)I_n$ in (5.10) and (5.11), and in Φ_n defined in Theorem 5.2. Similarly, replacing G_n° by $G_n - \frac{1}{n}\text{tr}(G_n)I_n$ in $\tau_n^2(\tilde{\beta}_n)$ given in Theorem 5.3 leads to $\tau_n^2(\hat{\beta}_n)$. The estimates of $\tau^2(\hat{\lambda}_n)$ and $\tau_n^2(\hat{\beta}_n)$ are obtained in the same way as those of $\tau^2(\tilde{\lambda}_n)$ and $\tau_n^2(\tilde{\beta}_n)$, and their consistency can be proved similarly to the results of Theorem 5.4.

5.4 Monte Carlo Study

Extensive Monte Carlo experiments were conducted to (i) investigate the finite sample behaviour of the original QML estimator $\hat{\lambda}_n$ and the ACQS estimator $\tilde{\lambda}_n$ proposed in this chapter, and their impacts on the estimators of β and σ^2, with respect to the changes in the sample size, spatial layouts, error distributions and the model parameters when the models are heteroskedastic; and (ii) compare the QML estimator and the ACQS estimator with the non-robust generalised method of moments (GMM) estimator of Lee (2001), the robust GMM (RGMM) estimator and the optimal RGMM (ORGMM) estimator of Lin and Lee (2010), two stage least squares (2SLS) estimator of Kelejian and Prucha (1998), and the root estimator (RE) of Jin and Lee (2012). We consider cases where the original QML estimator are robust against heteroskedasticity and the cases it is not.

The simulations are carried out based on the following data generation process (DGP):

$$Y_n = \lambda W_n Y_n + \iota_n \beta_0 + X_{1n}\beta_1 + X_{2n}\beta_2 + \epsilon_n,$$

where ι_n is an $n \times 1$ vector of ones corresponding to the intercept term, X_{1n} and X_{2n} are the $n \times 1$ vectors containing the values of two fixed regressors, and $\epsilon_n = \sigma H_n e_n$. The regression coefficients β is set to either $(3, 1, 1)'$ or $(.3, .1, .1)'$, σ is set to 1, λ takes values form $\{-0.5, -0.25, 0, 0.25, 0.5\}$ and n take values from

$\{100, 250, 500, 1000\}$. The ways of generating the values for (X_{1n}, X_{2n}), the spatial weights matrix W_n, the heteroskedasticity measure H_n, and the idiosyncratic errors e_n are described below. Each set of Monte Carlo results is based on $1,000$ Monte Carlo samples. The way regressors, weights matrices, error distribution and heteroskedasticity is generated are explained in Appendix B.

The GMM-type estimators are implemented by closely following Lin & Lee (2010). A GMM estimator is in general defined as a solution to the minimisation problem: $\min_{\theta \in \Theta} g_n'(\theta) a_n' a_n g_n(\theta)$ where $g_n(\theta) = \big(Q_n, P_{1n}\epsilon_n(\theta), \ldots, P_{mn}\epsilon_n(\theta)\big)' \epsilon_n(\theta)$ represents the linear and quadratic moment conditions, $Q_n = (X_n, W_n X_n)$ is the matrix of instrumental variables (IVs), and $a_n' a_n$ is the weighting matrix related to the distance function of the minimisation problem. The GMM estimator (Kelejian & Prucha, 1999; Lee, 2001) under homoskedastic disturbances can be defined using the usual moment condition, $P_n = \big(G_n - \frac{\text{tr}(G_n)}{n} I_n\big)$ and the IVs, $(G_n X_n \beta, X_n)$. For the RGMM estimator, the P_n matrix in the moment conditions changes to $G_n - \text{diag}(G_n)$. A first step GMM estimator with $P_n = W_n$ is used to evaluate G_n. The weighting matrices of the distance functions are computed using the variance formula of the iid case using residual estimates given by the first step GMM estimate. The ORGMM estimator is a variant of the RGMM estimator in which the weighting matrix is robust to unknown heteroskedasticity. The ORGMM estimator results given in the tables are computed using the RGMM estimator as the initial estimate to compute the standard error estimates and the instruments. Finally, the 2SLS estimator uses the same IV matrix Q_n. Lin and Lee (2010) gives a detailed comparison of the finite sample performance of ML estimator, GMM estimator, RGMM estimator, ORGMM estimator and 2SLS estimator for models with both homoskedastic and heteroskedastic errors. Our Monte Carlo experiments expand theirs by giving a detailed investigation on the effects of non-normality, spatial layouts as well as negative values for the spatial parameter. The

RE of Jin and Lee (2012) is also included.

To conserve space, only the partial results of QML estimator, ACQS estimator, RGMM estimator and ORGMM estimator are reported. The full set of results are available from the authors upon request. The GMM estimator and 2SLS estimator can perform very poorly. The root estimator performs equally well as the ACQS estimator when $|\lambda|$ is not large and n is not small but tends to give non-real roots otherwise. Tables 5.1-5.3 summarise the estimation results for λ and Tables 5.4-5.6 for β, where in each table, the Monte Carlo means, root mean square errors (rmse) and the standard errors (se) of the estimators are reported. To analyse the finite sample performance of the proposed OPG based robust standard error estimators, we also report the averaged se of the regular QML estimator when it is heteroskedasticity robust and the averaged se of the ACQS estimator based on Theorem 5.4. The experiments with $\beta = (0.3, 0.1, 0.1)$ represent cases where the stochastic component is relatively more dominant than the deterministic component of the model. This allows a comparison between the QML-type estimators and the GMM-type estimators when the model suffers from relatively more severe heteroskedasticity and the IVs are weaker. The main observations made from the Monte Carlo results are summarised as follows:

(i) ACQS estimator of λ performs well in all cases considered, and it generally outperforms all other estimators in terms of bias and rmse.[13] Further, in cases where QML estimator is consistent, ACQS estimator can be significantly less biased than QML estimator, and is as efficient as QML estimator.

(ii) RGMM estimator and ORGMM estimator of λ perform reasonably well

[13] A referee points out that under homoskedasticity, the GMM estimator can be as efficient as the ML estimator when errors are normal, and can be more efficient than the QML estimator when the errors are non-normal. See also Lee and Liu (2010). However, under heteroskedasticity, the latter is not observed from our extensive Monte Carlo Experiments. It would be interesting to carry out a theoretical comparison on the efficiency of the heteroskedasticity robust GMM-type and QML-type estimators, but such a study is clearly beyond the scope of this chapter.

when $\beta = (3, 1, 1)'$, but deteriorates significantly when $\beta = (.3, .1, .1)'$ and in this case GMM estimator and 2SLS estimator can be very erratic. In contrast, ACQS estimator is much less affected by the magnitude of β, and is less biased and more efficient than RGMM estimator and ORGMM estimator more significantly when $\beta = (.3, .1, .1)'$.

(iii) RE of λ performs equally well as ACQS estimator when $|\lambda|$ is not big and n is not small, but otherwise tends to give imaginary roots. Thus, when one encounters a super large dataset and the QML estimator or ACQS estimator run into computational difficulty, one may turn to RE and use its closed-form expression.

(iv) The GMM-type estimators can perform quite differently when the errors are normal as opposed to non-normal errors, especially when $\beta = (.3, .1, .1)'$. It is interesting to note that RGMM estimator often outperforms the ORGMM estimator.

(v) The OPG-based estimate of the robust standard errors of ACQS estimator of λ performs well in general with their values very close to their Monte Carlo counter parts.

(vi) Finally, the relative performance of various estimators of β is much less contrasting than that of various estimators of λ, although it can be seen that ACQS estimator of β is slightly more efficient than the competing RGMM estimator and ORGMM estimator.

5.5 Extension to More General Models

As discussed in the introduction and Remark 5.2 of Section 5.3.1, the ACQS estimation method can be extended to more general models (spatial or non-spatial) where there are two or more concentrated score elements that need to be adjusted to account for the unknown heteroskedasticity. One popular example is the SARAR(1,1) model, which extends the SAR model to include disturbances ϵ_n that follow a heteroskedastic SAR process. In this section, we first present a full set of 'feasible' results for the SARAR(1,1) model which takes the form,

$$Y_n = \lambda W_{1n} Y_n + X_n \beta + \epsilon_n, \quad \epsilon_n = \rho W_{2n} \epsilon_n + v_n, \tag{5.21}$$

where $v_{n,i} \sim inid(0, \sigma^2 h_{n,i})$ such that $\frac{1}{n}\sum_{i=1}^n h_{n,i} = 1$. Let $A_n(\lambda) = I_n - \lambda W_{1n}$ and $B_n(\rho) = I_n - \lambda W_{2n}$, then the concentrated Gaussian log-likelihood function for $\delta = (\lambda, \rho)'$ is,

$$\ell_n^c(\delta) = -\tfrac{n}{2}[\ln(2\pi) + 1] - \tfrac{n}{2}\ln(\hat{\sigma}_n^2(\delta)) + \ln|A_n(\lambda)| + \ln|B_n(\rho)|, \tag{5.22}$$

where $\hat{\sigma}_n^2(\delta) = \frac{1}{n} Y_n'(\delta) M_n(\rho) Y_n(\delta)$, $Y_n(\delta) = B_n(\rho) A_n(\lambda) Y_n$ and $M_n(\rho) = I_n - B_n(\rho) X_n [X_n' B_n'(\rho) B_n(\rho) X_n]^{-1} X_n' B_n'(\rho)$. Maximizing (5.22) gives the QML estimator $\hat{\delta}_n$ of δ, and thus the QML estimator of β as $\hat{\beta}_n \equiv \hat{\beta}_n(\hat{\delta}_n)$ where $\hat{\beta}_n(\delta) = [X_n' B_n'(\rho) B_n(\rho) X_n]^{-1} X_n' B_n'(\rho) Y_n(\delta)$, and the QML estimator of σ^2 as $\hat{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\hat{\delta}_n)$. The concentrated score function upon dividing by n is,

$$\tilde{\psi}_n(\delta) = \begin{cases} -\dfrac{1}{n}\mathrm{tr}(G_{1n}(\lambda)) + \dfrac{Y_n'(\delta) M_n(\rho) \bar{G}_{1n}(\delta) Y_n(\delta)}{Y_n'(\delta) M_n(\rho) Y_n(\delta)}, \\ -\dfrac{1}{n}\mathrm{tr}(G_{2n}(\rho)) + \dfrac{Y_n'(\delta) M_n(\rho) \bar{G}_{2n}(\rho) Y_n(\delta)}{Y_n'(\lambda) M_n(\rho) Y_n(\delta)}, \end{cases} \tag{5.23}$$

Table 5.1: Empirical Mean(rmse)[sd]{$\hat{\text{sd}}$} of Estimators of λ for SAR Model
Cases when Regular QML estimator is Consistent

λ_0	n	QML	ACQS	RGMM	ORGMM
		DGP 1: Constant Circular Neighbours (REG-1), $\beta_0 = (3, 1, 1)'$			
.50	100	.464 (.105)[.098]{.092}	.473(.117)[.114]{.099}	.469(.121)[.117]	.479(.132)[.130]
	250	.488(.061)[.060]{.064}	.492(.063)[.063]{.059}	.489(.064)[.063]	.494(.071)[.071]
	500	.494(.043)[.043]{.046}	.497(.043)[.043]{.042}	.495(.043)[.043]	.498(.048)[.048]
	1000	.497(.030)[.030]{.032}	.498(.030)[.030]{.029}	.498(.030)[.030]	.498(.033)[.033]
.25	100	.212(.133)[.127]{.115}	.230(.128)[.127]{.123}	.221(.132)[.129]	.232(.146)[.145]
	250	.233(.080)[.078]{.078}	.246(.081)[.081]{.079}	.242(.082)[.081]	.247(.090)[.090]
	500	.245(.052)[.052]{.054}	.245(.054)[.054]{.054}	.243(.054)[.054]	.244(.060)[.059]
	1000	.246(.041)[.041]{.040}	.247(.039)[.039]{.038}	.246(.039)[.039]	.247(.043)[.043]
.00	100	-.033(.153)[.149]{.142}	-.014(.150)[.149]{.142}	-.024(.156)[.154]	-.009(.172)[.172]
	250	-.017 (.090)[.089]{.089}	-.007(.091)[.091]{.089}	-.011(.092)[.092]	-.005(.102)[.102]
	500	-.006 (.063)[.063]{.062}	-.002(.061)[.061]{.064}	-.004(.061)[.061]	-.002(.069)[.069]
	1000	-.006(.046)[.046]{.046}	-.003(.043)[.043]{.045}	-.005(.043)[.043]	-.003(.047)[.047]
-.25	100	-.285(.155)[.151]{.149}	-.272(.171)[.169]{.167}	-.286(.176)[.173]	-.275(.200)[.198]
	250	-.266(.101)[.100]{.100}	-.258(.100)[.100]{.099}	-.264(.101)[.100]	-.260(.112)[.112]
	500	-.259(.070)[.070]{.072}	-.255(.070)[.070]{.070}	-.258(.070)[.070]	-.256(.077)[.076]
	1000	-.253(.050)[.050]{.050}	-.250(.050)[.050]{.049}	-.252(.050)[.050]	-.250(.055)[.055]
-.50	100	-.524(.172)[.170]{.179}	-.506(.172)[.172]{.162}	-.521(.175)[.174]	-.513(.195)[.194]
	250	-.515(.108)[.107]{.112}	-.505(.104)[.104]{.101}	-.511(.104)[.104]	-.507(.117)[.116]
	500	-.501(.075)[.075]{.080}	-.497(.075)[.075]{.073}	-.501(.075)[.075]	-.497(.084)[.084]
	1000	-.500(.054)[.054]{.058}	-.499(.051)[.051]{.051}	-.500(.051)[.051]	-.500(.057)[.057]
		DGP 2: Constant Circular Neighbours (REG-1), $\beta_0 = (3, 1, 1)'$			
.50	100	.465(.098)[.091]{.093}	.481(.107)[.105]{.099}	.475(.118)[.115]	.488(.142)[.141]
	250	.487(.062)[.061]{.063}	.494(.061)[.060]{.059}	.491(.061)[.061]	.495(.084)[.084]
	500	.494(.041)[.041]{.042}	.499(.042)[.042]{.040}	.497(.042)[.042]	.500(.059)[.059]
	1000	.498(.028)[.028]{.028}	.500(.028)[.028]{.029}	.499(.029)[.029]	.499(.041)[.041]
.25	100	.219(.129)[.126]{.124}	.238(.125)[.125]{.124}	.230(.128)[.127]	.251(.168)[.168]
	250	.236(.081)[.080]{.080}	.243(.080)[.079]{.079}	.239(.081)[.080]	.245(.108)[.108]
	500	.246(.056)[.056]{.059}	.250(.056)[.056]{.053}	.248(.056)[.056]	.251(.080)[.080]
	1000	.249(.039)[.039]{.041}	.251(.039)[.039]{.037}	.250(.039)[.039]	.250(.052)[.052]
.00	100	-.029(.146)[.143]{.139}	-.010(.143)[.143]{.139}	-.020(.150)[.148]	-.005(.209)[.209]
	250	-.011(.088)[.088]{.087}	-.003(.088)[.088]{.085}	-.008(.089)[.088]	.003(.122)[.122]
	500	-.005(.063)[.063]{.061}	-.008(.064)[.064]{.062}	-.010(.064)[.064]	-.004(.092)[.092]
	1000	-.003(.045)[.045]{.045}	-.001(.043)[.043]{.044}	-.003(.043)[.043]	.000(.060)[.060]
-.25	100	-.276(.158)[.155]{.145}	-.257(.156)[.156]{.153}	-.271(.160)[.159]	-.249(.223)[.223]
	250	-.268(.100)[.099]{.106}	-.261(.099)[.099]{.093}	-.266(.100)[.099]	-.260(.136)[.136]
	500	-.256(.073)[.073]{.077}	-.252(.073)[.073]{.069}	-.255(.074)[.073]	-.254(.102)[.102]
	1000	-.254(.050)[.050]{.050}	-.252(.049)[.049]{.048}	-.253(.050)[.049]	-.252(.068)[.068]
-.50	100	-.527(.155)[.153]{.163}	-.505(.154)[.154]{.154}	-.519(.158)[.157]	-.511(.221)[.221]
	250	-.505(.101)[.101]{.103}	-.500(.099)[.099]{.097}	-.506(.100)[.100]	-.502(.138)[.138]
	500	-.507(.075)[.075]{.077}	-.502(.072)[.072]{.072}	-.505(.072)[.072]	-.501(.103)[.103]
	1000	-.505(.050)[.049]{.049}	-.503(.050)[.049]{.050}	-.504(.050)[.050]	-.505(.071)[.071]
		DGP 3: Constant Circular Neighbours (REG-1), $\beta_0 = (3, 1, 1)'$			
.50	100	.474(.086)[.082]{.094}	.484(.096)[.095]{.089}	.476(.100)[.098]	.480(.149)[.148]
	250	.491(.057)[.056]{.054}	.497(.056)[.056]{.052}	.495(.076)[.076]	.499(.088)[.088]
	500	.493(.040)[.039]{.038}	.496(.040)[.039]{.038}	.494(.040)[.039]	.494(.067)[.067]
	1000	.496(.030)[.030]{.029}	.497(.029)[.028]{.027}	.497(.029)[.029]	.498(.045)[.045]
.25	100	.213(.124)[.119]{.110}	.231(.119)[.117]{.115}	.221(.125)[.122]	.233(.185)[.184]
	250	.240(.072)[.071]{.079}	.247(.071)[.070]{.067}	.242(.072)[.072]	.244(.116)[.116]
	500	.245(.050)[.050]{.052}	.247(.054)[.054]{.050}	.245(.055)[.054]	.245(.087)[.087]
	1000	.248(.037)[.037]{.038}	.250(.037)[.037]{.035}	.249(.037)[.037]	.250(.057)[.057]
.00	100	-.024(.124)[.122]{.116}	-.015(.140)[.140]{.143}	-.027(.148)[.145]	-.018(.221)[.220]
	250	-.010(.085)[.085]{.082}	-.002(.084)[.084]{.088}	-.007(.086)[.086]	-.002(.133)[.133]
	500	-.006(.059)[.058]{.060}	-.002(.058)[.058]{.058}	-.005(.059)[.059]	-.007(.101)[.101]
	1000	-.004(.045)[.044]{.044}	-.002(.042)[.042]{.041}	-.003(.043)[.043]	.000(.069)[.069]
-.25	100	-.276(.148)[.146]{.156}	-.258(.146)[.146]{.142}	-.272(.152)[.150]	-.261(.236)[.236]
	250	-.260(.093)[.092]{.101}	-.252(.093)[.093]{.096}	-.259(.094)[.093]	-.253(.153)[.153]
	500	-.256(.063)[.063]{.065}	-.254(.065)[.065]{.064}	-.256(.066)[.066]	-.251(.111)[.111]
	1000	-.254(.049)[.049]{.047}	-.250(.049)[.049]{.046}	-.252(.050)[.050]	-.251(.076)[.076]
-.50	100	-.514(.141)[.140]{.153}	-.508(.161)[.161]{.167}	-.526(.165)[.163]	-.513(.246)[.245]
	250	-.511(.092)[.091]{.098}	-.506(.097)[.097]{.091}	-.512(.099)[.098]	-.514(.155)[.154]
	500	-.503(.069)[.069]{.069}	-.499(.069)[.069]{.067}	-.503(.069)[.069]	-.498(.111)[.111]
	1000	-.503(.051)[.051]{.051}	-.501(.051)[.051]{.049}	-.503(.051)[.051]	-.505(.081)[.081]

Table 5.1: Cont'd

λ_0	n	QML	ACQS	RGMM	ORGMM
		DGP 1: Queen Contiguity (REG-1), $\beta_0 = (.3, .1, .1)'$			
.50	100	.447(.156)[.146]{.136}	.471(.147)[.144]{.148}	.463(.158)[.154]	.501(.207)[.207]
	250	.482(.081)[.079]{.088}	.495(.079)[.079]{.079}	.488(.081)[.080]	.499(.085)[.085]
	500	.489(.061)[.059]{.063}	.494(.056)[.056]{.056}	.491(.070)[.069]	.497(.071)[.071]
	1000	.496(.041)[.041]{.045}	.497(.042)[.042]{.040}	.495(.042)[.042]	.498(.043)[.043]
.25	100	.207(.170)[.165]{.155}	.231(.167)[.166]{.155}	.219(.172)[.169]	.240(.186)[.186]
	250	.232(.103)[.101]{.101}	.241(.102)[.102]{.099}	.234(.104)[.102]	.242(.106)[.106]
	500	.242(.072)[.072]{.072}	.249(.072)[.072]{.070}	.245(.072)[.072]	.250(.074)[.074]
	1000	.244(.050)[.050]{.052}	.247(.050)[.050]{.050}	.245(.050)[.050]	.247(.051)[.051]
.00	100	-.046(.192)[.186]{.173}	-.021(.188)[.187]{.174}	-.036(.195)[.192]	-.021(.205)[.204]
	250	-.019(.117)[.115]{.112}	-.008(.115)[.115]{.112}	-.017(.117)[.116]	-.010(.120)[.120]
	500	-.008(.080)[.080]{.079}	-.001(.080)[.080]{.080}	-.005(.080)[.080]	-.001(.082)[.082]
	1000	-.005(.058)[.058]{.057}	-.002(.058)[.058]{.057}	-.004(.058)[.058]	-.002(.059)[.059]
-.25	100	-.286(.199)[.195]{.192}	-.258(.198)[.198]{.193}	-.277(.205)[.204]	-.264(.218)[.217]
	250	-.272(.122)[.120]{.125}	-.258(.121)[.120]{.120}	-.268(.122)[.121]	-.265(.126)[.125]
	500	-.260(.089)[.088]{.089}	-.253(.089)[.089]{.086}	-.258(.089)[.089]	-.256(.090)[.090]
	1000	-.256(.063)[.063]{.064}	-.252(.063)[.063]{.061}	-.255(.063)[.063]	-.254(.064)[.064]
-.50	100	-.526(.194)[.192]{.201}	-.502(.194)[.194]{.187}	-.521(.197)[.196]	-.521(.214)[.213]
	250	-.513(.122)[.121]{.128}	-.501(.122)[.122]{.122}	-.513(.124)[.123]	-.514(.128)[.127]
	500	-.504(.087)[.087]{.088}	-.498(.088)[.088]{.087}	-.503(.088)[.088]	-.503(.089)[.089]
	1000	-.503(.063)[.063]{.061}	-.500(.063)[.063]{.063}	-.502(.063)[.063]	-.502(.064)[.064]
		DGP 2: Queen Contiguity (REG-1), $\beta_0 = (.3, .1, .1)'$			
.50	100	.455(.136)[.129]{.137}	.481(.129)[.128]{.123}	.470(.135)[.132]	.581(.354)[.345]
	250	.480(.087)[.083]{.100}	.493(.078)[.078]{.076}	.487(.080)[.079]	.533(.160)[.157]
	500	.490(.057)[.056]{.057}	.497(.056)[.056]{.054}	.495(.068)[.068]	.518(.088)[.086]
	1000	.496(.042)[.042]{.047}	.499(.042)[.042]{.039}	.498(.042)[.042]	.510(.053)[.052]
.25	100	.206(.171)[.166]{.155}	.233(.166)[.165]{.161}	.224(.180)[.178]	.308(.366)[.361]
	250	.222(.108)[.104]{.105}	.240(.097)[.096]{.094}	.232(.099)[.098]	.272(.139)[.137]
	500	.239(.072)[.071]{.076}	.246(.071)[.071]{.068}	.242(.072)[.071]	.259(.089)[.089]
	1000	.246(.050)[.050]{.050}	.245(.052)[.052]{.050}	.244(.053)[.052]	.257(.070)[.070]
.00	100	-.035(.177)[.174]{.165}	-.023(.184)[.182]{.188}	-.039(.191)[.187]	.002(.243)[.243]
	250	-.019(.116)[.115]{.109}	-.005(.115)[.114]{.106}	-.014(.117)[.116]	.016(.153)[.152]
	500	-.009(.081)[.080]{.078}	-.004(.081)[.081]{.077}	-.008(.082)[.081]	.012(.105)[.105]
	1000	-.004(.057)[.057]{.057}	-.002(.057)[.057]{.056}	-.005(.057)[.057]	.007(.069)[.069]
-.25	100	-.283(.185)[.182]{.190}	-.268(.186)[.185]{.186}	-.285(.192)[.189]	-.254(.251)[.251]
	250	-.270(.122)[.120]{.125}	-.256(.121)[.120]{.114}	-.267(.123)[.122]	-.253(.161)[.161]
	500	-.256(.085)[.084]{.085}	-.250(.085)[.085]{.082}	-.254(.085)[.085]	-.242(.106)[.106]
	1000	-.252(.063)[.063]{.060}	-.249(.063)[.063]{.060}	-.251(.063)[.063]	-.245(.078)[.078]
-.50	100	-.518(.195)[.194]{.204}	-.506(.188)[.187]{.180}	-.529(.193)[.190]	-.523(.255)[.254]
	250	-.513(.127)[.126]{.128}	-.501(.127)[.127]{.125}	-.512(.128)[.128]	-.513(.168)[.167]
	500	-.505(.088)[.088]{.084}	-.500(.089)[.089]{.085}	-.505(.089)[.088]	-.500(.110)[.110]
	1000	-.503(.063)[.063]{.060}	-.500(.063)[.063]{.061}	-.503(.063)[.063]	-.501(.077)[.077]
		DGP 3: Queen Contiguity (REG-1), $\beta_0 = (.3, .1, .1)'$			
.50	100	.453(.128)[.119]{.126}	.479(.120)[.118]{.109}	.470(.144)[.141]	.631(.463)[.444]
	250	.479(.079)[.076]{.072}	.492(.076)[.075]{.069}	.487(.079)[.077]	.583(.287)[.275]
	500	.486(.056)[.054]{.057}	.492(.054)[.054]{.049}	.489(.055)[.054]	.554(.206)[.198]
	1000	.494(.039)[.038]{.031}	.497(.039)[.038]{.037}	.496(.039)[.039]	.530(.107)[.103]
.25	100	.205(.151)[.144]{.146}	.232(.145)[.144]{.148}	.220(.154)[.151]	.354(.469)[.458]
	250	.231(.100)[.098]{.100}	.245(.098)[.098]{.095}	.237(.100)[.099]	.307(.277)[.271]
	500	.237(.071)[.070]{.072}	.244(.070)[.070]{.069}	.240(.071)[.070]	.306(.250)[.244]
	1000	.246(.049)[.049]{.055}	.248(.050)[.050]{.049}	.246(.051)[.050]	.271(.126)[.124]
.00	100	-.048(.164)[.157]{.159}	-.015(.169)[.168]{.164}	-.029(.175)[.172]	.057(.327)[.321]
	250	-.018(.106)[.104]{.104}	-.004(.104)[.104]{.099}	-.013(.107)[.106]	.038(.214)[.210]
	500	-.011(.077)[.076]{.075}	-.003(.077)[.076]{.071}	-.008(.077)[.077]	.032(.169)[.166]
	1000	-.004(.055)[.055]{.055}	-.001(.055)[.055]{.053}	-.003(.055)[.055]	.028(.132)[.129]
-.25	100	-.284(.170)[.167]{.179}	-.263(.169)[.168]{.163}	-.284(.175)[.172]	-.245(.283)[.283]
	250	-.268(.119)[.117]{.110}	-.254(.118)[.117]{.115}	-.265(.120)[.119]	-.220(.214)[.211]
	500	-.258(.081)[.081]{.083}	-.252(.081)[.081]{.079}	-.257(.081)[.081]	-.221(.176)[.174]
	1000	-.252(.059)[.059]{.054}	-.254(.059)[.059]{.056}	-.256(.059)[.059]	-.224(.151)[.148]
-.50	100	-.523(.176)[.175]{.189}	-.516(.182)[.182]{.187}	-.539(.192)[.188]	-.528(.312)[.311]
	250	-.514(.120)[.119]{.113}	-.501(.119)[.119]{.118}	-.513(.120)[.119]	-.501(.215)[.215]
	500	-.503(.085)[.085]{.084}	-.500(.085)[.085]{.088}	-.505(.085)[.085]	-.491(.172)[.172]
	1000	-.503(.063)[.063]{.061}	-.500(.063)[.063]{.059}	-.502(.063)[.063]	-.496(.150)[.150]

Table 5.2: Empirical Mean(rmse)[sd]{$\hat{\text{sd}}$} of Estimators of λ for SAR Model
Case I of Inconsistent QML estimator: Circular Neighbours (REG-1)

λ_0	n	QML	ACQS	RGMM	ORGMM
		DGP 1: $\beta_0 = (3,1,1)'$			
.50	100	.434(.119)[.100]	.481(.103)[.101]{.093}	.477(.107)[.104]	.483(.113)[.112]
	250	.458(.071)[.057]	.491(.059)[.059]{.057}	.489(.058)[.057]	.492(.061)[.060]
	500	.463(.056)[.043]	.496(.044)[.044]{.043}	.495(.043)[.043]	.496(.046)[.046]
	1000	.472(.040)[.028]	.500(.029)[.029]{.028}	.499(.028)[.028]	.500(.030)[.030]
.25	100	.197(.120)[.107]	.233(.116)[.115]{.115}	.226(.117)[.115]	.232(.127)[.125]
	250	.218(.077)[.070]	.242(.075)[.074]{.070}	.239(.073)[.072]	.242(.075)[.075]
	500	.222(.060)[.053]	.246(.057)[.057]{.054}	.245(.057)[.057]	.247(.061)[.060]
	1000	.225(.042)[.034]	.246(.037)[.036]{.035}	.245(.036)[.036]	.246(.038)[.038]
.00	100	-.023(.114)[.111]	-.009(.127)[.126]{.127}	-.015(.127)[.127]	-.006(.136)[.136]
	250	-.012(.073)[.072]	-.007(.081)[.080]{.078}	-.009(.080)[.079]	-.005(.084)[.084]
	500	-.005(.054)[.053]	-.002(.060)[.060]{.060}	-.003(.060)[.060]	-.001(.064)[.064]
	1000	-.002(.036)[.036]	-.001(.040)[.040]{.039}	-.002(.039)[.039]	-.001(.042)[.042]
-.25	100	-.249(.110)[.110]	-.271(.137)[.135]{.139}	-.271(.132)[.131]	-.270(.155)[.154]
	250	-.226(.072)[.068]	-.250(.082)[.081]{.080}	-.251(.076)[.076]	-.250(.081)[.081]
	500	-.224(.058)[.052]	-.252(.063)[.063]{.062}	-.252(.060)[.060]	-.251(.064)[.064]
	1000	-.225(.043)[.034]	-.252(.040)[.040]{.040}	-.252(.039)[.039]	-.252(.042)[.042]
-.50	100	-.449(.105)[.092]	-.494(.114)[.114]{.119}	-.492(.105)[.104]	-.498(.112)[.112]
	250	-.448(.079)[.059]	-.503(.076)[.076]{.076}	-.498(.065)[.065]	-.500(.070)[.070]
	500	-.444(.073)[.046]	-.506(.061)[.061]{.059}	-.505(.054)[.054]	-.506(.057)[.056]
	1000	-.444(.064)[.030]	-.501(.037)[.037]{.037}	-.500(.034)[.034]	-.501(.035)[.035]
		DGP 2: $\beta_0 = (3,1,1)'$			
.50	100	.438(.114)[.096]	.483(.098)[.097]{.089}	.477(.105)[.102]	.485(.130)[.129]
	250	.462(.066)[.054]	.495(.055)[.055]{.055}	.492(.053)[.053]	.496(.067)[.067]
	500	.467(.054)[.043]	.500(.044)[.044]{.042}	.498(.043)[.043]	.499(.057)[.057]
	1000	.473(.039)[.027]	.501(.028)[.028]{.028}	.500(.027)[.027]	.501(.034)[.034]
.25	100	.201(.123)[.113]	.236(.120)[.119]{.109}	.228(.122)[.120]	.235(.147)[.146]
	250	.219(.072)[.066]	.244(.070)[.070]{.069}	.242(.070)[.069]	.245(.087)[.087]
	500	.220(.059)[.051]	.244(.055)[.054]{.053}	.243(.054)[.054]	.247(.071)[.071]
	1000	.228(.040)[.033]	.248(.035)[.035]{.035}	.248(.035)[.034]	.249(.043)[.043]
.00	100	-.022(.116)[.114]	-.010(.131)[.131]{.129}	-.016(.129)[.128]	-.005(.159)[.158]
	250	-.010(.073)[.072]	-.005(.081)[.081]{.079}	-.008(.080)[.079]	-.004(.097)[.096]
	500	-.004(.051)[.051]	-.001(.058)[.058]{.058}	-.002(.057)[.057]	.001(.075)[.075]
	1000	-.003(.036)[.036]	-.002(.040)[.040]{.039}	-.002(.039)[.039]	-.001(.048)[.048]
-.25	100	-.239(.109)[.108]	-.257(.131)[.131]{.129}	-.256(.122)[.122]	-.248(.150)[.150]
	250	-.232(.071)[.069]	-.257(.083)[.082]{.079}	-.257(.077)[.077]	-.253(.093)[.093]
	500	-.223(.059)[.052]	-.251(.062)[.062]{.060}	-.251(.060)[.060]	-.247(.078)[.078]
	1000	-.222(.045)[.036]	-.249(.041)[.041]{.040}	-.249(.040)[.040]	-.249(.048)[.048]
-.50	100	-.452(.105)[.093]	-.499(.114)[.114]{.116}	-.495(.110)[.110]	-.496(.123)[.123]
	250	-.448(.080)[.061]	-.501(.073)[.073]{.073}	-.499(.066)[.066]	-.499(.079)[.079]
	500	-.438(.077)[.046]	-.500(.059)[.059]{.058}	-.498(.052)[.052]	-.497(.065)[.065]
	1000	-.444(.064)[.031]	-.501(.037)[.037]{.037}	-.502(.034)[.034]	-.502(.041)[.041]
		DGP 3: $\beta_0 = (3,1,1)'$			
.50	100	.445(.107)[.092]	.486(.087)[.086]{.079}	.482(.092)[.090]	.493(.144)[.144]
	250	.464(.066)[.055]	.495(.054)[.054]{.049}	.493(.054)[.053]	.497(.073)[.073]
	500	.467(.055)[.044]	.497(.041)[.041]{.039}	.496(.042)[.041]	.497(.060)[.060]
	1000	.473(.040)[.030]	.499(.027)[.027]{.026}	.499(.027)[.027]	.500(.037)[.037]
.25	100	.199(.116)[.105]	.230(.110)[.108]{.099}	.241(.068)[.067]	.245(.090)[.089]
	250	.219(.071)[.064]	.243(.069)[.068]{.063}	.241(.068)[.067]	.245(.090)[.089]
	500	.222(.058)[.050]	.244(.054)[.053]{.049}	.243(.053)[.053]	.242(.078)[.078]
	1000	.228(.040)[.033]	.248(.035)[.034]{.033}	.248(.034)[.034]	.250(.045)[.045]
.00	100	-.019(.107)[.105]	-.008(.120)[.120]{.119}	-.013(.119)[.119]	-.005(.164)[.164]
	250	-.008(.065)[.065]	-.003(.072)[.072]{.069}	-.006(.072)[.072]	-.003(.101)[.101]
	500	-.006(.051)[.050]	-.004(.057)[.057]{.054}	-.006(.058)[.058]	-.007(.089)[.089]
	1000	-.003(.035)[.034]	-.002(.038)[.038]{.037}	-.003(.038)[.038]	-.003(.053)[.053]
-.25	100	-.243(.102)[.102]	-.260(.123)[.123]{.120}	-.262(.118)[.117]	-.257(.157)[.156]
	250	-.230(.072)[.069]	-.250(.077)[.077]{.072}	-.251(.074)[.074]	-.248(.098)[.098]
	500	-.228(.055)[.050]	-.255(.058)[.058]{.056}	-.256(.058)[.057]	-.255(.083)[.083]
	1000	-.223(.044)[.035]	-.250(.039)[.039]{.038}	-.250(.039)[.039]	-.249(.052)[.052]
-.50	100	-.450(.107)[.095]	-.486(.110)[.109]{.112}	-.485(.105)[.104]	-.484(.125)[.123]
	250	-.450(.081)[.063]	-.502(.074)[.074]{.070}	-.498(.064)[.064]	-.496(.085)[.085]
	500	-.439(.081)[.053]	-.499(.061)[.061]{.059}	-.497(.051)[.051]	-.499(.069)[.069]
	1000	-.445(.066)[.037]	-.500(.038)[.038]{.036}	-.500(.034)[.034]	-.501(.044)[.044]

Table 5.2: Cont'd

λ_0	n	QML	ACQS	RGMM	ORGMM
			DGP 1: $\beta_0 = (.3, .1, .1)'$		
.50	100	.407(.154)[.123]	.474(.129)[.127]{.119}	.467(.148)[.144]	.499(.189)[.189]
	250	.437(.100)[.078]	.489(.080)[.079]{.075}	.485(.082)[.080]	.494(.083)[.083]
	500	.445(.076)[.053]	.494(.054)[.054]{.053}	.493(.069)[.069]	.497(.066)[.066]
	1000	.453(.060)[.037]	.499(.037)[.037]{.038}	.498(.037)[.037]	.500(.038)[.038]
.25	100	.174(.156)[.136]	.226(.155)[.153]{.149}	.213(.165)[.161]	.235(.195)[.194]
	250	.199(.101)[.087]	.238(.097)[.097]{.096}	.233(.100)[.098]	.241(.102)[.102]
	500	.208(.076)[.063]	.243(.069)[.069]{.068}	.240(.070)[.069]	.243(.070)[.070]
	1000	.213(.058)[.045]	.246(.049)[.049]{.048}	.245(.050)[.049]	.246(.050)[.050]
.00	100	-.041(.146)[.140]	-.023(.170)[.168]{.165}	-.040(.179)[.174]	-.026(.184)[.182]
	250	-.016(.096)[.095]	-.009(.114)[.113]{.117}	-.015(.115)[.114]	-.009(.116)[.116]
	500	-.008(.066)[.066]	-.004(.078)[.078]{.077}	-.008(.079)[.079]	-.005(.080)[.080]
	1000	-.004(.044)[.044]	-.002(.052)[.052]{.054}	-.003(.052)[.052]	-.002(.053)[.053]
-.25	100	-.240(.136)[.136]	-.270(.176)[.175]{.172}	-.292(.185)[.180]	-.291(.201)[.197]
	250	-.213(.095)[.087]	-.251(.110)[.110]{.111}	-.259(.111)[.111]	-.256(.114)[.114]
	500	-.210(.074)[.062]	-.252(.079)[.079]{.079}	-.256(.079)[.079]	-.255(.080)[080]
	1000	-.209(.060)[.044]	-.252(.055)[.055]{.056}	-.254(.055)[.055]	-.254(.056)[.056]
-.50	100	-.417(.149)[.124]	-.496(.164)[.164]{.159}	-.531(.202)[.199]	-.535(.213)[.210]
	250	-.413(.117)[.078]	-.504(.103)[.103]{.102}	-.512(.103)[.102]	-.516(.107)[.106]
	500	-.409(.107)[.056]	-.501(.073)[.073]{.073}	-.506(.073)[.073]	-.507(.074)[.074]
	1000	-.405(.103)[.039]	-.498(.051)[.051]{.052}	-.501(.051)[.051]	-.501(.051)[.051]
			DGP 2: $\beta_0 = (.3, .1, .1)'$		
.50	100	.416(.147)[.121]	.482(.123)[.121]{.119}	.475(.138)[.136]	.592(.342)[.329]
	250	.438(.101)[.080]	.490(.081)[.080]{.079}	.487(.090)[.089]	.528(.157)[.154]
	500	.448(.074)[.053]	.496(.053)[.053]{.052}	.494(.054)[.053]	.511(.068)[.067]
	1000	.452(.061)[.038]	.499(.038)[.038]{.037}	.498(.038)[.038]	.508(.047)[.047]
.25	100	.184(.152)[.137]	.236(.154)[.154]{.157}	.224(.165)[.163]	.304(.305)[.301]
	250	.203(.100)[.088]	.242(.097)[.097]{.091}	.236(.099)[.098]	.271(.149)[.147]
	500	.211(.073)[.062]	.246(.067)[.067]{.066}	.243(.068)[.068]	.264(.109)[.109]
	1000	.217(.055)[.044]	.250(.048)[.048]{.047}	.249(.048)[.048]	.258(.058)[.058]
.00	100	-.040(.144)[.139]	-.021(.171)[.169]{.164}	-.039(.180)[.176]	.014(.262)[.262]
	250	-.016(.091)[.089]	-.010(.107)[.107]{.104}	-.016(.109)[.108]	.008(.134)[.134]
	500	-.007(.063)[.063]	-.003(.075)[.075]{.074}	-.006(.075)[.075]	.008(.090)[.090]
	1000	-.003(.046)[.046]	-.001(.054)[.054]{.053}	-.003(.054)[.054]	.006(.066)[.066]
-.25	100	-.232(.133)[.131]	-.259(.169)[.169]{.159}	-.281(.180)[.177]	-.254(.266)[.266]
	250	-.216(.090)[.083]	-.254(.106)[.106]{.107}	-.262(.108)[.107]	-.249(.138)[.138]
	500	-.210(.073)[.061]	-.251(.077)[.077]{.077}	-.255(.077)[.077]	-.246(.088)[.088]
	1000	-.207(.063)[.046]	-.249(.057)[.057]{.055}	-.251(.057)[.057]	-.247(.067)[.067]
-.50	100	-.424(.148)[.127]	-.503(.163)[.163]{.160}	-.535(.191)[.187]	-.549(.246)[.241]
	250	-.410(.123)[.084]	-.499(.105)[.105]{.099}	-.507(.106)[.105]	-.513(.151)[.151]
	500	-.409(.108)[.058]	-.500(.071)[.071]{.072}	-.504(.071)[.071]	-.507(.086)[.086]
	1000	-.409(.100)[.041]	-.503(.050)[.050]{.051}	-.506(.051)[.050]	-.509(.063)[.062]
			DGP 3: $\beta_0 = (.3, .1, .1)'$		
.50	100	.416(.147)[.120]	.480(.118)[.116]{.099}	.473(.130)[.128]	.652(.453)[.426]
	250	.439(.096)[.074]	.490(.071)[.070]{.065}	.486(.073)[.071]	.572(.247)[.236]
	500	.449(.074)[.054]	.497(.050)[.050]{.048}	.495(.051)[.051]	.547(.189)[.184]
	1000	.453(.060)[.037]	.498(.034)[.034]{.035}	.497(.035)[.034]	.523(.104)[.101]
.25	100	.174(.153)[.133]	.224(.147)[.144]{.137}	.212(.156)[.152]	.335(.387)[.378]
	250	.210(.089)[.080]	.249(.087)[.087]{.083}	.243(.087)[.087]	.310(.245)[.237]
	500	.211(.072)[.061]	.244(.065)[.065]{.061}	.242(.066)[.065]	.283(.198)[.195]
	1000	.214(.057)[.044]	.247(.046)[.046]{.044}	.246(.047)[.046]	.266(.116)[.115]
.00	100	-.027(.135)[.133]	-.008(.161)[.160]{.153}	-.026(.172)[.170]	.077(.422)[.414]
	250	-.014(.087)[.086]	-.006(.103)[.103]{.099}	-.013(.105)[.104]	.052(.263)[.258]
	500	-.008(.059)[.058]	-.004(.070)[.070]{.069}	-.008(.071)[.070]	.026(.151)[.149]
	1000	-.003(.042)[.042]	-.001(.050)[.050]{.050}	-.003(.050)[.050]	.025(.116)[.114]
-.25	100	-.234(.131)[.130]	-.262(.172)[.172]{.179}	-.288(.184)[.180]	-.238(.295)[.295]
	250	-.218(.090)[.084]	-.254(.105)[.105]{.099}	-.262(.107)[.106]	-.223(.239)[.238]
	500	-.213(.073)[.063]	-.252(.076)[.076]{.071}	-.256(.077)[.076]	-.233(.161)[.160]
	1000	-.208(.062)[.046]	-.250(.055)[.055]{.053}	-.252(.055)[.055]	-.238(.128)[.127]
-.50	100	-.418(.151)[.127]	-.495(.158)[.158]{.151}	-.526(.178)[.176]	-.544(.304)[.301]
	250	-.411(.126)[.089]	-.503(.105)[.105]{.099}	-.511(.105)[.104]	-.508(.199)[.198]
	500	-.408(.113)[.066]	-.500(.073)[.073]{.069}	-.504(.072)[.072]	-.501(.156)[.156]
	1000	-.403(.109)[.049]	-.496(.051)[.051]{.049}	-.498(.051)[.051]	-.502(.129)[.129]

Table 5.3: Empirical Mean(rmse)[sd]{$\hat{\text{sd}}$} of Estimators of λ for SAR Model
Case II of Inconsistent QML estimator: Group Interaction (REG-2)

λ_0	n	QML	ACQS	RGMM	ORGMM
		DGP 1: $\beta_0 = (3, 1, 1)'$			
.50	100	.422(.124)[.096]	.478(.102)[.099]{.093}	.469(.109)[.105]	.470(.112)[.108]
	250	.461(.069)[.057]	.493(.059)[.059]{.056}	.488(.061)[.060]	.491(.065)[.064]
	500	.472(.047)[.037]	.497(.039)[.038]{.038}	.494(.039)[.039]	.496(.041)[.041]
	1000	.476(.037)[.028]	.499(.029)[.029]{.028}	.497(.029)[.029]	.498(.031)[.030]
.25	100	.159(.161)[.132]	.224(.142)[.140]{.139}	.210(.156)[.150]	.215(.162)[.158]
	250	.210(.087)[.078]	.244(.082)[.081]{.080}	.237(.085)[.084]	.242(.090)[.090]
	500	.223(.060)[.053]	.247(.056)[.056]{.055}	.243(.057)[.057]	.246(.061)[.061]
	1000	.232(.042)[.037]	.251(.039)[.039]{.040}	.249(.040)[.040]	.251(.043)[.043]
.00	100	-.079(.179)[.160]	-.023(.183)[.181]{.183}	-.035(.194)[.191]	-.026(.203)[.201]
	250	-.034(.100)[.094]	-.011(.103)[.103]{.102}	-.020(.107)[.105]	-.014(.112)[.111]
	500	-.018(.067)[.065]	-.006(.071)[.070]{.070}	-.013(.072)[.071]	-.009(.075)[.075]
	1000	-.011(.049)[.048]	-.005(.052)[.052]{.051}	-.009(.054)[.053]	-.007(.057)[.057]
-.25	100	-.317(.184)[.171]	-.285(.210)[.207]{.213}	-.300(.222)[.216]	-.291(.234)[.231]
	250	-.264(.109)[.108]	-.266(.126)[.124]{.123}	-.276(.128)[.125]	-.271(.134)[.132]
	500	-.247(.074)[.074]	-.258(.085)[.085]{.084}	-.265(.086)[.085]	-.262(.091)[.090]
	1000	-.235(.056)[.054]	-.254(.061)[.060]{.060}	-.257(.062)[.061]	-.255(.065)[.065]
-.50	100	-.532(.181)[.178]	-.534(.226)[.224]{.219}	-.546(.231)[.226]	-.543(.245)[.241]
	250	-.468(.120)[.116]	-.505(.146)[.146]{.144}	-.515(.143)[.142]	-.511(.151)[.150]
	500	-.460(.090)[.080]	-.507(.101)[.100]{.097}	-.511(.096)[.095]	-.509(.101)[.101]
	1000	-.448(.078)[.057]	-.501(.070)[.070]{.069}	-.505(.069)[.069]	-.503(.073)[.073]
		DGP 2: $\beta_0 = (3, 1, 1)'$			
.50	100	.437(.117)[.098]	.492(.099)[.098]{.089}	.487(.110)[.110]	.497(.126)[.126]
	250	.465(.066)[.056]	.499(.057)[.057]{.054}	.494(.060)[.059]	.504(.074)[.074]
	500	.471(.047)[.037]	.497(.038)[.038]{.038}	.494(.039)[.038]	.499(.050)[.050]
	1000	.477(.035)[.027]	.500(.028)[.028]{.028}	.498(.028)[.028]	.500(.036)[.036]
.25	100	.167(.155)[.130]	.230(.137)[.135]{.129}	.220(.151)[.148]	.235(.172)[.171]
	250	.211(.085)[.076]	.245(.079)[.079]{.077}	.236(.082)[.081]	.245(.100)[.100]
	500	.219(.060)[.051]	.243(.054)[.053]{.054}	.238(.055)[.054]	.245(.067)[.067]
	1000	.231(.042)[.038]	.251(.040)[.040]{.039}	.248(.040)[.040]	.250(.052)[.052]
.00	100	-.084(.181)[.160]	-.028(.179)[.176]{.169}	-.044(.195)[.190]	-.019(.228)[.227]
	250	-.031(.098)[.093]	-.008(.101)[.101]{.098}	-.018(.107)[.105]	-.005(.134)[.134]
	500	-.015(.068)[.067]	-.003(.073)[.073]{.069}	-.009(.074)[.074]	.001(.095)[.095]
	1000	-.008(.050)[.049]	-.002(.053)[.053]{.050}	-.005(.054)[.054]	.000(.069)[.069]
-.25	100	-.313(.178)[.167]	-.283(.206)[.203]{.211}	-.296(.215)[.210]	-.268(.259)[.258]
	250	-.262(.109)[.108]	-.263(.126)[.126]{.119}	-.272(.128)[.126]	-.256(.159)[.159]
	500	-.243(.072)[.072]	-.254(.082)[.082]{.082}	-.260(.081)[.081]	-.252(.101)[.101]
	1000	-.235(.055)[.053]	-.253(.060)[.060]{.060}	-.256(.061)[.061]	-.252(.080)[.080]
-.50	100	-.523(.182)[.181]	-.531(.241)[.239]{.230}	-.541(.237)[.233]	-.510(.284)[.283]
	250	-.471(.118)[.114]	-.510(.142)[.142]{.140}	-.517(.138)[.137]	-.497(.174)[.174]
	500	-.458(.092)[.082]	-.503(.101)[.101]{.095}	-.509(.098)[.097]	-.498(.121)[.121]
	1000	-.445(.079)[.057]	-.497(.068)[.068]{.069}	-.500(.068)[.068]	-.493(.090)[.089]
		DGP 3: $\beta_0 = (3, 1, 1)'$			
.50	100	.433(.115)[.094]	.484(.090)[.089]{.081}	.476(.110)[.107]	.485(.138)[.138]
	250	.469(.062)[.054]	.500(.053)[.053]{.050}	.495(.055)[.055]	.503(.076)[.076]
	500	.473(.046)[.037]	.497(.036)[.036]{.035}	.494(.037)[.037]	.496(.051)[.051]
	1000	.478(.035)[.027]	.500(.026)[.026]{.026}	.498(.027)[.027]	.502(.038)[.038]
.25	100	.173(.145)[.123]	.232(.125)[.124]{.114}	.221(.150)[.147]	.236(.187)[.186]
	250	.211(.086)[.077]	.243(.079)[.079]{.071}	.236(.084)[.083]	.247(.115)[.115]
	500	.225(.056)[.051]	.248(.052)[.052]{.051}	.244(.054)[.053]	.250(.078)[.078]
	1000	.228(.044)[.038]	.246(.039)[.039]{.038}	.244(.040)[.039]	.248(.056)[.056]
.00	100	-.078(.169)[.150]	-.026(.174)[.172]{.164}	-.044(.188)[.183]	-.019(.229)[.228]
	250	-.030(.098)[.093]	-.008(.102)[.102]{.099}	-.018(.107)[.106]	-.002(.145)[.145]
	500	-.017(.066)[.063]	-.005(.069)[.069]{.066}	-.012(.071)[.070]	-.005(.097)[.097]
	1000	-.007(.047)[.046]	-.001(.050)[.050]{.048}	-.005(.051)[.051]	-.003(.073)[.073]
-.25	100	-.305(.178)[.170]	-.270(.197)[.196]{.199}	-.291(.218)[.214]	-.262(.280)[.280]
	250	-.262(.104)[.103]	-.264(.123)[.122]{.119}	-.272(.124)[.122]	-.256(.173)[.173]
	500	-.248(.071)[.071]	-.259(.081)[.080]{.078}	-.265(.082)[.081]	-.256(.115)[.115]
	1000	-.234(.055)[.053]	-.251(.059)[.059]{.057}	-.255(.060)[.060]	-.249(.090)[.090]
-.50	100	-.535(.181)[.177]	-.530(.218)[.216]{.223}	-.555(.236)[.229]	-.528(.304)[.303]
	250	-.474(.118)[.115]	-.515(.148)[.147]{.139}	-.523(.142)[.141]	-.505(.195)[.195]
	500	-.457(.091)[.080]	-.504(.094)[.093]{.092}	-.509(.091)[.090]	-.500(.125)[.125]
	1000	-.449(.081)[.063]	-.502(.069)[.069]{.067}	-.505(.069)[.069]	-.498(.101)[.101]

Table 5.3: Cont'd

λ_0	n	QML	ACQS	RGMM	ORGMM
			DGP 1: $\beta_0 = (.3, .1, .1)'$		
.50	100	.364(.203)[.150]	.456(.148)[.141]{.129}	.419(.219)[.204]	.423(.234)[.220]
	250	.433(.105)[.080]	.487(.079)[.078]{.073}	.468(.095)[.090]	.469(.095)[.090]
	500	.450(.073)[.053]	.494(.053)[.053]{.051}	.482(.057)[.054]	.483(.057)[.054]
	1000	.460(.054)[.036]	.497(.036)[.036]{.036}	.491(.038)[.037]	.491(.038)[.037]
.25	100	.092(.246)[.188]	.193(.206)[.197]{.185}	.126(.269)[.239]	.127(.289)[.261]
	250	.178(.129)[.107]	.232(.114)[.112]{.109}	.203(.126)[.116]	.202(.127)[.117]
	500	.202(.084)[.069]	.242(.074)[.073]{.073}	.225(.079)[.075]	.225(.079)[.075]
	1000	.215(.059)[.048]	.246(.051)[.051]{.051}	.238(.053)[.051]	.238(.053)[.051]
.00	100	-.150(.258)[.211]	-.070(.257)[.247]{.233}	-.161(.331)[.289]	-.159(.346)[.307]
	250	-.060(.141)[.127]	-.028(.148)[.146]{.133}	-.066(.164)[.150]	-.066(.165)[.151]
	500	-.030(.090)[.085]	-.011(.097)[.097]{.093}	-.033(.104)[.099]	-.032(.104)[.099]
	1000	-.016(.059)[.057]	-.007(.065)[.065]{.066}	-.018(.068)[.066]	-.018(.069)[.066]
-.25	100	-.365(.241)[.212]	-.328(.294)[.283]{.272}	-.441(.381)[.330]	-.432(.409)[.366]
	250	-.260(.127)[.126]	-.264(.159)[.158]{.156}	-.308(.172)[.162]	-.309(.173)[.162]
	500	-.243(.093)[.093]	-.263(.116)[.115]{.110}	-.289(.123)[.117]	-.289(.123)[.117]
	1000	-.228(.071)[.068]	-.258(.084)[.084]{.088}	-.271(.087)[.085]	-.272(.088)[.085]
-.50	100	-.556(.216)[.209]	-.581(.312)[.301]{.299}	-.712(.409)[.350]	-.706(.404)[.347]
	250	-.464(.137)[.132]	-.526(.185)[.183]{.179}	-.576(.202)[.186]	-.579(.204)[.188]
	500	-.439(.113)[.095]	-.514(.129)[.128]{.124}	-.543(.137)[.130]	-.544(.138)[.131]
	1000	-.423(.101)[.066]	-.506(.089)[.089]{.088}	-.520(.092)[.090]	-.521(.092)[.090]
			DGP 2: $\beta_0 = (.3, .1, .1)'$		
.50	100	.361(.206)[.152]	.453(.150)[.143]{.137}	.426(.251)[.240]	.518(.396)[.396]
	250	.435(.103)[.080]	.489(.078)[.077]{.070}	.469(.085)[.079]	.510(.185)[.185]
	500	.453(.070)[.052]	.496(.050)[.050]{.049}	.485(.053)[.051]	.502(.113)[.113]
	1000	.460(.054)[.037]	.497(.036)[.036]{.035}	.492(.038)[.037]	.494(.042)[.042]
.25	100	.098(.241)[.187]	.197(.202)[.194]{.186}	.134(.269)[.242]	.230(.459)[.459]
	250	.176(.131)[.108]	.229(.116)[.114]{.109}	.199(.128)[.117]	.231(.219)[.218]
	500	.200(.086)[.070]	.239(.075)[.074]{.071}	.222(.080)[.075]	.234(.113)[.112]
	1000	.215(.062)[.052]	.246(.055)[.055]{.051}	.238(.057)[.055]	.239(.062)[.061]
.00	100	-.144(.254)[.209]	-.064(.257)[.249]{.241}	-.154(.314)[.273]	-.029(.573)[.573]
	250	-.052(.127)[.116]	-.017(.132)[.131]{.129}	-.054(.146)[.136]	-.015(.267)[.266]
	500	-.032(.091)[.085]	-.014(.098)[.097]{.090}	-.036(.105)[.099]	-.024(.119)[.116]
	1000	-.018(.063)[.060]	-.009(.069)[.069]{.065}	-.020(.072)[.069]	-.014(.082)[.081]
-.25	100	-.354(.235)[.211]	-.311(.283)[.276]{.265}	-.423(.348)[.302]	-.320(.534)[.529]
	250	-.264(.131)[.130]	-.268(.164)[.163]{.159}	-.312(.180)[.168]	-.278(.271)[.269]
	500	-.241(.090)[.089]	-.260(.110)[.109]{.106}	-.286(.117)[.111]	-.269(.136)[.135]
	1000	-.228(.067)[.064]	-.257(.078)[.078]{.077}	-.270(.081)[.078]	-.266(.092)[.091]
-.50	100	-.543(.218)[.214]	-.559(.308)[.302]{.296}	-.696(.424)[.376]	-.621(.616)[.604]
	250	-.468(.138)[.135]	-.532(.186)[.183]{.179}	-.583(.203)[.186]	-.563(.248)[.240]
	500	-.444(.113)[.098]	-.520(.129)[.128]{.122}	-.549(.138)[.129]	-.538(.161)[.156]
	1000	-.420(.104)[.066]	-.503(.086)[.086]{.087}	-.517(.088)[.086]	-.512(.101)[.100]
			DGP 3: $\beta_0 = (.3, .1, .1)'$		
.50	100	.378(.186)[.140]	.470(.131)[.127]{.114}	.439(.225)[.217]	.575(.428)[.421]
	250	.434(.100)[.076]	.487(.071)[.070]{.074}	.467(.080)[.073]	.536(.255)[.253]
	500	.450(.074)[.055]	.492(.052)[.051]{.049}	.481(.056)[.052]	.539(.229)[.226]
	1000	.460(.055)[.037]	.497(.035)[.035]{.033}	.491(.036)[.035]	.523(.168)[.167]
.25	100	.109(.217)[.165]	.210(.173)[.168]{.160}	.151(.252)[.232]	.286(.518)[.517]
	250	.183(.120)[.099]	.235(.103)[.102]{.099}	.207(.114)[.106]	.310(.398)[.394]
	500	.205(.081)[.067]	.243(.069)[.069]{.066}	.227(.074)[.070]	.287(.286)[.284]
	1000	.215(.058)[.046]	.246(.048)[.048]{.047}	.237(.050)[.048]	.265(.179)[.179]
.00	100	-.144(.241)[.194]	-.063(.235)[.227]{.199}	-.144(.329)[.296]	.056(.696)[.694]
	250	-.051(.123)[.112]	-.018(.130)[.129]{.119}	-.054(.144)[.133]	.094(.551)[.543]
	500	-.027(.084)[.079]	-.008(.091)[.090]{.089}	-.030(.098)[.093]	.032(.337)[.336]
	1000	-.015(.058)[.056]	-.006(.065)[.064]{.061}	-.017(.067)[.065]	.020(.210)[.209]
-.25	100	-.355(.231)[.205]	-.313(.273)[.265]{.250}	-.432(.357)[.307]	-.193(.780)[.778]
	250	-.267(.129)[.128]	-.272(.162)[.160]{.151}	-.317(.180)[.167]	-.183(.540)[.536]
	500	-.240(.087)[.086]	-.259(.106)[.106]{.100}	-.285(.114)[.108]	-.202(.376)[.373]
	1000	-.224(.068)[.063]	-.254(.075)[.075]{.073}	-.267(.078)[.076]	-.213(.253)[.251]
-.50	100	-.544(.209)[.204]	-.557(.290)[.284]{.279}	-.684(.447)[.407]	-.442(.904)[.903]
	250	-.467(.139)[.135]	-.526(.179)[.177]{.168}	-.577(.196)[.180]	-.464(.523)[.522]
	500	-.433(.119)[.099]	-.506(.123)[.123]{.119}	-.535(.130)[.125]	-.412(.483)[.475]
	1000	-.423(.107)[.074]	-.504(.086)[.086]{.083}	-.519(.088)[.086]	-.466(.257)[.255]

Table 5.4: Empirical Mean(rmse)[sd]{$\hat{\text{sd}}$} of Estimators of β for SAR Model
Cases of Consistent QML estimators

λ_0	n	β_0	QML	ACQS	RGMM	ORGMM
DGP 1: Constant Circular Neighbours (REG-1), $\beta_0 = (3, 1, 1)'$						
.5	100	3	3.220(.644)[.606]{.592}	3.166(.708)[.688]{.691}	3.192(.733)[.707]	3.129(.797)[.786]
		1	1.006(.131)[.131]{.123}	0.992(.153)[.152]{.143}	0.989(.152)[.152]	0.988(.152)[.152]
		1	1.003(.201)[.201]{.203}	0.990(.229)[.228]{.222}	0.983(.228)[.228]	0.981(.229)[.229]
	250	3	3.089(.396)[.386]{.392}	3.051(.388)[.385]{.369}	3.069(.395)[.389]	3.040(.437)[.435]
		1	0.999(.096)[.096]{.093}	0.999(.096)[.096]{.093}	0.996(.096)[.096]	0.996(.096)[.096]
		1	1.003(.138)[.138]{.134}	1.004(.149)[.149]{.144}	1.002(.149)[.149]	1.002(.149)[.149]
	500	3	3.039(.264)[.261]{.276}	3.019(.261)[.260]{.253}	3.030(.264)[.263]	3.013(.290)[.290]
		1	1.000(.068)[.068]{.068}	0.996(.070)[.070]{.070}	0.995(.070)[.070]	0.995(.070)[.070]
		1	0.999(.106)[.106]{.104}	0.998(.106)[.106]{.104}	0.997(.106)[.106]	0.997(.106)[.106]
-.5	100	3	3.047(.357)[.353]{.360}	3.011(.356)[.355]{.339}	3.041(.362)[.360]	3.024(.400)[.399]
		1	0.994(.130)[.130]{.123}	0.994(.157)[.157]{.149}	0.988(.157)[.157]	0.988(.158)[.158]
		1	0.995(.226)[.226]{.222}	0.996(.227)[.227]{.222}	0.988(.226)[.226]	0.987(.227)[.227]
	250	3	3.026(.221)[.220]{.230}	3.011(.220)[.220]{.214}	3.024(.221)[.220]	3.016(.246)[.245]
		1	0.999(.098)[.098]{.100}	0.995(.093)[.093]{.094}	0.992(.094)[.093]	0.992(.094)[.093]
		1	1.002(.130)[.130]{.135}	0.992(.143)[.143]{.144}	0.989(.143)[.143]	0.990(.144)[.143]
	500	3	3.001(.157)[.157]{.166}	2.993(.158)[.157]{.152}	3.000(.158)[.158]	2.993(.174)[.174]
		1	0.998(.067)[.067]{.068}	0.998(.067)[.067]{.070}	0.997(.067)[.067]	0.997(.067)[.067]
		1	0.999(.104)[.104]{.103}	0.999(.104)[.104]{.103}	0.997(.104)[.104]	0.998(.104)[.104]
DGP 2: Constant Circular Neighbours (REG-1), $\beta_0 = (3, 1, 1)'$						
.5	100	3	3.207(.597)[.560]{.570}	3.117(.641)[.631]{.645}	3.150(.706)[.690]	3.071(.843)[.840]
		1	1.007(.154)[.154]{.148}	1.007(.154)[.154]{.148}	1.003(.154)[.154]	1.003(.151)[.151]
		1	1.000(.207)[.207]{.198}	0.999(.220)[.220]{.211}	0.993(.220)[.220]	0.991(.217)[.217]
	250	3	3.078(.380)[.372]{.345}	3.041(.372)[.370]{.345}	3.057(.377)[.372]	3.029(.512)[.512]
		1	1.004(.096)[.096]{.092}	1.004(.096)[.096]{.092}	1.001(.095)[.095]	1.001(.095)[.095]
		1	0.993(.141)[.141]{.132}	1.010(.146)[.146]{.141}	1.007(.146)[.146]	1.007(.145)[.145]
	500	3	3.028(.254)[.253]{.229}	3.009(.252)[.252]{.245}	3.020(.254)[.253]	2.998(.357)[.357]
		1	1.001(.067)[.067]{.068}	0.996(.071)[.070]{.069}	0.995(.071)[.070]	0.995(.070)[.070]
		1	0.999(.100)[.100]{.097}	1.002(.108)[.108]{.103}	1.001(.108)[.108]	1.000(.108)[.108]
-.5	100	3	3.044(.326)[.323]{.310}	3.010(.324)[.324]{.316}	3.039(.331)[.329]	3.021(.450)[.449]
		1	0.997(.154)[.154]{.141}	0.999(.154)[.154]{.140}	0.992(.154)[.153]	0.993(.153)[.153]
		1	0.999(.235)[.235]{.217}	1.000(.235)[.235]{.218}	0.992(.234)[.234]	0.990(.231)[.231]
	250	3	3.012(.205)[.205]{.201}	2.997(.205)[.205]{.206}	3.010(.206)[.206]	3.002(.281)[.281]
		1	1.000(.097)[.097]{.093}	1.001(.097)[.097]{.093}	0.998(.097)[.097]	0.999(.097)[.097]
		1	0.997(.147)[.147]{.141}	0.998(.147)[.147]{.142}	0.994(.147)[.147]	0.995(.146)[.145]
	500	3	3.010(.148)[.148]{.101}	3.002(.148)[.148]{.150}	3.009(.148)[.148]	3.002(.207)[.207]
		1	1.001(.070)[.070]{.067}	0.995(.069)[.069]{.069}	0.994(.069)[.069]	0.994(.069)[.069]
		1	1.000(.104)[.104]{.103}	1.000(.104)[.104]{.103}	0.998(.104)[.104]	0.999(.103)[.103]
DGP 3: Constant Circular Neighbours (REG-1), $\beta_0 = (3, 1, 1)'$						
0.5	100	3	3.203(.624)[.591]{.664}	3.106(.589)[.579]{.497}	3.149(.617)[.599]	3.120(.903)[.895]
		1	0.998(.156)[.156]{.130}	0.997(.155)[.155]{.130}	0.994(.155)[.155]	0.993(.150)[.150]
		1	0.995(.240)[.240]{.207}	0.994(.240)[.239]{.207}	0.989(.240)[.240]	0.989(.228)[.227]
	250	3	3.057(.348)[.343]{.478}	3.020(.341)[.340]{.323}	3.028(.468)[.467]	3.001(.534)[.534]
		1	1.000(.094)[.094]{.088}	1.000(.094)[.094]{.087}	0.998(.094)[.094]	0.998(.094)[.094]
		1	0.999(.137)[.137]{.133}	0.998(.137)[.137]{.133}	0.995(.137)[.137]	0.995(.135)[.135]
	500	3	3.042(.243)[.240]{.382}	3.023(.240)[.239]{.231}	3.032(.242)[.240]	3.031(.406)[.405]
		1	1.001(.072)[.072]{.067}	1.000(.072)[.072]{.067}	0.999(.072)[.072]	0.999(.072)[.072]
		1	1.002(.106)[.106]{.099}	1.002(.106)[.106]{.099}	1.000(.106)[.106]	1.000(.105)[.105]
-0.5	100	3	3.048(.335)[.331]{.492}	3.013(.330)[.330]{.288}	3.048(.340)[.336]	3.016(.497)[.497]
		1	0.996(.154)[.154]{.131}	0.998(.154)[.154]{.131}	0.992(.153)[.153]	0.993(.151)[.151]
		1	0.997(.223)[.223]{.194}	0.999(.223)[.223]{.196}	0.992(.223)[.223]	0.991(.221)[.221]
	250	3	3.027(.207)[.205]{.355}	3.013(.205)[.205]{.194}	3.025(.209)[.207]	3.028(.322)[.321]
		1	1.001(.093)[.093]{.088}	1.002(.093)[.093]{.088}	0.999(.093)[.093]	1.000(.093)[.093]
		1	0.986(.146)[.145]{.134}	0.986(.146)[.145]{.134}	0.983(.146)[.145]	0.983(.144)[.143]
	500	3	3.007(.145)[.144]{.286}	3.000(.144)[.144]{.141}	3.006(.145)[.145]	2.997(.227)[.227]
		1	1.000(.072)[.072]{.068}	1.000(.072)[.072]{.068}	0.999(.072)[.072]	0.999(.072)[.072]
		1	1.001(.104)[.104]{.099}	1.001(.104)[.104]{.099}	0.999(.104)[.104]	1.000(.103)[.103]

Table 5.4: Cont'd

λ_0	n	β_0	QML	ACQS	RGMM	ORGMM
			DGP 1: Queen Contiguity (REG-1), $\beta_0 = (.3, .1, .1)'$			
.5	100	.3	.338(.154)[.149]{.139}	.323(.146)[.145]{.137}	.328(.154)[.151]	.306(.167)[.167]
		.1	.094(.163)[.163]{.159}	.094(.163)[.163]{.169}	.093(.162)[.162]	.092(.165)[.165]
		.1	.100(.204)[.204]{.195}	.100(.204)[.204]{.195}	.099(.202)[.202]	.100(.202)[.202]
	250	.3	.310(.082)[.081]{.082}	.303(.080)[.080]{.079}	.307(.081)[.081]	.300(.081)[.081]
		.1	.109(.096)[.096]{.096}	.109(.096)[.096]{.096}	.109(.096)[.095]	.108(.096)[.096]
		.1	.101(.139)[.139]{.134}	.096(.141)[.141]{.139}	.096(.141)[.141]	.096(.140)[.140]
	500	.3	.308(.060)[.059]{.059}	.304(.059)[.058]{.056}	.306(.064)[.064]	.302(.064)[.064]
		.1	.101(.067)[.067]{.068}	.101(.067)[.067]{.068}	.101(.067)[.067]	.100(.067)[.067]
		.1	.102(.100)[.100]{.098}	.102(.100)[.100]{.098}	.102(.100)[.100]	.101(.100)[.100]
-.5	100	.3	.306(.109)[.109]{.106}	.301(.108)[.108]{.104}	.305(.110)[.109]	.304(.110)[.110]
		.1	.100(.167)[.167]{.157}	.100(.168)[.168]{.159}	.099(.166)[.166]	.097(.168)[.168]
		.1	.087(.195)[.194]{.185}	.084(.199)[.198]{.189}	.082(.196)[.195]	.082(.196)[.195]
	250	.3	.303(.069)[.069]{.069}	.303(.069)[.069]{.068}	.305(.069)[.069]	.306(.069)[.069]
		.1	.097(.099)[.098]{.095}	.107(.100)[.100]{.095}	.106(.100)[.100]	.106(.100)[.099]
		.1	.096(.138)[.138]{.134}	.106(.138)[.138]{.133}	.105(.138)[.138]	.105(.138)[.138]
	500	.3	.301(.048)[.048]{.048}	.297(.049)[.049]{.048}	.298(.049)[.049]	.298(.049)[.049]
		.1	.100(.069)[.069]{.067}	.101(.069)[.069]{.067}	.101(.069)[.069]	.101(.069)[.069]
		.1	.100(.097)[.097]{.098}	.100(.097)[.097]{.098}	.100(.097)[.097]	.100(.097)[.097]
			DGP 2: Queen Contiguity (REG-1), $\beta_0 = (.3, .1, .1)'$			
.5	100	.3	.327(.136)[.133]{.128}	.311(.129)[.129]{.120}	.318(.134)[.133]	.251(.234)[.229]
		.1	.103(.161)[.161]{.153}	.103(.161)[.161]{.152}	.103(.161)[.161]	.102(.161)[.161]
		.1	.103(.194)[.194]{.189}	.094(.194)[.194]{.180}	.092(.193)[.193]	.093(.192)[.192]
	250	.3	.311(.080)[.079]{.087}	.304(.078)[.078]{.078}	.308(.079)[.079]	.280(.111)[.110]
		.1	.104(.095)[.095]{.093}	.108(.097)[.097]{.093}	.107(.097)[.097]	.106(.095)[.095]
		.1	.096(.130)[.130]{.132}	.096(.130)[.130]{.132}	.096(.129)[.129]	.096(.129)[.129]
	500	.3	.307(.057)[.057]{.064}	.305(.058)[.058]{.056}	.306(.064)[.063]	.292(.070)[.069]
		.1	.101(.069)[.069]{.067}	.101(.069)[.069]{.067}	.101(.069)[.069]	.100(.068)[.068]
		.1	.104(.102)[.102]{.098}	.094(.101)[.101]{.098}	.094(.101)[.101]	.092(.100)[.099]
-.5	100	.3	.306(.109)[.109]{.110}	.301(.108)[.108]{.103}	.306(.109)[.109]	.304(.111)[.111]
		.1	.104(.171)[.171]{.162}	.104(.172)[.172]{.164}	.103(.170)[.170]	.103(.159)[.159]
		.1	.101(.194)[.194]{.181}	.089(.194)[.194]{.181}	.088(.192)[.191]	.084(.181)[.180]
	250	.3	.300(.069)[.069]{.072}	.302(.067)[.067]{.066}	.304(.067)[.067]	.303(.070)[.070]
		.1	.103(.095)[.095]{.093}	.103(.095)[.095]{.093}	.102(.095)[.094]	.101(.092)[.092]
		.1	.101(.133)[.133]{.132}	.095(.138)[.138]{.130}	.094(.138)[.138]	.093(.133)[.133]
	500	.3	.299(.048)[.048]{.051}	.298(.048)[.048]{.048}	.299(.048)[.048]	.299(.049)[.049]
		.1	.102(.067)[.067]{.068}	.102(.067)[.067]{.068}	.101(.067)[.067]	.100(.066)[.066]
		.1	.099(.099)[.099]{.096}	.103(.103)[.103]{.098}	.103(.102)[.102]	.103(.101)[.101]
			DGP 3: Queen Contiguity (REG-1), $\beta_0 = (.3, .1, .1)'$			
.5	100	.3	.334(.136)[.131]{.089}	.318(.128)[.127]{.115}	.325(.138)[.136]	.211(.315)[.302]
		.1	.102(.164)[.164]{.146}	.102(.164)[.164]{.145}	.101(.164)[.164]	.103(.166)[.166]
		.1	.105(.214)[.214]{.178}	.105(.213)[.213]{.178}	.104(.212)[.212]	.104(.209)[.209]
	250	.3	.311(.079)[.078]{.060}	.304(.077)[.076]{.073}	.307(.078)[.078]	.239(.189)[.179]
		.1	.104(.100)[.100]{.088}	.104(.100)[.100]{.089}	.103(.100)[.100]	.101(.095)[.095]
		.1	.103(.140)[.140]{.127}	.103(.140)[.140]{.127}	.103(.140)[.140]	.105(.137)[.136]
	500	.3	.306(.056)[.055]{.046}	.302(.055)[.055]{.053}	.304(.055)[.055]	.258(.136)[.130]
		.1	.102(.068)[.068]{.064}	.102(.068)[.068]{.064}	.102(.068)[.068]	.101(.067)[.067]
		.1	.097(.099)[.099]{.093}	.097(.099)[.099]{.094}	.096(.099)[.099]	.094(.096)[.096]
-.5	100	.3	.308(.112)[.111]{.137}	.303(.110)[.110]{.098}	.307(.111)[.111]	.282(.118)[.117]
		.1	.101(.163)[.163]{.142}	.102(.163)[.163]{.144}	.101(.162)[.162]	.096(.144)[.144]
		.1	.096(.204)[.204]{.170}	.096(.205)[.204]{.173}	.094(.203)[.202]	.089(.179)[.178]
	250	.3	.305(.069)[.068]{.094}	.303(.068)[.068]{.065}	.305(.068)[.068]	.291(.078)[.078]
		.1	.094(.097)[.097]{.089}	.094(.097)[.097]{.090}	.094(.096)[.096]	.093(.090)[.089]
		.1	.102(.136)[.136]{.124}	.102(.136)[.136]{.124}	.102(.136)[.136]	.102(.128)[.128]
	500	.3	.302(.049)[.049]{.071}	.301(.049)[.049]{.047}	.302(.049)[.049]	.293(.057)[.057]
		.1	.102(.068)[.068]{.065}	.102(.068)[.068]{.065}	.102(.067)[.067]	.102(.065)[.065]
		.1	.096(.093)[.093]{.094}	.096(.093)[.093]{.094}	.096(.093)[.093]	.093(.089)[.089]

Table 5.5: Empirical Mean(rmse)[sd]{$\hat{\text{sd}}$} of Estimators of β for SAR Model
Case I of Inconsistent QML estimators: Circular Neighbours (REG-1)

λ_0	n	β_0	QML	ACQS	RGMM	ORGMM
			DGP 1: $\beta_0 = (3, 1, 1)'$			
.5	100	3	3.398(.598)[.719]	3.116(.596)[.607]{.594}	3.145(.641)[.624]	3.104(.679)[.671]
		1	1.001(.125)[.125]	0.997(.125)[.125]{.118}	0.993(.125)[.125]	0.993(.126)[.126]
		1	0.999(.190)[.190]	0.992(.189)[.189]{.188}	0.986(.188)[.187]	0.987(.187)[.187]
	250	3	3.254(.346)[.429]	3.055(.350)[.355]{.349}	3.067(.351)[.345]	3.048(.370)[.367]
		1	1.001(.076)[.076]	0.998(.076)[.076]{.073}	0.997(.076)[.076]	0.997(.076)[.076]
		1	1.011(.125)[.125]	1.004(.124)[.124]{.119}	1.002(.124)[.124]	1.002(.124)[.124]
	500	3	3.219(.263)[.342]	3.024(.265)[.266]{.262}	3.030(.266)[.264]	3.021(.281)[.280]
		1	1.006(.054)[.055]	1.000(.054)[.054]{.056}	0.999(.054)[.054]	0.999(.055)[.055]
		1	1.008(.090)[.090]	1.002(.089)[.089]{.089}	1.001(.089)[.089]	1.001(.089)[.089]
-.5	100	3	2.897(.206)[.231]	2.986(.259)[.259]{.270}	2.981(.232)[.231]	2.993(.245)[.245]
		1	1.003(.127)[.127]	0.999(.127)[.127]{.120}	0.996(.127)[.127]	0.995(.127)[.127]
		1	1.014(.191)[.191]	1.003(.192)[.192]{.194}	0.996(.192)[.192]	0.993(.192)[.192]
	250	3	2.898(.134)[.169]	3.010(.177)[.177]{.166}	3.000(.146)[.146]	3.003(.154)[.154]
		1	1.005(.072)[.073]	0.996(.072)[.072]{.074}	0.995(.073)[.073]	0.995(.073)[.073]
		1	1.001(.122)[.122]	0.996(.121)[.121]{.119}	0.995(.121)[.121]	0.995(.121)[.121]
	500	3	2.887(.101)[.152]	3.011(.136)[.137]{.135}	3.009(.115)[.115]	3.011(.121)[.120]
		1	1.003(.055)[.055]	1.000(.055)[.055]{.055}	0.999(.055)[.055]	0.999(.055)[.055]
		1	1.002(.089)[.089]	0.995(.089)[.089]{.088}	0.994(.089)[.089]	0.993(.089)[.089]
			DGP 2: $\beta_0 = (3, 1, 1)'$			
.5	100	3	3.374(.572)[.683]	3.104(.568)[.578]{.563}	3.136(.623)[.607]	3.092(.767)[.762]
		1	1.009(.122)[.122]	1.005(.122)[.122]{.115}	1.001(.122)[.122]	1.001(.121)[.121]
		1	0.995(.193)[.193]	0.988(.192)[.193]{.182}	0.982(.192)[.191]	0.983(.192)[.191]
	250	3	3.229(.325)[.397]	3.030(.327)[.328]{.330}	3.045(.321)[.318]	3.021(.399)[.399]
		1	1.000(.073)[.073]	0.997(.073)[.073]{.072}	0.995(.074)[.073]	0.995(.074)[.074]
		1	1.013(.118)[.118]	1.006(.117)[.118]{.117}	1.004(.118)[.118]	1.005(.118)[.118]
	500	3	3.200(.261)[.329]	3.003(.262)[.262]{.259}	3.013(.265)[.264]	3.005(.343)[.343]
		1	1.007(.054)[.055]	1.001(.054)[.054]{.055}	1.000(.054)[.054]	1.000(.054)[.054]
		1	1.006(.089)[.089]	1.001(.088)[.088]{.087}	1.000(.088)[.088]	1.000(.088)[.088]
-.5	100	3	2.907(.209)[.229]	3.002(.260)[.260]{.273}	2.992(.239)[.239]	2.994(.265)[.265]
		1	0.997(.125)[.125]	0.993(.124)[.124]{.119}	0.990(.125)[.124]	0.991(.124)[.124]
		1	1.016(.198)[.199]	1.003(.199)[.199]{.195}	0.997(.200)[.200]	0.998(.199)[.199]
	250	3	2.892(.135)[.173]	3.000(.168)[.168]{.161}	2.995(.145)[.145]	2.996(.169)[.168]
		1	1.010(.075)[.076]	1.001(.075)[.075]{.072}	1.000(.076)[.076]	1.000(.076)[.076]
		1	0.996(.122)[.122]	0.991(.121)[.121]{.116}	0.989(.121)[.121]	0.990(.121)[.121]
	500	3	2.875(.101)[.161]	2.997(.133)[.133]{.129}	2.994(.113)[.113]	2.991(.137)[.137]
		1	1.007(.056)[.057]	1.004(.056)[.056]{.055}	1.003(.056)[.056]	1.003(.056)[.056]
		1	1.010(.090)[.090]	1.002(.090)[.090]{.088}	1.001(.090)[.090]	1.001(.090)[.090]
			DGP 3: $\beta_0 = (3, 1, 1)'$			
.5	100	3	3.330(.573)[.661]	3.087(.509)[.516]{.475}	3.110(.558)[.547]	3.035(.882)[.881]
		1	1.002(.123)[.123]	0.998(.123)[.123]{.109}	0.994(.122)[.122]	0.994(.119)[.118]
		1	1.011(.193)[.193]	1.005(.191)[.191]{.173}	1.001(.190)[.190]	1.003(.189)[.189]
	250	3	3.215(.347)[.408]	3.032(.321)[.323]{.305}	3.040(.329)[.326]	3.018(.440)[.439]
		1	1.003(.071)[.071]	1.001(.071)[.071]{.068}	0.999(.072)[.072]	0.999(.071)[.071]
		1	1.010(.121)[.121]	1.003(.120)[.120]{.111}	1.002(.120)[.120]	1.002(.120)[.120]
	500	3	3.194(.274)[.336]	3.016(.246)[.246]{.236}	3.019(.254)[.253]	3.015(.362)[.362]
		1	1.006(.055)[.055]	1.001(.055)[.055]{.052}	1.000(.055)[.055]	1.000(.055)[.055]
		1	1.004(.086)[.086]	0.999(.086)[.086]{.084}	0.998(.086)[.086]	0.998(.086)[.086]
-.5	100	3	2.898(.193)[.219]	2.971(.241)[.243]{.263}	2.969(.226)[.223]	2.964(.264)[.262]
		1	1.008(.122)[.122]	1.005(.121)[.121]{.110}	1.003(.121)[.121]	1.003(.120)[.120]
		1	1.013(.192)[.192]	1.004(.192)[.192]{.174}	0.998(.193)[.193]	1.000(.193)[.193]
	250	3	2.900(.131)[.165]	3.006(.174)[.174]{.156}	2.998(.146)[.146]	2.993(.184)[.184]
		1	1.003(.077)[.077]	0.995(.076)[.077]{.071}	0.994(.077)[.077]	0.995(.078)[.078]
		1	1.009(.123)[.124]	1.004(.123)[.123]{.114}	1.003(.123)[.123]	1.004(.122)[.122]
	500	3	2.879(.105)[.161]	2.999(.136)[.136]{.121}	2.996(.112)[.112]	2.998(.146)[.146]
		1	1.001(.058)[.058]	0.998(.058)[.058]{.053}	0.997(.059)[.059]	0.997(.059)[.058]
		1	1.006(.095)[.096]	0.998(.095)[.095]{.086}	0.997(.095)[.095]	0.997(.095)[.095]

Table 5.6: Empirical Mean(rmse)[sd]{$\hat{\text{sd}}$} of Estimators of β for SAR Model
Case II of Inconsistent QML estimators: Group Interaction (REG-2)

λ_0	n	β_0	QML	ACQS	RGMM	ORGMM
			DGP 1: $\beta_0 = (3, 1, 1)'$			
.5	100	3	3.493(.795)[.623]	3.146(.645)[.628]{.599}	3.207(.698)[.667]	3.196(.714)[.687]
		1	1.131(.253)[.217]	1.036(.221)[.218]{.205}	1.043(.237)[.233]	1.043(.239)[.235]
		1	1.096(.272)[.254]	1.015(.245)[.244]{.247}	1.019(.260)[.260]	1.019(.262)[.261]
	250	3	3.239(.423)[.348]	3.041(.358)[.355]{.349}	3.074(.375)[.367]	3.054(.397)[.394]
		1	1.059(.160)[.149]	1.008(.149)[.149]{.142}	1.012(.151)[.151]	1.008(.155)[.155]
		1	1.058(.160)[.149]	1.007(.149)[.149]{.139}	1.011(.152)[.151]	1.008(.155)[.155]
	500	3	3.173(.291)[.234]	3.017(.237)[.236]{.239}	3.038(.245)[.242]	3.027(.258)[.256]
		1	1.045(.101)[.090]	1.002(.090)[.090]{.091}	1.006(.091)[.091]	1.003(.093)[.093]
		1	1.045(.106)[.096]	1.004(.096)[.096]{.099}	1.008(.097)[.096]	1.005(.099)[.099]
-.5	100	3	3.070(.388)[.382]	3.075(.489)[.483]{.480}	3.104(.493)[.482]	3.097(.521)[.512]
		1	1.011(.168)[.168]	1.011(.183)[.182]{.202}	1.009(.190)[.190]	1.009(.194)[.194]
		1	1.019(.230)[.229]	1.020(.247)[.246]{.245}	1.016(.243)[.242]	1.015(.245)[.245]
	250	3	2.938(.251)[.243]	3.015(.308)[.307]{.301}	3.033(.296)[.294]	3.025(.312)[.310]
		1	0.980(.129)[.127]	0.997(.135)[.135]{.134}	0.998(.136)[.136]	0.997(.139)[.139]
		1	0.982(.127)[.125]	1.000(.134)[.134]{.131}	1.001(.134)[.134]	1.001(.136)[.136]
	500	3	2.918(.189)[.170]	3.013(.216)[.215]{.204}	3.023(.202)[.200]	3.017(.212)[.212]
		1	0.976(.082)[.078]	1.001(.087)[.087]{.083}	1.002(.083)[.083]	1.001(.085)[.085]
		1	0.976(.086)[.083]	1.000(.088)[.088]{.092}	1.001(.087)[.087]	0.999(.089)[.089]
			DGP 2: $\beta_0 = (3, 1, 1)'$			
.5	100	3	3.397(.746)[.631]	3.057(.622)[.620]{.654}	3.088(.693)[.688]	3.027(.786)[.786]
		1	1.106(.239)[.214]	1.012(.213)[.213]{.198}	1.009(.234)[.234]	0.998(.255)[.255]
		1	1.084(.277)[.264]	1.003(.252)[.252]{.239}	0.999(.275)[.275]	0.989(.285)[.285]
	250	3	3.211(.408)[.349]	3.006(.349)[.349]{.333}	3.036(.366)[.364]	2.979(.450)[.449]
		1	1.045(.152)[.146]	0.993(.145)[.144]{.141}	0.996(.148)[.148]	0.984(.165)[.165]
		1	1.046(.153)[.145]	0.993(.144)[.144]{.138}	0.997(.148)[.148]	0.984(.163)[.162]
	500	3	3.172(.287)[.229]	3.016(.230)[.230]{.235}	3.036(.238)[.235]	3.005(.303)[.303]
		1	1.049(.102)[.090]	1.005(.090)[.090]{.091}	1.009(.091)[.091]	1.001(.105)[.105]
		1	1.046(.110)[.100]	1.005(.101)[.101]{.099}	1.008(.101)[.101]	1.001(.112)[.112]
-.5	100	3	3.055(.397)[.394]	3.073(.520)[.515]{.508}	3.096(.508)[.499]	3.031(.598)[.597]
		1	1.016(.174)[.173]	1.020(.197)[.196]{.218}	1.019(.197)[.196]	1.004(.214)[.214]
		1	1.004(.225)[.225]	1.009(.246)[.246]{.260}	1.001(.241)[.241]	0.991(.248)[.248]
	250	3	2.939(.247)[.239]	3.018(.301)[.300]{.392}	3.031(.286)[.284]	2.992(.357)[.357]
		1	0.986(.128)[.127]	1.006(.136)[.136]{.133}	1.005(.137)[.137]	0.997(.149)[.148]
		1	0.986(.123)[.122]	1.005(.132)[.131]{.130}	1.006(.130)[.130]	0.997(.140)[.140]
	500	3	2.912(.195)[.174]	3.003(.216)[.216]{.200}	3.015(.206)[.205]	2.993(.253)[.253]
		1	0.976(.081)[.078]	1.000(.085)[.085]{.083}	1.002(.083)[.083]	0.996(.091)[.091]
		1	0.982(.090)[.088]	1.005(.093)[.093]{.092}	1.007(.094)[.093]	1.002(.100)[.100]
			DGP 3: $\beta_0 = (3, 1, 1)'$			
.5	100	3	3.430(.770)[.638]	3.111(.586)[.575]{.516}	3.166(.723)[.704]	3.101(.899)[.893]
		1	1.111(.242)[.215]	1.024(.206)[.205]{.188}	1.028(.245)[.244]	1.015(.277)[.277]
		1	1.081(.281)[.269]	1.009(.256)[.256]{.234}	1.010(.276)[.276]	0.998(.289)[.289]
	250	3	3.188(.390)[.342]	3.002(.321)[.321]{.309}	3.030(.343)[.342]	2.980(.468)[.467]
		1	1.044(.157)[.151]	0.997(.149)[.149]{.135}	0.999(.153)[.153]	0.988(.174)[.174]
		1	1.042(.148)[.143]	0.994(.139)[.139]{.132}	0.997(.143)[.143]	0.986(.163)[.162]
	500	3	3.164(.287)[.235]	3.016(.220)[.219]{.220}	3.038(.231)[.228]	3.023(.314)[.313]
		1	1.042(.101)[.092]	1.002(.087)[.087]{.087}	1.006(.090)[.090]	1.003(.108)[.108]
		1	1.045(.111)[.101]	1.007(.099)[.098]{.094}	1.010(.099)[.098]	1.007(.112)[.111]
-.5	100	3	3.076(.397)[.390]	3.066(.477)[.473]{.564}	3.122(.516)[.501]	3.061(.656)[.653]
		1	1.016(.172)[.171]	1.014(.187)[.186]{.194}	1.020(.194)[.193]	1.007(.214)[.214]
		1	0.996(.224)[.224]	0.994(.238)[.238]{.233}	0.993(.235)[.235]	0.982(.245)[.245]
	250	3	2.952(.245)[.240]	3.036(.319)[.317]{.285}	3.053(.304)[.299]	3.015(.407)[.407]
		1	0.986(.136)[.135]	1.005(.143)[.142]{.130}	1.006(.142)[.142]	0.998(.155)[.155]
		1	0.987(.122)[.121]	1.007(.133)[.132]{.127}	1.008(.132)[.131]	1.000(.147)[.147]
	500	3	2.916(.186)[.166]	3.011(.203)[.202]{.195}	3.022(.194)[.192]	3.002(.262)[.262]
		1	0.975(.083)[.079]	1.000(.084)[.084]{.082}	1.002(.083)[.083]	0.997(.095)[.095]
		1	0.976(.091)[.088]	1.000(.093)[.093]{.090}	1.001(.093)[.093]	0.997(.099)[.099]

where $\bar{G}_{1n}(\delta) = B_n(\rho)G_{1n}(\lambda)B_n^{-1}(\rho)$, $\bar{G}_{2n}(\rho) = G_{2n}(\rho)M_n(\rho)$, $G_{1n}(\lambda) = W_{1n}A_n^{-1}(\lambda)$, and $G_{2n}(\rho) = W_{2n}B_n^{-1}(\rho)$. Using similar arguments as given in Section 5.3, we have, after some algebraic manipulations, the following adjusted concentrated score function,

$$\tilde{\psi}_n^*(\delta) = \begin{cases} \dfrac{Y_n'(\delta)M_n(\rho)\bar{G}_{1n}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)M_n(\rho)Y_n(\delta)}, \\ \dfrac{Y_n'(\delta)M_n(\rho)\bar{G}_{2n}^{\circ}(\rho)Y_n(\delta)}{Y_n'(\delta)M_n(\rho)Y_n(\delta)}, \end{cases} \tag{5.24}$$

where $\bar{G}_{rn}^{\circ}(\delta) = \bar{G}_{rn}(\delta) - \text{diag}(M_n(\rho))^{-1}\text{diag}[M_n(\rho)\bar{G}_{rn}(\delta)], r = 1, 2$.

The ACQS estimator of δ is defined as $\tilde{\delta}_n = \arg\{\tilde{\psi}_n^*(\delta) = 0\}$, and the ACQS estimators of β and σ^2 are $\tilde{\beta}_n \equiv \hat{\beta}_n(\tilde{\delta}_n)$ and $\tilde{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\tilde{\delta}_n)$. To the best of our knowledge, the three-step estimator of Kelejian and Prucha (2010) may be the only heteroskedasticity robust estimator for the SARAR(1,1) model available in the literature.[14] Thus, it would be of a great interest to investigate and compare the finite sample properties of the three-step estimator and the proposed ACQS estimator estimator for the SARAR(1,1) model. For brevity, Table 5.7 presents a small set of Monte Carlo results that serve such purposes, and more results are available from the authors. Both the reported and unreported Monte Carlo results show that the proposed ACQS estimator has an excellent finite sample performance, and it outperforms the three-step estimator of Kelejian and Prucha (2010) from a combined consideration in terms of bias, consistency and efficiency.[15]

For heteroskedasticity robust inferences based on the SARAR(1,1) model, one needs the feasible heteroskedasticity robust estimators of the asymptotic variances of $\tilde{\delta}$ and $\tilde{\beta}_n$. Under an extended set of regularity conditions and using the mul-

[14] Arraiz, et al. (2010) provide some additional details for this estimator including some Monte Carlo results.

[15] A more rigorous comparison may be interesting but beyond the scope of this chapter. The robust GMM approach of Lin and Lee (2010) may lead to a more efficient estimator than does the three-step approach of Kelejian and Prucha (2010), but from Lin and Lee (2010) it is not clear how to extend their robust GMM estimation approach for the SAR to the general SARAR(1,1) model.

tivariate CLT for linear-quadratic forms of Kelejian and Prucha (2010, Appendix A), we can show that as $n \to \infty$,

$$\sqrt{n}(\tilde{\delta}_n - \delta_0) \xrightarrow{D} N\big(0, \lim_{n\to\infty} \tau_n^2(\tilde{\delta}_n)\big), \quad \text{and} \ \ \tau_n^2(\tilde{\delta}_n) = \Phi_n^{-1}\tau_n^2(\tilde{\psi}_n^*)\Phi_n^{-1}, \qquad (5.25)$$

where Φ_n equals to $-\mathrm{E}[\frac{\partial}{\partial \delta_0'}\tilde{\psi}^*(\delta_0)]$ or its first-order term, and $\tau_n^2(\tilde{\psi}_n^*)$ is the first-order terms of $\mathrm{Var}[\sqrt{n}\tilde{\psi}^*(\delta_0)]$. Both Φ_n and $\tau_n^2(\tilde{\psi}_n^*)$ possess analytical expressions but are not needed for practical applications as the former can be estimated consistently by $\tilde{\Phi}_n = -\frac{\partial}{\partial \delta_0'}\tilde{\psi}^*(\delta_0)|_{\delta_0=\tilde{\delta}_n}$, and the latter by the following OPG estimator:

$$\tilde{\tau}_n^2(\tilde{\psi}_n^*) = \textstyle\sum_{i=1}^n \tilde{\epsilon}_{n,i}^2 \tilde{\Upsilon}_{n,i}\tilde{\Upsilon}_{n,i}', \qquad (5.26)$$

where $\tilde{\Upsilon}_{n,i} = (\tilde{\zeta}_{1n,i}+\tilde{p}_{1n,ii}\tilde{\epsilon}_{n,i}+\tilde{c}_{1n,i},\ \tilde{\zeta}_{2n,i}+\tilde{p}_{2n,ii}\tilde{\epsilon}_{n,i}+\tilde{c}_{2n,i})'$, $\tilde{\zeta}_{rn} = (P_{rn}^{u\prime}+P_{rn}^{l})\tilde{\epsilon}_n$, $r = 1, 2$, $\tilde{\epsilon}_n = Y(\tilde{\delta}_n) - B_n(\tilde{\rho})X_n\tilde{\beta}_n$, and P_{rn} and c_{rn} are defined in the following asymptotic representation:

$$\sqrt{n}\tilde{\psi}_n^* = \begin{cases} \frac{1}{\sqrt{n}\sigma_0^2}\big(\epsilon_n' P_{1n}\epsilon_n + c_{1n}'\epsilon_n\big) + o_p(1), \\ \frac{1}{\sqrt{n}\sigma_0^2}\big(\epsilon_n' P_{2n}\epsilon_n + c_{2n}'\epsilon_n\big) + o_p(1), \end{cases} \qquad (5.27)$$

where $P_{rn} = M_n\bar{G}_{rn}^{\circ}$ and $c_{rn} = M_n\bar{G}_{rn}^{\circ}B_nX_n\beta_0$, $r = 1, 2$, with $p_{rn,ii}$, P_{rn}^u and P_{rn}^l denoting, respectively, the diagonal elements, the upper and lower triangular matrices of P_{rn}.

With the asymptotic results for $\tilde{\delta}_n$, one can easily derive the asymptotic results for $\tilde{\beta}_n$. Under a similar set of regularity conditions, we can show that as $n \to \infty$,

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) \xrightarrow{D} N\big(0, \lim_{n\to\infty}(X_n'B_n'B_nX_n)^{-1}X_n'B_n'\mathbb{A}_nB_nX_n(X_n'B_n'B_nX_n)^{-1}\big), \qquad (5.28)$$

where,

$\mathbb{A}_n = n\sigma_0^2 H_n + \tau_{n,11}^2(\tilde{\delta}_n)\eta_n\eta_n' + 2\sqrt{n}(\sigma_0^{-2}P_{1n}^d s_n + H_n c_{1n}, \sigma_0^{-2}P_{2n}^d s_n + H_n c_{2n})\Phi_n^{-1}(\eta_n, 0_n)'$, $s_n = \mathrm{E}(\epsilon_n^3)$, $P_{rn}^d = \mathrm{diag}(P_{rn})$, $\eta_n = B_n G_{1n} X_n \beta_0$, $\tau_{n,11}^2(\tilde{\delta}_n)$ is the top-right corner element of $\tau_n^2(\tilde{\delta}_n)$, and 0_n is an $n \times 1$ vector of 0's. With the estimates $\tilde{\Phi}_n$ and $\tilde{\tau}_n^2(\tilde{\psi}_n^*)$ defined above, the estimates $\tilde{s}_n = \tilde{\epsilon}_n^3$ and $\tilde{H}_n = \tilde{\sigma}_n^{-2}\mathrm{diag}(\tilde{\epsilon}_n^2)$ of s_n and H_n, and the plug-in estimates for the remaining quantities, a consistent estimate for $\tau_n^2(\tilde{\beta}_n^*)$ follows.

The proposed methods can be further extended. For example, the SARAR(p,q), which contains spatial lags of order p and spatial autoregressive errors of order q, can be dealt with in a similar manner as for the SARAR(1,1) model. To have an idea on how our methods can be extended to the SARAR(p,q) model, note that the Gaussian likelihood takes an identical form as (5.23) for SARAR(1,1), except that now $A_n(\lambda) = I_n - \sum_{j=1}^p \lambda_j W_{1,jn}$ and $B_n(\rho) = I_n - \sum_{j=1}^q \rho_j W_{2,jn}$, $\lambda = \{\lambda_1, \ldots, \lambda_p\}$ and $\rho = \{\rho_1, \ldots, \rho_q\}$, see Lee and Liu (2010). Thus, the concentrated scores and their adjustment can be found in a similar manner, resulting ACQS estimators for the SARAR(p,q) model that are robust against unknown heteroskedasticity.[16] Moving further, our methods can be applied to give heteroskedasticity robust estimator for the fixed effects spatial panel data model. As argued in the introduction, heteroskedasticity is common particularly in spatial models. This makes it more desirable to develop heteroskedasticity robust inference methods for these models. The methods proposed in this chapter shed much light on these intriguing research problems. However, formal studies on these models, including detailed proofs of the results (5.25)-(5.28) and the proofs of consistency of the variance estimates therein, are beyond the scope of this chapter, and will be pursued in future research.

[16]Lee and Liu (2010) proposed efficient GMM estimation of this model under homoskedasticity assumption. Badinger and Egger (2011) extend the estimation strategy of Kelejian and Prucha (2010) to give a heteroskedasticity robust three-step estimator of the SARAR(p,q) model, where some Monte Carlo results are presented under a SARAR(3,3) model and some special spatial weight matrices.

Table 5.7: Empirical Mean(rmse)[sd] of Estimators of λ and ρ for SARAR(1,1) Model
Case I of Inconsistent QML estimators: Circular Neighbours (REG-1)

Par	QML-λ	ACQS-λ	KP-λ	QML-ρ	ACQS-ρ	KP-ρ
			`DGP 1:` $\beta_0 = (3, 1, 1)'$			
1-1	.470(.141)[.138]	.472(.197)[.195]	.578(.219)[.204]	.409(.195)[.172]	.446(.237)[.231]	.335(.341)[.299]
	.484(.080)[.078]	.482(.118)[.117]	.528(.109)[.105]	.445(.116)[.102]	.488(.140)[.139]	.479(.180)[.179]
	.487(.065)[.064]	.489(.097)[.097]	.515(.093)[.092]	.454(.088)[.075]	.491(.110)[.109]	.512(.156)[.156]
	.490(.043)[.042]	.495(.060)[.059]	.505(.057)[.057]	.458(.066)[.051]	.497(.070)[.070]	.533(.103)[.097]
1-2	.372(.173)[.116]	.418(.233)[.218]	.494(.143)[.143]	-.307(.249)[.158]	-.505(.252)[.239]	-.507(.244)[.244]
	.411(.109)[.063]	.488(.095)[.094]	.501(.072)[.072]	-.324(.202)[.100]	-.502(.153)[.153]	-.492(.150)[.150]
	.400(.112)[.050]	.498(.071)[.071]	.498(.060)[.060]	-.305(.208)[.072]	-.504(.126)[.125]	-.476(.121)[.119]
	.421(.084)[.030]	.502(.047)[.047]	.499(.035)[.035]	-.321(.186)[.051]	-.506(.109)[.108]	-.470(.083)[.078]
2-1	.280(.144)[.141]	.250(.200)[.200]	.333(.239)[.224]	.374(.208)[.165]	.441(.225)[.217]	.358(.307)[.272]
	.292(.095)[.086]	.253(.128)[.127]	.297(.133)[.124]	.399(.140)[.097]	.470(.135)[.131]	.464(.176)[.172]
	.293(.080)[.067]	.252(.106)[.106]	.276(.105)[.101]	.408(.119)[.075]	.491(.109)[.107]	.499(.146)[.146]
	.287(.057)[.043]	.250(.064)[.064]	.259(.064)[.064]	.421(.093)[.049]	.494(.065)[.065]	.524(.092)[.089]
2-2	.113(.189)[.130]	.233(.188)[.163]	.235(.186)[.186]	-.330(.231)[.156]	-.582(.269)[.249]	-.507(.259)[.259]
	.156(.120)[.074]	.239(.131)[.131]	.248(.092)[.092]	-.337(.188)[.095]	-.503(.209)[.209]	-.484(.151)[.150]
	.140(.125)[.059]	.248(.099)[.099]	.247(.079)[.079]	-.319(.193)[.069]	-.510(.115)[.114]	-.484(.117)[.116]
	.164(.093)[.036]	.250(.052)[.052]	.250(.045)[.045]	-.332(.175)[.047]	-.501(.102)[.101]	-.475(.080)[.076]
3-1	.082(.168)[.147]	.015(.210)[.209]	.080(.236)[.222]	.335(.239)[.172]	.428(.241)[.230]	.367(.292)[.260]
	.090(.124)[.086]	.012(.126)[.125]	.047(.131)[.123]	.373(.160)[.097]	.472(.127)[.124]	.467(.165)[.161]
	.094(.116)[.067]	.006(.099)[.099]	.025(.103)[.100]	.380(.140)[.072]	.495(.093)[.091]	.502(.131)[.131]
	.082(.093)[.043]	.001(.062)[.062]	.009(.064)[.063]	.397(.114)[.048]	.496(.059)[.059]	.526(.088)[.083]
3-2	-.104(.163)[.125]	-.027(.171)[.148]	-.027(.196)[.194]	-.353(.208)[.147]	-.488(.208)[.187]	-.485(.250)[.250]
	-.078(.109)[.076]	-.023(.152)[.150]	-.006(.108)[.108]	-.356(.170)[.091]	-.489(.117)[.116]	-.481(.148)[.147]
	-.086(.106)[.062]	.000(.123)[.123]	-.001(.096)[.096]	-.343(.170)[.065]	-.501(.102)[.102]	-.478(.120)[.117]
	-.071(.081)[.040]	.001(.060)[.059]	-.001(.055)[.055]	-.350(.156)[.045]	-.502(.106)[.104]	-.473(.081)[.077]
4-1	-.126(.183)[.135]	-.219(.194)[.192]	-.189(.210)[.201]	.323(.246)[.170]	.430(.239)[.228]	.395(.258)[.235]
	-.132(.144)[.082]	-.240(.106)[.105]	-.224(.117)[.114]	.363(.169)[.099]	.478(.112)[.109]	.485(.150)[.149]
	-.119(.148)[.068]	-.247(.090)[.090]	-.236(.095)[.093]	.365(.155)[.075]	.490(.085)[.085]	.510(.118)[.118]
	-.131(.127)[.043]	-.247(.057)[.057]	-.242(.055)[.055]	.376(.134)[.049]	.492(.056)[.055]	.520(.079)[.076]
4-2	-.303(.130)[.119]	-.300(.217)[.215]	-.279(.208)[.206]	-.395(.168)[.131]	-.484(.224)[.228]	-.488(.221)[.221]
	-.288(.084)[.075]	-.272(.155)[.154]	-.260(.122)[.121]	-.384(.145)[.086]	-.487(.200)[.199]	-.475(.152)[.150]
	-.289(.069)[.057]	-.249(.106)[.106]	-.255(.098)[.098]	-.378(.137)[.061]	-.508(.101)[.100]	-.472(.115)[.112]
	-.284(.050)[.037]	-.244(.056)[.056]	-.253(.058)[.058]	-.381(.126)[.043]	-.506(.104)[.103]	-.471(.081)[.076]
5-1	-.357(.192)[.128]	-.458(.165)[.160]	-.449(.169)[.161]	.320(.244)[.164]	.438(.201)[.191]	.413(.215)[.197]
	-.373(.146)[.071]	-.491(.082)[.082]	-.481(.088)[.086]	.362(.169)[.097]	.483(.101)[.100]	.488(.126)[.126]
	-.352(.159)[.057]	-.496(.068)[.068]	-.493(.073)[.073]	.357(.160)[.072]	.491(.074)[.074]	.509(.097)[.097]
	-.374(.131)[.037]	-.499(.041)[.041]	-.498(.045)[.045]	.377(.132)[.047]	.497(.047)[.047]	.526(.069)[.064]
5-2	-.478(.104)[.101]	-.523(.180)[.179]	-.518(.189)[.188]	-.437(.140)[.125]	-.490(.215)[.214]	-.491(.217)[.217]
	-.480(.069)[.066]	-.513(.128)[.126]	-.511(.111)[.111]	-.424(.113)[.084]	-.491(.146)[.145]	-.474(.142)[.140]
	-.472(.057)[.050]	-.498(.109)[.109]	-.501(.093)[.093]	-.429(.092)[.059]	-.507(.107)[.106]	-.474(.113)[.110]
	-.478(.040)[.033]	-.499(.054)[.053]	-.502(.056)[.056]	-.424(.086)[.042]	-.500(.077)[.076]	-.470(.078)[.073]

Note: (i) The DGP used: $Y_n = \lambda W_n Y_n + \iota_n \beta_0 + X_{1n}\beta_1 + X_{2n}\beta_2 + \epsilon_n, \quad \epsilon_n = \rho W_n \epsilon_n + v_n$.
(ii) Par = i-j, where '$i = 1, 2, 3, 4, 5$' represents '$\lambda = .5, .25, 0, -.25, -.5$'; '$j = 1, 2$' represents '$\rho = .5, -.5$'.
Under each Par setting, $n = 100, 250, 500, 1000$, corresponding to the four rows.

(iii) KP denotes Kelejian and Prucha's (2010) three-step estimator.

Table 5.7: Cont'd

Par	λ_{QML}	λ_{MQML}	λ_{GS}	ρ_{QML}	ρ_{MQML}	ρ_{GS}
DGP 2: $\beta_0 = (3, 1, 1)'$						
1-1	.471(.138)[.135]	.473(.188)[.186]	.576(.216)[.202]	.404(.195)[.170]	.443(.235)[.228]	.329(.340)[.294]
	.487(.085)[.084]	.489(.123)[.122]	.529(.114)[.110]	.440(.119)[.103]	.477(.144)[.142]	.466(.180)[.177]
	.490(.064)[.063]	.494(.096)[.096]	.521(.092)[.089]	.454(.090)[.077]	.489(.113)[.113]	.513(.158)[.157]
	.493(.042)[.041]	.499(.059)[.059]	.508(.056)[.056]	.455(.068)[.051]	.493(.070)[.070]	.526(.099)[.095]
1-2	.380(.173)[.124]	.474(.153)[.138]	.495(.141)[.141]	-.318(.253)[.175]	-.498(.266)[.252]	-.509(.246)[.246]
	.408(.114)[.067]	.483(.101)[.100]	.495(.072)[.072]	-.320(.212)[.113]	-.492(.200)[.199]	-.483(.156)[.155]
	.399(.116)[.056]	.497(.063)[.063]	.497(.060)[.060]	-.303(.215)[.085]	-.504(.124)[.123]	-.476(.125)[.123]
	.422(.085)[.035]	.502(.038)[.038]	.500(.035)[.035]	-.324(.185)[.056]	-.506(.100)[.100]	-.473(.079)[.074]
2-1	.285(.143)[.139]	.262(.194)[.194]	.349(.241)[.220]	.370(.214)[.170]	.432(.232)[.222]	.345(.317)[.276]
	.280(.092)[.087]	.244(.126)[.126]	.282(.127)[.123]	.407(.136)[.099]	.478(.134)[.132]	.478(.177)[.176]
	.292(.081)[.069]	.253(.105)[.105]	.275(.106)[.103]	.411(.116)[.074]	.483(.103)[.102]	.502(.145)[.145]
	.285(.056)[.043]	.247(.064)[.064]	.257(.063)[.062]	.424(.091)[.049]	.497(.065)[.064]	.529(.096)[.092]
2-2	.120(.186)[.134]	.235(.198)[.175]	.227(.184)[.182]	-.337(.235)[.169]	-.490(.271)[.254]	-.504(.263)[.263]
	.157(.121)[.078]	.247(.109)[.108]	.249(.094)[.094]	-.336(.194)[.104]	-.492(.125)[.124]	-.485(.153)[.152]
	.141(.125)[.062]	.251(.101)[.101]	.247(.081)[.081]	-.318(.197)[.076]	-.502(.129)[.127]	-.479(.121)[.119]
	.163(.096)[.039]	.252(.043)[.043]	.248(.047)[.046]	-.334(.174)[.053]	-.507(.101)[.109]	-.474(.083)[.079]
3-1	.086(.171)[.148]	.033(.210)[.207]	.085(.234)[.218]	.339(.231)[.166]	.419(.235)[.221]	.371(.282)[.250]
	.082(.121)[.089]	.007(.124)[.124]	.037(.130)[.125]	.376(.161)[.103]	.472(.128)[.125]	.476(.166)[.164]
	.092(.114)[.068]	.005(.096)[.096]	.022(.102)[.100]	.382(.138)[.071]	.490(.088)[.087]	.502(.124)[.124]
	.081(.092)[.044]	.001(.061)[.061]	.009(.061)[.060]	.395(.116)[.050]	.493(.058)[.058]	.525(.088)[.085]
3-2	-.087(.156)[.129]	-.026(.205)[.201]	-.012(.199)[.199]	-.367(.196)[.144]	-.479(.211)[.197]	-.492(.238)[.238]
	-.078(.109)[.077]	-.021(.138)[.137]	-.007(.109)[.109]	-.356(.173)[.096]	-.487(.195)[.195]	-.479(.152)[.150]
	-.088(.108)[.062]	-.005(.106)[.106]	-.009(.092)[.091]	-.344(.171)[.071]	-.508(.115)[.115]	-.474(.120)[.117]
	-.070(.080)[.039]	.008(.057)[.057]	.000(.054)[.054]	-.352(.156)[.049]	-.508(.107)[.105]	-.472(.080)[.075]
4-1	-.132(.186)[.144]	-.214(.201)[.198]	-.185(.215)[.205]	.329(.238)[.165]	.428(.227)[.215]	.393(.245)[.221]
	-.132(.148)[.090]	-.237(.119)[.118]	-.219(.120)[.116]	.366(.170)[.105]	.477(.128)[.126]	.481(.150)[.149]
	-.119(.149)[.072]	-.246(.088)[.088]	-.236(.096)[.095]	.368(.153)[.076]	.491(.082)[.082]	.508(.114)[.114]
	-.133(.125)[.045]	-.248(.055)[.055]	-.244(.057)[.056]	.382(.129)[.051]	.497(.054)[.054]	.527(.079)[.074]
4-2	-.297(.132)[.123]	-.287(.206)[.205]	-.280(.207)[.205]	-.401(.175)[.144]	-.462(.232)[.235]	-.492(.233)[.233]
	-.288(.083)[.074]	-.278(.107)[.104]	-.261(.115)[.115]	-.387(.146)[.092]	-.479(.157)[.156]	-.480(.151)[.150]
	-.289(.071)[.060]	-.253(.100)[.100]	-.253(.102)[.102]	-.381(.135)[.064]	-.508(.117)[.117]	-.480(.118)[.116]
	-.283(.050)[.038]	-.252(.053)[.053]	-.254(.058)[.058]	-.381(.127)[.045]	-.509(.107)[.106]	-.470(.082)[.076]
5-1	-.373(.186)[.135]	-.472(.154)[.151]	-.459(.163)[.158]	.327(.249)[.180]	.445(.205)[.198]	.414(.223)[.206]
	-.380(.144)[.080]	-.492(.082)[.082]	-.485(.090)[.089]	.369(.167)[.103]	.484(.104)[.102]	.492(.127)[.127]
	-.355(.158)[.063]	-.498(.066)[.066]	-.491(.068)[.067]	.362(.156)[.074]	.495(.070)[.070]	.510(.093)[.092]
	-.375(.132)[.042]	-.498(.043)[.043]	-.498(.044)[.044]	.377(.133)[.051]	.495(.048)[.048]	.524(.068)[.063]
5-2	-.473(.120)[.117]	-.520(.191)[.190]	-.518(.198)[.198]	-.436(.156)[.143]	-.483(.226)[.226]	-.483(.229)[.229]
	-.476(.072)[.068]	-.505(.109)[.108]	-.507(.112)[.112]	-.427(.113)[.086]	-.485(.146)[.146]	-.477(.146)[.144]
	-.472(.059)[.052]	-.502(.107)[.107]	-.500(.096)[.096]	-.431(.093)[.062]	-.503(.107)[.107]	-.481(.109)[.107]
	-.480(.040)[.034]	-.498(.063)[.063]	-.507(.057)[.057]	-.422(.089)[.042]	-.505(.075)[.074]	-.469(.081)[.074]

5.6 Conclusion

This chapter looks at heteroskedasticity robust QML-type estimation for spatial autoregressive (SAR) models. We provide clear conditions for the regular QML estimator to be consistent even when the disturbances suffer from heteroskedasticity of unknown form. When these conditions are violated, the regular QML estimator becomes inconsistent and in this case we suggest a ACQS estimator by making a simple adjustment to the score function so that it becomes robust to unknown heteroskedasticity. This method is proven to work well in the simulation studies and was shown to be robust to many situations including, deteriorated signal strength as well as non-normal errors (besides the unknown heteroskedasticity). To provide inference methods robust to heteroskedasticity and non-normality, OPG-based estimators of the variances of QML estimator and ACQS estimator are introduced. Monte Carlo results show that the proposed ACQS estimator for the SAR model and the associated robust variance estimator work very well in finite samples.

The proposed methodology (adjusting score for achieving heteroskedasticity robustness for parameter estimation and finding a suitable OPG for achieving heteroskedasticity robustness for variance estimation) has a great potential to be extended to more general models, not necessarily the spatial models, thus paving a simple way for developing heteroskedasticity robust inference methods for applied researchers.

CHAPTER 6

A General Method for Heteroskedasticity Robust Inferences of Spatial Econometric Models

6.1 Introduction

The spatial econometric literature has come a long way in filling many theoretical gaps in handling estimation and inference related issues of spatial models.[1] Of the many challenges, estimation and inference in the presence heteroskedasticity is important since heteroskedasticity may occur more naturally in spatial models due to peer interaction as a result of unobservable heterogeneous characteristics of spatial units that does not interact with each other (Glaeser, 1996). Heteroskedasticity will be a conspicuous problem if this peer interaction is misspecified (Anselin, 1988b). The effect of heteroskedasticity in spatial models appeared as early as in Anselin (1988b) which provides tests for spatial dependence in the presence

[1] See Cliff and Ord (1972,1973, 1981), Ord (1975), Anselin (1988b, 2003), Anselin and Bera (1998), Le Sage and Pace (2009) for some pioneering theoretical work on spatial models.

of heteroskedasticity. However, estimation in the presence of heteroskedasticity was not considered until recent times.[2] Most of these available methodologies concentrate on specific spatial models and provide heteroskedasticity robust estimators for that particular spatial model. Of the existing heteroskedasticity robust estimators, the robust generalised method of moments (GMM) estimator based on quadratic moments given in Lin and Lee (2010) is popular due to its efficiency and ease of implementation. However, this estimator is given for a spatial autoregressive model with one lagged dependent variable (SLD model), may not readily carry over its desirable properties when more complex spatial dependencies are present. The robust three step estimator combining two stage least squares (2SLS) and GMM given in Kelejian and Prucha (2010) is also popular due to its simplicity and versatility as it has been widely extended to many complex spatial econometric models. However, this estimator, especially the 2SLS estimator may lack efficiency as its estimation does not consider the reduced form of the model but rather only the deterministic part.

In the previous chapter we introduce an adjusted concentrated quasi score (ACQS) estimator which has the advantage of ease of implementation and enjoys the efficiency aspect as it is based on the quasi maximum likelihood function. However, Chapter 5 treatment focuses on a SLD model. Since spatial interaction can enter a model in many forms, a unified method that is robust to heteroskedasticity, but is also simple to implement and efficient will be useful. This chapter looks at a general methodology that can handle heteroskedasticity of unknown form in a wide class of spatial models and still perform well in both finite and large samples. The proposed ACQS estimator is arrived at by adjusting the concentrated quasi score function for the spatial parameters to make it robust against

[2] See Kelejian and Prucha (2007, 2010), Le Sage (1997), Lin and Lee (2010), Arraiz et al. (2010), Badinger and Egger (2011, 2015), Jin and Lee (2012), Baltagi and Yang (2013), Doğan and Taşpinar (2014) for related work.

unknown heteroskedasticity in a general spatial econometric model. In order to conduct heteroskedasticity robust inferences, we also propose an outer-product-of-gradient (OPG) method for estimating the variance of the ACQS estimator. We establish the consistency and asymptotic normality of the proposed estimator. The general method that we propose is discussed in detail two interesting applications: (a) higher order spatial autoregressive model with SLD and spatial autoregressive error dependent (SED) variables, also known as the SARAR(p, q) model, and (b) fixed effects spatial panel data (SPD) model with SARAR(1,1).

For the SPD model, a key feature is the effect of the transformation of variables in order to eliminate the incidental parameters driven by the fixed effects. The resulting transformed spatial model has the added complication of the disturbances being dependent in addition to heteroskedastic. This fact in particular causes a problem in applying the OPG method to estimate the standard errors. The problem is tackled by making use of the special properties of the transformed model.

Monte Carlo results show that the ACQS estimator is easy to implement, computationally as simple as the regular QML estimator and is effective in attaining consistency under unknown heteroskedasticity while limiting the compromise on the efficiency aspect of the usual QML estimator. As the ACQS estimator captures the extra variability coming from the estimation of the regression coefficients and the average of error variance, the ACQS estimator generally outperforms the regular QML estimator when the latter is indeed consistent. As discussed in detail later, heteroskedasticity does not affect the consistency of the QML estimator of the covariate effects in general. However, since the estimate of the standard errors depends on the estimate of the spatial parameters, direct inferences based on standard $t-$statistics are likely to be affected by heteroskedasticity. In this case we propose refined $t-$statistics for covariate effects which are robust.

The rest of the chapter is organised as follows. In the next section we illustrate the general derivation of the robust ACQS estimator and the method of robust estimation of the standard errors. Section 6.3 gives the application of the methods in a SARAR(p, q) model and Section 6.4 gives a detailed application of the methods in a fixed effects spatial panel data model. Section 6.5 gives Monte Carlo results and Section 6.6 concludes the chapter. All accompanying lemmas and proofs of theorems are given in the Appendices.

6.2 General Method for Heteroskedasticity Robust Estimation

Consider the general model,

$$f\left(Y_n, X_n, W_{1n}, \ldots, W_{kn}; \vartheta, \lambda\right) = \epsilon_n, \tag{6.1}$$

with a dependent variable Y_n conditional on a set of independent variables X_n and spatial weight matrices $W_{1n}, \ldots, W_{kn}$. Parameter vector ϑ denotes the parameters of the model and λ denotes the spatial parameters. ϵ_n is an $n \times 1$ vector of model errors, uncorrelated with mean 0 and variances $\sigma^2 h_i$ where $h_i > 0, \forall i = 1, \ldots, n$ and $\sum_{i=1}^{n} h_i = n$. Clearly in the case of identically distributed errors, $h_i = 1, \forall i$. Popular spatial regression models can be written in this form. For example, the pth order spatial autoregressive lagged dependent model $Y_n = \sum_{r=1}^{p} \lambda_r W_{rn} Y_n + X_n \beta + \epsilon_n$ can be written in the form, $C_n\left(\lambda_1, \ldots, \lambda_p\right) Y_n - X_n \beta = \epsilon_n$ where $C_n\left(\lambda_1, \ldots, \lambda_p\right) = I_n - \sum_{r=1}^{p} \lambda_r W_{rn}$ and I_n is an $n \times n$ identity matrix. The qth order spatial error dependent model, $Y_n = X_n \beta + u_n$ with $u_n = \sum_{r=1}^{q} \lambda_r W_{rn} u_n + \epsilon_n$ can be written as $C_n\left(\lambda_1, \ldots, \lambda_q\right)\left(Y_n - X_n \beta\right) = \epsilon_n$. Combining these two models gives a SARAR(p, q) model that can be written in the form, $C_{2n}(\rho_1, \ldots, \rho_q)[C_{1n}(\lambda_1, \ldots, \lambda_p) Y_n - X_n\beta] = \epsilon_n$. Fixed effects spatial panel data models can also be written in the form given in (6.1) after eliminating the fixed effects using a suitable transformation.

6.2.1 Robust estimation of model parameters

When the errors are homoskedastic, the quasi maximum likelihood estimators of (ϑ, λ) of the model given in (6.1) are consistent as illustrated time and again in the literature.[3] When the errors are heteroskedastic, the QML estimator of ϑ remains consistent while that of λ becomes inconsistent in general (See Lin and Lee (2010), Liu and Yang (2015b)). As such it makes sense to use the concentrated quasi log-likelihood function for estimation by concentrating out ϑ from the Gaussian log-likelihood function derived under homoskedasticity. This makes it easier to make an adjustment to the concentrated quasi log-likelihood based estimating equation to make it robust against unknown heteroskedasticity when using it to estimate the spatial parameters λ. There are other advantages of using a concentrated quasi log-likelihood function as opposed to a full quasi log-likelihood function: (i) the dimensionality of the optimisation problem is greatly reduced when applying a numerical optimisation method to find the estimate, and (ii) the additional variability coming from the estimation of ϑ is captured by the concentrated quasi log-likelihood function. Once an estimate for the spatial parameters λ is derived from the concentrated quasi log-likelihood function, the estimator for ϑ can be defined as $\hat{\vartheta}_n = \vartheta_n(\lambda)$

The usual first order condition from the concentrated quasi log-likelihood function gives the following concentrated quasi score function,

[3] See among others, Lee (2004), Liu and Yang (2015a), Jin and Lee (2013), Lee and Yu (2010a).

$$\psi_n(\lambda) = \begin{cases} \frac{1}{n\hat{\sigma}_n^2(\lambda)} Y_n'(\lambda)\mathcal{A}_{1n}(\lambda) Y_n(\lambda) \\ \vdots \\ \frac{1}{n\hat{\sigma}_n^2(\lambda)} Y_n'(\lambda)\mathcal{A}_{kn}(\lambda) Y_n(\lambda) \end{cases} \tag{6.2}$$

where $\hat{\sigma}_n^2(\lambda)$ is the QML estimator for σ_0^2, $\mathcal{A}_{rn}(\lambda), r = 1, \ldots k$ is an $n \times n$ non-stochastic matrix that involve X_n, and W_{rn} and the spatial parameters λ, and $Y_n(\lambda)$ is an $n \times 1$ vector that involve Y_n, W_{rn} and λ. For the pth order spatial autoregressive model, $\mathcal{A}_{rn}(\lambda) = M_n\left[G_{rn}(\lambda_1, \ldots, \lambda_p) - \frac{1}{n}\mathrm{tr}\left(G_{rn}(\lambda_1, \ldots, \lambda_p)\right)\right]$ where $M_n = I_n - X_n\left(X_n'X_n\right)X_n'$, $G_{rn}(\lambda_1, \ldots, \lambda_p) = W_{rn}C_n^{-1}(\lambda_1, \ldots, \lambda_p)$, and $Y_n(\lambda) = C_n(\lambda_1, \ldots, \lambda_p) Y_n$ for $r = 1, \ldots, p$. The QML estimator for the spatial parameter is derived as the solution to the estimating equation as, $\hat{\lambda}_n = \arg\{\psi_n(\lambda) = 0\}$ and the QML estimator for ϑ is defined as $\hat{\vartheta}_n = \vartheta(\hat{\lambda}_n)$.

For $\hat{\lambda}_n$ to be consistent under unknown heteroskedasticity, it is necessary that $\mathrm{E}(\psi_n(\lambda_0))$ equals to or tends to zero (See van der Vaart, 1998, ch. 5). However, this condition is not necessarily satisfied if the errors are heteroskedastic. To observe this consider the score function, given in (6.2), which can be written as a linear quadratic form in ϵ_n, evaluated at the true parameters[4] as follows:

$$Q_n(\epsilon_n) = \begin{cases} \frac{1}{\epsilon_n'\mathcal{M}_n\epsilon_n}\left(\epsilon_n'\mathcal{A}_{1n}\epsilon_n + \epsilon_n' b_{1n}\right) \\ \vdots \\ \frac{1}{\epsilon_n'\mathcal{M}_n\epsilon_n}\left(\epsilon_n'\mathcal{A}_{kn}\epsilon_n + \epsilon_n' b_{kn}\right) \end{cases} \tag{6.3}$$

where $\mathcal{M}_n$ is the orthogonal projection matrix of the model given in (6.1) that involves the regressors X_n and the spatial parameters λ. The following basic set of regularity conditions are required in the set up of the general asymptotic theory.

[4]Quantities evaluated at the true parameters are denoted with a suppressed function argument.

In general we have, $\frac{1}{n}\mathrm{E}(\epsilon_n^{'}\mathcal{M}_n\epsilon_n) = \sigma_0^2 + o(1)$. Hence, in order to attain consistency, we consider the numerator of (6.2).[5] Note $\mathrm{E}(\epsilon_n^{'}\mathcal{A}_{rn}\epsilon_n) = \sigma_0^2\mathrm{tr}\,(H_n\mathcal{A}_{rn})$ where $H_n = \mathrm{diag}\,(h_{n,1}, \ldots, h_{n,n})$. A potentially robust estimator can be attained by adjusting the concentrated quasi score function by replacing $\mathcal{A}_{rn}$ with $\mathcal{A}_{rn}^{\Diamond} = \mathcal{A}_{rn} - \mathrm{tr}(\mathcal{A}_{rn})I_n$, where I_n is the $n \times n$ identity matrix. However, when the errors are heteroskedastic, $\mathrm{E}(\epsilon_n^{'}\mathcal{A}_{rn}^{\Diamond}\epsilon_n)$ will not be necessarily zero (Lin and Lee, 2010) since the ith component of $\mathcal{A}_{rn}^{\Diamond}\epsilon_n$ is correlated with the corresponding component of ϵ_n unless the ith diagonal element of $\mathcal{A}_{rn}^{\Diamond}$ is zero. Thus to attain a robust estimator we need to adjust the concentrated quasi score function as follows:

$$\psi_n^*(\lambda) = \begin{cases} \frac{1}{n\hat{\sigma}_n^2(\lambda)} Y_n^{'}(\lambda)\,\mathcal{A}_{1n}^{\circ}(\lambda)\,Y_n(\lambda) \\ \vdots \\ \frac{1}{n\hat{\sigma}_n^2(\lambda)} Y_n^{'}(\lambda)\,\mathcal{A}_{kn}^{\circ}(\lambda)\,Y_n(\lambda) \end{cases} \tag{6.6}$$

where $\mathcal{A}_{kn}^{\circ}(\lambda) = \mathcal{A}_{kn}(\lambda) - \mathrm{diag}\,(\mathcal{A}_{kn}(\lambda))$. The adjusted concentrated quasi score (ACQS) estimator is defined as, $\tilde{\lambda}_n = \arg\{\psi_n^*(\lambda) = 0\}$[6] and the adjusted score

[5] Using the moment expansion for the ratio of the quadratic terms (Lieberman, 1994) of $\mathrm{E}(Q_n(\epsilon_n))$ is,

$$\mathrm{E}\left(\frac{\epsilon_n^{'}\mathcal{A}_{rn}\epsilon_n}{\epsilon_n^{'}\mathcal{M}_n\epsilon_n}\right) = \frac{\mathrm{E}(\epsilon_n^{'}\mathcal{A}_{rn}\epsilon_n)}{\mathrm{E}\,(\epsilon_n^{'}\mathcal{M}_n\epsilon_n)} + T_{1n} + T_{2n} + \ldots \tag{6.4}$$

for $r = 1, \ldots, k$, where

$$T_{1n} = \frac{\mathrm{E}(\epsilon_n^{'}\mathcal{A}_{rn}\epsilon_n)\kappa_2}{[\mathrm{E}\,(\epsilon_n^{'}\mathcal{M}_n\epsilon_n)]^3} - \frac{\kappa_{11}}{[\mathrm{E}\,(\epsilon_n^{'}\mathcal{M}_n\epsilon_n)]^2} = O\left(\frac{1}{n}\right), \tag{6.5}$$

κ_p is the pth cumulant of $\epsilon_n^{'}\mathcal{M}_n\epsilon_n$ and κ_{pq} is the generalised cumulant of the product of $(\epsilon_n^{'}\mathcal{A}_{rn}\epsilon_n)^p$ and $(\epsilon_n^{'}\mathcal{M}_n\epsilon_n)^q$. (One can show that the higher order terms of the expansion are increasingly of smaller order.)

[6] Maximum likelihood based estimators such as this broadly fall into the umbrella of estimators known as M-estimators in the literature, which stand for maximum likelihood type estimators. Such an estimator can be either a solution to a maximisation function or a root of an estimating equation. The proposed robust estimator falls into the latter which is also known as the Z(ero)-estimator in van der Vaart (1998). For detailed discussions, see Huber (1964, 1981), van der Vaart (1998). Yang (2016) gives a more recent treatment of the M-estimator in a spatial dynamic panel data setting.

estimator for ϑ is defined as $\tilde{\vartheta}_n = \hat{\vartheta}_n \left(\tilde{\lambda}_n\right)$.[7]

Adjusting the score function as in (6.6) is similar in spirit, to the adjustment made to the moment conditions given in Lin and Lee (2010) in order to arrive at a robust GMM estimator for the spatial parameter of the spatial lagged dependent variable model. Adjusting the score function can also be viewed as adjusting a moment condition, however, in a quasi maximum likelihood environment the usual GMM considerations such as the availability of valid instruments, over/under identification (and the associated implications) and optimal feasible weighting matrix is not present. More importantly in a spatial model set up, when using a GMM estimator, one also need to pay special attention to the fact that the GMM estimator may give a spatial parameter estimate outside of the parameter set Λ which ensures that $C_n^{-1}(\lambda)$ is well defined. Lin and Lee shows that their robust GMM estimator using the adjusted moment conditions can lead to an estimate as efficient as the QML estimator, however, it is not clear how similar robust moment conditions can be constructed for more advanced spatial models such as the SARAR(1,1) model. For such models, the robust 3-step 2SLS/GMM estimator pioneered by Kelejian and Prucha (2010), can be used, however, the liaison with the 2SLS method comes at a cost of reduced efficiency. In contrast the adjusted quasi score estimator that we suggest has the versatility of being able to extend itself to a wide array of spatial models and estimate all the spatial parameters together while leaning on the efficiency of QML method and optimal robust method of moments (Lin and Lee, 2010).

The idea of adjusting the score function in order to attain a robust consistent estimator has been appreciated in the past, (Alvarez, and Arellano, 2004), however had been dormant until recent times. We believe that the ideas presented in this

[7]Further, in order to improve finite sample performance of the adjustment, one may also need to consider the impact of the matrix $\mathcal{M}_n$ on the limiting behaviour of the adjusted score. More on this will be discussed in the following two sections

chapter has wider applications beyond spatial models. For example, non-linear models such as limited dependent variable models and even SPD models with dynamic parameters are interesting future avenues for research.

6.2.2 Robust Estimation for VC Matrix

In order to conduct robust inference on the parameter estimates, an estimate of the standard errors are required which usually involve the estimation of the variance of the adjusted score function (6.2). However, the first order variance of (6.2) contains the second, third and fourth order moments of $\epsilon_{n,i}$ which vary across i. As such a simple White type estimator (White, 1980) is not suitable which makes it infeasible to estimate the variance of the score. In this case, we recommend the use of the outer product of the gradients (OPG) of the decomposed numerator of the adjusted score (Baltagi and Yang, 2013). The idea is to decompose the numerator of (6.3) into a sum of uncorrelated terms by writing $\mathcal{A}^{\circ}_{rn}$ as the sum of an upper, lower and a diagonal matrices as, $\mathcal{A}^{\circ}_{rn} = \mathcal{A}^{\circ u}_{rn} + \mathcal{A}^{\circ l}_{rn} + \mathcal{A}^{\circ d}_{rn}$ where $\mathcal{A}^{\circ u}_{rn}$, $\mathcal{A}^{\circ l}_{rn}$, and $\mathcal{A}^{\circ d}_{rn}$ are respectively, an upper triangular, a lower triangular, and the diagonal matrices of $\mathcal{A}^{\circ}_{rn}$. Let, $\Upsilon_i = \{\zeta_{rn,i} + a^{\circ}_{rn,ii}\epsilon_{n,i} + b_{rn,i};\ r = 1, \ldots, k\}'$, where $\zeta_{rn,i} = (\mathcal{A}^{\circ u'}_{rn} + \mathcal{A}^{\circ l}_{rn})\epsilon_n$, and $a^{\circ}_{rn,ii}$ are the diagonal elements of $\mathcal{A}^{\circ}_{rn}$ for $i = 1, \ldots, n$. In order to apply the OPG method to estimate the variance of the score, (6.3) must be written as the sum of n uncorrelated terms as follows, $Q_n(\epsilon_n) = \sum_{i=1}^{n} \epsilon_{n,i}\Upsilon_i$.

While independence of ϵ_n is sufficient to guarantee that $\epsilon_{n,i}\Upsilon_i$ are uncorrelated for each i, it may not be the case for some types of spatial models. For example, when we consider a spatial panel model with fixed effects, it is more suitable to transform the variables to eliminate the fixed effects to avoid the incidental parameter problem. In this case $f(\cdot)$ denotes a spatial model of transformed variables and the resulting $\epsilon_{n,i}$ will no longer be independent although uncorrelated and as a result $\epsilon_{n,i}\Upsilon_i$ will fail to be uncorrelated. However, the OPG method will still be applicable if $\epsilon_{n,i}\Upsilon_i$ are asymptotically uncorrelated. Consequently we consider two cases in the applications: $(i)\,\epsilon_{n,i}$ are independent, and $(ii)\,\epsilon_{n,i}$ are not independent but uncorrelated.

6.3 Higher Order SARAR Model

In this section we apply the general ideas presented in Section 6.2 to give a detailed treatment of the SARAR(p,q) model.[8] We first present asymptotic results for the quasi maximum likelihood (QML) estimators with iid errors, and then we study the robustness of the QML estimators when homoskedasticity assumption is violated. The latter in part motivates the new estimator, the adjusted concentrated quasi score (ACQS) estimators, which is robust to unknown heteroskedasticity. The consistency and asymptotic normality of the ACQS estimators are established. We introduce heteroskedasticity robust standard errors for the parameter estimates to give a set of robust inference methods. Extensive Monte Carlo experiments are conducted, and the results show excellent performance of the proposed estimators. The proposed methods are simple and can be easily adopted by the applied researchers. The results presented in this chapter contain, as special cases, the results of Jin and Lee (2013) for the QML estimations and the results of Chapter 5 for the ACQS estimators. Extensive Monte Carlo experiments are conducted, and the results show excellent performance of the proposed estimators. Compared to the QML estimator, the proposed estimator has a better finite sample performance and is robust against heteroskedasticity.

[8]For related works see Badinger and Egger (2011) which extends the robust three step 2SLS/GMM estimator given in Kelejian and Prucha (2010) to the SARAR(p,q) model, and Lee and Liu (2010) which consider the efficient GMM estimation of SARAR(p,q) under homoskedasticity. Liu and Yang (2015b) present heteroskedasticity robust adjusted concentrated quasi score estimators for SARAR(1,0) and SARAR(1,1), but the asymptotic properties of the ACQS estimators for the SARAR(1,1) model are not given.

The SARAR(p, q) model we consider is,

$$Y_n = \sum_{j=1}^{p} \lambda_i W_{jn} Y_n + X_n \beta + u_n \text{ where } u_n = \sum_{k=1}^{q} \rho_k M_{kn} u_n + \epsilon_n, \qquad (6.7)$$

where Y_n is an $n \times 1$ vector of observations on the dependent variable, X_n is an $n \times k$ matrix of observations on the exogenous regressors with the $k \times 1$ regression parameter vector β, u_n is the $n \times 1$ vector of disturbances, W_{jn} and M_{kn} are the $n \times n$ spatial weights matrices which summarises the higher order spatial dependence of the dependent variable and the disturbances with the associated spatial parameters $\lambda = (\lambda_1, \ldots, \lambda_p)'$ and $\rho = (\rho_1, \ldots, \rho_q)'$. The $W_{jn}'s$ are different and so are the $M_k's$ to account for different levels of spatial dependence, but some (or all) of the W_{jn}'s can be the same as some (or all) of the M_k's.

6.3.1 Robustness of QML estimator against unknown heteroskedasticity

We now examine the properties of the QML estimator of the SARAR(p, q) model when the errors are iid, and their robustness when they are inid. We derive the asymptotic distribution of the QML estimator when the errors are iid and give its asymptotic variance. When the errors are inid we provide conditions under which the regular QML estimator is consistent. The reduced form equation of model (6.7) is,

$$Y_n = A_n^{-1}(\lambda)[X_n \beta + B_n^{-1}(\rho) \epsilon_n], \qquad (6.8)$$

where, $A_n(\lambda) = I_n - \sum_{j=1}^{p} \lambda_i W_{jn}$ and $B_n(\rho) = I_n - \sum_{k=1}^{q} \rho_k M_{kn}$. Let $\lambda = (\lambda_1, \ldots, \lambda_p)'$, $\rho = (\rho_1, \ldots, \rho_q)'$, $\delta = (\lambda', \rho')'$, and $\theta = (\beta', \sigma^2, \lambda', \rho')'$. Let θ_0 denote the true parameter vector. The Gaussian log-likelihood function for θ is,

$$\ell_n(\theta) = -\tfrac{n}{2} \ln(2\pi) - \tfrac{n}{2} \ln(\sigma^2) + \ln|A_n(\lambda)| + \ln|B_n(\rho)| - \tfrac{1}{2\sigma^2} \epsilon_n'(\beta, \delta) \epsilon_n(\beta, \delta), \qquad (6.9)$$

where $\epsilon_n(\beta,\delta) = B_n(\rho)[A_n(\lambda)Y_n - X_n\beta]$. Maximizing (6.9) gives the maximum likelihood estimates of θ_0 if the disturbances are indeed Gaussian, otherwise QML estimators. Let $X_n(\rho) = B_n(\rho)X_n$ and $Y_n(\delta) = B_n(\rho)A_n(\lambda)Y_n$. Given δ, we have the constrained QML estimators for β_0 and σ_0^2 as follows,

$$\hat{\beta}(\delta) = [X_n'(\rho)X_n(\rho)]^{-1}X_n'(\rho)Y_n(\delta) \text{ and } \hat{\sigma}^2(\delta) = Y_n'(\delta)M_n(\rho)Y_n(\delta), \tag{6.10}$$

where $\mathcal{M}_n(\rho) = I_n - X_n(\rho)[X_n'(\rho)X_n(\rho)]^{-1}X_n'(\rho)$. Concentrating out $\hat{\beta}(\delta)$ and $\hat{\sigma}^2(\delta)$ from the quasi log-likelihood function we get the concentrated quasi log-likelihood function for δ as,

$$\ell_n^c(\delta) = -\tfrac{n}{2}\ln(2\pi) - \tfrac{n}{2}\ln[\hat{\sigma}^2(\delta)] + \ln|A_n(\lambda)| + \ln|B_n(\rho)|, \tag{6.11}$$

Maximizing $\ell_n^c(\delta)$ gives the unconstrained QML estimator $\hat{\delta}_n$ of δ, and thus the QML estimators of β and σ^2 as $\hat{\beta}_n \equiv \hat{\beta}(\hat{\delta}_n)$ and $\hat{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\hat{\delta}_n)$. Denote $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\delta}_n)'$, the QML estimator of the parameter vector θ.

Consider the score function derived from (6.9) evaluated at the true parameter θ_0,

$$\frac{\partial}{\partial\theta}\ell_n(\theta_0) = \begin{cases} \frac{1}{\sigma_0^2}X_n'B_n'\epsilon_n \\ \frac{1}{2\sigma_0^4}(\epsilon_n'\epsilon_n - n\sigma^2) \\ \frac{1}{\sigma_0^2}\left(\epsilon_n'\bar{B}_{jn}\epsilon_n - \sigma^2\text{tr}(F_{jn}) + \epsilon_n'\eta_{jn}\right), & j = 1,\dots,p \\ \frac{1}{\sigma_0^2}\left(\epsilon_n'G_{kn}\epsilon_n - \sigma^2\text{tr}(G_{kn})\right), & k = 1,\dots,q \end{cases} \tag{6.12}$$

where $F_{jn} = W_{jn}A_n^{-1}$, $G_{kn} = M_{kn}B_n^{-1}$, $\bar{B}_{jn} = B_nF_{jn}B_n^{-1}$ and $\eta_{jn} = B_nF_{jn}X_n\beta_0$.

Some additional regularity conditions are required to establish the asymptotic properties of the QML estimators.

Assumption 6.1: *The true spatial parameter vector λ_0 is in the interior of a* *compact parameter set* Λ.[9]

Assumption 6.2: *The errors* $\{\epsilon_{n,i}\}$ *are independent such that* $E(\epsilon_{n,i}) = 0$, $Var(\epsilon_{n,i}) = \sigma_0^2 h_{n,i}$, *where* $h_{n,i} > 0$ *and* $\sum_{i=1}^{n} h_{n,i} = n$.[10] *Further,* $E|\epsilon_{n,i}|^{4+\eta} < c$ *for some* $\eta > 0$ *and constant* c *for all* n *and* i.

Assumption 6.3: *The elements of the* $n \times k$ *regressor matrix* X_n *are uniformly bounded for all* n, X_n *has the full rank* k, *and* $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ *exists and is non-singular.*

Assumption 6.4: *The spatial weights matrices* W_{rn} *are uniformly bounded in both row and column sums and their diagonal elements are zero.*

Assumption 6.5: *The matrix* A_n *and* B_n *are non-singular and* A_n^{-1} *and* B_n^{-1} *are uniformly bounded in absolute value in both row and column sums. Further,* $A_n^{-1}(\lambda)$ *and* $B_n^{-1}(\rho)$ *are uniformly bounded in either row or column sums, uniformly in* $\delta \in \Delta$.

Assumption 6.6: *Either* (a) : $\lim_{n\to\infty} \mathcal{H}_n(\rho)$ *is non-singular* $\forall \rho$ *and* $\lim_{n\to\infty} \mathcal{Q}_{1n}(\rho) \neq 0$ *for* $\rho \neq \rho_0$; *or* (b) : $\lim_{n\to\infty} \mathcal{Q}_{2n}(\delta) \neq 0$ *for* $\delta \neq \delta_0$, *where*

$$\mathcal{H}_n(\rho) = \frac{1}{n}(X_n, F_{1n}X_n\beta_0, \ldots, F_{pn}X_n\beta_0)' B_n'(\rho) B_n(\rho)(X_n, F_{1n}X_n\beta_0, \ldots, F_{pn}X_n\beta_0),$$

$$\mathcal{Q}_{1n}(\rho) = \frac{1}{n}\left(\ln|\sigma_0^2 B_n^{-1\prime} B_n^{-1}| - \ln|\sigma_n^2(\rho) B_n^{-1\prime}(\rho) B_n^{-1}(\rho)|\right),$$

$$\mathcal{Q}_{2n}(\delta) = \frac{1}{n}\left(\ln|\sigma_0^2 B_n^{-1\prime} A_n^{-1\prime} A_n^{-1} B_n^{-1}| - \ln|\sigma_n^2(\delta) B_n^{-1\prime}(\rho) A_n^{-1\prime}(\lambda) A_n^{-1}(\lambda) B_n^{-1}(\rho)|\right),$$

$$\sigma_n^2(\rho) = \frac{\sigma_0^2}{n}\mathrm{tr}[B_n^{-1\prime} B_n'(\rho) B_n(\rho) B_n^{-1}],\ \sigma_n^2(\delta)$$

[9] The parameter space Λ must be such that the reduced form of (6.1) is well defined and the Jacobian terms of the quasi log-likelihood function is non-singular $\forall \lambda \in \Lambda$. For a general W_{rn}, Lee and Liu (2010) shows that since, $||\sum_{r=1}^{k} \lambda_r W_{rn}|| \leq (\sum_{r=1}^{k} |\lambda_r|) \cdot \max_{r=1,\ldots,p} ||W_{rn}||$, where $||\cdot||$ denote the matrix norm, a viable parameter space for λ_r is s.t. $\sum_{r=1}^{k} |\lambda_r| < (\max_{r=1,\ldots,k} ||W_{rn}||)^{-1}$. When W_{rn} are row normalised, we have $\max_{r=1,\ldots,k} ||W_{rn}|| = 1$ s.t. $\sum_{r=1}^{k} |\lambda_r| < 1$. When W_{rn} are not row-normalised, we can use the relation $||W_{rn}|| = \prod_{i=1}^{n} w_{ji}$, where w_{ji} are the eigenvalues of W_{rn} to avoid the need to compute the determinants of W_{rn}. However, Elhorst et al. (2012) argues that this parametrisation is too restrictive and gives an alternative procedure to determine the exact boundaries which depends on the specification of W_{rn}. Also see Kelejian and Prucha (2010) and Le Sage (2009) for related discussions of the parameter space of the spatial parameters.

[10] For generality, we allow $h_{n,i}$ to depend on n for each i. This parametrisation, is a non-parametric version of Breusch and Pagan (1979), allows the estimation of the average scale parameter.

$= \frac{\sigma_0^2}{n}\text{tr}[B_n^{-1\prime}A_n^{-1\prime}A_n'(\lambda)B_n'(\rho)B_n(\rho)A_n(\lambda)A_n^{-1}B_n^{-1}].$

Assumption 6.2 is the heteroskedasticity condition of the errors. Clearly $h_{n,i} = 1, \forall i$ when the errors are homoskedastic. The parameter σ_0^2 is the average scale parameter of $\epsilon_{n,i}$ and $h_{n,i}$ denote the heteroskedasticity parameters. The uniform boundedness conditions in Assumption 6.4 limit the spatial dependence to a manageable level. Assumption 6.5 is a standard assumption in spatial econometric literature introduced by Kelejian and Prucha (1999) and allows a uniform boundedness property in the asymptotic formulation. Assumption 6.6 is the identification condition introduced by Lee (2004) adjusted to suit the SARAR(p,q) model.

Jin and Lee (2013) presents asymptotic results for the QML estimators of the SARAR(1,1) model under homoskedasticity. However, the asymptotic distribution for the SARAR(p,q) model is not given in the literature. The following Theorem fills in this gap.

Theorem 6.1 *Under Assumptions 6.1-6.6 and further assuming that* $h_{n,i} = 1\ \forall i$, *the QML estimator* $\hat{\theta}_n$ *is consistent and asymptotically normal with*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0,\ \Sigma^{-1}\Omega\Sigma^{-1}),$$

where $\Sigma = \lim_{n\to\infty} -\frac{1}{n}E\left(\frac{\partial^2}{\partial\theta\partial\theta'}\ell_n(\theta_0)\right)$ *and* $\Omega = \lim_{n\to\infty} \frac{1}{n}E\left(\frac{\partial}{\partial\theta}\ell_n(\theta_0)\frac{\partial}{\partial\theta'}\ell_n(\theta_0)\right)$, *where the elements of these matrices are:*

$\Sigma_{n,\beta\beta} = \frac{1}{\sigma_0^2}X_n'B_n'B_nX_n,$ $\Sigma_{n,\sigma^2\sigma^2} = \frac{n}{2\sigma_0^4},$

$\Sigma_{n,\lambda_j\lambda_{j'}} = \frac{1}{\sigma_0^2}\eta_{jn}'\eta_{j'n} + tr(\bar{B}_{jn}^s\bar{B}_{j'n}),$ $\Sigma_{n,\rho_k\rho_{k'}} = tr(G_{k'n}^sG_{kn}),$

$\Sigma_{n,\beta,\sigma^2} = 0_{k\times 1},$ $\Sigma_{n,\beta\lambda_j} = \frac{1}{\sigma_0^2}X_n'B_n'\eta_{jn},$

$\Sigma_{n,\beta\rho_k} = 0_{k\times 1},$ $\Sigma_{n,\sigma^2\lambda_j} = \frac{1}{\sigma_0^2}tr(\bar{B}_{jn}),$

$\Sigma_{n,\sigma^2\rho_k} = \frac{1}{\sigma_0^2}tr(G_{kn})$ *and* $\Sigma_{n,\lambda_j\rho_k} = tr(G_{kn}^s\bar{B}_{jn}),$

for $j, j' = 1,\ldots,p$ *and* $k, k' = 1,\ldots,q$, *where,* $G_{kn} = M_{kn}B_n^{-1}$, $\bar{B}_{jn} = B_nF_{jn}B_n^{-1}$,

$F_{jn} = W_{jn}A_n^{-1}$, $\eta_{jn} = B_n F_{jn} X_n \beta_0$, $G^s_{kn} = G_{kn} + G^{'}_{kn}$, $\bar{B}^s_{jn} = \bar{B}_{jn} + \bar{B}^{'}_{jn}$.

Let $\Gamma_n = \Omega_n - \Sigma_n$, then the following elements representing block matrices make up $n\Gamma_n$,

$n\Gamma_{n,11} = 0_{k\times k}$, $\quad n\Gamma_{n,22} = \frac{n\kappa}{4\sigma_0^4}$,

$n\Gamma_{n,33} = \kappa \bar{b}^{'}_{jn}\bar{b}_{j'n} + \frac{2\gamma}{\sigma_0}\bar{b}^{'}_{jn}\eta_{j'n}$, $\quad n\Gamma_{n,44} = \kappa g^{'}_{kn} g_{k'n}$,

$n\Gamma_{n,12} = \frac{\gamma}{2\sigma_0^3} X^{'}_n B^{'}_n \iota_n$, $\quad n\Gamma_{n,13} = \frac{\gamma}{\sigma_0} X^{'}_n B^{'}_n \bar{b}_{jn}$,

$n\Gamma_{n,14} = \frac{\gamma}{\sigma_0} X^{'}_n B^{'}_n g_{kn}$, $\quad n\Gamma_{n,23} = \frac{\kappa}{2\sigma_0^2} tr(\bar{B}_{jn}) + \frac{\gamma}{2\sigma_0^3}\iota^{'}_n \eta_{jn}$,

$n\Gamma_{n,24} = \frac{\kappa}{2\sigma_0^2} tr(G_{kn})$ *and* $\quad n\Gamma_{n,34} = \kappa g^{'}_{kn}\bar{b}_{jn} + \frac{\gamma}{\sigma_0} g^{'}_{kn}\eta_{jn}$,

for $j, j' = 1, \ldots, p$ and $k, k' = 1, \ldots, q$, where, ι_n is an $n \times 1$ vector of ones, $g_{kn} = diagv(G_{kn})$, $\bar{b}_{jn} = diagv(\bar{B}_n)$, and γ and κ are respectively, the measures of skewness and kurtosis of ϵ_n.

Now suppose we relax the condition that $h_{n,i} = 1\ \forall i$. Does $\hat{\theta}_n$ continue to be consistent? Let $H_n = \text{diag}(h_{n,1}, \ldots, h_{n,n})$. For any extremum estimator such as $\hat{\theta}_n$ to be consistent, it is necessary that $\text{plim}_{n\to\infty}\frac{1}{n}\psi_n(\theta_0) = 0$. This is satisfied for the derivatives w.r.t. β and σ^2, however it may not be the case for the derivatives w.r.t. λ_j and ρ_k. Note that,

$$\begin{aligned}
\frac{1}{n}\frac{\partial}{\partial\lambda_j}\ell_n(\theta_0) &= \frac{1}{n}\text{tr}(H_n\bar{B}_{jn} - F_{jn}) + o_p(1),\ j = 1, \ldots, p \\
&= \frac{1}{n}\text{tr}(H_n\bar{B}_{jn} - \bar{B}_{jn}) + o_p(1) \\
&= \frac{1}{n}\sum_{i=1}^n (h_{n,i} - 1)(\bar{b}_{jn} - \text{E}(\bar{b}_{jn})) + o_p(1) \\
&= \text{Cov}(h_{n,i}, \bar{b}_{jn}) + o_p(1),\ \text{and} \qquad (6.13) \\
\frac{1}{n}\frac{\partial}{\partial\rho_k}\ell_n(\theta_0) &= \frac{1}{n}\text{tr}(H_n G_{kn} - G_{kn}) + o_p(1),\ k = 1, \ldots, q \\
&= \frac{1}{n}\sum_{i=1}^n (h_{n,i} - 1)(g_{kn} - \text{E}(g_{kn})) + o_p(1) \\
&= \text{Cov}(h_{n,i}, g_{kn}) + o_p(1), \qquad (6.14)
\end{aligned}$$

where $\bar{b}_{jn} = \text{diagv}(\bar{B}_n)$ and $g_{kn} = \text{diagv}(G_{kn})$. As such, for $\hat{\theta}_n$ to be consistent, it is necessary that as $n \to \infty$, $\text{Cov}(h_{n,i}, \bar{b}_{jn}) = 0$ and $\text{Cov}(h_{n,i}, g_{kn}) = 0$. However, to establish consistency more conditions are necessary. In practice it is more likely that it is impossible to check whether these necessary conditions are satisfied or now given that heteroskedasticity is of unknown form. Hence we move to an alternative robust estimator directly.[11]

6.3.2 Adjusted Concentrated Quasi Score Estimation

We illustrate adjusted concentrated quasi score (ACQS) method for estimating the parameters in the SARAR(p, q) described in a previous section. The Monte Carlo results confirms the excellent performance of the new estimator for both finite and large samples. For robust inferences concerning the spatial or regression parameters, we introduce estimators of the variances of the ACQS estimators based on the OPG of the score function.

The method

Consider the concentrated quasi score function of δ, when β and σ^2 are concentrated out from (6.9),

$$\tilde{\psi}_n(\delta) = \begin{cases} \dfrac{Y_n'(\delta)\mathcal{M}_n(\rho)[\bar{B}_{jn}(\delta) - \frac{1}{n}\text{tr}(F_{jn}(\lambda))I_n]Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)}, & j = 1, \ldots, p \\ \dfrac{Y_n'(\delta)\mathcal{M}_n(\rho)[\bar{G}_{kn}(\rho) - \frac{1}{n}\text{tr}(G_{kn}(\rho))]Y_n(\delta)}{Y_n'(\lambda)\mathcal{M}_n(\rho)Y_n(\delta)}, & k = 1, \ldots, q \end{cases} \tag{6.15}$$

where $\bar{B}_{jn} = B_n F_{jn} B_n^{-1}$ and for $\bar{G}_{kn}(\rho) = G_{kn}(\rho)\mathcal{M}_n(\rho)$.

Following the ideas given before, we can adjust the numerator of the concentrated quasi score function to attain the desired probability limit of zero at the true parameters under heteroskedasticity. In order to ensure $\tilde{\psi}_n(\delta)$ tend to zero in

[11] The exact form of these additional necessary conditions and the consistency of the QML estimator is established in the FE-SPD model given in the next section. A milder form of this results are also given in Chapter 5 for the SLD model.

expectation so that $\hat{\delta}_n = \arg\{\tilde{\psi}_n(\delta) = 0\}$ is consistent we adjust the concentrated quasi score function as follows,

$$\tilde{\psi}_n^*(\delta) = \begin{cases} \dfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)}, & j = 1, \ldots, p \\ \dfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}^{\circ}(\rho)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)}, & k = 1, \ldots, q \end{cases} \tag{6.16}$$

where for $j = 1, \ldots, p$, $\bar{B}_{jn}^{\circ}(\delta) = \bar{B}_{jn}(\delta) - \text{diag}(\mathcal{M}_n(\rho))^{-1}\text{diag}[\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)]$ and $k = 1, \ldots, q$, $\bar{G}_{kn}^{\circ}(\rho) = \bar{G}_{kn}(\rho) - \text{diag}(\mathcal{M}_n(\rho))^{-1}\text{diag}[\mathcal{M}_n(\rho)\bar{G}_{kn}(\rho)]$.

The ACQS estimator of δ_0 is defined to be $\tilde{\delta}_n = \arg\{\tilde{\psi}_n^*(\delta) = 0\}$ and we can show that this estimator is consistent and asymptotically normal in the presence of unknown heteroskedasticity.

Remark 6.1 *As explained in the general section, the adjustment suggested here is in line with the ideas of Lin and Lee (2010)'s adjustment to the moment conditions of the SARAR(1,0) model so that the resulting GMM estimator is robust. They suggest to restrict a broader class of quadratic moments of the form* $\epsilon_n' P_n \epsilon_n$ *where* $\text{tr}(P_n) = 0$ *to* $\text{diag}(P_n) = 0$. *Thus although the condition* $\text{tr}(P_n) = 0$ *is sufficient to ensure that the moment condition satisfies* $E(\epsilon_n' P_n \epsilon_n) = 0$ *the latter condition also makes it robust. However, it is not clear how we can extend these ideas to construct a GMM estimator for the SARAR*(p, q) *model. To attain the ACQS estimator, if we apply a similar adjustment on the full score function* (6.12)*, where we replace* $\sigma^2\text{tr}(F_{jn})$ *with* $\epsilon_n'\text{diag}(F_{jn})\epsilon_n$ *and replace* $\sigma^2\text{tr}(G_{kn})$ *with* $\epsilon_n'\text{diag}(G_{kn})\epsilon_n$ *we still get an asymptotically valid adjustment that satisfies the conditions given in* (6.14) *and* (6.14). *However, this adjustment has poor small sample performance since it does not take into account the variation stemming from the estimation of the other parameters in the score function.*

Remark 6.2 *For the component of* (6.15) *w.r.t.* λ, *we replace* $\frac{1}{n}tr(F_{jn}(\lambda))I_n$ *with* $diag(\mathcal{M}_n(\rho))^{-1}diag[\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)]$, *instead of* $\text{diag}(\bar{B}_{jn}(\delta))$ *(in the spirit of*

Lin and Lee, 2010). Although the placement $\mathrm{diag}(\bar{B}_{jn}(\delta))$ *is asymptotically valid, finite sample performance, once again, maybe poor. Given* $\mathrm{E}(Y_n'\mathcal{M}_n\bar{B}_{jn}Y_n) = \sigma_n^2\mathrm{tr}(H_n\mathcal{M}_n\bar{B}_{jn}) = \sigma_n^2\mathrm{tr}(H_n\mathrm{diag}(\mathcal{M}_n\bar{B}_{jn}))$*, if we were to replace* $\frac{1}{n}tr(F_{jn}(\lambda))I_n$ *with* $\mathrm{diag}(\mathcal{M}_n\bar{B}_{jn})$ *it ignores the effect of* $\mathcal{M}_n$ *since* $\mathrm{E}(Y_n'\mathcal{M}_n\mathrm{diag}(\mathcal{M}_n\bar{B}_{jn})Y_n) = \sigma_n^2\mathrm{tr}(H_n\mathcal{M}_n\mathrm{diag}(\mathcal{M}_n\bar{B}_{jn}) \neq \mathrm{E}(Y_n'\mathcal{M}_n\bar{B}_{jn}Y_n)$*. A similar argument can be given to the adjustment applied to the component of the score with respect to* ρ*.*

Remark 6.3 *The adjustment applied to* (6.16) *is in line with the heteroskedasticity robust LM test of Baltagi and Yang (2013) with finite sample corrections.*

Remark 6.4 *Due to the non-linear manner in which the spatial parameters enter the quasi log-likelihood function, the resulting QML estimators are biased in finite samples. However, an estimator derived from an estimating function with an expected value of zero leads to a potentially bias corrected estimator. Thus the proposed estimator is not only robust for heteroskedasticity and non-normality, but also performs well in finite samples. Thus combined with a robust estimator for the standard errors we have an improved basis for inference related matters for the SARAR*(p,q) *model.*

Asymptotic properties of the ACQS estimator

In order to establish the asymptotic distribution of the ACQS estimators of the SARAR(p,q) model we need to adjust the identification condition given in Assumption 6.6 to suit the new model as follows,

Assumption 6.6*: $\lim_{n\to\infty}\mathcal{R}_{1n}(\delta) \neq 0$ *and* $\lim_{n\to\infty}\mathcal{R}_{2n}(\delta) \neq 0$, $\forall\delta \neq \delta_0$*, where*

$$\mathcal{R}_{1n}(\delta) = \tfrac{1}{n}\beta_0'X_n'B_n'^{-1}A_n'^{-1}\Psi_{jn}(\delta)A_n^{-1}B_n^{-1}X_n\beta_0 + \tfrac{\sigma_0^2}{n}\mathrm{tr}\left(H_nB_n'^{-1}A_n'^{-1}\Psi_{jn}(\delta)A_n^{-1}B_n'^{-1}\right),$$

$$\mathcal{R}_{2n}(\delta) = \tfrac{\sigma_0^2}{n}\mathrm{tr}\left(H_nB_n'^{-1}A_n'^{-1}\Omega_{kn}(\delta)A_n^{-1}B_n'^{-1}\right),$$

$$\Psi_{jn}(\delta) = A_n'(\lambda)B_n'(\rho)[\bar{B}_{jn}(\delta) - \mathrm{diag}(\bar{B}_{jn}(\delta))]B_n(\rho)A_n(\lambda),$$

$\Psi_{kn}(\delta) = A'_n(\lambda)B'_n(\rho)[\bar{G}_{kn}(\rho) - \text{diag}(\bar{G}_{kn}(\rho))]B_n(\rho)A_n(\lambda)$, *for* $j = 1, \ldots, p$ *and* $k = 1, \ldots, q$.

The normalised and adjusted concentrated quasi score function evaluated at δ_0 takes the form,

$$\sqrt{n}\tilde{\psi}^*_n = \begin{cases} \frac{1}{\sqrt{n}\sigma_0^2}\left(\epsilon'_n P_{jn}\epsilon_n + c'_{jn}\epsilon_n\right) + o_p(1), & j = 1, \ldots, p \\ \frac{1}{\sqrt{n}\sigma_0^2}\left(\epsilon'_n Q_{kn}\epsilon_n + c'_{kn}\epsilon_n\right) + o_p(1), & k = 1, \ldots, q \end{cases} \tag{6.17}$$

where $P_{jn} = \mathcal{M}_n\bar{B}^\circ_{jn}$, $Q_{kn} = \mathcal{M}_n\bar{G}^\circ_{kn}\mathcal{M}_n$, $c_{jn} = \mathcal{M}_n\bar{B}^\circ_{jn}B_nX_n\beta_0$ and $c_{kn} = \mathcal{M}_n\bar{G}^\circ_{kn}B_nX_n\beta_0$. Let $\tau^2_n(\tilde{\psi}^*_n) = \text{Var}(\sqrt{n}\tilde{\psi}^*_n)$. Using the central limit theorem for linear-quadratic forms of Kelejian and Prucha (2001), we give the following theorem:

Theorem 6.2 *Under Assumptions 6.1-6.5 and R6*, the ACQS estimator* $\tilde{\delta}_n$ *is consistent and asymptotically normal, i.e., as* $n \to \infty$, $\tilde{\delta}_n \xrightarrow{p} \delta_0$, *and*

$$\sqrt{n}(\tilde{\delta}_n - \delta_0) \xrightarrow{D} N\left(0,\ \lim_{n\to\infty} \tau^2_n(\tilde{\delta}_n)\right),$$

where $\tau^2_n(\tilde{\delta}_n) = \Phi_n^{-1}\tau^2_n(\tilde{\psi}^*_n)\Phi_n^{-1}$, *where* $\Phi_n = -E\left(\frac{\partial}{\partial\delta'}\tilde{\psi}^*_n(\delta_0)\right)$ *or it's first order term given by,*

$\Phi_{n,11} = \frac{1}{n}tr\left[H_n\left(\bar{B}^\circ_{jn}\bar{B}_{j'n} + \bar{B}'_{j'n}\bar{B}^\circ_{jn} - \dot{\bar{B}}^\circ_{jj',n}\right)\right] + \frac{1}{n\sigma_0^2}c'_{jn}\eta_{j'n}$,

$\Phi_{n,12} = \frac{1}{n}tr\left[H_n\left(\bar{B}^\circ_{jn}G_{kn} + \bar{G}'_{kn}\bar{B}^\circ_{jn} - \dot{\bar{B}}^\circ_{jk,n} - G_{kn}\bar{B}^\circ_{jn} + \bar{G}_{kn}\bar{B}^\circ_{jn}\right)\right] + \frac{1}{n\sigma_0^2}c'_{jn}\eta_{kn}$,

$\Phi_{n,21} = \frac{1}{n}tr\left[H_n\left(\bar{G}^\circ_{kn}\bar{B}_{jn} + \bar{B}'_{jn}\bar{G}^\circ_{kn}\right)\right] + \frac{1}{n\sigma_0^2}c'_{kn}\eta_{jn}$ *and*

$\Phi_{n,22} = \frac{1}{n}tr\left[H_n\left(\bar{G}^\circ_{kn}G_{k'n} + \bar{G}'_{k'n}\bar{G}^\circ_{kn} - \dot{\bar{G}}^\circ_{kk',n} - G_{k'n}\bar{G}^\circ_{kn} + \bar{G}_{k'n}\bar{G}^\circ_{kn}\right)\right]$,

$\quad + \frac{1}{n\sigma_0^2}c'_{kn}\eta_{k'n}$ *where*

$c_{jn} = \mathcal{M}_n\bar{B}^\circ_{jn}B_nX_n\beta_0$, $c_{kn} = \mathcal{M}_n\bar{G}^\circ_{kn}B_nX_n\beta_0$,

$\eta_{jn} = B_nF_{jn}X_n\beta_0$, $\eta_{kn} = \bar{G}_{kn}B_nX_n\beta_0$,

$$\dot{\bar{B}}^\circ_{jj',n} = \text{E}\left[\frac{\partial}{\partial\lambda}\bar{B}^\circ_{jn}(\delta)\big|_{\delta=\delta_0}\right]$$
$$= B_nF_{jn}F_{j'n}B_n^{-1} - diag(\mathcal{M}_n)^{-1}diag[\mathcal{M}_nB_nF_{jn}F_{j'n}B_n^{-1}],$$

$$\dot{\bar{B}}^{\circ}_{jk,n} = \mathrm{E}\left[\frac{\partial}{\partial\rho}\bar{B}^{\circ}_{jn}(\delta)\Big|_{\delta=\delta_0}\right]$$
$$= -G_{kn}\bar{B}_{jn} - \bar{B}_{jn}G_{kn} - diag(\mathcal{M}_n^{-1}\dot{\mathcal{M}}_{kn}\mathcal{M}_n^{-1})\,diag[\mathcal{M}_n\bar{B}_{jn}]$$
$$-diag(\mathcal{M}_n)^{-1}diag[\dot{\mathcal{M}}_{kn}\bar{B}_{jn}],$$
$$\dot{\bar{G}}^{\circ}_{kn} = G_{kn}\bar{G}_{kn} + G_{kn}\dot{\mathcal{M}}_{kn} - diag(\mathcal{M}_n)^{-1}diag(\dot{\mathcal{M}}_{kn}\bar{G}_{kn})$$
$$-diag(\mathcal{M}_n)^{-1}diag(\mathcal{M}_nG_{kn}\dot{\mathcal{M}}_{kn}) - diag(\mathcal{M}_n)^{-1}diag(\mathcal{M}_nG_{kn}\bar{G}_{kn})$$
$$+diag(\mathcal{M}_n^{-1}\dot{\mathcal{M}}_{kn}\mathcal{M}_n^{-1})\,diag(\mathcal{M}_n\bar{G}_{kn}),$$
$$\dot{\mathcal{M}}_{kn} = \mathcal{M}_nG_{kn}\mathcal{P}_n + \mathcal{P}_nG'_{kn}\mathcal{M}_n \text{ and } \mathcal{P}_n = I_n - \mathcal{M}_n.$$

Given $\tilde{\delta}_n$, the ACQS estimator for β_0 is

$$\hat{\beta}_n(\tilde{\delta}_n) \equiv \tilde{\beta}_n = (X_n'\tilde{B}_n'\tilde{B}_nX_n)^{-1}X_n'\tilde{B}_n'\tilde{B}_n\tilde{A}_nY_n.$$

A Taylor expansion of $\hat{\beta}_n(\delta)$ around $\delta = \delta_0$ gives $\hat{\beta}_n(\delta) = \hat{\beta}_n(\delta_0) + \dot{\beta}_n(\delta_0)(\delta - \delta_0) + O_p(\frac{1}{n})$, where $\dot{\beta}(\delta_0) = \frac{\partial}{\partial\delta}\hat{\beta}_n\Big|_{\delta=\delta_0}$. This can be used to derive the asymptotic distribution of $\tilde{\beta}_n$.

Theorem 6.3 *Under Assumptions 6.1-6.5 and 6.6*, the ACQS estimator* $\tilde{\beta}_n$ *is consistent and asymptotically normal, i.e., as* $n \to \infty$, $\tilde{\beta}_n \xrightarrow{p} \beta_0$, $\tilde{\sigma}_n^2 \xrightarrow{p} \sigma_0^2$ *and*

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) \xrightarrow{D} N\left(0,\ \lim_{n\to\infty}(X_n'B_n'B_nX_n)^{-1}X_n'B_n'\mathbb{A}_nB_nX_n(X_n'B_n'B_nX_n)^{-1}\right),$$

where

$\mathbb{A}_n = n\sigma_0^2H_n + \eta_{jn}\tau_{n,j}^2(\tilde{\lambda}_n)\eta_{jn}' + 2\sqrt{n}(\sigma_0^{-2}P_{jn}^ds_n + H_nc_{jn},\ \sigma_0^{-2}Q_{kn}^ds_n + H_nc_{kn})\Phi^{-1}(\eta_{jn}, 0_n)'$, $\tau_{n,j}^2(\tilde{\delta}_n)$ *is the VC matrix block corresponding to* λ_j, $s_n = \mathrm{E}(\epsilon_n^3)$, $P_{jn}^d = \mathrm{diag}(P_{jn})$, $Q_{kn}^d = \mathrm{diag}(Q_{kn})$ *for* $j = 1,\ldots,p$ *and* $k = 1,\ldots,q$ *and* 0_n *is an* $n \times 1$ *vector of zeros.*[12]

[12] The limiting distribution of $\tilde{\sigma}_n^2$ can be easily derived. However, it is of limited use as any inference on $\tilde{\sigma}_n^2$ requires the consistent estimation of $\frac{1}{n}\sum_{i=1}^n \mathrm{Var}(\epsilon_{n,i}^2) = \frac{\sigma^4}{n}\sum_{i=1}^n(\kappa_{n,i} + h_{n,i}^2)$ which cannot be done.

Robust estimation of VC matrix

In order to apply the OPG method to estimate $\tau_n^2(\tilde{\psi}_n^*)$, write the numerator of (6.17) as,

$$R_n(\epsilon_n) = \begin{cases} \epsilon_n' P_{jn}\epsilon_n + c_{jn}'\epsilon_n, & j = 1, \dots, p \\ \epsilon_n' Q_{kn}\epsilon_n + c_{kn}'\epsilon_n, & k = 1, \dots, q \end{cases} \tag{6.18}$$

where both P_{jn} and Q_{kn} can be decomposed as a sum of upper triangular, lower triangular and diagonal matrices. Then we can estimate the variance of the score as,

$$\tilde{\tau}_n^2(\tilde{\psi}_n^*) = \tfrac{1}{n^2\tilde{\sigma}_n^4}\textstyle\sum_{i=1}^n \tilde{\epsilon}_{n,i}^2 \tilde{\Upsilon}_{n,i}\tilde{\Upsilon}_{n,i}' \tag{6.19}$$

where, $\tilde{\Upsilon}_{n,i} = \begin{pmatrix} \tilde{\zeta}_{jn,i} + \tilde{p}_{jn,ii}\tilde{\epsilon}_{n,i} + \tilde{c}_{jn,i}, \; j = 1, \dots, p \\ \tilde{\xi}_{kn,i} + \tilde{q}_{kn,ii}\tilde{\epsilon}_{n,i} + \tilde{c}_{kn,i}, \; k = 1, \dots, q \end{pmatrix}$, $\tilde{\zeta}_{jn} = (\tilde{P}_{jn}^{u\prime} + \tilde{P}_{jn}^{l})\tilde{\epsilon}_n$, $\tilde{\xi}_{jn} = (Q_{kn}^{u\prime} + Q_{kn}^{l})\epsilon_n$, and $\tilde{p}_{jn,ii}$ and $\tilde{q}_{kn,ii}$ are respectively the diagonal elements of P_{jn} and Q_{kn}. The estimator of $\tau_n^2(\tilde{\delta}_n)$ is defined as $\tilde{\tau}_n^2(\tilde{\delta}_n) = \tilde{\Phi}_n^{-1}\tilde{\tau}_n^2(\tilde{\psi}_n^*)\tilde{\Phi}_n^{-1}$, where $\tilde{\Phi}_n$ is the plug-in estimator of Φ_n. Now using estimates $\tilde{\Phi}_n$, $\tilde{\tau}_n^2(\tilde{\psi}_n^*)$, $\tilde{\tau}_n^2(\tilde{\delta}_n)$, $\tilde{s}_n = \tilde{\epsilon}_n^3$, $\tilde{H}_n = \tilde{\sigma}_n^{-2}\text{diag}(\tilde{\epsilon}_n^2)$, $\tilde{\mathbb{A}}_n$ and plug-in estimates for other quantities, we have a consistent estimator for $\tau_n^2(\tilde{\beta}_n)$, $\tilde{\tau}_n^2(\tilde{\beta}_n) = (X_n'\tilde{B}_n'\tilde{B}_nX_n)^{-1}X_n'\tilde{B}_n'\tilde{\mathbb{A}}_n\tilde{B}_nX_n(X_n'\tilde{B}_n'\tilde{B}_nX_n)^{-1}$. We give the following Theorem.

Theorem 6.4 *If Assumptions 6.1-6.5 and 6.6* hold, then we have as* $n \to \infty$, $\tilde{\tau}_n^2(\tilde{\delta}_n) - \tau_n^2(\tilde{\delta}_n) \xrightarrow{p} 0$ *and* $\tilde{\tau}_n^2(\tilde{\beta}_n) - \tau_n^2(\tilde{\beta}_n) \xrightarrow{p} 0$.

For the plug-in estimator of $\tilde{\Phi}_n$, one could use $-\frac{\partial}{\partial\delta_0^\tau}\tilde{\psi}^*(\delta_0)|_{\delta_0=\tilde{\delta}}$ or its first order term of $-\text{E}(\frac{\partial}{\partial\delta_0^\tau}\tilde{\psi}^*(\delta_0))$ using the expressions given at the end of Theorem 6.2. Given the results of Theorem 6.2, we have that this estimator is a consistent estimator of Φ_n.

6.3.3 Monte Carlo Results

We consider a SARAR(3,3) specification with $W_j = M_j$ for $j = 1, 2, 3$. As such we have the following DGP:

$$Y_n = \sum_{j=1}^{3} \lambda_j W_{jn} Y_n + \beta_0 \iota_n + \beta_1 X_{1n} + \beta_2 X_{2n} + u_n, \quad u_n = \sum_{j=1}^{3} \rho_j W_{jn} u_n + \epsilon_n, \tag{6.20}$$

where ι_n is an $n \times 1$ vector of ones corresponding to the intercept, X_{1n} and X_{2n} are the $n \times 1$ vectors containing the values of two fixed regressors generated as random draws from a standard normal distribution and $\epsilon_n = \sigma H_n e_n$. The regression coefficients β is set to $(3, 1, 1)'$, σ is set to 1 and n take values from $\{100, 250, 500\}$. The ways of generating the values for the spatial weights matrix W_{jn}, the heteroskedasticity measure H_n, and the idiosyncratic errors e_n are described below. Each set of Monte Carlo results is based on $1,000$ Monte Carlo samples.

We use two different spatial layouts: (i) Circular Neighbours and (ii) Queen Contiguity. In (i), neighbours occur in the positions immediately ahead and behind a particular spatial unit. In (i) the initial weights matrix we consider W_{0n} has 2, 4, 6, 8 and 10 neighbours with equal proportion. Then we decompose W_{0n} into three distinct matrices s.t. $W_{1n} + W_{2n} + W_{3n} = W_{0n}$, where W_{1n} contains 2 and 4 band of neighbours of W_{0n}, W_{2n} contains 6 and 8 band of neighbours of W_{0n} and W_{3n} contains 10 band of neighbours of W_{0n}. Details on how weight matrices are generated and how other stochastic and non-stochastic quantities are generated are given in Appendix B.

Given row normalised weights matrices, the parameter space for λ and ρ must satisfy, $0 \leq \sum_{j=1}^{3} |\lambda_j| < 1$ and $0 \leq \sum_{j=1}^{3} |\rho_j| < 1$. We follow the following parameter constellations:

Constellation	λ_1	λ_2	λ_3	ρ_1	ρ_2	ρ_3
1	.5	.3	.1	.4	.2	.1
2	.4	.2	.1	.5	.3	.1
3	.2	.2	.2	.2	.2	.2
4	0	0	0	0	0	0
5	.5	.3	.1	0	0	0
6	0	0	0	.5	.3	.1

In case 1, the spatial dependence in the dependent variable Y is at least as strong as the spatial dependence in the disturbance u_n while the opposite holds in case 2. In case 3 the spatial dependence is equal. Cases 4-6 are sub-models developed in SARAR(p, q) beginning with no spatial dependence in case 4 which is the general linear regression model, SARAR(3,0) in case 5 and SARAR(0,3) in case 6. Partial results of these experiments are summarised in Tables 6.1 and 6.2 with additional results available upon request.

From the Monte Carlo results, we observe that the ACQS estimator of δ performs well in all cases, and it generally outperforms QML estimator in terms of bias and rmse. Further, in the case where QML estimator is consistent, ACQS estimator can be less biased than QML estimator, and is as efficient as QML estimator. The relative performance of various estimators of β is much less contrasting than that of various estimators of δ, although it can be seen that ACQS estimator of β is slightly less biased and more efficient than the QML estimator.

Table 1a. Empirical Mean(rmse)[sd] of Estimators of δ of SARAR(3,3). Case when the regular QML estimator is inconsistent under heteroskedasticity. $n = 100$, $\beta = (3, 1, 1)'$, $\sigma = 1$, `Circular Neighbours, REG-1`

δ_0	QML	ACQS	QML	ACQS	QML	ACQS
	Normal Errors		Mixed Normal Errors		Log-Normal Errors	
.5	.322[.142](.131)	.420[.181](.178){.179}	.334[.153](.142)	.453[.171](.169){.159}	.339[.204](.196)	.441[.169](.165){.167}
.3	.172[.183](.170)	.275[.156](.146){.151}	.187[.195](.183)	.280[.161](.154){.158}	.174[.191](.181)	.290[.139](.135){.137}
.1	-.069[.132](.126)	.043[.191](.187){.189}	-.052[.132](.125)	.061[.193](.192){.193}	-.055[.127](.118)	.052[.184](.177){.180}
.4	.366[.120](.111)	.372[.148](.146){.131}	.258[.179](.163)	.391[.182](.181){.176}	.275[.180](.168)	.362[.183](.181){.178}
.2	.063[.125](.122)	.189[.180](.180){.180}	.059[.154](.140)	.169[.117](.116){.116}	.058[.122](.118)	.179[.162](.161){.162}
.1	.058[.171](.165)	.096[.110](.101){.106}	-.046[.180](.168)	.045[.173](.171){.173}	-.063[.174](.167)	.090[.158](.156){.164}
.4	.289[.143](.139)	.387[.196](.184){.190}	.296[.145](.132)	.321[.182](.178){.180}	.215[.127](.111)	.396[.154](.153){.153}
.2	.142[.176](.167)	.144[.159](.153){.164}	.042[.189](.180)	.187[.156](.154){.155}	.053[.186](.170)	.183[.148](.139){.144}
.1	.020[.119](.104)	.084[.184](.177){.181}	-.022[.112](.108)	.055[.177](.174){.175}	-.038[.203](.193)	.074[.157](.157){.157}
.5	.354[.189](.172)	.479[.143](.136){.139}	.348[.132](.127)	.462[.149](.146){.139}	.378[.116](.102)	.459[.140](.137){.124}
.3	.190[.134](.129)	.254[.178](.169){.153}	.184[.133](.125)	.283[.173](.168){.170}	.173[.143](.130)	.294[.148](.143){.141}
.1	.019[.126](.115)	.085[.194](.194){.194}	-.020[.134](.123)	.078[.193](.193){.193}	-.097[.117](.107)	.093[.175](.175){.175}
.2	.023[.138](.125)	.165[.166](.162){.164}	.100[.186](.178)	.175[.167](.164){.174}	.088[.141](.136)	.143[.169](.163){.152}
.2	.061[.177](.163)	.180[.126](.125){.125}	.050[.190](.173)	.183[.116](.115){.115}	.035[.180](.179)	.162[.129](.126){.128}
.2	.062[.176](.161)	.177[.122](.121){.121}	.048[.185](.177)	.179[.112](.111){.112}	.032[.189](.176)	.161[.121](.118){.119}
.2	.035[.199](.182)	.168[.166](.164){.177}	.003[.149](.132)	.178[.117](.114){.128}	.068[.151](.144)	.167[.152](.146){.148}
.2	.001[.208](.194)	.192[.153](.146){.149}	.005[.190](.187)	.183[.130](.130){.130}	.009[.209](.199)	.150[.139](.136){.138}
.2	.005[.183](.177)	.197[.182](.179){.180}	.009[.168](.153)	.185[.163](.155){.160}	.020[.185](.174)	.171[.173](.173){.176}
.0	-.170[.145](.135)	-.016[.176](.171){.173}	-.177[.159](.147)	-.050[.181](.176){.166}	-.175[.148](.137)	-.058[.152](.146){.149}
.0	-.126[.164](.152)	-.035[.133](.131){.145}	-.125[.179](.167)	-.029[.131](.130){.130}	-.125[.171](.169)	-.045[.120](.115){.117}
.0	-.129[.166](.153)	-.036[.126](.124){.125}	-.127[.183](.171)	-.027[.128](.127){.127}	-.126[.181](.179)	-.040[.117](.113){.115}
.0	-.105[.205](.195)	-.029[.170](.168){.169}	-.162[.151](.146)	-.029[.104](.104){.104}	-.184[.189](.175)	-.056[.166](.157){.161}
.0	-.151[.181](.173)	-.060[.116](.110){.113}	-.138[.187](.173)	-.053[.121](.116){.118}	-.138[.182](.178)	-.028[.198](.197){.189}

Table 6.1b. Empirical Mean(rmse)[sd] of Estimators of δ of SARAR(3,3). Case when the regular QML estimator is inconsistent under heteroskedasticity. $n = 250$, $\beta = (3, 1, 1)'$, $\sigma = 1$, `Circular Neighbours, REG-1`

δ_0	QML	ACQS	QML	ACQS	QML	ACQS
	Normal Errors		Mixed Normal Errors		Log-Normal Errors	
.5	.343[.098](.087)	.505[.121](.119){.120}	.346[.109](.091)	.503[.106](.105){.107}	.360[.142](.130)	.503[.100](.100){.105}
.3	.113[.114](.107)	.273[.110](.095){.102}	.114[.159](.113)	.289[.084](.081){.083}	.116[.127](.114)	.284[.073](.072){.087}
.1	-.056[.092](.082)	.105[.118](.116){.117}	-.052[.097](.080)	.104[.130](.125){.127}	-.033[.085](.077)	.081[.123](.116){.119}
.4	.271[.081](.076)	.395[.105](.099){.102}	.274[.136](.109)	.385[.123](.114){.118}	.255[.110](.107)	.387[.121](.114){.117}
.2	.084[.088](.077)	.206[.111](.111){.114}	.075[.091](.089)	.209[.075](.075){.074}	.076[.085](.073)	.202[.103](.103){.102}
.1	-.087[.118](.107)	.086[.067](.067){.067}	-.073[.131](.105)	.104[.105](.101){.103}	-.036[.121](.108)	.091[.098](.097){.098}
.4	.226[.094](.082)	.384[.123](.117){.120}	.221[.096](.082)	.388[.128](.124){.126}	.237[.098](.088)	.393[.105](.099){.102}
.2	.086[.119](.108)	.196[.087](.083){.085}	.075[.136](.124)	.193[.095](.089){.092}	.078[.139](.128)	.206[.085](.080){.082}
.1	-.010[.071](.062)	.103[.125](.116){.121}	-.010[.074](.065)	.102[.121](.118){.119}	-.004[.139](.121)	.075[.104](.098){.101}
.5	.320[.118](.101)	.471[.077](.076){.088}	.352[.097](.084)	.479[.101](.094){.097}	.322[.077](.069)	.508[.098](.098){.087}
.3	.194[.098](.085)	.265[.107](.105){.106}	.101[.088](.078)	.280[.105](.104){.105}	.197[.102](.090)	.307[.076](.075){.076}
.1	-.090[.083](.072)	.084[.140](.132){.136}	-.085[.081](.070)	.106[.150](.145){.148}	-.016[.076](.069)	.091[.110](.110){.111}
.2	.045[.091](.073)	.185[.107](.107){.104}	.030[.124](.111)	.172[.105](.103){.104}	.035[.090](.079)	.197[.106](.106){.105}
.2	.080[.118](.105)	.202[.074](.074){.079}	.069[.111](.102)	.190[.073](.072){.073}	.073[.119](.106)	.201[.082](.082){.081}
.2	.079[.111](.106)	.201[.077](.077){.077}	.064[.123](.117)	.186[.074](.073){.074}	.072[.121](.118)	.202[.070](.070){.075}
.2	.043[.120](.115)	.193[.102](.101){.104}	.060[.091](.082)	.204[.074](.072){.073}	.080[.107](.092)	.204[.102](.099){.100}
.2	.025[.138](.116)	.193[.098](.098){.094}	.038[.134](.119)	.186[.088](.084){.086}	.023[.141](.117)	.183[.089](.081){.085}
.2	.056[.129](.113)	.176[.120](.114){.117}	.065[.107](.098)	.207[.101](.097){.099}	.049[.124](.116)	.204[.110](.100){.105}
.0	-.151[.109](.093)	-.014[.108](.108){.110}	-.145[.121](.096)	-.012[.111](.110){.111}	-.149[.099](.084)	-.019[.098](.097){.098}
.0	-.108[.106](.096)	.000[.081](.081){.083}	-.105[.110](.096)	-.001[.087](.087){.083}	-.106[.124](.113)	-.009[.083](.076){.079}
.0	-.109[.115](.105)	-.001[.071](.071){.079}	-.105[.117](.107)	.000[.085](.085){.081}	-.108[.136](.116)	-.012[.072](.072){.073}
.0	-.102[.134](.101)	-.018[.112](.106){.109}	-.112[.095](.088)	-.018[.076](.068){.072}	-.182[.129](.119)	-.017[.093](.091){.092}
.0	-.121[.122](.116)	-.025[.076](.071){.073}	-.116[.122](.111)	-.033[.073](.072){.075}	-.124[.130](.118)	-.029[.126](.124){.125}
.0	-.145[.128](.112)	-.058[.103](.102){.108}	-.141[.134](.120)	-.025[.105](.098){.101}	-.142[.109](.095)	-.037[.105](.102){.103}

Table 6.1c. Empirical Mean(rmse)[sd] of Estimators of δ of SARAR(3,3). Case when the regular QML estimator is inconsistent under heteroskedasticity. $n = 500$, $\beta = (3, 1, 1)'$, $\sigma = 1$, `Circular Neighbours, REG-1`

δ_0	QML	ACQS	QML	ACQS	QML	ACQS
	Normal Errors		Mixed Normal Errors		Log-Normal Errors	
.5	.304[.068](.054)	.496[.089](.088){.080}	.333[.088](.070)	.497[.064](.063){.076}	.322[.098](.089)	.498[.061](.060){.075}
.3	.110[.090](.077)	.292[.063](.063){.067}	.124[.130](.106)	.294[.064](.064){.070}	.120[.094](.086)	.299[.064](.064){.061}
.1	-.015[.065](.054)	.098[.097](.097){.084}	-.096[.063](.054)	.099[.096](.096){.086}	-.088[.059](.046)	.094[.099](.094){.096}
.4	.225[.087](.055)	.410[.069](.063){.066}	.296[.088](.073)	.398[.085](.083){.084}	.287[.088](.076)	.405[.089](.082){.086}
.2	.000[.061](.051)	.208[.082](.081){.080}	.083[.081](.070)	.197[.076](.074){.075}	.085[.063](.058)	.291[.079](.077){.078}
.1	-.053[.103](.090)	.198[.048](.048){.047}	-.096[.097](.080)	.103[.078](.078){.077}	-.075[.084](.077)	.104[.046](.041){.043}
.4	.256[.072](.066)	.392[.087](.085){.086}	.248[.064](.056)	.394[.082](.082){.080}	.260[.052](.047)	.393[.035](.035){.035}
.2	.006[.098](.087)	.198[.058](.058){.058}	.100[.099](.088)	.193[.070](.067){.069}	.100[.103](.093)	.197[.065](.063){.064}
.1	-.070[.052](.044)	.100[.098](.096){.097}	-.054[.077](.066)	.100[.079](.075){.077}	-.047[.110](.090)	.091[.072](.071){.071}
.5	.384[.081](.076)	.509[.076](.070){.073}	.390[.066](.055)	.505[.089](.087){.088}	.386[.080](.050)	.507[.092](.090){.091}
.3	.182[.073](.057)	.303[.095](.092){.093}	.183[.072](.059)	.303[.071](.070){.071}	.183[.069](.053)	.303[.051](.050){.065}
.1	-.099[.065](.055)	.101[.088](.088){.087}	-.076[.081](.071)	.199[.069](.065){.067}	-.078[.052](.042)	.094[.074](.074){.078}
.2	.046[.056](.047)	.198[.075](.074){.075}	.047[.092](.083)	.192[.099](.099){.099}	.047[.076](.067)	.192[.075](.071){.073}
.2	.084[.075](.063)	.196[.060](.060){.056}	.083[.075](.063)	.199[.075](.075){.075}	.080[.083](.070)	.200[.080](.080){.080}
.2	.083[.075](.063)	.195[.058](.057){.054}	.083[.076](.064)	.197[.075](.075){.075}	.080[.071](.068)	.198[.082](.082){.082}
.2	.084[.098](.087)	.196[.076](.073){.075}	.088[.057](.047)	.198[.059](.059){.052}	.099[.077](.067)	.195[.042](.040){.041}
.2	.042[.101](.083)	.191[.063](.063){.067}	.043[.101](.083)	.198[.049](.047){.048}	.036[.097](.086)	.192[.046](.043){.045}
.2	.077[.101](.079)	.196[.082](.081){.081}	.070[.079](.064)	.197[.065](.063){.064}	.073[.081](.078)	.190[.043](.041){.042}
.0	-.115[.073](.061)	.000[.098](.098){.098}	-.120[.078](.066)	-.007[.092](.091){.080}	-.117[.074](.064)	-.003[.061](.061){.067}
.0	-.101[.080](.060)	.002[.060](.060){.059}	-.104[.071](.061)	-.003[.079](.078){.078}	-.102[.076](.066)	.003[.083](.083){.083}
.0	-.100[.069](.059)	.002[.056](.056){.056}	-.105[.072](.062)	-.004[.077](.076){.077}	-.101[.077](.067)	.001[.079](.079){.079}
.0	-.134[.100](.077)	-.009[.063](.063){.063}	-.122[.089](.070)	.000[.057](.057){.046}	-.113[.087](.079)	-.003[.050](.047){.048}
.0	-.112[.098](.082)	-.005[.056](.054){.055}	-.113[.090](.079)	-.005[.049](.047){.048}	-.119[.092](.079)	-.002[.048](.044){.046}
.0	-.118[.102](.081)	-.001[.079](.078){.076}	-.117[.094](.083)	-.009[.074](.074){.070}	-.125[.062](.049)	-.005[.033](.030){.031}

Table 6.2a. Empirical Mean(rmse)[sd] of Estimators of δ of SARAR(3,3). Case when the regular QML estimator is consistent under heteroskedasticity. $n = 100$, $\beta = (3, 1, 1)'$, $\sigma = 1$, `Queen Contiguity, REG-1`

δ_0	QML	ACQS	QML	ACQS	QML	ACQS
	Normal Errors		Mixed Normal Errors		Log-Normal Errors	
.5	.411[.120](.110)	.463[.110](.108){.112}	.494[.110](.109)	.484[.110](.101){.105}	.451[.117](.107)	.490[.111](.104){.107}
.3	.271[.120](.112)	.250[.128](.126){.127}	.303[.114](.113)	.272[.123](.120){.121}	.262[.204](.200)	.232[.105](.104){.105}
.1	.096[.148](.145)	.085[.140](.140){.142}	.065[.147](.143)	.076[.137](.135){.139}	.084[.143](.137)	.075[.134](.133){.135}
.4	.394[.138](.138)	.392[.126](.126){.132}	.397[.138](.138)	.382[.124](.123){.130}	.405[.130](.130)	.380[.110](.109){.109}
.2	.200[.192](.192)	.204[.195](.194){.193}	.201[.182](.181)	.203[.174](.172){.173}	.199[.183](.183)	.204[.171](.170){.176}
.1	.043[.146](.140)	.044[.129](.125){.127}	.094[.142](.137)	.071[.120](.119){.119}	.064[.132](.130)	.073[.119](.118){.119}
.4	.361[.116](.113)	.381[.110](.109){.111}	.395[.107](.100)	.387[.112](.109){.111}	.386[.106](.103)	.378[.104](.103){.104}
.2	.182[.122](.121)	.173[.124](.123){.122}	.176[.109](.106)	.175[.110](.109){.108}	.187[.107](.107)	.182[.109](.108){.109}
.1	.089[.142](.141)	.088[.143](.142){.143}	.082[.143](.142)	.076[.140](.139){.141}	.083[.135](.134)	.097[.133](.131){.133}
.5	.469[.141](.133)	.464[.122](.112){.117}	.487[.135](.129)	.452[.107](.101){.104}	.475[.119](.113)	.495[.201](.196){.198}
.3	.291[.183](.181)	.264[.168](.166){.173}	.293[.172](.164)	.276[.158](.156){.160}	.264[.183](.171)	.284[.163](.160){.165}
.1	.066[.130](.127)	.065[.109](.107){.117}	.081[.128](.126)	.061[.117](.115){.120}	.061[.124](.123)	.087[.103](.102){.103}
.2	.166[.106](.104)	.178[.108](.107){.106}	.195[.111](.107)	.172[.109](.107){.109}	.163[.113](.109)	.182[.102](.101){.105}
.2	.193[.133](.125)	.185[.141](.135){.138}	.193[.123](.122)	.152[.124](.119){.122}	.184[.124](.117)	.152[.125](.120){.121}
.2	.165[.168](.163)	.164[.175](.170){.172}	.152[.165](.161)	.161[.166](.163){.164}	.185[.144](.138)	.164[.139](.136){.137}
.2	.195[.150](.146)	.164[.149](.147){.146}	.162[.144](.141)	.176[.139](.137){.139}	.161[.141](.138)	.195[.129](.126){.127}
.2	.174[.171](.170)	.194[.178](.178){.174}	.188[.177](.176)	.185[.167](.167){.171}	.166[.164](.161)	.180[.160](.159){.161}
.2	.185[.118](.114)	.179[.114](.113){.115}	.194[.118](.114)	.167[.112](.110){.112}	.187[.187](.185)	.171[.178](.176){.180}
.0	-.005[.105](.105)	.001[.103](.103){.104}	-.019[.105](.104)	-.016[.105](.105){.104}	-.011[.103](.103)	-.012[.110](.110){.110}
.0	-.049[.129](.124)	-.055[.147](.141){.144}	-.040[.119](.116)	-.039[.135](.132){.133}	-.047[.120](.115)	-.048[.138](.133){.135}
.0	-.032[.174](.173)	-.035[.182](.180){.181}	-.036[.178](.176)	-.033[.180](.178){.179}	-.027[.144](.143)	-.027[.154](.153){.154}
.0	-.070[.170](.161)	-.048[.161](.156){.159}	-.050[.155](.150)	-.027[.147](.146){.148}	-.047[.136](.131)	-.017[.120](.119){.125}
.0	-.031[.108](.107)	.004[.107](.107){.107}	-.032[.171](.169)	-.004[.173](.173){.171}	-.027[.172](.170)	-.003[.179](.179){.179}
.0	-.026[.139](.135)	-.018[.134](.134){.136}	-.029[.129](.128)	.000[.129](.129){.128}	-.053[.204](.199)	-.021[.194](.193){.196}

Table 6.2b. Empirical Mean(rmse)[sd] of Estimators of δ of SARAR(3,3). Case when the regular QML estimator is consistent under heteroskedasticity. $n = 250$, $\beta = (3, 1, 1)'$, $\sigma = 1$, `Queen Contiguity, REG-1`

δ_0	QML	ACQS	QML	ACQS	QML	ACQS
	Normal Errors		Mixed Normal Errors		Log-Normal Errors	
.5	.493[.068](.068)	.493[.070](.070){.071}	.494[.065](.065)	.496[.064](.064){.065}	.496[.068](.068)	.493[.068](.061){.064}
.3	.306[.074](.074)	.275[.086](.086){.080}	.290[.079](.079)	.276[.076](.076){.078}	.290[.127](.127)	.296[.077](.070){.066}
.1	.087[.099](.096)	.090[.084](.083){.090}	.075[.095](.090)	.094[.087](.087){.089}	.092[.082](.082)	.090[.082](.082){.085}
.4	.403[.081](.081)	.404[.083](.083){.082}	.400[.087](.087)	.408[.071](.071){.079}	.482[.084](.082)	.406[.063](.063){.069}
.2	.165[.123](.121)	.209[.121](.121){.122}	.204[.115](.115)	.205[.108](.108){.109}	.203[.114](.114)	.206[.105](.100){.111}
.1	.088[.084](.083)	.077[.081](.080){.081}	.106[.081](.081)	.084[.080](.080){.075}	.094[.090](.090)	.081[.072](.072){.075}
.4	.375[.072](.070)	.381[.063](.062){.066}	.387[.067](.065)	.376[.064](.062){.064}	.396[.068](.068)	.393[.064](.064){.066}
.2	.189[.072](.071)	.184[.079](.070){.071}	.197[.068](.068)	.188[.068](.068){.069}	.189[.061](.060)	.171[.060](.068){.069}
.1	.082[.086](.085)	.086[.096](.090){.088}	.098[.098](.098)	.084[.087](.087){.089}	.087[.088](.088)	.073[.083](.081){.084}
.5	.472[.091](.088)	.473[.075](.072){.074}	.473[.083](.080)	.480[.069](.068){.066}	.460[.082](.077)	.466[.126](.122){.125}
.3	.288[.125](.123)	.280[.106](.105){.110}	.276[.105](.105)	.288[.095](.095){.100}	.264[.109](.107)	.300[.102](.102){.104}
.1	.087[.088](.086)	.086[.066](.065){.074}	.096[.076](.076)	.091[.072](.072){.074}	.065[.079](.077)	.107[.068](.063){.065}
.2	.197[.070](.070)	.186[.069](.069){.067}	.177[.072](.069)	.187[.063](.060){.065}	.172[.072](.069)	.177[.069](.067){.068}
.2	.196[.074](.074)	.181[.088](.088){.087}	.194[.071](.071)	.177[.072](.072){.071}	.187[.077](.077)	.188[.074](.073){.075}
.2	.171[.106](.103)	.187[.108](.107){.109}	.197[.107](.107)	.197[.106](.106){.104}	.162[.094](.090)	.177[.083](.082){.086}
.2	.186[.097](.097)	.197[.094](.093){.095}	.202[.087](.087)	.198[.088](.088){.087}	.196[.086](.086)	.195[.078](.078){.081}
.2	.187[.102](.102)	.186[.116](.116){.110}	.192[.113](.113)	.186[.110](.110){.111}	.193[.104](.104)	.190[.106](.105){.102}
.2	.187[.080](.072)	.176[.077](.075){.074}	.188[.078](.077)	.183[.062](.061){.069}	.203[.112](.112)	.191[.119](.111){.114}
.0	-.005[.065](.065)	-.007[.069](.069){.067}	-.009[.068](.068)	-.010[.062](.062){.065}	-.011[.073](.064)	-.012[.069](.068){.070}
.0	-.014[.079](.078)	-.014[.086](.085){.082}	-.017[.072](.071)	-.019[.090](.089){.084}	-.016[.074](.074)	-.017[.082](.081){.077}
.0	-.008[.107](.107)	-.006[.115](.115){.111}	-.011[.113](.112)	-.009[.113](.112){.113}	-.014[.100](.099)	-.012[.096](.096){.097}
.0	-.023[.102](.101)	-.011[.093](.092){.097}	-.018[.100](.099)	-.005[.107](.100){.094}	-.020[.087](.086)	-.009[.079](.079){.082}
.0	-.013[.070](.070)	-.008[.071](.065){.067}	-.021[.118](.110)	-.005[.105](.105){.107}	-.017[.106](.105)	-.002[.114](.114){.113}

Table 6.2c. Empirical Mean(rmse)[sd] of Estimators of δ of SARAR(3,3)
Case when the regular QML estimator is consistent under heteroskedasticity
$n = 500,\ \beta = (3, 1, 1)',\ \sigma = 1,$ `Queen Contiguity, REG-1`

δ_0	QML	ACQS	QML	ACQS	QML	ACQS
	Normal Errors		Mixed Normal Errors		Log-Normal Errors	
.5	.494[.041](.041)	.496[.050](.050){.050}	.493[.042](.042)	.499[.040](.040){.041}	.490[.048](.048)	.493[.045](.045){.046}
.3	.298[.058](.058)	.295[.053](.053){.055}	.295[.054](.054)	.291[.059](.059){.056}	.295[.081](.081)	.292[.044](.044){.047}
.1	.091[.068](.068)	.090[.061](.061){.064}	.091[.064](.064)	.099[.064](.064){.064}	.090[.067](.067)	.092[.055](.055){.060}
.4	.408[.062](.062)	.401[.056](.056){.059}	.406[.066](.066)	.392[.058](.058){.058}	.403[.054](.054)	.399[.050](.050){.049}
.2	.201[.082](.082)	.208[.085](.085){.086}	.206[.081](.081)	.199[.078](.078){.079}	.205[.083](.083)	.204[.078](.078){.079}
.1	.108[.062](.062)	.096[.054](.054){.057}	.103[.068](.068)	.098[.053](.053){.053}	.108[.052](.052)	.105[.059](.059){.055}
.4	.391[.057](.057)	.399[.047](.047){.050}	.395[.044](.044)	.398[.044](.044){.044}	.392[.042](.042)	.396[.048](.048){.046}
.2	.197[.058](.058)	.195[.058](.058){.058}	.191[.049](.049)	.195[.045](.045){.047}	.193[.047](.047)	.198[.046](.046){.047}
.1	.096[.066](.066)	.092[.062](.062){.064}	.100[.068](.068)	.095[.063](.063){.063}	.094[.064](.064)	.094[.056](.056){.059}
.5	.497[.058](.058)	.499[.052](.052){.052}	.493[.057](.057)	.500[.046](.046){.047}	.496[.053](.053)	.494[.081](.081){.089}
.3	.297[.085](.085)	.304[.071](.071){.078}	.296[.080](.080)	.305[.073](.073){.071}	.309[.071](.071)	.307[.079](.079){.075}
.1	.099[.052](.052)	.103[.041](.041){.047}	.095[.060](.060)	.102[.053](.053){.054}	.099[.057](.057)	.100[.041](.041){.046}
.2	.194[.045](.045)	.196[.042](.042){.044}	.197[.050](.050)	.190[.048](.048){.049}	.197[.044](.044)	.191[.040](.040){.042}
.2	.198[.059](.059)	.199[.062](.062){.062}	.194[.057](.057)	.195[.051](.051){.054}	.191[.053](.053)	.195[.055](.055){.054}
.2	.199[.078](.078)	.193[.072](.072){.075}	.196[.074](.074)	.199[.070](.070){.072}	.199[.069](.069)	.196[.060](.060){.061}
.2	.198[.061](.061)	.198[.063](.063){.062}	.196[.070](.070)	.196[.069](.069){.069}	.195[.061](.061)	.194[.056](.056){.057}
.2	.190[.074](.074)	.193[.086](.086){.078}	.197[.077](.077)	.199[.078](.078){.078}	.191[.077](.077)	.191[.075](.075){.076}
.2	.194[.052](.052)	.196[.054](.054){.053}	.190[.058](.058)	.193[.047](.047){.050}	.196[.083](.083)	.194[.074](.074){.079}
.0	.000[.044](.044)	.000[.045](.045){.045}	-.004[.043](.043)	-.005[.045](.045){.044}	-.007[.044](.044)	-.006[.042](.042){.043}
.0	-.013[.054](.054)	-.013[.064](.064){.065}	-.005[.055](.055)	-.005[.055](.055){.055}	-.019[.053](.053)	-.019[.055](.055){.054}
.0	-.009[.073](.073)	-.010[.085](.085){.081}	-.016[.071](.071)	-.016[.082](.082){.080}	-.009[.062](.062)	-.008[.061](.061){.062}
.0	-.012[.070](.070)	-.006[.071](.071){.071}	-.013[.068](.068)	-.007[.068](.068){.068}	-.011[.052](.052)	-.006[.050](.050){.051}
.0	-.005[.042](.042)	.001[.045](.045){.043}	-.013[.074](.074)	-.005[.074](.074){.074}	-.004[.074](.074)	-.003[.084](.084){.080}
.0	-.008[.067](.067)	.000[.069](.069){.068}	-.006[.055](.055)	.001[.055](.055){.055}	-.009[.085](.085)	-.003[.084](.084){.084}

6.4 Fixed Effects Spatial Panel Data Model

Spatial panel data (SPD) models are popular since these models allow a location related dependence structure to be attached to the conventional panel model in terms of spatial dependence or spatial heterogeneity (Anselin et al., 2008),[13] with a wide practical applicability.[14] It allows robustness as fixed effects are allowed to depend on included regressors and forms a platform for different random effects models to be enveloped in.[15] In particular short spatial panels (large number of spatial units over a short time span) seems to be the prevalent setting. In this chapter, we consider the spatial panel model (SPD) with fixed effects when the model suffers from heteroskedasticity of unknown form with a special focus on the short panel case. In the model we consider, spatial correlation appear both in the dependent variable and the disturbance term.

The same array of heteroskedasticity robust estimation techniques for parameters of cross sectional models are unavailable for heteroskedastic SPD models with fixed effects.[16] In terms of heteroskedasticity robust estimation, Moscone and Tosetti (2011), extends the robust GMM estimation methods for the pure cross sectional spatial model, given in Kelejian and Prucha (2010) and Lin and Lee (2010), to the spatial panel framework where they consider unknown heteroskedasticity in a panel model with only spatial error dependence. Badinger and Egger (2015) considers a higher order spatial panel model with heteroskedastic error components and gives a three step 2SLS/GMM robust estimator extend-

[13]See Anselin (2001), Baltagi et al. (2003, 2013), Elhorst (2003), Kapoor et al. (2007), and Lee and Yu (2010b, 2012, 2015) for some related works. Lee and Yu (2010b, 2015) provide surveys of the evolution of SPD models in general.

[14]See Baltagi et al. (2016), Hsieh and Lee (2014), Kelejian and Piras (2014) among others for some recent empirical studies.

[15]See Lee and Yu (2012) for a survey of the fixed effects SPD model vs. the random effects SPD model and Mutl and Pfaffermayr (2011) for another comparison.

[16]General estimation and inference related issues for SPD models with homoskedastic disturbances have been considered in, among others, Baltagi et al. (2003, 2013), Fingleton (2008), Kapoor et al. (2007), Lee and Yu (2010a), and Robinson and Rossi (2015a,b).

ing the methods in Kapoor et al. (2007). However, like in the cross sectional case, this estimator may lack efficiency compared to a pure GMM estimator or a ML based estimator. Since ML based methods provides the most efficient estimate, a ML based method that is simple to implement and robust to unknown heteroskedasticity for this model is useful.[17]

The line-up for this chapter is as follows. First we introduce the spatial panel data model allowing both spatial lag and spatial error and individual-specific effects. QML estimation based on the transformed likelihood function is considered along with conditions under which the usual QML estimator can be consistent even under unknown heteroskedasticity. Next we introduce the ACQS estimator that is generally robust against unknown heteroskedasticity and non-normality, and provide methods for robust inferences. We also provide details of the Monte Carlo experiment conducted for this model.

6.4.1 Robustness of QML estimator against unknown heteroskedasticity

In this section we outline briefly the QML estimation of the one-way fixed effects panel data model with a spatial autoregressive lagged dependent variable and a spatial autoregressive error structure (SARAR) where the truly idiosyncratic component is first set to be independent and identically distributed (iid) as given in Lee and Yu (2010a). Then, we examine the properties of this QML estimator when the errors are independent but not identically distributed (inid). We provide

[17] When the disturbances are homoskedastic in a spatial panel model with fixed effects, Lee and Yu (2010a) show that direct estimation of all the parameters in the model (including the fixed effects parameters), yields consistent QML estimators (QML estimators) for all the parameters when the number of spatial units (n) becomes large, except the QML estimator for the variance parameter when the time dimension (T) is small. However, upon transformation of the model, QML estimators for all the parameters become consistent irrespective of the size of T and the estimates other than the variance estimate are identical to those from the direct approach. However, Lee and Yu (2010a) does not consider heteroskedasticity.

conditions under which the QML estimator is robust against heteroskedasticity of unknown form, and derive asymptotic distribution of this robust QML estimator.

Consider the SARAR panel data model with individual fixed effects,

$$Y_{nt} = \lambda_0 W_{1n} Y_{nt} + X_{nt}\beta_0 + \mathbf{c}_n + U_{nt}, \quad U_{nt} = \rho_0 W_{2n} U_{nt} + V_{nt}, \quad t = 1, \ldots, T, \tag{6.21}$$

for $t = 1, 2, ..., T$, where, $Y_{nt} = (y_{1t}, y_{2t}, ..., y_{nt})'$ is the vector of dependent variables, X_{nt} represents the $n \times k$ matrix containing the values of k non-stochastic time varying regressors, $V_{nt} = (v_{1t}, v_{2t}, ..., v_{nt})'$ is the vector of disturbances where, $v_{it} \sim iid(0, \sigma^2)$. β is a $k \times 1$ vector of regression coefficients, λ and ρ are the spatial parameters with W_{1n} and W_{2n} being the respective $n \times n$ non-stochastic spatial weights matrices which may or may not be the same in practice. $\mathbf{c}_n$ is the an $n \times 1$, time invariant vector of individual fixed effects.

The individual fixed effects in this model cause the incidental parameter problem (Neyman and Scott, 1948) which can be avoided by transformation of variables. Lee and Yu (2010a) suggests a transformation using the orthonormal matrix of the deviation from the time mean operator, $J_T = I_T - \frac{1}{T}\iota_T\iota_T'$.[18] The orthonormal matrix is given as $[F_{T,T-1}, \frac{1}{\sqrt{T}}\iota_T]$, where $F_{T,T-1}$ is the $T \times T-1$ eigenvector matrix corresponding to the eigenvalues 1 and $\frac{1}{\sqrt{T}}\iota_T$ corresponds to the eigenvalue 0.[19] Let a typical transformed variable be, $Z^*_{n1}, ..., Z^*_{n,T-1} = (Z_{n1}, ..., Z_{n,T})F_{T,T-1}$ then the original model can be written as, $Y^*_{nt} = \lambda W_{1n} Y^*_{nt} + X^*_{nt}\beta + U^*_{nt}$, where $U^*_{nt} = \rho W_{2n} U^*_{nt} + V^*_{nt}$ for $t = 1, ..., T-1$. The effective sample size after the transformations is $N = n(T-1)$. Upon stacking the vectors and matrices, we have the

[18]Lee and Yu (2010a) illustrates a QML estimator using either a direct approach or a transformation approach. In the direct approach parameters are jointly estimated along with the individual effects. The transformation approach eliminates the individual effects and thus removes the problem of incidental parameters when n becomes large, and yields consistent estimates when either n or T is large.

[19]Transformation using orthonormal matrix of J_T, allows the resulting transformed disturbances to be independent over t. However if the variables are transformed using a time mean operator, then the transformed disturbances fail to be independent over t.

following compact form of the model,

$$\mathbf{Y}_N = \lambda_0 \mathbf{W}_{1N}\mathbf{Y}_N + \mathbf{X}_N\beta_0 + \mathbf{U}_N, \quad \mathbf{U}_N = \rho_0\mathbf{W}_{2N}\mathbf{U}_N + \mathbf{V}_N, \quad t = 1, \ldots, T-1, \tag{6.22}$$

where, $\mathbf{Y}_N = (Y^{*\prime}_{n1}, \ldots, Y^{*\prime}_{n,T-1})'$, $\mathbf{U}_N = (U^{*\prime}_{n1}, \ldots, U^{*\prime}_{n,T-1})'$, $\mathbf{V}_N = (V^{*\prime}_{n1}, \ldots, V^{*\prime}_{n,T-1})'$, $\mathbf{X}_N = (X^{*\prime}_{n1}, \ldots, X^{*\prime}_{n,T-1})'$, and $\mathbf{W}_{rN} = I_{T-1} \otimes W_{rn},\ r = 1, 2$. The transformed errors, $\{V^*_{nt}\}$ are uncorrelated for all n and t. Using the identity $(V^{*\prime}_{n1}, \ldots, V^{*\prime}_{n,T-1})' = (F'_{T,T-1} \otimes I_n)(V'_{n1}, \ldots, V'_{nT})'$, we have that, $\mathrm{E}[(V^{*\prime}_{n1}, \ldots, V^{*\prime}_{n,T-1})'(V^*_{n1}, \cdots, V^*_{n,T-1})] = \sigma^2_0(F'_{T,T-1} \otimes I_n)(F_{T,T-1} \otimes I_n) = \sigma^2_0 I_N$. Hence, $\{v^*_{it}\}$ are iid $N(0, \sigma^2_0)$ if the original errors $\{v_{it}\}$ are iid $N(0, \sigma^2_0)$. However, if $\{v_{it}\}$ is non-normal, then the transformed errors may fail to be independent even though they are uncorrelated.

Let $\delta = (\lambda, \rho)'$, $\zeta = (\beta', \delta')'$, $\theta = (\zeta', \sigma^2)'$. Then the Gaussian log-likelihood is,

$$\ell_N(\theta) = -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{A}_{1N}(\lambda)| + \ln|\mathbf{A}_{2N}(\rho)| - \tfrac{1}{2\sigma^2}\mathbf{V}'_N(\zeta)\mathbf{V}_N(\zeta), \tag{6.23}$$

where $\mathbf{V}_N(\zeta) = \mathbf{A}_{2N}(\rho)[\mathbf{A}_{1N}(\lambda)\mathbf{Y}_N - \mathbf{X}_N\beta]$, $\mathbf{A}_{1N}(\lambda) = I_N - \lambda\mathbf{W}_{1N}$, and $\mathbf{A}_{2N}(\rho) = I_N - \rho\mathbf{W}_{2N}$. Let $\mathbf{Y}_N(\delta) = \mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{Y}_N$ and $\mathbf{X}_N(\rho) = \mathbf{A}_{2N}(\rho)\mathbf{X}_N$. The constrained QML estimators of β and σ^2, given δ is:

$$\hat{\beta}_N(\delta) = [\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)\mathbf{Y}_N(\delta), \tag{6.24}$$

$$\hat{\sigma}^2_N(\delta) = \tfrac{1}{N}\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)\mathbf{Y}_N(\delta), \tag{6.25}$$

where $\mathbf{M}_N(\rho) = I_N - \mathbf{X}_N(\rho)[\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)$. Substituting $\hat{\beta}_N(\delta)$ and $\hat{\sigma}^2_N(\delta)$ into (6.23) gives the concentrated quasi log likelihood function of (δ):

$$\ell^c_N(\delta) = -\tfrac{N}{2}(\ln(2\pi) + 1) + \ln|\mathbf{A}_{1N}(\lambda)| + \ln|\mathbf{A}_{2N}(\rho)| - \tfrac{N}{2}\ln\tilde{\sigma}^2_N(\delta). \tag{6.26}$$

Maximizing (6.26) gives the unconstrained QML estimator $\hat{\delta}_N$ of δ_0 and thus the unconstrained QML estimators of β_0 and σ_0^2 as $\hat{\beta}_N(\hat{\delta}_N)$ and $\hat{\sigma}_N^2(\hat{\delta}_N)$.[20] Lee and Yu (2010a) show that $\hat{\theta}_N$ is $\sqrt{N}$-consistent when the errors are iid. Next we turn to some further issues that can be considered. First we examine conditions under which the regular QML estimator $\hat{\theta}_N$ remains consistent when errors are heteroskedastic. From a practical point of view, it may not be possible to validate these conditions for a given dataset, especially considering that heteroskedasticity is of unknown form. Hence, secondly we give methods to adjust the regular QML estimator $\hat{\theta}_N$ so that it becomes generally consistent under unknown heteroskedasticity. However, inferences based on these estimates are not possible without having a consistent estimate of the standard errors. As such, thirdly we give methods for estimating standard errors required for inference based on an outer product of the gradient method.

Lin and Lee (2010) shows that QML estimator of the usual cross sectional SAR model without a time varying index, is inconsistent when the errors are heteroskedastic when a certain necessary condition is violated. In this section we show that the violation of a similar condition will render the QML estimates for the for the parameters of the model given in (6.21) inconsistent in general. As such, it is at least theoretically possible to find situations where the original QML estimates are consistent even when the disturbances are heteroskedastic. Suppose now we have disturbances that are independent but not identically distributed (inid), i.e., $v_{it} \sim inid(0, \sigma^2 h_i),\ i = 1, \dots, n,\ t = 1, \dots, T$, where $\frac{1}{n}\sum_{i=1}^{n} h_i = 1$ and $h_i > 0$.[21] Consider the score function derived from (6.23),

[20] The computation of the two determinant terms can be simplified using, $|\mathbf{A}_{1N}(\lambda)| = |I_{n-1} - \lambda W_{1n}^*|^{T-1} = \left(\frac{1}{1-\lambda}|I_n - \lambda W_{1n}|\right)^{T-1} = \left(\frac{1}{1-\lambda}\prod_{i=1}^{n}(1-\lambda\omega_{1i})\right)^{T-1}$, where ω_{1i} are the eigenvalues of W_{1n}. A similar expression can be derived for $|\mathbf{B}_N(\rho)|$. Refer Lee and Yu (2010a) and Griffith (1988).

[21] Note that σ^2 is the average of $\mathrm{Var}(v_{it})$. Under homoskedasticity, $h_i = 1, \forall i$.

$$\frac{d}{d\theta}\ell_N(\theta) = \begin{cases} \frac{1}{\sigma^2}\mathbf{V}_N'(\zeta)\mathbf{X}_N(\rho), \\ \frac{1}{2\sigma^4}\left[\mathbf{V}_N'(\zeta)\mathbf{V}_N(\zeta) - N\sigma^2\right], \\ \frac{1}{\sigma^2}\mathbf{V}_N'(\zeta)\mathbf{A}_{2N}(\rho)\mathbf{W}_{1N}\mathbf{Y}_N - \text{tr}(\mathbf{G}_{1N}(\lambda)), \\ \frac{1}{\sigma^2}\mathbf{V}_N'(\zeta)\mathbf{W}_{2N}[\mathbf{A}_{1N}(\lambda)\mathbf{Y}_N - \mathbf{X}_N\beta] - \text{tr}(\mathbf{G}_{2N}(\rho)), \end{cases} \tag{6.27}$$

where $\mathbf{G}_{rN} = \mathbf{W}_{rN}\mathbf{A}_{rN}^{-1},\ r = 1, 2$. The necessary condition for consistency is satisfied for the components with respect to β and σ^2, however, it is not always true for the components with respect to the spatial parameters λ and ρ when the disturbances are heteroskedastic. Consider,[22]

$$\begin{aligned} \frac{d}{d\lambda}\frac{1}{N}\ell_N(\theta_0) &= \frac{1}{N\sigma_0^2}\mathbf{V}_N'\bar{\mathbf{G}}_{1N}\mathbf{V}_N - \frac{1}{N}\text{tr}(\mathbf{G}_{1N}) + o_p(1) \\ &= \frac{1}{N\sigma_0^2}\mathbf{V}_N'\left(\bar{\mathbf{G}}_{1N} - \frac{1}{N}\text{tr}(\mathbf{G}_{1N})I_N\right)\mathbf{V}_N + o_p(1) \\ &= \frac{1}{N}\left(\text{tr}(\mathbf{H}_N\bar{\mathbf{G}}_{1N}) - \frac{1}{N}\text{tr}(\bar{\mathbf{G}}_{1N})\text{tr}(\mathbf{H}_N)\right) + o_p(1) \\ &= \text{Cov}(\bar{g}_{1n,ii}, h_i) + o_p(1) \text{ and} \\ \frac{d}{d\rho}\frac{1}{N}\ell_N(\theta_0) &= \frac{1}{N\sigma_0^2}\mathbf{V}_N'\mathbf{G}_{2N}\mathbf{V}_N - \frac{1}{N}\text{tr}(\mathbf{G}_{2N}) + o_p(1) \\ &= \frac{1}{N\sigma_0^2}\mathbf{V}_N'\left((\mathbf{G}_{2N}) - \frac{1}{N}\text{tr}(\mathbf{G}_{2N})I_N\right)\mathbf{V}_N + o_p(1) \\ &= \frac{1}{N}\left(\text{tr}(\mathbf{H}_N\mathbf{G}_{2N}) - \frac{1}{N}\text{tr}(\mathbf{G}_{2N})\text{tr}(\mathbf{H}_N)\right) + o_p(1) \\ &= \text{Cov}(g_{2n,ii}, h_i) + o_p(1), \end{aligned}$$

where $\bar{\mathbf{G}}_{1N} = \mathbf{A}_{2N}\mathbf{G}_{1N}\mathbf{A}_{2N}^{-1}$ and $\mathbf{G}_{2N}$ are block diagonal matrices, $\bar{g}_{1n,ii}$ and $g_{2n,ii}$ are the diagonal elements of the block matrices of $\bar{\mathbf{G}}_{1n}$ and $\mathbf{G}_{2n}$ respectively, and $\mathbf{H}_N = I_{T-1} \otimes H_n$, where $H_n = \text{diag}(h_1, \ldots, h_n)$. Hence, $\text{plim}_{N\to\infty}\frac{1}{N}\frac{\partial}{\partial\theta}\ell_N(\theta_0) =$

[22] Note that all the quantities defined at the true parameter is represented with a suppressed variable notation, e.g., $\mathbf{A}_{1N} \equiv \mathbf{A}_{1N}(\lambda_0)$ and $\mathbf{G}_{1N} \equiv \mathbf{G}_{1N}(\lambda_0)$ and so on. In addition diag(.) denotes the symmetric matrix formed by the diagonal elements of a matrix, diagv(.) denotes the vector formed by the diagonal elements of a matrix and tr(.) denotes the trace of a matrix.

0 if $\text{Cov}(\bar{g}_{1n,ii}, h_i)$ and $\text{Cov}(g_{2n,ii}, h_i)$ goes to 0 in the limit. Several cases can be identified as cases where this condition is satisfied: (i) Lin and Lee (2010) shows that this condition is satisfied when almost all the diagonal elements of the matrix G_{rn} are equal[23], (ii) the condition is also related to the variability of the number of neighbours, i.e., $\text{Var}(\bar{g}_{1n}) \to 0$ and $\text{Var}(g_{2n}) \to 0$, boils down to $\text{Var}(k_{rn}) \to 0$, where k_{rn} is the vector of number of neighbours for each spatial unit[24] and (iii) when heteroskedasticity arises due to reasons unrelated to the number of neighbours[25]. Hence it seems necessary to investigate the asymptotic results when the QML estimators are robust. We refresh the regularity conditions as follows:

Assumption 6.7: *The true spatial parameters δ_0 is in the interior of a compact set Δ.*

Assumption 6.8: *$V_{nt} \sim (0, \sigma_0^2 H_n)$, where $H_n = \text{diag}(h_1, \ldots, h_n)$, such that $\frac{1}{n}\sum_{i=1}^{n} h_i = 1$ and $h_i > 0, \forall i$ and $E|v_{it}|^{4+\eta} < c$ for some $\eta > 0$ and constant c for all n and t.*

Assumption 6.9: *The elements of the regressor matrix $\mathbf{X}_N$ are non-stochastic and uniformly bounded and $\lim_{N\to\infty} \frac{1}{N}\mathbf{X}'_N\mathbf{X}_N$ exists and is non-singular.*

Assumption 6.10: *The spatial weights matrices W_{rn}, $r = 1, 2$ are uniformly bounded in absolute value in both row and column sums and the diagonal elements are zero.*

Assumption 6.11: *The matrices $\mathbf{A}_{rN}$ are non-singular and $\mathbf{A}_{rN}^{-1}$ is uniformly bounded in absolute value in both row and column sums. Further, $\mathbf{A}_{1N}^{-1}(\lambda)$ and $\mathbf{A}_{2N}^{-1}(\rho)$ are uniformly bounded in either row or column sums, uniformly in $\delta \in \Delta$.*

[23] such as the case where, (a) the group sizes are the same in a Group Interaction weights matrix, (b) where there are equal number of neighbours before and after in a Circular world or (c) when the spatial weights matrix is sparse

[24] This is because the diagonal elements of the higher order powers of W_{rn} inversely relate to k_{rn}, see Anselin (2003). For example, when W_{rn} is row-normalised and symmetric, $\text{diag}(W_{rn}^2) = \{k_{rn,i}^{-1}\}$. $\text{Var}(k_{rn}) = o(1)$ can be seen to be true for many popular spatial layouts such as Rook, Queen, group interactions, etc., see Yang (2010).

[25] such as when heteroskedasticity is a function of the exogenous regressors X_{tn}.

Assumption 6.12:

Either (a): $\lim_{n\to\infty} \mathcal{H}_N(\rho)$ *is non-singular* $\forall\rho$ *and* $\lim_{n\to\infty} \mathcal{Q}_{1n}(\rho) \neq 0$ *for* $\rho \neq \rho_0$*; or (b)* $\lim_{n\to\infty} \mathcal{Q}_{2n}(\delta) \neq 0$ *for* $\delta \neq \delta_0$*, where*

$\mathcal{H}_N(\rho) = \frac{1}{N}(\mathbf{X}_N,\, \mathbf{G}_{1N}\mathbf{X}_N\beta_0)'\mathbf{A}'_{2N}\mathbf{A}_{2N}(\mathbf{X}_N,\, \mathbf{G}_{1N}\mathbf{X}_N\beta_0)$,

$\mathcal{Q}_{1N}(\rho) = \frac{1}{N}(\ln|\sigma_0^2\mathbf{A}'^{-1}_{2N}\mathbf{A}^{-1}_{2N}| - \ln|\sigma_N^2(\rho)\mathbf{A}'^{-1}_{2N}(\rho)\mathbf{A}^{-1}_{2N}(\rho)|)$

$\mathcal{Q}_{2N}(\delta) = \frac{1}{N}(\ln|\sigma_0^2\mathbf{A}'^{-1}_{2N}\mathbf{A}'^{-1}_{1N}\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N}| - \ln|\sigma_N^2(\delta)\mathbf{A}'^{-1}_{2N}(\rho)\mathbf{A}'^{-1}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}(\lambda)\mathbf{A}^{-1}_{2N}(\rho)|)$

$\sigma_N^2(\delta) = \frac{1}{N}\sigma_0^2\mathrm{tr}(\mathbf{H}_N\mathbf{A}'^{-1}_{2N}\mathbf{A}'^{-1}_{1N}\mathbf{A}'_{1N}(\lambda)\mathbf{A}'_{2N}(\rho)\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N})$ *and*

$\sigma_N^2(\rho) = \sigma_N^2(\delta)|_{\lambda=\lambda_0}$.

Assumption 6.13: *n is large and T is finite or large.*[26]

Let, $\mathbf{Q}_{1N} = \bar{\mathbf{G}}'_{1N}\bar{\mathbf{G}}_{1N}$, $\mathbf{Q}_{2N} = \mathbf{G}'_{2N}\mathbf{G}_{2N}$, $\mathbf{Q}_{3N} = \mathbf{G}_{2N}\bar{\mathbf{G}}_{1N}$.

Assumption 6.14: *The covariance between the vector of diagonal elements of* $\mathbf{H}_N$ *and the vectors of diagonal elements of,* $\bar{\mathbf{G}}_{1N}$, $\mathbf{G}_{2N}$, $\mathbf{Q}_{1N}$, $\mathbf{Q}_{2N}$, $\mathbf{Q}_{3N}$, $\mathbf{Q}'_{3N}\mathbf{Q}_{3N}$, $\mathbf{Q}_{2N}\bar{\mathbf{G}}_{1N}$, $\mathbf{G}'_{2N}\bar{\mathbf{G}}_{1N}$, $\bar{\mathbf{G}}'_{1N}\mathbf{Q}_{3N}$ *is zero.*

Theorem 6.5 *Under Assumptions 6.7-6.14, as* $N \to \infty$, $\hat{\theta}_N \xrightarrow{p} \theta_0$.

6.4.2 Adjusted Concentrated Quasi Score Estimation

In this section we look at an alternative robust estimator when the necessary conditions for consistency of the QML estimator, $\lim_{n\to\infty}\mathrm{Cov}(\bar{g}_{1n,ii}, h_i) = 0$ and $\lim_{n\to\infty}\mathrm{Cov}(g_{2n,ii}, h_i) = 0$, are violated. This can happen when h_i is proportional to the number of neighbours k_i for each spatial unit and $\lim_{n\to\infty}\mathrm{Var}(k_n) \neq 0$. However, even when these conditions hold for a given dataset, it is impossible to check as the form of heteroskedasticity is unknown. Inspired by Lin and Lee (2010), and Kelejian and Prucha (2010), Moscone and Tosetti (2010) proposed a heteroskedasticity robust GMM estimator for a spatial panel data model with

[26] The case of finite n and large T is of less interest for two reasons: (a) the incidental parameter problem does not arise and (b) the problem of heteroskedasticity is not as severe as the varying measures of h_i, skewness and kurtosis can be consistently estimated. When $T > n$, the spatial structure is less influential as the weight matrix can be estimated explicitly using the T observations for each n.

one way fixed effects and spatial autoregressive errors (with no extension to include a spatial lagged dependent variable). Also building on the works of Kelejian and Prucha (2010), Badinger and Egger (2015), proposed a robust GMM based estimator for a higher order spatial autoregressive panel data model with heteroskedasticity and error components. In this chapter, a robust quasi maximum likelihood estimator (QML estimator) is proposed, to estimate the parameters of the model given in (6.21) which includes a spatial lagged dependent variable as well as a spatial error dependent variable. The proposed estimator is defined by adjusting the concentrated quasi score function for δ. The method can be easily extended to a two way fixed effects model or a higher order spatial panel data model. We further introduce a method for estimating its robust standard errors.

The method

As evident from the analysis given after (6.27), the inconsistency of the QML estimator of the model parameters of (6.21) is caused by elements of the score function derived with respect to the spatial parameters, failing to reach a probability limit of zero. As such we look at an adjustment to the score function that allows it to reach the required probability limit. Although it is possible to adjust the full score function (6.27) to attain a robust estimator, with a decent asymptotic performance, the finite sample performance is diluted by the fact that the full score function does not take into account the variability caused by estimating the other model parameters β_0 and σ_0^2. As such adjusting the concentrated quasi score function is desirable in attaining both finite sample as well as asymptotic performance in the robust estimator, since the concentrated quasi score function captures the variability coming from estimating β_0 and σ_0^2.

The averaged concentrated quasi score function derived by taking the derivatives of the concentrated quasi log-likelihood function (6.26) with respect to δ:

$$\tilde{\psi}_N(\delta) = \begin{cases} \dfrac{\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)[\bar{\mathbf{G}}_{1N}(\delta) - \frac{1}{n}\text{tr}(\mathbf{G}_{1N}(\lambda))I_N]\mathbf{Y}_N(\delta)}{\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)\mathbf{Y}_N(\delta)}, \\ \dfrac{\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)[\bar{\mathbf{G}}_{2N}(\rho) - \frac{1}{n}\text{tr}(\mathbf{G}_{2N}(\rho))I_N]\mathbf{Y}_N(\delta)}{\mathbf{Y}'_N(\lambda)\mathbf{M}_N(\rho)\mathbf{Y}_N(\delta)}, \end{cases} \tag{6.28}$$

where $\bar{\mathbf{G}}_{1N}(\delta) = \mathbf{A}_{2N}(\rho)\mathbf{G}_{1N}(\lambda)\mathbf{A}_{2N}^{-1}(\rho)$ and $\bar{\mathbf{G}}_{2N}(\rho) = \mathbf{G}_{2N}(\rho)\mathbf{M}_N(\rho)$.

Using $\tilde{\psi}_N(\delta)$, the regular QML estimator is defined as, $\hat{\delta}_N = \arg\{\tilde{\psi}_N(\delta) = 0\}$. For $\hat{\delta}_N$ to be consistent under unknown heteroskedasticity, it is necessary that $\text{E}(\tilde{\psi}_N)$ equals or tends to zero, see van der Vaart (1998, ch. 5). However, this does not hold if there exists unknown heteroskedasticity and the conditions of Theorem 6.5 are violated. In other words, a condition required to attain consistency is, $\text{E}[\mathbf{Y}'_N\mathbf{M}_N(\bar{\mathbf{G}}_{rN} - \frac{1}{n}\text{tr}(\mathbf{G}_{rN})I_N)\mathbf{Y}_N]$ equals or tends to zero for $r = 1, 2$. Thus, we can adjust (6.28) to ensure that it is zero in expectation. To that effect, note, $\text{E}(\mathbf{Y}'_N\mathbf{M}_N\bar{\mathbf{G}}_{rN}\mathbf{Y}_N) = \sigma_0^2\text{tr}(\mathbf{H}_N\mathbf{M}_N\bar{\mathbf{G}}_{rN}) = \sigma_0^2\text{tr}(\mathbf{H}_N\text{diag}(\mathbf{M}_N\bar{\mathbf{G}}_{rN}))$. Hence, a possible way to go is to replace $\frac{1}{n}\text{tr}(\mathbf{G}_{rN})$ of (6.28) with $\text{diag}(\mathbf{M}_N\bar{\mathbf{G}}_{rN})$. However, this introduces an additional $\mathbf{M}_N$ to $\text{E}(\mathbf{Y}'_N\mathbf{M}_N\bar{\mathbf{G}}_{rN}\mathbf{Y}_N)$ and hence the final adjustment made is of the form $\text{diag}(\mathbf{M}_N)^{-1}\text{diag}(\mathbf{M}_N\bar{\mathbf{G}}_{rN})$.

Thus, we have the adjusted concentrated quasi score function,

$$\tilde{\psi}^*_N(\delta) = \begin{cases} \dfrac{\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}^\circ_{1N}(\delta)\mathbf{Y}_N(\delta)}{\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)\mathbf{Y}_N(\delta)}, \\ \dfrac{\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}^\circ_{2N}(\rho)\mathbf{Y}_N(\delta)}{\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)\mathbf{Y}_N(\delta)}, \end{cases} \tag{6.29}$$

where $\bar{\mathbf{G}}^\circ_{rN}(\delta) = \bar{\mathbf{G}}_{rN}(\delta) - \text{diag}(\mathbf{M}_N)^{-1}(\rho)\text{diag}[\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{rN}(\delta)]$, $r = 1, 2$ and we define the adjusted concentrated quasi score (ACQS) estimator as,

$$\tilde{\delta}_N = \arg\{\tfrac{d}{d\delta}\ell^{c*}_N(\delta) = 0\}. \tag{6.30}$$

Once a heteroskedasticity robust estimator for δ_0 is obtained, heteroskedasticity robust estimators for β_0 and σ_0^2 follows from $\tilde{\beta}_N \equiv \hat{\beta}_N(\tilde{\delta}_N)$ and $\tilde{\sigma}_N^2 \equiv \hat{\sigma}_N^2(\tilde{\delta}_N)$. Hence the ACQS estimators of the model (6.21) are computationally as simple as the QML estimators while being generally consistent under unknown heteroskedasticity and preserving the nature of being robust against non-normality.

Asymptotic properties of the ACQS estimator

In this section we derive the asymptotic properties of the ACQS estimator. To do so, first, in order to ensure that the adjusted estimating equation given in (6.29) uniquely identifies δ_0, the Assumption 6.12 needs to be adjusted as follows:

Assumption 6.12*: $\forall \delta \neq \delta_0$,

(i) $\lim_{N\to\infty} \frac{1}{N}[\beta_0' \mathbf{X}_N' \mathbf{A}_{2N}'^{-1} \mathbf{A}_{1N}'^{-1} \mathbf{\Omega}_{1N}(\delta) \mathbf{A}_{1N}^{-1} \mathbf{A}_{2N}^{-1} \mathbf{X}_N \beta_0$
$+\sigma_0^2 \mathrm{tr}(\mathbf{H}_N \mathbf{A}_{2N}'^{-1} \mathbf{A}_{1N}'^{-1} \mathbf{\Omega}_{1N}(\delta) \mathbf{A}_{1N}^{-1} \mathbf{A}_{2N}^{-1})] \neq 0$ *and*

(ii) $\lim_{N\to\infty} \frac{1}{N}[\sigma_0^2 \mathrm{tr}(\mathbf{H}_N \mathbf{A}_{2N}'^{-1} \mathbf{A}_{1N}'^{-1} \mathbf{\Omega}_{2N}(\delta) \mathbf{A}_{1N}^{-1} \mathbf{A}_{2N}^{-1})] \neq 0$, *where*

$\mathbf{\Omega}_{1N}(\delta) = \mathbf{A}_{1n}'(\lambda) \mathbf{A}_{2n}'(\rho)[\bar{\mathbf{G}}_{1N}(\delta) - \mathrm{diag}(\bar{\mathbf{G}}_{1N}(\delta))] \mathbf{A}_{2N}(\rho) \mathbf{A}_{1N}(\lambda)$ *and*

$\mathbf{\Omega}_{2N}(\delta) = \mathbf{A}_{1n}'(\lambda) \mathbf{A}_{2n}'(\rho)[\bar{\mathbf{G}}_{2N}(\rho) - \mathrm{diag}(\bar{\mathbf{G}}_{2N}(\rho))] \mathbf{A}_{2N}(\rho) \mathbf{A}_{1N}(\lambda)$.

Asymptotic normality of the ACQS estimators can be established using CLT for linear quadratic forms given in Kelejian and Prucha (2001). Consider the normalised and adjusted concentrated quasi score function at δ_0,

$$\sqrt{N}\tilde{\psi}_N^* = \begin{cases} \frac{1}{\sqrt{N}\sigma_0^2}\left(\mathbf{V}_N' \mathbf{B}_{1N} \mathbf{V}_N + \mathbf{c}_{1N}' \mathbf{V}_N\right) + o_p(1), \\ \frac{1}{\sqrt{N}\sigma_0^2}\left(\mathbf{V}_N' \mathbf{B}_{2N} \mathbf{V}_N + \mathbf{c}_{2N}' \mathbf{V}_N\right) + o_p(1), \end{cases} \tag{6.31}$$

where $\mathbf{B}_{rN} = \mathbf{M}_N \bar{\mathbf{G}}_{rN}^\circ$, $\mathbf{c}_{rN} = \mathbf{M}_N \bar{\mathbf{G}}_{rN}^\circ \mathbf{X}_N(\rho_0)\beta_0$ and $\mathrm{diag}(\mathbf{B}_{rN}) = 0_{n\times n}$ by construction for $r = 1, 2$. As $\hat{\sigma}_N^2(\lambda_0) = \frac{1}{N}\mathbf{V}_N' \mathbf{M}_N \mathbf{V}_N = \frac{1}{N}\mathrm{E}(\mathbf{V}_N' \mathbf{M}_N \mathbf{V}_N) + o_p(1) = \frac{\sigma_0^2}{N}\mathrm{tr}(\mathbf{H}_N \mathbf{M}_N) + o_p(1) = \sigma_0^2 + o_p(1)$, and it follows that $\hat{\sigma}_N^{-2}(\lambda_0) = \sigma_0^{-2} + o_p(1)$.

Let $\Sigma_N(\cdot)$ denote the first-order variance-covariance of a normalised quantity, for example, $\Sigma_N(\tilde{\psi}_N^*)$ is the first-order term of $\mathrm{Var}(\sqrt{N}\tilde{\psi}_N^*)$. By (6.31) and Lemma A.3, we derive $\Sigma_N(\tilde{\psi}_N^*)$ which has the following components,

$$\begin{aligned}
\Sigma_{N,11}(\tilde{\psi}_N^*) &= \tfrac{1}{N}\textstyle\sum_{i=1}^n\sum_{t=1}^T(b_{1N,it}^2h_i^2\kappa_i + \tfrac{2}{\sigma_0^4}b_{1N,it}c_{1N,it}s_i) + \tfrac{1}{n\sigma_0^2}\mathbf{c}'_{1N}\mathbf{H}_N\mathbf{c}_{1N} \\
&\quad + \tfrac{1}{N}\mathrm{tr}[\mathbf{H}_N\mathbf{B}_{1N}(\mathbf{H}_N\mathbf{B}_{1N} + \mathbf{H}_N\mathbf{B}'_{1N})], \\
\Sigma_{N,12}(\tilde{\psi}_N^*) &= \Sigma_{N,21}(\tilde{\psi}_N^*) \\
&= \tfrac{1}{N}\textstyle\sum_{i=1}^n\sum_{t=1}^T(b_{1N,it}b_{2N,it}h_i^2\kappa_i + \tfrac{2}{\sigma_0^4}b_{2N,it}c_{1N,it}s_i + \tfrac{2}{\sigma_0^4}b_{1N,it}c_{2N,it}s_i) \\
&\quad + \tfrac{1}{n\sigma_0^2}\mathbf{c}'_{1N}\mathbf{H}_N\mathbf{c}_{2N} + \tfrac{1}{N}\mathrm{tr}[\mathbf{H}_N\mathbf{B}_{1N}(\mathbf{H}_N\mathbf{B}_{2N} + \mathbf{H}_N\mathbf{B}'_{2N})] \text{ and} \\
\Sigma_{N,22}(\tilde{\psi}_N^*) &= \tfrac{1}{N}\textstyle\sum_{i=1}^n\sum_{t=1}^T(b_{2N,it}^2h_i^2\kappa_i + \tfrac{2}{\sigma_0^4}b_{2N,it}c_{2N,it}s_i) + \tfrac{1}{n\sigma_0^2}\mathbf{c}'_{2N}\mathbf{H}_N\mathbf{c}_{2N} \\
&\quad + \tfrac{1}{N}\mathrm{tr}[\mathbf{H}_N\mathbf{B}_{2N}(\mathbf{H}_N\mathbf{B}_{2N} + \mathbf{H}_N\mathbf{B}'_{2N})],
\end{aligned}$$

where $b_{rN,it}$ are the diagonal elements of $(F_{T,T-1}\otimes I_n)\mathbf{B}_{rN}(F'_{T,T-1}\otimes I_n)$, c_{rN} are the elements of the vector $\mathbf{c}'_{rN}(F'_{T,T-1}\otimes I_n)$, $s_i = \mathrm{E}(v_{it}^3)$ and κ_i is the measure of excess kurtosis of v_{it}. The score function $\sqrt{N}\tilde{\psi}_N^*$ can be rewritten as a linear quadratic form of the original disturbances, $\{V_{it}\}$ and by Assumptions 6.9-6.11, we also have that $(F_{T,T-1}\otimes I_n)\mathbf{B}_{rN}(F'_{T,T-1}\otimes I_n)$ is uniformly bounded in row and column sums. Then by the multivariate CLT for linear-quadratic forms given in Lemma A.3, we have,

$$\sqrt{N}\tilde{\psi}_N^* \xrightarrow{D} \mathcal{N}\Big(\mathbf{0},\ \lim_{N\to\infty}\Sigma_N(\tilde{\psi}_N^*)\Big). \tag{6.32}$$

This leads to the following theorem regarding the asymptotic properties of the ACQS estimator $\tilde{\delta}_N$ of the spatial parameter δ_0.

Theorem 6.6 *Under Assumptions 6.7-6.11 and 6.12*, the ACQS estimator* $\tilde{\delta}_N$ *is consistent and asymptotically normal, i.e., as* $N\to\infty$, $\tilde{\delta}_N \xrightarrow{p} \delta_0$ *and*

$$\sqrt{N}(\tilde{\delta}_N - \delta_0) \xrightarrow{D} \mathcal{N}\Big(\mathbf{0},\ \lim_{N\to\infty}\Sigma_N(\tilde{\delta}_N)\Big),$$

where $\Sigma_N(\tilde{\delta}_N) = \Phi_N^{-1}\Sigma_N(\tilde{\psi}_N^*)\Phi_N^{-1}$ *and* $\Phi_N = -\mathrm{E}[\frac{\partial}{\partial \delta_0'}\tilde{\psi}^*(\delta_0)]$ *or its first-order term, has the following components:*

$$\Phi_{N,11} = \tfrac{1}{N}\mathrm{tr}[\mathbf{H}_N(\bar{\mathbf{G}}_{1N}^{\circ}\bar{\mathbf{G}}_{1N} + \bar{\mathbf{G}}_{1N}'\bar{\mathbf{G}}_{1N}^{\circ} - \dot{\bar{\mathbf{G}}}_{1N}^{\circ})] + \tfrac{1}{N\sigma_0^2}\mathbf{c}_{1N}'\eta_{1N},$$

$$\Phi_{N,12} = \tfrac{1}{N}\mathrm{tr}[\mathbf{H}_N(\bar{\mathbf{G}}_{1N}^{\circ}\mathbf{G}_{2N} + \bar{\mathbf{G}}_{2N}'\bar{\mathbf{G}}_{1N}^{\circ} - \dot{\bar{\mathbf{G}}}_{2N}^{\circ} - \mathbf{G}_{2N}\bar{\mathbf{G}}_{1N}^{\circ} + \bar{\mathbf{G}}_{2N}\bar{\mathbf{G}}_{1N}^{\circ})] + \tfrac{1}{N\sigma_0^2}\mathbf{c}_{1N}'\eta_{2N}$$

$$\Phi_{N,21} = \tfrac{1}{N}\mathrm{tr}[\mathbf{H}_N(\bar{\mathbf{G}}_{2N}^{\circ}\bar{\mathbf{G}}_{1N} + \bar{\mathbf{G}}_{1N}'\bar{\mathbf{G}}_{2N}^{\circ})] + \tfrac{1}{N\sigma_0^2}\mathbf{c}_{2N}'\eta_{1N} \text{ and}$$

$$\Phi_{N,22} = \tfrac{1}{N}\mathrm{tr}[\mathbf{H}_N(\bar{\mathbf{G}}_{2N}^{\circ}\mathbf{G}_{2N} + \mathbf{G}_{2N}'\bar{\mathbf{G}}_{2N}^{\circ} - \dot{\bar{\mathbf{G}}}_{2N}^{\circ} - \mathbf{G}_{2N}\bar{\mathbf{G}}_{2N}^{\circ} + \bar{\mathbf{G}}_{2N}\bar{\mathbf{G}}_{2N}^{\circ})] + \tfrac{1}{N\sigma_0^2}\mathbf{c}_{2N}'\eta_{2N}, \text{ where}$$

$$\eta_{1N} = \mathbf{A}_{2N}\mathbf{G}_{1N}\mathbf{X}_N\beta,\ \eta_{2N} = \bar{\mathbf{G}}_{2N}\mathbf{A}_{2N}\mathbf{X}_N\beta,$$

$$\dot{\bar{\mathbf{G}}}_{1N}^{\circ} = \bar{\mathbf{G}}_{1N}^2 - \mathrm{diag}(\mathbf{M}_N)^{-1}\mathrm{diag}(\mathbf{M}_N\bar{\mathbf{G}}_{1N}^2),$$

$$\begin{aligned}\dot{\bar{\mathbf{G}}}_{2N}^{\circ} = {} & \mathbf{G}_{2N}\bar{\mathbf{G}}_{2N} + \mathbf{G}_{2N}\dot{\mathbf{M}}_N - \mathrm{diag}(\mathbf{M}_N)^{-1}\mathrm{diag}(\mathbf{M}_N\mathbf{G}_{2N}\bar{\mathbf{G}}_{2N}) \\ & -\mathrm{diag}(\mathbf{M}_N^{-1})\mathrm{diag}(\mathbf{M}_N\mathbf{G}_{2N}\dot{\mathbf{M}}_N) - \mathrm{diag}(\mathbf{M}_N)^{-1}\mathrm{diag}(\dot{\mathbf{M}}_N\bar{\mathbf{G}}_{2N}) \\ & +\mathrm{diag}(\mathbf{M}_N^{-1}\dot{\mathbf{M}}_N\mathbf{M}_N^{-1})\mathrm{diag}(\mathbf{M}_N\bar{\mathbf{G}}_{2N}),\end{aligned}$$

$$\mathbf{P}_N = I_N - \mathbf{M}_N \text{ and } \dot{\mathbf{M}}_N = \mathbf{M}_N\mathbf{G}_{2N}\mathbf{P}_N + \mathbf{P}_N\mathbf{G}_{2N}'\mathbf{M}_N.$$

As in the previous section, we consider δ_0 and β_0 to be the main parameters of interest. A Taylor expansion of $\hat{\beta}_N(\hat{\delta}_0)$ around $\delta = \delta_0$ gives $\hat{\beta}_N(\hat{\delta}_N) = \hat{\beta}_N(\delta_0) + \dot{\beta}_N(\delta_0)(\hat{\delta}_N - \delta_0) + O_p(\frac{1}{N})$, where $\dot{\beta}(\delta_0) = \frac{\partial}{\partial\delta}\hat{\beta}_N\Big|_{\delta=\delta_0}$. This is used to derive the asymptotic distribution of $\tilde{\beta}_N$.

Theorem 6.7 *Under Assumptions 6.7-6.11 and 6.12*, the ACQS estimator* $\tilde{\beta}_N$ *is consistent and asymptotically normal, i.e., as* $n \to \infty$, $\tilde{\beta}_N \xrightarrow{p} \beta_0$, *and*

$$\sqrt{n}(\tilde{\beta}_N - \beta_0) \xrightarrow{D} \mathcal{N}\big(0,\ \lim_{N\to\infty}\Sigma_N(\tilde{\beta}_N)\big),$$

where, $\Sigma_N(\tilde{\beta}_N) = [\mathbf{X}_N'(\rho_0)\mathbf{X}_N(\rho_0)]^{-1}\mathbf{X}_N'(\rho_0)\mathbb{A}_N\mathbf{X}_N(\rho_0)[\mathbf{X}_N'(\rho_0)\mathbf{X}_N(\rho_0)]^{-1}$, $\mathbb{A}_N = 2\sqrt{N}(\sigma_0^{-2}\mathbf{B}_{1N}^d\mathbf{S}_N + \mathbf{H}_N\mathbf{c}_{1N},\ \sigma_0^{-2}\mathbf{B}_{2N}^d\mathbf{S}_N + \mathbf{H}_N\mathbf{c}_{2N})\Phi_N^{-1}(\eta_N, \mathbf{0}_N)' + \Sigma_{N,11}(\tilde{\delta}_N)\eta_N\eta_N' + N\sigma_0^2\mathbf{H}_N$, *where* $\mathbf{B}_{rN}^d = \mathrm{diag}(\mathbf{B}_{rN})$, $\mathbf{S}_N = \mathrm{E}(\mathbf{V}_N^3)$, $\Sigma_{N,11}(\tilde{\delta}_N)$ *is the top-right corner* *element of* $\Sigma_N(\tilde{\delta}_N)$*, and* 0_N *is an* $N \times 1$ *vector of 0's.*[27]

For asymptotically valid inferences using Theorems 6.6 and 6.7, we need a consistent estimator of $\Sigma_N(\tilde{\psi}^*_N)$ and a consistent estimate of $\Sigma_N(\tilde{\beta}_N)$. As before a plug-in estimator will work for Φ_N, but not for $\Sigma_N(\tilde{\psi}^*_N)$ as it requires the estimation of higher-order moments. Hence we look at an alternative method.

Robust estimation of VC matrix

The first-order variance of the adjusted score, $\Sigma_N(\tilde{\psi}^*_N)$, contains second, third and fourth moments of V_{it} which vary across i, and hence a simple White-type estimator (White, 1980) may not be suitable, which in turn makes $\Sigma_N(\tilde{\delta}_N)$ infeasible.[28] To overcome this difficulty, we follow the idea of Baltagi and Yang (2013) to decompose the numerator of the adjusted score into a sum of uncorrelated terms, and then use the outer product of gradients (OPG) to estimate the variance of this score function which in turn leads to a consistent estimate of $\Sigma(\tilde{\delta}_N)$. To illustrate the method, denote the numerator of (6.31) by,

$$\mathbf{Q}_N(\mathbf{V}_N) = \begin{cases} \mathbf{V}'_N \mathbf{B}_{1N} \mathbf{V}_N + \mathbf{c}'_{1N} \mathbf{V}_N, \\ \mathbf{V}'_N \mathbf{B}_{2N} \mathbf{V}_N + \mathbf{c}'_{2N} \mathbf{V}_N. \end{cases} \tag{6.33}$$

[27] As before, the limiting distribution of $\tilde{\sigma}^2_N$ can be easily derived. However, it is of little use as any inference on $\tilde{\sigma}^2_N$ requires the consistent estimation of $\frac{1}{N}\sum_{i=1}^{n} \mathrm{Var}(\mathbf{V}^2_{N,i}) = \frac{\sigma^4}{N}\sum_{i=1}^{N}(\kappa_{N,i} + h^2_{N,i})$ which cannot be done.

[28] Stock and Watson (2008) illustrated, that the White estimator used to estimate the variance-covariance matrix of the disturbances of a fixed effects panel data model when T is fixed, is inconsistent. Further, they gave a consistent heteroskedasticity robust estimate for the variance-covariance matrix. Their method was later extended to the spatial panel data model by Badinger and Egger (2015) under a GMM estimation setting. However, compared to their approach, the technique we recommend is much easier to implement and does not rely on the estimation of higher order moments of the disturbances. Kelejian and Prucha (2007) gives a non-parametric HAC estimation in a spatial framework concentrating on the cross sectional model. However, this method was never extended to the present framework.

$\mathbf{Q}_N$ is not a sum of uncorrelated components, but can be made to be so (Baltagi and Yang, 2013). First decompose the non-stochastic matrices $\mathbf{B}_{rN}$ as,

$$\mathbf{B}_{rN} = \mathbf{B}^u_{rN} + \mathbf{B}^l_{rN} + \mathbf{B}^d_{rN}, \tag{6.34}$$

where $\mathbf{B}^u_{rN}$, $\mathbf{B}^l_{rN}$ and $\mathbf{B}^d_{rN}$ are respectively, an upper triangular, a lower triangular and the diagonal matrices of $\mathbf{B}_{rN}$. Let $\Upsilon_g = (\zeta_{1N,g} + b_{1N,gg}V^*_g + c_{1N,g},\ \zeta_{2N,g} + b_{2N,gg}V^*_g + c_{2N,g})'$, where $\zeta_{rN} = (\mathbf{B}^{u\prime}_{rN} + \mathbf{B}^l_{rN})\mathbf{V}_N$, $b_{rN,gg}$ are the diagonal elements of $\mathbf{B}^d_{rN}$ and V^*_g are the transformed errors, for $r = 1, 2$ and $g = 1, \ldots, N$. To apply the OPG method to estimate the variance of the score function, $\mathbf{Q}_N(\mathbf{V}_N)$ has to be written as a sum of N uncorrelated terms. The usual method is to write $\mathbf{Q}_N(\mathbf{V}_N)$ as,

$$\mathbf{Q}_N(\mathbf{V}_N) = \textstyle\sum_{g=1}^N V^*_g \Upsilon_g. \tag{6.35}$$

However, since the transformed errors are not necessarily independent, $\{V^*_g \Upsilon_g\}$ may fail to be uncorrelated if the higher order moments (specifically third and fourth order) of the original errors are non-zero.

We first look at the case where the original errors are inid Gaussian so that the transformed errors will also have the same properties. In that case $\{V^*_g \Upsilon_g\}$ will be uncorrelated for $g = 1, \ldots, N$. Then the OPG of (6.35) gives a consistent estimator of the variance of the score as follows,

$$\tilde{\Sigma}_N(\tilde{\psi}^*_N) = \tfrac{1}{N\tilde{\sigma}^4_N} \textstyle\sum_{g=1}^N \tilde{V}^{*2}_g \tilde{\Upsilon}_g \tilde{\Upsilon}'_g, \tag{6.36}$$

where $\tilde{\mathbf{V}}_{nt}$ are the residuals computed from the ACQS estimators $\tilde{\theta}_N = (\tilde{\beta}'_N, \tilde{\sigma}^2_N, \tilde{\delta}_N)'$ and $\tilde{\Upsilon}_g = (\tilde{\zeta}_{1N,g} + \tilde{b}_{1N,gg}\tilde{V}^*_g + \tilde{c}_{1N,g},\ \tilde{\zeta}_{2N,g} + \tilde{b}_{2N,gg}\tilde{V}^*_g + \tilde{c}_{2N,g})'$. Using (6.36), define the estimator of $\Sigma_N(\tilde{\delta}_N)$ as,

$$\tilde{\Sigma}_N(\tilde{\delta}_N) = \tilde{\Phi}^{-1}_N \tilde{\Sigma}_N(\tilde{\psi}^*_N) \tilde{\Phi}^{-1}_N, \tag{6.37}$$

where Φ_N can be estimated by the plug-in estimator of $-\frac{d}{\delta_0}\tilde{\psi}_N^*|_{\delta_0=\tilde{\delta}_N}$ or the 1st-order term $-\mathrm{E}(\frac{d}{d\delta_0}\tilde{\psi}_N^*)$ using the terms given in Theorem 6.6 and $\tilde{\mathbf{H}}_N = \frac{1}{\tilde{\sigma}_N^2}\mathrm{diag}(\tilde{V}_{11}^{*2}, \ldots, \tilde{V}_N^{*2})$. We give the following theorem.

Theorem 6.8 *If Assumptions 6.7-6.11 and 6.12* and 6.13 hold, and in addition if $V_{nt} \sim N(0, \sigma_0^2 H_n)$ then we have as $N \to \infty$, $\tilde{\Sigma}_N(\tilde{\delta}_N) - \Sigma_N(\tilde{\delta}_N) \xrightarrow{p} 0$.*

Now suppose the disturbances are not Gaussian. In this case we no longer have $\{V_g^* \Upsilon_g\}$ to be uncorrelated for $g = 1, \ldots, N$ so that the OPG of (6.35) is not a valid estimator of the variance of the score. However, as shown in the Appendix G, when T is finite, we have,

$$\mathrm{Var}[\mathbf{Q}_N(\mathbf{V}_N)] = \textstyle\sum_{g=1}^N \mathrm{E}(V_g^{*2}\Upsilon_g\Upsilon_g') + o_p(1). \tag{6.38}$$

where the covariance term, $2\sum_{g=2}^N \sum_{h=1}^{g-1} \mathrm{E}(V_g^* \Upsilon_{g,1} V_h^* \Upsilon_{h,1})$ becomes asymptotically negligible, so that the estimator given in (6.36) is an asymptotically valid estimator.

Theorem 6.9 *If Assumptions 6.7-6.11 and 6.12* hold, and in addition if T is finite, then we have, as $N \to \infty$, $\tilde{\Sigma}_N(\tilde{\delta}_N) - \Sigma_N(\tilde{\delta}_N) \xrightarrow{p} 0$.*

Given the consistent estimator for $\Sigma_N(\tilde{\delta}_N)$, a consistent estimator for $\Sigma_N(\tilde{\beta}_N)$ can be given as,

$$\tilde{\Sigma}_N(\tilde{\beta}_N) = (\mathbf{X}_N'\tilde{\mathbf{A}}_{2N}'\tilde{\mathbf{A}}_{2N}\mathbf{X}_N)^{-1} X_N'\tilde{\mathbf{A}}_{2N}'\tilde{\mathbb{A}}_N\tilde{\mathbf{A}}_{2N}\mathbf{X}_N(\mathbf{X}_N'\tilde{\mathbf{A}}_{2N}'\tilde{\mathbf{A}}_{2N}\mathbf{X}_N)^{-1}, \tag{6.39}$$

where $\tilde{\mathbb{A}}_N = 2\sqrt{N}(\tilde{\sigma}_0^{-2}\tilde{\mathbf{B}}_{1N}^d\tilde{\mathbf{S}}_N + \tilde{\mathbf{H}}_N\tilde{\mathbf{c}}_{1N},\ \tilde{\sigma}_0^{-2}\tilde{\mathbf{B}}_{2N}^d\tilde{\mathbf{S}}_N + \tilde{\mathbf{H}}_N\tilde{\mathbf{c}}_{2N})\tilde{\Phi}_N^{-1}(\tilde{\eta}_N, \mathbf{0}_N)' + \tilde{\Sigma}_{N,11}^2(\tilde{\delta}_N)\tilde{\eta}_N\tilde{\eta}_N' + N\tilde{\sigma}_0^2\tilde{\mathbf{H}}_N$ and $\tilde{\mathbf{S}}_N = \tilde{\mathbf{V}}_N^3$. We give the following corollary.

Corollary 6.2 *Under the conditions in Theorem 6.8, $\tilde{\Sigma}_N(\tilde{\beta}_N) - \Sigma_N(\tilde{\beta}_N) \xrightarrow{p}$ 0.*

6.4.3 Monte Carlo Results

Extensive Monte Carlo experiments were run to investigate the finite sample performance of the QML estimator $\hat{\delta}_N$ and the ACQS estimator $\tilde{\delta}_N$ proposed in this chapter, and their impacts on the estimators of β_0 and σ_0^2, with respect to changes in the sample size, spatial layouts, error distributions and the model parameters when the disturbances are heteroskedastic. We consider cases where the QML estimator are robust against heteroskedasticity and the cases it is not. The simulations are carried out based on the following data generation process:

$$Y_{nt} = \lambda_0 W_n Y_{nt} + X_{1,nt}\beta_1 + X_{2,nt}\beta_2 + \mathbf{c}_n + U_{nt}, \quad U_{nt} = \rho_0 W_n U_{nt} + V_{nt}, \quad t = 1, 2, 3,$$

where $X_{1,nt}$ and $X_{2,nt}$ are fixed regressors and $V_{nt} = \sigma H_n e_{nt}$. Regression coefficients β is set to $(3, 1, 1)'$, σ is set to 1, δ takes values from $\{-0.5, -0.25, 0, 0.25, 0.5\}$, n take values from $\{50, 100, 250, 500\}$ and T is initially set to be 3. The ways of generating the regressors, the spatial weights matrix W_n, the heteroskedasticity measure H_n, and the idiosyncratic errors e_{nt} are described in Appendix B. Each set of Monte Carlo results is based on $5,000$ Monte Carlo samples.

Tables 6.3-6.5 summarise partial estimation results for δ, where in each table, the Monte Carlo means, root mean square errors (rmse) and the standard errors (se) of the estimators are reported. To analyse the finite sample performance of the proposed OPG based robust standard error estimators, we also report the averaged se of the regular QML estimator when it is heteroskedasticity robust and the averaged se of the ACQS estimator based on Theorem 6.9. The main observations made from the Monte Carlo results are summarised as follows:

(i) For the case where QML estimator is consistent such as in Queen contiguity given in Table 6.1, both estimators show less bias. In addition ACQS estimator can be significantly less biased than QML estimator and is as efficient as QML estimator.

(ii) For the cases where the original QML estimator is inconsistent as given in Tables 6.2-6.3, ACQS estimator provides a useful consistent alternative with significantly less bias with little or no impact on the efficiency.

(iii) The OPG-based estimates of the robust standard errors of λ_0 and ρ_0 performs well with their values very close to their Monte Carlo counterparts.

(iv) As the theory suggest, the QML estimate for the covariate effects remains consistent under heteroskedasticity. The ACQS estimator for the covariate effects (unreported for brevity) performs well as well.

(v) A second set of results with large T relative to n was carried out by setting $T = 15$ and $n = 20$. The results (unreported for brevity) show that the ACQS estimator for δ_0 and the OPG based estimate for the standard errors continue to perform well.

The case of large T relative to n is of particular interest especially due to the effect it may have on the performance of the OPG estimate of the standard errors. The validity of the OPG estimator of the variance of the score function depends on the condition that the terms $\{V_g^* \Upsilon_g\}$ for $g = 1, \ldots, N$ are asymptotically uncorrelated. For the cases of Gaussian errors or finite T, we show that the OPG estimator is valid. However, when T gets large the viability of the OPG method is questionable. However, our Monte Carlo results suggest that even for large T the estimator works well. The case where $T > n$ was not considered as in this case alternative models than the one given in (6.21) is more suitable.

6.5 Conclusion

In this chapter we consider robust estimation and inference for spatial regression models where the disturbances are heteroskedastic of unknown form. In contrast to the available methods in the literature that focuses only on specific spatial regression model, in this chapter we provide a likelihood based method of robust estimation that can be widely applied to a wide class of spatial regression models. In addition, the likelihood based method has the advantage of being efficient. The method proposed works by making a deliberate adjustment to the concentrated quasi score function and hence, is named the adjusted concentrated quasi score (ACQS) estimator.

In order to facilitate robust inference, we also provide a means of estimating the standard errors based on the outer product of the gradient of the adjusted concentrated quasi score function.

These techniques are illustrated using a SARAR(p, q) model and a fixed effects spatial panel data model with a spatial lagged dependent variable and spatial error dependent variable of order one. The related asymptotic theory is given where consistency and the asymptotic distribution is given for the ACQS estimators. The OPG method for estimating the standard errors are also given for these two models along with the consistency of the estimator.

Extensive Monte Carlo experiments were carried out to evaluate the performance of the proposed methods in the context of the spatial models considered. The results are very promising, some of which are presented here.

Table 6.3a. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is consistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Queen Contiguity, REG-1, DGP 1`

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.474[.202](.200)	.490[.209](.207){.190}	.452[.239](.234)	.449[.244](.238){.234}
		.25	.462[.190](.186)	.470[.195](.191){.180}	.225[.266](.265)	.221[.268](.267){.266}
		.00	.468[.166](.163)	.470[.168](.165){.158}	-.017[.275](.274)	-.021[.273](.272){.279}
		-.25	.469[.150](.147)	.472[.151](.148){.149}	-.257[.271](.271)	-.258[.267](.267){.271}
		-.50	.472[.138](.135)	.476[.138](.136){.129}	-.501[.271](.271)	-.500[.267](.267){.270}
	.00	.50	.007[.234](.234)	-.001[.239](.239){.239}	.443[.232](.225)	.443[.234](.227){.226}
		.25	-.002[.218](.218)	-.004[.222](.222){.220}	.204[.260](.256)	.202[.261](.257){.250}
		.00	-.008[.213](.213)	-.007[.216](.216){.210}	-.036[.281](.279)	-.038[.279](.277){.273}
		-.25	-.010[.199](.198)	-.008[.200](.200){.200}	-.278[.286](.284)	-.277[.283](.282){.280}
		-.50	-.018[.190](.190)	-.013[.192](.191){.200}	-.508[.281](.281)	-.503[.278](.278){.280}
	-.50	.50	-.469[.225](.223)	-.480[.226](.225){.215}	.450[.211](.205)	.450[.211](.205){.192}
		.25	-.475[.222](.221)	-.480[.224](.223){.220}	.196[.252](.246)	.194[.252](.245){.239}
		.00	-.484[.222](.221)	-.485[.223](.223){.219}	-.049[.277](.273)	-.001[.275](.271){.268}
		-.25	-.487[.218](.218)	-.486[.220](.220){.217}	-.288[.286](.284)	-.274[.284](.281){.281}
		-.50	-.489[.219](.219)	-.490[.221](.221){.221}	-.532[.288](.287)	-.521[.285](.284){.280}
100	.50	.50	.472[.169](.167)	.470[.169](.166){.151}	.485[.179](.178)	.490[.177](.177){.172}
		.25	.474[.144](.142)	.474[.143](.140){.150}	.244[.194](.194)	.250[.191](.191){.200}
		.00	.481[.119](.118)	.481[.118](.117){.118}	-.005[.196](.196)	-.003[.192](.192){.195}
		-.25	.486[.099](.097)	.490[.098](.097){.093}	-.253[.193](.193)	-.249[.190](.190){.192}
		-.50	.487[.087](.086)	.490[.087](.086){.083}	-.504[.186](.185)	-.498[.183](.183){.185}
	.00	.50	-.003[.189](.189)	-.002[.188](.188){.190}	.474[.168](.166)	.479[.167](.164){.162}
		.25	-.008[.177](.177)	-.008[.176](.176){.169}	.231[.194](.193)	.229[.191](.190){.189}
		.00	-.009[.165](.165)	-.008[.164](.164){.154}	-.018[.209](.208)	-.014[.205](.204){.199}
		-.25	-.011[.152](.152)	-.011[.151](.150){.143}	-.256[.210](.210)	-.252[.206](.206){.200}
		-.50	-.011[.143](.143)	-.011[.143](.142){.135}	-.499[.207](.207)	-.494[.204](.204){.199}
	-.50	.50	-.486[.181](.181)	-.485[.180](.179){.174}	.474[.151](.149)	.469[.151](.148){.148}
		.25	-.495[.174](.174)	-.500[.172](.172){.169}	.228[.181](.180)	.230[.179](.177){.177}
		.00	-.494[.173](.173)	-.493[.171](.171){.170}	-.022[.202](.201)	-.023[.199](.197){.196}
		-.25	-.501[.169](.169)	-.500[.167](.167){.162}	-.263[.212](.212)	-.261[.208](.208){.208}
		-.50	-.501[.169](.169)	-.500[.167](.167){.160}	-.510[.216](.216)	-.504[.211](.211){.214}
250	.50	.50	.486[.118](.118)	.490[.121](.120){.119}	.489[.128](.127)	.490[.130](.130){.128}
		.25	.486[.098](.097)	.488[.099](.098){.096}	.248[.134](.134)	.250[.135](.135){.133}
		.00	.487[.081](.080)	.490[.081](.080){.078}	.001[.135](.135)	.000[.134](.134){.132}
		-.25	.490[.068](.068)	.500[.068](.068){.066}	-.247[.128](.128)	-.250[.128](.128){.127}
		-.50	.493[.059](.059)	.500[.059](.059){.058}	-.500[.122](.122)	-.500[.121](.121){.121}
	.00	.50	.005[.139](.139)	.000[.141](.141){.140}	.482[.116](.115)	.485[.117](.116){.113}
		.25	.001[.127](.127)	.000[.129](.129){.128}	.234[.135](.134)	.240[.136](.135){.132}
		.00	-.007[.115](.114)	-.007[.115](.115){.115}	-.006[.141](.141)	-.004[.141](.141){.140}
		-.25	-.006[.105](.104)	-.005[.105](.105){.105}	-.255[.141](.141)	-.254[.141](.141){.140}
		-.50	-.005[.098](.098)	-.004[.098](.098){.097}	-.502[.136](.136)	-.500[.136](.136){.136}
	-.50	.50	-.486[.127](.127)	-.491[.128](.127){.127}	.481[.100](.098)	.484[.099](.098){.096}
		.25	-.490[.126](.126)	-.493[.126](.126){.126}	.233[.122](.121)	.240[.122](.121){.121}
		.00	-.493[.125](.125)	-.500[.126](.125){.124}	-.014[.141](.140)	-.013[.141](.140){.140}
		-.25	-.497[.123](.123)	-.497[.123](.123){.121}	-.260[.149](.148)	-.258[.148](.148){.146}
		-.50	-.500[.118](.118)	-.500[.118](.118){.118}	-.505[.148](.148)	-.502[.147](.147){.146}
500	.50	.50	.492[.082](.082)	.500[.083](.083){.083}	.497[.089](.089)	.497[.089](.089){.088}
		.25	.494[.066](.066)	.495[.066](.066){.064}	.250[.095](.095)	.250[.095](.095){.092}
		.00	.496[.052](.052)	.500[.052](.052){.052}	-.001[.093](.093)	.000[.093](.093){.092}
		-.25	.497[.045](.045)	.500[.045](.045){.045}	-.251[.088](.088)	-.250[.088](.088){.088}
		-.50	.497[.041](.041)	.500[.041](.041){.040}	-.501[.086](.086)	-.500[.086](.086){.085}
	.00	.50	.002[.095](.095)	.001[.095](.095){.095}	.492[.078](.078)	.492[.078](.077){.076}
		.25	-.003[.088](.088)	-.003[.088](.087){.087}	.246[.092](.092)	.246[.092](.092){.091}
		.00	-.002[.079](.079)	-.002[.078](.078){.078}	-.004[.098](.098)	-.003[.097](.097){.097}
		-.25	-.002[.071](.071)	-.002[.071](.071){.071}	-.253[.098](.098)	-.251[.097](.097){.097}
		-.50	-.001[.067](.067)	-.001[.067](.067){.067}	-.503[.096](.096)	-.500[.095](.095){.095}
	-.50	.50	-.497[.086](.086)	-.500[.086](.086){.086}	.494[.065](.065)	.500[.065](.065){.065}
		.25	-.498[.087](.087)	-.500[.087](.087){.086}	.244[.085](.085)	.243[.085](.085){.083}
		.00	-.499[.085](.085)	-.499[.084](.084){.084}	-.004[.096](.096)	-.001[.096](.096){.094}
		-.25	-.502[.082](.082)	-.500[.082](.082){.082}	-.252[.102](.102)	-.252[.101](.101){.101}
		-.50	-.502[.081](.081)	-.501[.080](.080){.080}	-.502[.101](.101)	-.500[.100](.100){.101}

Table 6.3b. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is consistent under heteroskedasticity
$T = 3, \beta = (1,1)', \sigma = 1$, Queen Contiguity, REG-1, DGP 2

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.475[.201](.200)	.472[.208](.206){.220}	.451[.239](.234)	.450[.243](.237){.237}
		.25	.467[.183](.180)	.467[.187](.184){.173}	.227[.255](.254)	.230[.256](.255){.258}
		.00	.469[.165](.162)	.470[.167](.164){.160}	-.016[.268](.267)	-.012[.266](.265){.265}
		-.25	.469[.152](.148)	.480[.152](.149){.140}	-.255[.268](.268)	-.255[.264](.264){.260}
		-.50	.471[.143](.140)	.480[.143](.140){.145}	-.503[.269](.269)	-.500[.264](.264){.259}
	.00	.50	.008[.229](.229)	.001[.233](.233){.217}	.446[.223](.217)	.445[.225](.218){.199}
		.25	-.003[.213](.213)	-.004[.216](.216){.215}	.208[.250](.246)	.206[.250](.246){.244}
		.00	-.010[.206](.205)	-.009[.207](.207){.205}	-.034[.269](.267)	-.032[.267](.264){.266}
		-.25	-.012[.193](.193)	-.010[.195](.195){.192}	-.271[.275](.274)	-.270[.271](.270){.267}
		-.50	-.019[.193](.193)	-.012[.193](.193){.196}	-.510[.281](.281)	-.505[.275](.275){.276}
	-.50	.50	-.469[.225](.223)	-.480[.226](.224){.217}	.448[.211](.205)	.447[.211](.204){.196}
		.25	-.481[.223](.222)	-.484[.224](.223){.210}	.201[.251](.246)	.200[.249](.244){.246}
		.00	-.487[.217](.216)	-.487[.218](.217){.210}	-.041[.274](.271)	-.042[.271](.268){.265}
		-.25	-.494[.216](.216)	-.492[.218](.217){.200}	-.279[.282](.281)	-.272[.279](.277){.277}
		-.50	-.499[.216](.216)	-.495[.216](.216){.210}	-.516[.283](.283)	-.512[.278](.278){.274}
100	.50	.50	.473[.167](.165)	.473[.165](.163){.148}	.483[.177](.176)	.482[.174](.173){.169}
		.25	.473[.144](.141)	.480[.140](.138){.133}	.246[.193](.193)	.250[.189](.189){.189}
		.00	.479[.123](.121)	.480[.121](.119){.110}	-.001[.199](.199)	.000[.194](.194){.191}
		-.25	.487[.101](.100)	.487[.100](.099){.092}	-.252[.192](.192)	-.248[.188](.188){.187}
		-.50	.487[.091](.090)	.487[.091](.090){.090}	-.501[.185](.185)	-.495[.182](.182){.182}
	.00	.50	.000[.191](.191)	.000[.188](.188){.188}	.472[.169](.166)	.470[.167](.163){.162}
		.25	-.006[.173](.173)	-.005[.170](.170){.164}	.229[.191](.190)	.227[.188](.186){.184}
		.00	-.010[.163](.162)	-.009[.161](.160){.152}	-.011[.200](.200)	-.011[.196](.196){.197}
		-.25	-.012[.151](.151)	-.011[.148](.148){.141}	-.255[.205](.205)	-.252[.199](.199){.199}
		-.50	-.010[.143](.143)	-.010[.142](.141){.140}	-.504[.205](.205)	-.500[.199](.199){.199}
	-.50	.50	-.488[.181](.181)	-.486[.179](.179){.169}	.476[.151](.149)	.480[.150](.147){.143}
		.25	-.494[.177](.177)	-.500[.174](.174){.165}	.226[.183](.181)	.223[.180](.178){.173}
		.00	-.499[.174](.174)	-.497[.171](.171){.160}	-.015[.201](.201)	-.012[.197](.196){.192}
		-.25	-.498[.173](.173)	-.497[.171](.170){.159}	-.264[.213](.213)	-.262[.208](.208){.199}
		-.50	-.503[.169](.169)	-.500[.167](.167){.157}	-.506[.214](.214)	-.501[.209](.209){.200}
250	.50	.50	.485[.119](.118)	.484[.122](.121){.119}	.493[.128](.128)	.500[.130](.130){.127}
		.25	.485[.099](.098)	.486[.100](.099){.095}	.251[.132](.132)	.250[.133](.133){.132}
		.00	.489[.080](.079)	.499[.080](.079){.076}	.001[.132](.132)	.000[.132](.132){.130}
		-.25	.491[.066](.065)	.493[.066](.065){.065}	-.248[.126](.126)	-.250[.125](.125){.125}
		-.50	.492[.060](.059)	.500[.060](.059){.058}	-.498[.124](.124)	-.499[.124](.124){.120}
	.00	.50	.005[.143](.142)	.000[.144](.144){.140}	.481[.119](.117)	.484[.119](.118){.112}
		.25	.000[.129](.129)	-.001[.130](.130){.127}	.235[.136](.135)	.237[.136](.136){.130}
		.00	-.007[.117](.117)	-.006[.118](.117){.115}	-.006[.143](.143)	-.005[.143](.142){.140}
		-.25	-.008[.106](.106)	-.007[.106](.106){.105}	-.249[.141](.141)	-.250[.141](.141){.140}
		-.50	-.008[.100](.100)	-.007[.100](.100){.100}	-.496[.140](.140)	-.500[.139](.139){.135}
	-.50	.50	-.490[.127](.126)	-.500[.127](.127){.125}	.485[.097](.096)	.490[.097](.096){.094}
		.25	-.491[.130](.130)	-.500[.130](.130){.126}	.233[.125](.124)	.240[.125](.124){.120}
		.00	-.498[.126](.126)	-.499[.126](.126){.123}	-.011[.140](.139)	-.010[.139](.139){.136}
		-.25	-.498[.123](.123)	-.498[.123](.123){.120}	-.261[.149](.149)	-.254[.149](.148){.143}
		-.50	-.502[.118](.118)	-.500[.118](.118){.117}	-.507[.147](.147)	-.504[.146](.146){.144}
500	.50	.50	.493[.082](.082)	.500[.083](.082){.080}	.496[.089](.089)	.496[.089](.089){.088}
		.25	.494[.066](.065)	.495[.066](.065){.064}	.251[.093](.093)	.250[.093](.093){.092}
		.00	.497[.053](.053)	.500[.053](.053){.052}	-.003[.093](.093)	-.002[.092](.092){.091}
		-.25	.496[.046](.046)	.500[.046](.046){.045}	-.251[.090](.090)	-.250[.090](.090){.089}
		-.50	.498[.040](.040)	.500[.040](.040){.040}	-.503[.085](.085)	-.500[.084](.084){.084}
	.00	.50	.003[.094](.094)	.003[.094](.094){.094}	.492[.078](.077)	.500[.077](.077){.077}
		.25	-.002[.087](.087)	-.001[.087](.087){.086}	.244[.093](.092)	.244[.092](.092){.090}
		.00	-.003[.080](.080)	-.003[.079](.079){.078}	-.002[.098](.098)	-.002[.097](.097){.096}
		-.25	.000[.072](.072)	.000[.072](.072){.071}	-.255[.100](.100)	-.254[.099](.099){.097}
		-.50	-.002[.066](.066)	-.002[.066](.066){.066}	-.501[.094](.094)	-.500 [.094](.094){.094}
	-.50	.50	-.497[.087](.087)	-.497[.087](.086){.086}	.494[.065](.065)	.493[.066](.065){.065}
		.25	-.500[.087](.087)	-.499[.087](.087){.086}	.246[.084](.084)	.250[.083](.083){.083}
		.00	-.500[.084](.084)	-.499[.084](.084){.084}	-.004[.094](.094)	-.005[.093](.093){.093}
		-.25	-.499[.085](.084)	-.498[.084](.084){.082}	-.255[.103](.103)	-.252[.102](.102){.100}
		-.50	-.502[.082](.082)	-.501[.081](.081){.080}	-.502[.104](.104)	-.500[.103](.103){.101}

Table 6.3c. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is consistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, Queen Contiguity, REG-1, DGP 3

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.475[.194](.193)	.480[.200](.198){.195}	.456[.228](.223)	.453[.231](.226){.221}
		.25	.466[.182](.179)	.470[.187](.184){.188}	.228[.250](.249)	.230[.251](.249){.252}
		.00	.466[.172](.168)	.468[.173](.169){.153}	-.009[.265](.264)	-.011[.261](.261){.263}
		-.25	.471[.149](.146)	.473[.149](.146){.140}	-.256[.265](.265)	-.255[.260](.260){.263}
		-.50	.475[.140](.138)	.477[.140](.138){.130}	-.500[.258](.258)	-.495[.253](.253){.252}
	.00	.50	.014[.225](.225)	.008[.228](.228){.210}	.442[.228](.220)	.449[.229](.221){.199}
		.25	-.005[.215](.215)	-.005[.216](.216){.196}	.207[.250](.246)	.204[.249](.245){.246}
		.00	-.015[.203](.203)	-.013[.203](.202){.188}	-.026[.262](.261)	-.029[.258](.256){.255}
		-.25	-.015[.195](.194)	-.012[.194](.193){.184}	-.270[.275](.274)	-.269[.269](.268){.267}
		-.50	-.018[.188](.187)	-.015[.187](.186){.180}	-.508[.268](.268)	-.503[.262](.262){.264}
	-.50	.50	-.467[.222](.220)	-.480[.221](.219){.200}	.448[.209](.203)	.450[.208](.201){.190}
		.25	-.477[.222](.221)	-.480[.223](.221){.199}	.201[.242](.237)	.199[.241](.236){.234}
		.00	-.487[.214](.214)	-.490[.214](.213){.199}	-.036[.268](.265)	-.038[.264](.261){.259}
		-.25	-.491[.209](.209)	-.490[.209](.208){.198}	-.285[.273](.270)	-.250[.268](.266){.269}
		-.50	-.498[.214](.214)	-.500[.213](.213){.197}	-.519[.280](.280)	-.515[.274](.274){.270}
100	.50	.50	.478[.162](.160)	.480[.158](.156){.144}	.484[.170](.170)	.482[.168](.167){.164}
		.25	.475[.145](.143)	.480[.140](.138){.137}	.244[.189](.189)	.250[.184](.184){.184}
		.00	.480[.124](.123)	.480[.122](.120){.107}	.001[.189](.189)	.002[.184](.184){.185}
		-.25	.486[.104](.103)	.490[.103](.101){.090}	-.254[.187](.187)	-.249[.182](.182){.179}
		-.50	.487[.090](.089)	.486[.091](.089){.084}	-.499[.180](.180)	-.491[.176](.176){.177}
	.00	.50	-.001[.188](.188)	.001[.184](.184){.170}	.475[.166](.164)	.471[.163](.160){.163}
		.25	-.013[.173](.172)	-.010[.167](.166){.160}	.235[.183](.182)	.240[.177](.177){.173}
		.00	-.011[.162](.162)	-.010[.156](.155){.150}	-.009[.195](.195)	-.009[.188](.188){.187}
		-.25	-.007[.153](.153)	-.007[.147](.147){.140}	-.263[.204](.204)	-.258[.196](.196){.191}
		-.50	-.010[.143](.143)	-.010[.139](.139){.130}	-.506[.203](.203)	-.500[.195](.195){.191}
	-.50	.50	-.491[.180](.179)	-.490[.174](.173){.160}	.476[.150](.148)	.480[.146](.143){.145}
		.25	-.493[.176](.176)	-.490[.171](.171){.155}	.226[.180](.178)	.230[.175](.173){.174}
		.00	-.496[.173](.173)	-.500[.167](.167){.155}	-.019[.198](.197)	-.021[.191](.190){.187}
		-.25	-.500[.171](.171)	-.498[.164](.164){.150}	-.260[.214](.213)	-.259[.203](.203){.194}
		-.50	-.501[.170](.170)	-.500[.164](.164){.150}	-.509[.215](.215)	-.500[.205](.205){.199}
250	.50	.50	.489[.118](.118)	.490[.119](.119){.120}	.489[.127](.126)	.490[.128](.127){.129}
		.25	.485[.102](.100)	.486[.102](.101){.100}	.248[.137](.137)	.250[.137](.137){.137}
		.00	.487[.082](.081)	.489[.082](.082){.080}	.003[.133](.133)	.001[.133](.133){.130}
		-.25	.493[.064](.064)	.495[.064](.063){.063}	-.250[.125](.125)	-.250[.123](.123){.120}
		-.50	.493[.058](.058)	.496[.058](.057){.056}	-.500[.120](.120)	-.500[.118](.118){.114}
	.00	.50	.008[.142](.142)	.004[.142](.142){.140}	.479[.121](.119)	.490[.120](.118){.113}
		.25	-.004[.131](.131)	-.004[.130](.130){.130}	.240[.135](.135)	.240[.135](.134){.130}
		.00	-.006[.117](.117)	-.006[.117](.117){.113}	-.007[.143](.142)	-.006[.142](.142){.139}
		-.25	-.005[.107](.107)	-.004[.106](.106){.101}	-.257[.143](.143)	-.255[.141](.141){.143}
		-.50	-.010[.099](.098)	-.008[.097](.097){.094}	-.495[.136](.136)	-.495[.133](.133){.130}
	-.50	.50	-.488[.130](.130)	-.491[.128](.127){.128}	.483[.101](.099)	.485[.099](.098){.098}
		.25	-.491[.131](.131)	-.500[.129](.129){.124}	.233[.127](.126)	.235[.125](.124){.120}
		.00	-.501[.128](.128)	-.500[.126](.126){.120}	-.010[.142](.141)	-.010[.140](.140){.140}
		-.25	-.495[.123](.123)	-.500[.122](.122){.117}	-.262[.147](.147)	-.261[.146](.145){.140}
		-.50	-.502[.123](.123)	-.501[.121](.121){.120}	-.504[.153](.153)	-.501[.150](.150){.149}
500	.50	.50	.496[.082](.082)	.500[.081](.081){.078}	.494[.089](.089)	.494[.088](.088){.086}
		.25	.493[.065](.064)	.494[.064](.064){.063}	.251[.092](.092)	.251[.091](.091){.090}
		.00	.496[.053](.053)	.500[.053](.053){.051}	-.003[.092](.092)	-.002[.092](.092){.089}
		-.25	.497[.045](.045)	.497[.044](.044){.044}	-.251[.088](.088)	-.250[.087](.087){.086}
		-.50	.498[.041](.040)	.498[.040](.040){.040}	-.501[.086](.086)	-.499[.085](.085){.082}
	.00	.50	.002[.096](.096)	.001[.094](.094){.093}	.492[.077](.077)	.500[.076](.076){.075}
		.25	-.004[.092](.092)	-.002[.090](.090){.090}	.246[.094](.094)	.250[.093](.093){.089}
		.00	-.003[.081](.081)	-.002[.080](.080){.080}	-.001[.101](.101)	-.001[.099](.099){.100}
		-.25	-.001[.072](.072)	-.001[.072](.072){.070}	-.253[.098](.098)	-.250[.097](.097){.095}
		-.50	-.001[.067](.067)	-.001[.066](.066){.065}	-.502[.095](.095)	-.500[.093](.093){.092}
	-.50	.50	-.498[.088](.088)	-.500[.087](.087){.084}	.495[.067](.067)	.494[.066](.065){.063}
		.25	-.498[.087](.087)	-.500[.086](.086){.084}	.243[.085](.084)	.242[.084](.083){.080}
		.00	-.500[.087](.087)	-.499[.085](.085){.082}	-.004[.096](.096)	-.006[.095](.095){.092}
		-.25	-.503[.084](.084)	-.500[.082](.082){.080}	-.250[.102](.102)	-.250[.100](.100){.098}
		-.50	-.499[.084](.084)	-.500[.081](.081){.080}	-.503[.104](.104)	-.500[.101](.101){.100}

Table 6.4a. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is inconsistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Circular Neighbours, REG-1, DGP 1`

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.486[.124](.123)	.485[.165](.164){.208}	.422[.181](.164)	.444[.220](.213){.218}
		.25	.451[.123](.112)	.476[.144](.142){.144}	.229[.172](.171)	.213[.236](.233){.239}
		.00	.435[.123](.104)	.480[.126](.124){.127}	.043[.179](.174)	-.026[.241](.240){.229}
		-.25	.418[.129](.100)	.480[.116](.114){.115}	-.142[.198](.166)	-.267[.233](.232){.232}
		-.50	.405[.137](.099)	.479[.112](.110){.115}	-.321[.241](.161)	-.493[.219](.219){.226}
	.00	.50	.075[.150](.130)	.014[.172](.172){.177}	.375[.201](.157)	.445[.200](.193){.189}
		.25	.033[.123](.119)	.006[.161](.160){.164}	.162[.183](.161)	.195[.226](.220){.219}
		.00	-.007[.114](.113)	-.004[.157](.157){.154}	-.020[.162](.160)	-.037[.240](.237){.237}
		-.25	-.036[.119](.114)	-.001[.156](.156){.150}	-.193[.165](.155)	-.277[.241](.240){.232}
		-.50	-.072[.136](.116)	-.010[.158](.157){.151}	-.359[.201](.143)	-.504[.223](.223){.206}
	-.50	.50	-.390[.152](.104)	-.481[.117](.115){.117}	.368[.199](.149)	.457[.163](.157){.157}
		.25	-.401[.143](.103)	-.480[.126](.124){.121}	.127[.208](.167)	.202[.207](.201){.201}
		.00	-.421[.128](.100)	-.480[.137](.136){.134}	-.078[.182](.165)	-.047[.233](.228){.207}
		-.25	-.443[.117](.102)	-.478[.151](.149){.171}	-.258[.152](.152)	-.288[.237](.234){.379}
		-.50	-.478[.106](.104)	-.485[.161](.160){.155}	-.426[.156](.137)	-.523[.226](.225){.282}
100	.50	.50	.485[.096](.095)	.490[.133](.132){.136}	.447[.129](.117)	.481[.154](.153){.153}
		.25	.459[.093](.083)	.483[.106](.105){.109}	.245[.123](.123)	.237[.163](.163){.162}
		.00	.443[.095](.076)	.486[.088](.087){.086}	.053[.134](.123)	-.005[.165](.165){.165}
		-.25	.435[.095](.069)	.490[.075](.074){.073}	-.142[.161](.120)	-.258[.161](.161){.161}
		-.50	.428[.097](.065)	.491[.068](.067){.072}	-.332[.202](.112)	-.495[.148](.148){.101}
	.00	.50	.082[.129](.099)	.006[.140](.140){.142}	.382[.161](.110)	.467[.142](.138){.140}
		.25	.036[.099](.092)	.000[.129](.129){.131}	.174[.137](.114)	.221[.161](.158){.159}
		.00	-.002[.088](.088)	-.003[.122](.122){.119}	-.011[.116](.115)	-.019[.173](.172){.170}
		-.25	-.039[.092](.083)	-.011[.114](.113){.113}	-.183[.129](.110)	-.256[.171](.170){.170}
		-.50	-.068[.107](.083)	-.010[.110](.109){.113}	-.356[.176](.102)	-.498[.155](.155){.160}
	-.50	.50	-.359[.166](.088)	-.487[.101](.100){.100}	.364[.174](.108)	.477[.114](.112){.112}
		.25	-.381[.144](.082)	-.487[.105](.105){.105}	.121[.175](.118)	.220[.149](.146){.146}
		.00	-.409[.120](.079)	-.489[.110](.110){.107}	-.081[.144](.118)	-.029[.171](.168){.168}
		-.25	-.441[.095](.075)	-.493[.113](.113){.114}	-.257[.108](.108)	-.269[.175](.174){.174}
		-.50	-.479[.077](.074)	-.498[.119](.119){.120}	-.421[.125](.097)	-.504[.168](.168){.162}
250	.50	.50	.490[.059](.058)	.491[.086](.086){.083}	.458[.082](.071)	.494[.099](.099){.100}
		.25	.461[.065](.052)	.495[.067](.066){.066}	.255[.078](.077)	.242[.108](.108){.108}
		.00	.441[.076](.048)	.495[.055](.055){.055}	.066[.102](.077)	-.003[.107](.107){.107}
		-.25	.427[.086](.045)	.495[.050](.049){.050}	-.124[.148](.076)	-.251[.105](.105){.105}
		-.50	.418[.093](.043)	.496[.046](.046){.047}	-.318[.195](.070)	-.497[.093](.093){.093}
	.00	.50	.086[.107](.063)	.003[.090](.090){.090}	.393[.127](.069)	.489[.085](.084){.084}
		.25	.040[.070](.058)	-.001[.085](.085){.085}	.183[.098](.072)	.241[.103](.103){.103}
		.00	.000[.055](.055)	.001[.080](.080){.080}	-.006[.073](.073)	-.011[.114](.113){.113}
		-.25	-.037[.066](.054)	.000[.078](.078){.078}	-.177[.100](.069)	-.256[.113](.113){.113}
		-.50	-.075[.092](.054)	-.003[.076](.076){.080}	-.347[.166](.064)	-.500[.102](.102){.102}
	-.50	.50	-.370[.141](.053)	-.495[.057](.057){.060}	.374[.143](.068)	.491[.067](.066){.066}
		.25	-.384[.127](.051)	-.497[.061](.061){.061}	.129[.143](.075)	.239[.088](.088){.088}
		.00	-.407[.105](.048)	-.497[.066](.065){.065}	-.078[.107](.073)	-.009[.103](.103){.103}
		-.25	-.436[.080](.047)	-.495[.073](.073){.073}	-.259[.067](.066)	-.258[.111](.110){.111}
		-.50	-.476[.053](.048)	-.497[.084](.084){.084}	-.422[.099](.060)	-.502[.113](.113){.113}
500	.50	.50	.492[.039](.038)	.497[.054](.054){.054}	.460[.063](.048)	.497[.066](.066){.066}
		.25	.464[.050](.034)	.498[.043](.043){.043}	.257[.053](.053)	.246[.072](.072){.072}
		.00	.445[.064](.033)	.498[.038](.038){.038}	.064[.084](.054)	-.003[.076](.076){.076}
		-.25	.430[.076](.031)	.498[.034](.034){.034}	-.125[.136](.053)	-.252[.074](.074){.074}
		-.50	.419[.086](.029)	.497[.032](.032){.032}	-.319[.187](.049)	-.499[.066](.066){.070}
	.00	.50	.078[.089](.042)	.000[.059](.059){.060}	.401[.110](.048)	.495[.057](.056){.056}
		.25	.037[.053](.038)	.000[.054](.054){.054}	.188[.079](.048)	.246[.067](.067){.067}
		.00	-.001[.037](.037)	-.001[.053](.053){.053}	.000[.051](.051)	-.002[.076](.076){.076}
		-.25	-.036[.052](.037)	-.001[.052](.052){.053}	-.176[.088](.048)	-.252[.078](.078){.078}
		-.50	-.073[.082](.038)	-.002[.053](.053){.053}	-.348[.158](.044)	-.499[.072](.072){.072}
	-.50	.50	-.377[.129](.036)	-.497[.039](.039){.040}	.380[.129](.048)	.494[.047](.046){.046}
		.25	-.389[.116](.035)	-.498[.041](.041){.041}	.136[.126](.052)	.245[.060](.060){.060}
		.00	-.409[.096](.033)	-.497[.043](.043){.043}	-.074[.090](.051)	-.005[.070](.070){.070}
		-.25	-.438[.070](.033)	-.496[.049](.049){.049}	-.258[.049](.048)	-.257[.079](.078){.078}
		-.50	-.477[.040](.033)	-.498[.057](.057){.060}	-.422[.088](.042)	-.502[.078](.078){.080}

Table 6.4b. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is inconsistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Circular Neighbours, REG-1, DGP 2`

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.483[.125](.124)	.482[.163](.162){.163}	.432[.177](.163)	.454[.215](.210){.211}
		.25	.456[.119](.111)	.479[.142](.140){.174}	.230[.171](.169)	.216[.231](.229){.228}
		.00	.438[.122](.105)	.482[.126](.124){.123}	.038[.177](.173)	-.029[.241](.240){.240}
		-.25	.420[.132](.105)	.479[.119](.117){.122}	-.143[.202](.172)	-.264[.237](.237){.232}
		-.50	.406[.139](.102)	.479[.111](.109){.107}	-.328[.238](.165)	-.500[.218](.218){.276}
	.00	.50	.054[.135](.123)	.010[.148](.148){.150}	.388[.194](.158)	.450[.192](.185){.185}
		.25	.030[.119](.115)	.013[.145](.144){.144}	.168[.185](.166)	.196[.225](.218){.219}
		.00	-.004[.113](.113)	.002[.146](.146){.142}	-.024[.166](.164)	-.044[.240](.236){.236}
		-.25	-.030[.116](.112)	.001[.147](.147){.131}	-.206[.162](.156)	-.287[.238](.235){.238}
		-.50	-.065[.134](.117)	-.014[.153](.152){.153}	-.366[.202](.151)	-.503[.230](.230){.241}
	-.50	.50	-.396[.151](.110)	-.484[.117](.116){.116}	.376[.196](.152)	.461[.163](.158){.158}
		.25	-.406[.144](.109)	-.481[.127](.126){.121}	.135[.203](.167)	.207[.202](.198){.204}
		.00	-.420[.130](.103)	-.476[.135](.133){.131}	-.076[.180](.164)	-.048[.230](.224){.227}
		-.25	-.445[.118](.104)	-.480[.151](.150){.183}	-.257[.151](.151)	-.290[.236](.233){.262}
		-.50	-.475[.110](.107)	-.483[.164](.164){.103}	-.425[.161](.143)	-.523[.230](.229){.297}
100	.50	.50	.486[.095](.094)	.484[.130](.129){.122}	.445[.128](.116)	.477[.151](.149){.115}
		.25	.461[.092](.083)	.485[.105](.104){.102}	.240[.125](.125)	.232[.165](.164){.201}
		.00	.446[.094](.077)	.488[.088](.087){.087}	.048[.133](.125)	-.011[.167](.167){.167}
		-.25	.434[.098](.072)	.487[.076](.075){.075}	-.139[.165](.122)	-.249[.160](.160){.159}
		-.50	.430[.097](.067)	.492[.067](.067){.067}	-.338[.200](.117)	-.502[.144](.144){.144}
	.00	.50	.079[.131](.105)	.005[.143](.143){.146}	.389[.158](.112)	.472[.139](.136){.137}
		.25	.036[.099](.092)	.000[.128](.128){.130}	.177[.135](.114)	.225[.159](.156){.156}
		.00	.000[.086](.086)	-.001[.118](.118){.118}	-.012[.114](.114)	-.021[.169](.168){.168}
		-.25	-.037[.091](.083)	-.009[.111](.111){.111}	-.184[.126](.108)	-.258[.166](.166){.166}
		-.50	-.064[.107](.085)	-.007[.110](.110){.111}	-.360[.177](.108)	-.501[.158](.158){.153}
	-.50	.50	-.363[.167](.096)	-.488[.103](.102){.102}	.365[.177](.114)	.476[.117](.115){.115}
		.25	-.384[.144](.086)	-.487[.105](.104){.104}	.126[.173](.120)	.220[.147](.144){.144}
		.00	-.411[.120](.081)	-.490[.108](.108){.108}	-.075[.139](.117)	-.024[.166](.164){.160}
		-.25	-.441[.098](.078)	-.491[.117](.117){.117}	-.257[.109](.108)	-.271[.177](.176){.176}
		-.50	-.479[.078](.075)	-.497[.120](.120){.124}	-.420[.126](.098)	-.504[.166](.166){.160}
250	.50	.50	.490[.059](.058)	.491[.086](.085){.103}	.456[.084](.072)	.491[.100](.100){.104}
		.25	.460[.067](.054)	.493[.068](.068){.068}	.256[.078](.078)	.244[.108](.108){.108}
		.00	.441[.077](.049)	.495[.056](.056){.056}	.064[.102](.079)	-.005[.109](.109){.109}
		-.25	.427[.087](.048)	.495[.050](.050){.050}	-.124[.148](.078)	-.252[.104](.104){.104}
		-.50	.419[.093](.046)	.496[.046](.046){.046}	-.320[.195](.075)	-.499[.093](.093){.093}
	.00	.50	.085[.107](.065)	.003[.091](.091){.091}	.393[.128](.070)	.488[.085](.084){.084}
		.25	.038[.070](.059)	-.002[.086](.086){.086}	.183[.099](.073)	.241[.103](.103){.103}
		.00	-.001[.055](.055)	-.001[.079](.079){.079}	-.005[.071](.071)	-.009[.110](.110){.110}
		-.25	-.039[.067](.054)	-.004[.077](.077){.077}	-.177[.101](.069)	-.253[.111](.111){.111}
		-.50	-.074[.092](.055)	-.002[.077](.076){.080}	-.349[.165](.067)	-.501[.103](.103){.103}
	-.50	.50	-.371[.142](.059)	-.496[.058](.058){.060}	.375[.144](.070)	.490[.067](.066){.066}
		.25	-.383[.129](.055)	-.494[.063](.063){.063}	.129[.144](.078)	.238[.089](.088){.088}
		.00	-.405[.107](.050)	-.494[.066](.065){.065}	-.080[.108](.073)	-.013[.103](.102){.102}
		-.25	-.435[.081](.048)	-.493[.073](.073){.074}	-.258[.068](.067)	-.259[.112](.112){.112}
		-.50	-.477[.053](.048)	-.499[.083](.083){.083}	-.422[.099](.060)	-.501[.109](.109){.107}
500	.50	.50	.491[.040](.039)	.496[.055](.055){.055}	.460[.063](.050)	.497[.067](.067){.067}
		.25	.464[.051](.036)	.498[.044](.044){.044}	.256[.053](.053)	.246[.073](.073){.073}
		.00	.445[.064](.033)	.499[.038](.037){.037}	.064[.084](.055)	-.004[.075](.075){.075}
		-.25	.431[.077](.033)	.498[.035](.035){.035}	-.124[.137](.055)	-.250[.074](.074){.074}
		-.50	.420[.086](.031)	.498[.032](.032){.032}	-.320[.188](.054)	-.500[.066](.066){.070}
	.00	.50	.080[.091](.044)	.002[.060](.060){.060}	.400[.111](.048)	.494[.057](.057){.057}
		.25	.037[.053](.038)	-.001[.054](.054){.055}	.188[.080](.050)	.247[.068](.068){.068}
		.00	-.001[.037](.037)	-.001[.052](.052){.052}	-.002[.050](.050)	-.003[.075](.075){.075}
		-.25	-.036[.051](.037)	-.001[.052](.052){.052}	-.176[.088](.049)	-.252[.077](.077){.077}
		-.50	-.072[.082](.038)	-.002[.053](.053){.053}	-.349[.158](.047)	-.499[.071](.071){.071}
	-.50	.50	-.378[.129](.039)	-.498[.039](.039){.040}	.382[.128](.050)	.496[.047](.046){.046}
		.25	-.390[.116](.037)	-.498[.041](.041){.041}	.136[.126](.054)	.245[.061](.061){.061}
		.00	-.411[.095](.035)	-.499[.044](.044){.044}	-.072[.088](.051)	-.003[.070](.070){.070}
		-.25	-.438[.070](.034)	-.497[.050](.050){.050}	-.256[.048](.048)	-.254[.078](.078){.078}
		-.50	-.477[.040](.033)	-.498.057](.057){.060}	-.423[.088](.043)	-.502[.077](.077){.080}

Table 6.4c. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is inconsistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Circular Neighbours, REG-1, DGP 3`

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.480[.129](.128)	.480[.160](.158){.154}	.435[.175](.163)	.459[.204](.200){.208}
		.25	.461[.126](.120)	.483[.139](.138){.139}	.224[.177](.175)	.212[.227](.224){.224}
		.00	.439[.131](.116)	.476[.132](.130){.130}	.035[.181](.178)	-.020[.238](.237){.237}
		-.25	.429[.132](.111)	.478[.124](.122){.125}	-.155[.203](.180)	-.261[.237](.237){.231}
		-.50	.416[.138](.109)	.478[.124](.122){.183}	-.334[.245](.180)	-.496[.234](.234){.266}
	.00	.50	.052[.138](.128)	.016[.148](.147){.146}	.392[.192](.159)	.449[.188](.181){.201}
		.25	.022[.123](.121)	.007[.140](.140){.140}	.171[.186](.168)	.199[.211](.205){.205}
		.00	-.001[.113](.113)	.006[.141](.141){.141}	-.029[.165](.163)	-.048[.231](.226){.229}
		-.25	-.028[.120](.116)	-.003[.145](.145){.142}	-.207[.167](.161)	-.281[.234](.232){.280}
		-.50	-.055[.132](.120)	-.009[.148](.148){.157}	-.377[.200](.158)	-.511[.224](.224){.300}
	-.50	.50	-.401[.159](.124)	-.480[.128](.126){.124}	.382[.197](.158)	.460[.168](.163){.169}
		.25	-.413[.145](.116)	-.481[.126](.125){.125}	.140[.201](.168)	.202[.201](.195){.192}
		.00	-.428[.135](.114)	-.479[.137](.135){.135}	-.064[.181](.169)	-.042[.223](.219){.219}
		-.25	-.448[.121](.110)	-.481[.146](.145){.147}	-.252[.162](.162)	-.287[.235](.232){.221}
		-.50	-.478[.112](.110)	-.489[.159](.158){.195}	-.419[.171](.151)	-.520[.230](.229){.210}
100	.50	.50	.486[.097](.096)	.486[.124](.123){.121}	.446[.129](.117)	.476[.146](.144){.140}
		.25	.461[.096](.087)	.482[.107](.105){.106}	.243[.125](.125)	.239[.162](.162){.162}
		.00	.451[.093](.079)	.488[.089](.089){.106}	.041[.132](.126)	-.012[.165](.164){.201}
		-.25	.436[.103](.081)	.485[.082](.081){.085}	-.145[.168](.131)	-.249[.161](.161){.160}
		-.50	.433[.102](.077)	.489[.082](.081){.080}	-.341[.208](.134)	-.497[.157](.157){.120}
	.00	.50	.070[.130](.110)	.006[.138](.138){.139}	.394[.158](.117)	.469[.135](.132){.138}
		.25	.029[.103](.099)	.000[.125](.125){.125}	.180[.139](.121)	.222[.157](.154){.154}
		.00	-.002[.096](.096)	-.002[.121](.121){.124}	-.011[.122](.122)	-.021[.169](.168){.170}
		-.25	-.031[.095](.089)	-.005[.111](.111){.109}	-.192[.130](.117)	-.264[.168](.167){.167}
		-.50	-.060[.108](.090)	-.010[.111](.110){.131}	-.363[.181](.118)	-.498[.159](.159){.148}
	-.50	.50	-.369[.172](.111)	-.483[.110](.109){.110}	.369[.178](.121)	.472[.120](.117){.114}
		.25	-.393[.147](.100)	-.487[.106](.105){.105}	.136[.170](.126)	.223[.145](.143){.146}
		.00	-.417[.124](.092)	-.490[.111](.110){.109}	-.069[.138](.120)	-.025[.163](.161){.165}
		-.25	-.446[.102](.087)	-.494[.112](.111){.110}	-.249[.117](.117)	-.265[.170](.169){.169}
		-.50	-.476[.088](.085)	-.493[.123](.123){.121}	-.422[.133](.108)	-.512[.169](.169){.170}
250	.50	.50	.488[.063](.062)	.490[.086](.086){.083}	.457[.086](.074)	.492[.099](.098){.100}
		.25	.462[.067](.055)	.494[.067](.067){.067}	.255[.077](.076)	.245[.103](.103){.103}
		.00	.444[.078](.054)	.495[.056](.056){.056}	.060[.102](.082)	-.005[.106](.106){.106}
		-.25	.431[.088](.055)	.496[.053](.053){.050}	-.132[.148](.089)	-.254[.107](.107){.107}
		-.50	.420[.097](.055)	.495[.050](.050){.049}	-.324[.201](.096)	-.499[.097](.097){.097}
	.00	.50	.080[.107](.071)	.004[.091](.091){.091}	.398[.127](.076)	.488[.083](.082){.082}
		.25	.036[.071](.061)	-.001[.085](.085){.085}	.186[.099](.075)	.241[.102](.101){.100}
		.00	.000[.058](.058)	.000[.079](.079){.079}	-.004[.073](.073)	-.007[.108](.108){.108}
		-.25	-.035[.068](.058)	-.002[.077](.077){.077}	-.182[.101](.074)	-.257[.111](.111){.110}
		-.50	-.070[.092](.059)	-.005[.074](.074){.074}	-.353[.167](.080)	-.498[.102](.102){.102}
	-.50	.50	-.374[.147](.076)	-.494[.063](.063){.060}	.380[.145](.082)	.491[.070](.069){.069}
		.25	-.391[.128](.067)	-.495[.060](.060){.060}	.138[.140](.083)	.240[.086](.086){.086}
		.00	-.410[.109](.062)	-.495[.066](.065){.065}	-.073[.107](.079)	-.011[.101](.100){.099}
		-.25	-.440[.084](.059)	-.496[.075](.075){.075}	-.256[.072](.072)	-.260[.110](.110){.110}
		-.50	-.476[.059](.053)	-.497[.082](.082){.085}	-.424[.103](.068)	-.505[.111](.111){.116}
500	.50	.50	.492[.040](.039)	.498[.054](.054){.054}	.458[.065](.049)	.494[.065](.065){.061}
		.25	.464[.052](.037)	.497[.044](.044){.044}	.256[.054](.054)	.246[.072](.072){.072}
		.00	.446[.066](.037)	.498[.038](.038){.038}	.062[.085](.058)	-.002[.075](.075){.075}
		-.25	.432[.078](.039)	.498[.036](.036){.036}	-.128[.138](.064)	-.252[.075](.075){.075}
		-.50	.423[.087](.041)	.498[.032](.032){.032}	-.323[.191](.074)	-.499[.067](.067){.067}
	.00	.50	.076[.090](.048)	.002[.059](.059){.060}	.404[.110](.053)	.494[.055](.055){.055}
		.25	.036[.055](.042)	.001[.056](.056){.056}	.189[.081](.053)	.245[.069](.068){.068}
		.00	-.002[.038](.038)	-.001[.053](.053){.053}	-.002[.050](.050)	-.003[.075](.075){.075}
		-.25	-.034[.052](.039)	.000[.052](.052){.052}	-.180[.087](.052)	-.254[.077](.077){.077}
		-.50	-.069[.081](.042)	-.002[.052](.052){.052}	-.353[.159](.059)	-.500[.072](.072){.072}
	-.50	.50	-.380[.132](.055)	-.498[.040](.040){.040}	.385[.129](.058)	.496[.046](.046){.046}
		.25	-.393[.117](.049)	-.498[.041](.041){.041}	.139[.127](.062)	.244[.061](.061){.060}
		.00	-.413[.098](.045)	-.498[.044](.044){.044}	-.070[.090](.056)	-.005[.070](.070){.070}
		-.25	-.439[.072](.039)	-.497[.049](.049){.049}	-.254[.049](.048)	-.253[.075](.075){.075}
		-.50	-.477[.045](.038)	-.498[.058](.058){.059}	-.423[.091](.049)	-.503[.077](.077){.080}

Table 6.5a. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is inconsistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Group Interaction, REG-2, DGP 1`

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.473[.161](.158)	.482[.184](.165){.173}	.416[.214](.197)	.474[.278](.172){.170}
		.25	.431[.169](.154)	.433[.295](.284){.271}	.218[.229](.227)	.253[.239](.237){.250}
		.00	.416[.162](.139)	.456[.210](.206){.205}	.030[.243](.241)	-.012[.257](.241){.240}
		-.25	.409[.156](.126)	.473[.163](.161){.152}	-.150[.272](.253)	-.239[.245](.243){.236}
		-.50	.404[.150](.115)	.479[.139](.137){.138}	-.316[.310](.249)	-.462[.144](.142){.130}
	.00	.50	.130[.228](.188)	-.073[.260](.197){.193}	.324[.259](.190)	.540[.232](.200){.205}
		.25	.042[.186](.181)	-.059[.157](.152){.151}	.123[.250](.215)	.243[.269](.254){.253}
		.00	-.022[.178](.177)	-.052[.131](.130){.138}	-.057[.228](.221)	-.011[.241](.239){.228}
		-.25	-.056[.180](.171)	-.033[.165](.163){.151}	-.234[.233](.232)	-.238[.237](.218){.217}
		-.50	-.085[.182](.161)	-.026[.176](.161){.155}	-.393[.254](.230)	-.522[.236](.218){.215}
	-.50	.50	-.186[.380](.213)	-.492[.334](.334){.344}	.263[.308](.196)	.484[.231](.228){.236}
		.25	-.305[.277](.197)	-.519[.286](.285){.298}	.048[.294](.214)	.229[.236](.234){.239}
		.00	-.389[.222](.192)	-.522[.235](.234){.245}	-.144[.266](.224)	-.013[.240](.238){.230}
		-.25	-.447[.190](.182)	-.515[.211](.211){.228}	-.319[.235](.225)	-.239[.243](.241){.253}
		-.50	-.498[.178](.178)	-.519[.185](.184){.199}	-.477[.223](.222)	-.503[.230](.209){.205}
100	.50	.50	.483[.114](.112)	.490[.126](.125){.127}	.445[.144](.133)	.489[.121](.121){.126}
		.25	.446[.120](.107)	.464[.170](.167){.163}	.248[.157](.157)	.258[.143](.140){.142}
		.00	.431[.119](.097)	.477[.126](.124){.118}	.057[.180](.171)	-.049[.127](.225){.124}
		-.25	.425[.114](.085)	.487[.101](.100){.110}	-.127[.216](.177)	-.231[.127](.127){.124}
		-.50	.420[.111](.077)	.500[.089](.089){.090}	-.307[.265](.181)	-.576[.135](.133){.125}
	.00	.50	.131[.188](.135)	-.062[.233](.233){.247}	.360[.188](.126)	.546[.120](.119){.125}
		.25	.045[.141](.134)	-.036[.155](.152){.157}	.154[.175](.146)	.220[.139](.123){.122}
		.00	-.010[.128](.128)	-.023[.120](.120){.148}	-.035[.161](.158)	-.061[.170](.163){.162}
		-.25	-.045[.128](.120)	-.018[.166](.157){.178}	-.206[.169](.163)	-.231[.198](.192){.176}
		-.50	-.071[.135](.115)	-.013[.157](.157){.162}	-.373[.210](.167)	-.557[.132](.132){.133}
	-.50	.50	-.181[.354](.152)	-.503[.235](.235){.247}	.293[.244](.128)	.544[.198](.189){.190}
		.25	-.309[.237](.141)	-.503[.266](.254){.253}	.086[.216](.141)	.260[.121](.122){.122}
		.00	-.387[.177](.136)	-.502[.248](.247){.251}	-.111[.192](.156)	-.062[.164](.157){.166}
		-.25	-.446[.142](.132)	-.513[.221](.221){.228}	-.289[.161](.157)	-.232[.190](.181){.170}
		-.50	-.489[.128](.128)	-.504[.120](.120){.126}	-.454[.168](.162)	-.521[.132](.131){.140}
250	.50	.50	.489[.068](.067)	.499[.114](.112){.123}	.464[.087](.080)	.492[.115](.115){.121}
		.25	.462[.071](.060)	.495[.076](.076){.073}	.259[.093](.093)	.258[.127](.126){.124}
		.00	.453[.069](.051)	.496[.058](.057){.057}	.058[.117](.101)	-.017[.112](.102){.102}
		-.25	.447[.070](.045)	.497[.049](.049){.049}	-.132[.159](.107)	-.256[.157](.156){.155}
		-.50	.442[.071](.041)	.498[.045](.045){.046}	-.313[.217](.109)	-.502[.108](.106){.102}
	.00	.50	.115[.141](.082)	-.015[.169](.167){.160}	.391[.132](.074)	.486[.109](.108){.114}
		.25	.039[.086](.077)	-.011[.124](.124){.125}	.184[.107](.085)	.250[.102](.102){.102}
		.00	-.004[.074](.074)	-.008[.103](.103){.099}	-.012[.098](.097)	-.011[.102](.101){.104}
		-.25	-.031[.074](.067)	-.005[.087](.087){.087}	-.196[.113](.099)	-.257[.139](.126){.122}
		-.50	-.051[.082](.064)	-.004[.080](.080){.081}	-.369[.167](.103)	-.502[.117](.103){.106}
	-.50	.50	-.240[.276](.091)	-.501[.178](.178){.180}	.348[.168](.071)	.494[.095](.094){.101}
		.25	-.340[.181](.085)	-.500[.146](.146){.151}	.126[.150](.084)	.253[.107](.106){.108}
		.00	-.405[.124](.080)	-.502[.127](.127){.124}	-.076[.121](.095)	-.023[.104](.104){.103}
		-.25	-.453[.091](.078)	-.503[.114](.114){.111}	-.261[.098](.098)	-.239[.105](.105){.105}
		-.50	-.488[.073](.072)	-.502[.102](.102){.104}	-.431[.122](.100)	-.524[.170](.168){.167}
500	.50	.50	.492[.049](.048)	.499[.077](.077){.078}	.468[.064](.056)	.495[.081](.081){.082}
		.25	.466[.054](.042)	.496[.051](.051){.051}	.261[.066](.065)	.249[.087](.087){.088}
		.00	.457[.057](.036)	.498[.039](.039){.039}	.059[.093](.072)	-.012[.100](.099){.098}
		-.25	.450[.059](.032)	.498[.033](.033){.033}	-.130[.141](.074)	-.251[.106](.106){.109}
		-.50	.447[.060](.028)	.500[.030](.030){.031}	-.313[.202](.077)	-.501[.102](.101){.102}
	.00	.50	.115[.130](.059)	-.007[.120](.120){.123}	.396[.116](.052)	.499[.076](.076){.078}
		.25	.041[.069](.056)	-.005[.088](.088){.088}	.188[.086](.060)	.250[.086](.086){.086}
		.00	-.001[.052](.052)	-.003[.071](.071){.071}	-.006[.068](.067)	-.009[.097](.096){.096}
		-.25	-.029[.056](.048)	-.004[.060](.060){.061}	-.190[.092](.070)	-.251[.106](.105){.107}
		-.50	-.046[.064](.044)	-.001[.054](.054){.056}	-.366[.152](.071)	-.501[.107](.107){.107}
	-.50	.50	-.239[.269](.066)	-.500[.132](.132){.135}	.351[.157](.051)	.498[.068](.068){.068}
		.25	-.342[.170](.061)	-.502[.107](.107){.109}	.133[.132](.060)	.250[.083](.082){.083}
		.00	-.408[.109](.058)	-.504[.091](.091){.090}	-.067[.095](.067)	-.010[.096](.096){.094}
		-.25	-.452[.072](.054)	-.500[.078](.078){.078}	-.255[.069](.069)	-.256[.106](.105){.105}
		-.50	-.488[.054](.053)	-.500[.073](.073){.073}	-.424[.103](.070)	-.501[.117](.117){.118}

Table 6.5b. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is inconsistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Group Interaction, REG-2, DGP 2`

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.467[.182](.179)	.538[.180](.162){.172}	.419[.221](.206)	.542[.233](.232){.241}
		.25	.433[.177](.164)	.443[.185](.175){.155}	.220[.238](.237)	.257[.233](.231){.244}
		.00	.417[.170](.148)	.454[.213](.208){.204}	.029[.251](.250)	-.011[.214](.200){.205}
		-.25	.408[.160](.131)	.472[.159](.156){.144}	-.153[.274](.256)	-.242[.238](.220){.224}
		-.50	.404[.154](.121)	.477[.139](.137){.148}	-.316[.324](.266)	-.460[.236](.223){.211}
	.00	.50	.114[.241](.213)	-.075[.229](.263){.270}	.333[.260](.199)	.540[.231](.229){.234}
		.25	.029[.202](.200)	-.063[.135](.135){.142}	.131[.252](.222)	.248[.237](.232){.238}
		.00	-.021[.193](.192)	-.046[.131](.131){.136}	-.062[.242](.233)	-.012[.241](.236){.227}
		-.25	-.058[.186](.177)	-.033[.161](.159){.154}	-.238[.242](.242)	-.238[.240](.220){.225}
		-.50	-.082 [.185](.166)	-.027[.228](.226){.228}	-.400[.261](.241)	-.519[.243](.241){.247}
	-.50	.50	-.208[.380](.242)	-.494[.272](.271){.270}	.278[.301](.203)	.539[.304](.283){.374}
		.25	-.324[.278](.215)	-.521[.279](.278){.286}	.061[.290](.220)	.137[.235](.233){.255}
		.00	-.402[.227](.205)	-.525[.239](.238){.228}	-.137[.269](.231)	-.013[.239](.237){.267}
		-.25	-.454[.198](.192)	-.520[.213](.212){.230}	-.311[.242](.234)	-.239[.242](.241){.233}
		-.50	-.495[.186](.186)	-.518[.187](.186){.182}	-.475[.234](.232)	-.463[.243](.240){.244}
100	.50	.50	.480[.118](.116)	.465[.127](.126){.119}	.449[.143](.134)	.473[.121](.120){.145}
		.25	.445[.122](.109)	.466[.171](.167){.158}	.248[.157](.157)	.251[.141](.138){.123}
		.00	.433[.119](.098)	.482[.127](.126){.119}	.055[.178](.169)	-.051[.166](.162){.160}
		-.25	.428[.114](.088)	.495[.101](.101){.105}	-.133[.215](.180)	-.232[.206](.198){.193}
		-.50	.422[.112](.080)	.494[.089](.088){.099}	-.305[.271](.187)	-.514[.232](.131){.133}
	.00	.50	.127[.194](.146)	-.052[.229](.225){.240}	.362[.191](.131)	.460[.155](.150){.160}
		.25	.041[.144](.138)	-.037[.153](.150){.166}	.160[.174](.148)	.223[.135](.131){.121}
		.00	-.012[.130](.130)	-.026[.121](.120){.120}	-.034[.163](.159)	-.008[.168](.161){.171}
		-.25	-.045[.130](.122)	-.018[.136](.127){.143}	-.210[.172](.167)	-.234[.194](.187){.197}
		-.50	-.065[.132](.115)	-.008[.151](.149){.147}	-.382[.207](.170)	-.530[.132](.131){.159}
	-.50	.50	-.196[.349](.171)	-.514[.235](.234){.249}	.306[.234](.131)	.451[.187](.181){.190}
		.25	-.312[.244](.156)	-.516[.225](.219){.219}	.088[.221](.151)	.248[.130](.123){.121}
		.00	-.390[.181](.143)	-.513[.247](.247){.250}	-.108[.191](.158)	-.063[.161](.154){.157}
		-.25	-.447[.148](.138)	-.510[.223](.223){.247}	-.289[.167](.162)	-.224[.195](.186){.176}
		-.50	-.489[.131](.131)	-.504[.201](.201){.227}	-.454[.170](.163)	-.517[.171](.164){.168}
250	.50	.50	.489[.069](.068)	.497[.117](.113){.111}	.464[.088](.080)	.493[.116](.116){.120}
		.25	.462[.071](.060)	.497[.075](.074){.073}	.257[.094](.094)	.254[.127](.126){.124}
		.00	.453[.070](.052)	.499[.057](.057){.057}	.058[.117](.102)	-.017[.141](.140){.138}
		-.25	.448[.070](.047)	.498[.049](.049){.049}	-.136[.159](.110)	-.257[.109](.108){.105}
		-.50	.443[.071](.042)	.498[.044](.044){.046}	-.316[.217](.114)	-.502[.107](.107){.108}
	.00	.50	.112[.140](.084)	-.016[.107](.106){.108}	.391[.132](.074)	.498[.106](.105){.108}
		.25	.038[.088](.079)	-.011[.103](.103){.102}	.185[.109](.088)	.255[.125](.124){.125}
		.00	-.004[.073](.073)	-.008[.101](.100){.099}	-.011[.097](.096)	-.016[.110](.109){.106}
		-.25	-.032[.075](.068)	-.006[.087](.087){.087}	-.196[.113](.100)	-.257[.105](.105){.105}
		-.50	-.051[.081](.063)	-.004[.076](.075){.081}	-.369[.168](.105)	-.502[.107](.107){.107}
	-.50	.50	-.244[.275](.100)	-.501[.101](.101){.099}	.348[.170](.075)	.489[.097](.095){.101}
		.25	-.345[.177](.086)	-.501[.146](.145){.140}	.130[.148](.086)	.253[.118](.117){.116}
		.00	-.406[.124](.081)	-.503[.124](.124){.124}	-.073[.118](.093)	-.019[.134](.132){.131}
		-.25	-.453[.089](.076)	-.503[.105](.105){.102}	-.260[.100](.100)	-.257[.105](.105){.105}
		-.50	-.489[.075](.074)	-.504[.103](.103){.104}	-.432[.122](.101)	-.506[.107](.107){.107}
500	.50	.50	.493[.048](.048)	.499[.076](.076){.077}	.468[.066](.057)	.499[.082](.082){.083}
		.25	.467[.054](.042)	.497[.051](.051){.051}	.261[.066](.065)	.254[.088](.088){.088}
		.00	.457[.056](.036)	.499[.039](.039){.039}	.060[.094](.072)	-.011[.099](.098){.098}
		-.25	.451[.059](.033)	.500[.034](.034){.034}	-.133[.140](.077)	-.256[.108](.107){.107}
		-.50	.447[.061](.029)	.499[.030](.030){.031}	-.313[.203](.079)	-.500[.118](.118){.121}
	.00	.50	.114[.129](.060)	-.008[.117](.114){.113}	.397[.115](.053)	.496[.076](.075){.080}
		.25	.041[.070](.056)	-.004[.089](.088){.088}	.190[.086](.062)	.249[.087](.087){.087}
		.00	-.002[.052](.052)	-.003[.070](.070){.070}	-.007[.069](.068)	-.008[.098](.097){.097}
		-.25	-.028[.056](.048)	-.003[.061](.061){.061}	-.192[.091](.070)	-.256[.106](.106){.107}
		-.50	-.047[.065](.045)	-.002[.054](.054){.056}	-.366[.152](.073)	-.501[.114](.114){.113}
	-.50	.50	-.241[.269](.072)	-.500[.134](.134){.130}	.352[.158](.054)	.497[.069](.068){.065}
		.25	-.342[.170](.062)	-.501[.105](.105){.109}	.133[.132](.062)	.254[.084](.083){.083}
		.00	-.407[.109](.058)	-.500[.089](.089){.089}	-.066[.093](.066)	-.001[.094](.094){.094}
		-.25	-.453[.072](.054)	-.500[.078](.078){.078}	-.251[.069](.069)	-.251[.104](.104){.104}
		-.50	-.487[.053](.052)	-.500[.071](.071){.073}	-.426[.103](.071)	-.501[.117](.116){.117}

Table 6.5c. Empirical Mean[rmse](sd){$\hat{\text{sd}}$} of Estimators of λ and ρ, 1FE-SPD Model with SARAR
Case when the regular QML estimator is inconsistent under heteroskedasticity
$T = 3, \beta = (1, 1)', \sigma = 1$, `Group Interaction, REG-2, DGP 3`

n	λ	ρ	QML-λ	ACQS-λ	QML-ρ	ACQS-ρ
50	.50	.50	.444[.253](.247)	.538[.264](.246){.248}	.428[.225](.214)	.543[.231](.230){.236}
		.25	.421[.239](.225)	.460[.233](.234){.231}	.218[.257](.255)	.266[.237](.230){.234}
		.00	.413[.211](.192)	.452[.224](.219){.234}	.028[.269](.267)	-.088[.240](.239){.258}
		-.25	.409[.195](.172)	.467[.192](.188){.156}	-.162[.298](.285)	-.258[.244](.243){.245}
		-.50	.413[.168](.144)	.480[.164](.162){.149}	-.349[.333](.297)	-.462[.233](.219){.210}
	.00	.50	.084[.285](.273)	-.072[.241](.241){.199}	.347[.265](.217)	.441[.230](.228){.259}
		.25	.014[.244](.244)	-.051[.239](.235){.234}	.132[.266](.238)	.244[.235](.234){.233}
		.00	-.033[.235](.232)	-.043[.255](.223){.233}	-.058[.261](.255)	-.012[.240](.239){.236}
		-.25	-.055[.205](.198)	-.034[.254](.252){.246}	-.245[.265](.265)	-.237[.242](.240){.244}
		-.50	-.076[.199](.184)	-.024[.223](.222){.210}	-.411[.276](.261)	-.462[.242](.239){.261}
	-.50	.50	-.235[.384](.278)	-.498[.341](.341){.360}	.293[.301](.218)	.540[.228](.225){.230}
		.25	-.346[.302](.260)	-.514[.336](.336){.322}	.075[.299](.243)	.244[.233](.232){.248}
		.00	-.411[.250](.234)	-.513[.262](.259){.242}	-.127[.275](.244)	-.013[.236](.236){.234}
		-.25	-.464[.219](.216)	-.516[.296](.299){.292}	-.307[.252](.245)	-.282[.240](.237){.231}
		-.50	-.500[.204](.204)	-.511[.270](.269){.268}	-.473[.257](.255)	-.528[.242](.237){.227}
100	.50	.50	.462[.204](.200)	.463[.227](.226){.254}	.452[.163](.156)	.472[.121](.121){.125}
		.25	.441[.161](.150)	.462[.176](.172){.176}	.247[.172](.172)	.243[.190](.163){.150}
		.00	.433[.145](.129)	.476[.136](.133){.136}	.044[.199](.194)	-.041[.173](.168){.177}
		-.25	.432[.123](.103)	.487[.106](.105){.104}	-.147[.227](.202)	-.233[.204](.198){.196}
		-.50	.429[.117](.093)	.492[.099](.099){.116}	-.329[.275](.216)	-.508[.225](.217){.231}
	.00	.50	.099[.235](.213)	-.051[.232](.231){.236}	.373[.194](.147)	.546[.199](.194){.196}
		.25	.028[.190](.188)	-.033[.251](.248){.249}	.163[.192](.171)	.230[.139](.134){.123}
		.00	-.016[.163](.162)	-.023[.203](.201){.200}	-.030[.179](.176)	-.005[.163](.157){.170}
		-.25	-.042[.144](.137)	-.016[.175](.175){.194}	-.214[.183](.179)	-.253[.196](.188){.198}
		-.50	-.064[.142](.127)	-.013[.157](.157){.180}	-.385[.224](.193)	-.540[.232](.231){.230}
	-.50	.50	-.223[.353](.218)	-.506[.234](.234){.232}	.315[.237](.148)	.463[.192](.184){.187}
		.25	-.337[.263](.206)	-.511[.286](.286){.275}	.100[.227](.170)	.287[.133](.125){.120}
		.00	-.406[.202](.179)	-.510[.227](.225){.224}	-.099[.199](.173)	-.068[.162](.154){.168}
		-.25	-.454[.171](.165)	-.506[.176](.174){.161}	-.282[.183](.180)	-.232[.191](.182){.196}
		-.50	-.492[.155](.155)	-.501[.135](.134){.147}	-.453[.192](.186)	-.458[.216](.206){.192}
250	.50	.50	.485[.104](.103)	.487[.115](.114){.124}	.464[.094](.087)	.490[.114](.114){.113}
		.25	.464[.080](.071)	.500[.078](.078){.072}	.256[.097](.097)	.257[.108](.107){.103}
		.00	.456[.073](.058)	.496[.057](.057){.056}	.052[.129](.118)	-.016[.104](.104){.104}
		-.25	.449[.073](.052)	.498[.049](.049){.048}	-.140[.162](.120)	-.258[.106](.105){.110}
		-.50	.445[.075](.051)	.500[.044](.044){.044}	-.326[.224](.141)	-.506[.107](.107){.107}
	.00	.50	.100[.156](.119)	-.018[.169](.168){.168}	.398[.136](.091)	.499[.108](.107){.110}
		.25	.032[.106](.101)	-.013[.126](.126){.123}	.189[.115](.097)	.254[.124](.123){.118}
		.00	-.006[.082](.082)	-.008[.099](.099){.098}	-.012[.101](.100)	-.017[.108](.107){.104}
		-.25	-.029[.083](.078)	-.004[.088](.087){.087}	-.200[.119](.108)	-.257[.105](.105){.105}
		-.50	-.047[.087](.073)	-.003[.081](.081){.080}	-.379[.172](.123)	-.507[.108](.107){.107}
	-.50	.50	-.263[.274](.137)	-.502[.174](.173){.180}	.357[.168](.089)	.489[.095](.094){.102}
		.25	-.356[.186](.118)	-.504[.145](.145){.146}	.135[.151](.097)	.253[.115](.113){.113}
		.00	-.416[.134](.104)	-.505[.126](.126){.126}	-.067[.125](.106)	-.020[.109](.105){.107}
		-.25	-.455[.102](.091)	-.502[.106](.105){.107}	-.256[.108](.108)	-.258[.105](.105){.104}
		-.50	-.487[.084](.083)	-.502[.103](.103){.106}	-.432[.131](.112)	-.508[.107](.106){.107}
500	.50	.50	.490[.075](.074)	.498[.077](.077){.078}	.470[.067](.060)	.500[.080](.080){.080}
		.25	.467[.057](.046)	.500[.054](.054){.050}	.261[.068](.067)	.254[.088](.088){.087}
		.00	.457[.060](.043)	.500[.043](.043){.040}	.057[.096](.078)	-.010[.098](.097){.096}
		-.25	.451[.061](.038)	.500[.034](.034){.034}	-.134[.145](.087)	-.251[.108](.108){.108}
		-.50	.448[.064](.037)	.500[.030](.030){.031}	-.317[.210](.103)	-.501[.117](.116){.120}
	.00	.50	.105[.139](.092)	-.009[.122](.121){.122}	.402[.117](.063)	.500[.075](.075){.079}
		.25	.039[.076](.065)	-.003[.081](.081){.084}	.191[.088](.065)	.250[.086](.085){.085}
		.00	-.002[.061](.061)	-.003[.070](.070){.070}	-.006[.072](.071)	-.006[.097](.095){.094}
		-.25	-.026[.058](.052)	-.001[.060](.060){.060}	-.196[.094](.077)	-.253[.106](.105){.105}
		-.50	-.046[.066](.047)	-.002[.054](.054){.054}	-.370[.155](.084)	-.501[.116](.115){.113}
	-.50	.50	-.255[.267](.106)	-.500[.131](.131){.135}	.359[.156](.066)	.500[.067](.065){.064}
		.25	-.350[.172](.086)	-.500[.106](.106){.108}	.139[.132](.070)	.254[.082](.082){.082}
		.00	-.411[.114](.071)	-.500[.090](.090){.089}	-.063[.095](.071)	-.001[.094](.093){.092}
		-.25	-.454[.080](.065)	-.500[.078](.077){.077}	-.251[.074](.074)	-.252[.106](.105){.106}
		-.50	-.486[.060](.058)	-.500[.071](.071){.071}	-.426[.106](.076)	-.501[.116](.115){.115}

CHAPTER 7

Conclusions

In this study, we provide asymptotically refined and heteroskedasticity robust inferences for spatial linear and panel regression models, based on the QML or the ACQS approaches. We recommend refinements through bias correcting the QML estimators, bias correcting the t-ratios for covariates effects, and improving tests for spatial effects. We also provide heteroskedasticity-robust inferences by adjusting the quasi score functions so that it goes to zero in expectation at the true parameter by design. Thus the resulting estimator is consistent even when the model suffers from heteroskedasticity.

These methods are illustrated using several popular spatial linear and panel regression models including the linear regression models with spatial error dependence (SED), spatial lag dependence (SLD), or both SED and SLD (SARAR), the linear regression models with higher-order spatial effects, SARAR(p, q), and the fixed effects panel data models with SED or SLD or both. Asymptotic properties of the new estimators and the new inferential statistics are examined. The methods are also implemented in extensive Monte Carlo experiments which show excellent performance.

Implementation of the methodologies discussed in this thesis to *dynamic* spatial panel data models will be an interesting avenue of research. Especially when the time dimension is short, how the initial observation affect asymptotic refinements and the ACQS estimator need to be explored. We wish to pursue this issue in future.

Bibliography

[1] Alvarez, J., Arellano, M., 2004. Robust likelihood estimation of dynamic panel data models. Working Paper 0421, CEMFI.

[2] Amemiya, T., 1985. *Advanced Econometrics.* Cambridge, Massachusetts: Harvard University Press.

[3] Anselin, L., 1980. *Estimation methods for spatial autoregressive structures: A study in spatial econometrics.* Regional Science Dissertation and Monograph Series 8.

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

[5] Anselin, L., 2001. *Spatial Econometrics. A Companion to Theoretical Econometrics, edited by Badi H. Baltagi.* Blackwell Publishing.

[6] Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26, 153-166.

[7] Anselin, L., Bera, A. K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: *Handbook of Applied Economic Statistics, edited by Aman Ullah and David E. A. Giles.* New York: Marcel Dekker.

[8] Anselin, L., Le Gallo, J., Jayet, J., 2008. Spatial panel econometrics. In: Mátyás, L., Sevestre, P. (Eds.), The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice. Springer-Verlag, Berlin Heidelberg.

[9] Arellano, M., Hahn, J., 2005. Understanding bias in non-linear panel models: some recent developments. In: R. Blundell, W. K. Newey and T. Persson (Eds.), *Advances in Econometrics and Economics: Theory and Applications, IXth World Congress, Volume II,* Econometric Society Monographs.

[10] Arraiz, I., Drukker, D. M., Kelejian H. H., Prucha, I. R., 2010. A spatial Cliff-Ord type model with heteroskedastic innovations: small and large sample results. *Journal of Regional Science*, 50, 592-614.

[11] Badinger, H., Egger, P., 2011. Estimation of higher-order spatial autoregressive cross-section models with heteroskedastic disturbances. *Papers in Regional Science* 90, 213-235.

[12] Badinger, H., Egger, P., 2015. Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroskedastic disturbances. *Spatial Economic Analysis* 10:1, 11-35.

[13] Baltagi, B., Egger, P., Pfaffermayr, M., 2007. Estimating models of complex FDI: are there third country effects? *Journal of Econometrics* 140, 260-281.

[14] Baltagi, B., Egger, P., Pfaffermayr, M., 2013. A generalised spatial panel data model with random effects. *Econometric Reviews* 32, 650-685.

[15] Baltagi, B., Egger, P., Kesina, M., 2016. Firm level productivity spillovers in China's chemical industry: a spatial Hausman-Taylor approach. *Journal of Applied Econometrics* 31, 214-248.

[16] Baltagi, B., Song, S. H., Kon, W., 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117, 123-150.

[17] Baltagi, B., Egger, P., Pfaffermayr, M., 2013. A generalised spatial panel data model with random effects. *Econometric Reviews* 32, 650-685.

[18] Baltagi, B. H., Yang, Z. L., 2013a. Standardised LM tests for spatial error dependence in linear or panel regressions. *The Econometrics Journal* 16, 103-134.

[19] Baltagi, B. H., Yang, Z. L., 2013b. Heteroskedasticity and non-normality robust LM tests for spatial dependence. *Regional Science and Urban Economics* 43, 725-739.

[20] Bao, Y., 2013. Finite sample bias of the QML estimator in spatial autoregressive models. *Econometric Theory* 29, 68-88.

[21] Bao, Y., Ullah, A., 2007. Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* 137, 396-413.

[22] Born, B., Breitung, J., 2011. Simple regression based tests for spatial dependence. *The Econometrics Journal*. 14, 330-342.

[23] Breusch, T., Pagan, A., 1979. A simple test for heteroskedasticity and random coefficient variation. *Econometrica* 47, 1287-1294.

[24] Bun, M. J. G., Carree, M. A., 2005. Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics* 23, 200-210.

[25] Burridge, P. 1980. On the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society* B 42, 107-8.

[26] Case, T., 1991. Spatial patterns in household demand. *Econometrica.* 59, 953-965.

[27] Cliff, A., Ord, J. K., 1972. Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* 4, 267-284.

[28] Cliff, A. D., Ord, J. K., 1973. Spatial Autocorrelation. London: Pion.

[29] Cliff, A. D., Ord, J. K., 1981. *Spatial Process, Models and Applications*. London: Pion.

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

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

[32] Davidson , J., 1994. *Stochastic Limit Theory*. Oxford University Press.

[33] Davidson, R., Flachaire, E., 2008. The wild bootstrap, tamed at last. *Journal of Econometrics* 146, 162169.

[34] Doğan, O., Taşpinar, S., 2014. Spatial autoregressive models with unknown heteroskedasticity: a comparison of Bayesian and robust GMM approach. *Regional Science and Urban Economics* 45, 1-21.

[35] Efron, B., 1979. Bootstrap method: another look at the jackknife. *The Annals of Statistics* 7, 1-26.

[36] Elhorst, J. P., 2003. Specification and estimation of spatial panel data models. *International Regional Science Review* 26(3), 244-268.

[37] Elhorst, J. P., Lacombe, D. J., Piras, G., 2012. On model specification and parameter space definitions in higher order spatial econometrics models. *Regional Science and Urban Economics* 42, 211-220.

[38] Fingleton, B., 2008. A generalised method of moments estimator for a spatial model with moving average errors, with application to real estate prices. *Empirical Economics* 33, 35-57.

[39] Fleming, M. M., 2004. Techniques for Estimating Spatially Dependent Discrete Choice Models *Advances in Spatial Econometrics, edited by Luc Anselin, Raymond J. G. M. Florax and Sergio J. Rey.* Springer Berlin Heidelberg.

[40] Glaeser, E. L., Sacerdote, B., Scheinkman, J. A., 1996. Crime and social interactions. *Quarterly Journal of Economics* 111, 507-548.

[41] Griffith, D. A. 1988. *Advanced Spatial Statistics.* Kluwer, Dordrecht, the Netherlands.

[42] Hahn J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639-1657.

[43] Hahn J., Moon, H. R., 2006. Reducing bias in ML estimator in a dynamic panel data model. *Econometric Theory* 22, 499-512.

[44] Hahn J., Newey, W., 2004. Jackknife and analytical bias reduction for non-linear panel models. *Econometrica* 72, 1295-1319.

[45] Hanushek, E. A., Kain, J. F., Markman, J. M., Rivkin, S. G., 2003. Does peer ability affect student achievement? *Journal of Applied Econometrics* 18, 527-544.

[46] Hsieh, C. S., Lee, L. F., 2014. A social interactions model with endogenous friendship and selectivity. *Forthcoming in Journal of Applied Econometrics.*

[47] Huber, P. J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics.* 35, 73-101.

[48] Huber, P. J., 1981. Robust Statistics. *Wiley*, New York.

[49] Jin, F., Lee, L. F., 2012. Approximated likelihood and root estimators for spatial interaction in spatial autoregressive models. *Regional Science and Urban Economics.* 42 446-458.

[50] Jin, F., Lee, L. F., 2013. Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. *Regional Science and Urban Economics* 43, 590-616.

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

[52] Kelejian H. H., Piras, G., 2014. An extension of the J-test to spatial panel data framework. *Forthcoming in Journal of Applied Econometrics.*

[53] Kelejian H. H., Prucha, I. R., 1998. A generalised spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. *Journal of Real Estate Finance and Economics* 17, 99-121.

[54] Kelejian H. H., Prucha, I. R., 1999. A generalised moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.

[55] Kelejian H. H., Prucha, I. R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219-257.

[56] Kelejian H. H., Prucha, I. R., 2004. Instrumental variable estimation of a spatial autoregressive models with autoregressive disturbances: Large and small sample results. In *Spatial and Spatiotemporal Econometrics, Advances in Econometrics edited by LeSage, J., and Pace, R. K.* New York: Elsevier 18, 163-198.

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

[58] Kelejian H. H., Prucha, I. R., 2010 Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157 53-67.

[59] Kelejian H. H., 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* 72 297-312.

[60] Kendall, M. G., Stuart, A., 1969. *Advanced Theory of Statistics*, Vol. 1, 3rd Ed. Griffin, London.

[61] Kiviet, J. F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 53-78.

[62] Lancaster, T., 2000. Incidental parameter problem since 1948. *Journal of Econometrics* 95, 391-413.

[63] Lee, L. F., 2001. Generalised method of moments estimation of spatial autoregressive processes - Unpublished manuscript.

[64] Lee, L. F., 2002. Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. *Econometric Theory* 18, 252-277.

[65] Lee, L. F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive models with autoregressive disturbances. *Econometric Reviews* 22, 307-335.

[66] Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.

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

[68] Lee, L. F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26 187-230.

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

[70] Lee, L. F., Yu, J., 2010b. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154, 165-185.

[71] Lee, L. F., Yu, J., 2012. Spatial panels: random components vs. fixed effects. *International Economic Review* 53, 1369-1412.

[72] Lee, L. F., Yu, J., 2015. Spatial panel data models. In: Baltagi, B. H. (Eds.), The Oxford Handbook of Panel Data. Oxford University Press, New York, NY, pp. 363-400.

[73] Lee, L. F., Yu, J., 2016. Identification of spatial Durbin models. *Journal of Applied Econometrics* 31, 133-162.

[74] Le Sage, J., 1997. Bayesian estimation of spatial autoregressive models. *International Regional Science Review* 20, 113-129.

[75] Le Sage, J., Pace, R.K., 2009. *Introduction to Spatial Econometrics*. CRC Press, Taylor & Francis Group, London.

[76] Lieberman, O. (1994). A Laplace approximation to the moments of a ratio of quadratic forms. *Biometrika* 81, 681-90.

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

[78] Liu, R.Y., 1988. Bootstrap procedures under some non-I.I.D. models. *Annals of Statistics* 16, 1696-1708.

[79] Liu, S. F., Yang, Z. L., 2015a. Asymptotic distribution and finite-sample bias correction of QML estimators for spatial error dependence model. *Econometrics* 3, 376-411.

[80] Liu, S. F., Yang, Z. L., 2015b. Adjusted QML Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity and Nonnormality. *Regional Science and Urban Economics* {52}, 50-70.

[81] Liu, S. F., Yang, Z. L., 2015c. Improved inferences for spatial regression models. *Regional Science and Urban Economics* 55, 55-67.

[82] Magnus, J. R., 1978. Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics*, 7, 281-312.

[83] Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21, 255

[84] Martellosio, F., 2010. Power properties of invariant tests for spatial autocorrelation in linear regression. *Econometric Theory* 26, 152-186.

[85] Moscone, F., Tosetti, E., 2011. GMM estimation of spatial panels with fixed effects and unknown heteroskedasticity. *Regional Science and Urban Economics.* 41 487-497.

[86] Mutl, J., Pfaffermayr, M., 2011. The Hausman test in a Cliff and Ord panel model. *Econometrics Journal* 14, 48-76.

[87] Neyman, J., Scott, E. L., 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16, 1-32.

[88] Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 39, 383-396.

[89] Ord, J., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.

[90] Pinkse, J. 1998. Asymptotic properties of the Moran and related tests and a test for spatial correlation in probit models. Department of Economics, University of British Columbia.

[91] Pinkse, J., Slade, M. E., Brett, C., 2002. Spatial price competition: a semiparametric approach. *Econometrica* 70, 1111-1153.

[92] Rao, C. R., 1973. Linear Statistical Inference and Its Applications. New York: John Wiley and Sons.

[93] Rilstone, P., Srivastava, V. K., Ullah, A., 1996. The second-order bias and mean squared error of non-linear estimators. *Journal of Econometrics* 75, 369-395.

[94] Robinson, P. M., Rossi, F., 2014a. Refined tests for spatial correlation. *Econometric Theory*, 31 1249-1280.

[95] Robinson, P. M., Rossi, F., 2014b. Improved Lagrange multiplier tests in spatial autoregressions. *Econometrics Journal* 17, 139-164.

[96] Stock, J. H., Watson, M. W., 2008. Heterokedasticity robust standard errors for fixed effects panel data regression. *Econometrica*, 76, 155-174.

[97] Su, L. J., Yang Z. L., 2015. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185, 230-258.

[98] van der Vaart, A. W., 1998. *Asymptotic Statistics.* Cambridge University Press.

[99] White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 817-838.

[100] White, H., 1994. *Estimation, Inference and Specification Analysis.* New York: Cambridge University Press.

[101] Yang, Z. L., 2010. A robust LM test for spatial error components. *Regional Science and Urban Economics* 40, 299-310.

[102] Yang, Z. L., 2015a. LM tests of spatial dependence based on bootstrap critical values. *Journal of Econometrics* 185, 33-39.

[103] Yang, Z. L., 2015b. A general method for third-order bias and variance corrections on a non-linear estimator. *Journal of Econometrics* 186, 178-200.

[104] Yang, Z. L., 2016. Unified M-estimation of fixed effects spatial dynamic models with short panels.

[105] Yu, J., de Jong, R., Lee, L. F., 2008. Quasi maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146, 118-134.

[106] Yu, J., de Jong, R., Lee, L. F., 2012. Estimation of spatial dynamic panel data with fixed effects: The case of spatial cointegration. *Journal of Econometrics* 167, 16-37.

[107] Yu, J., Lee, L. F., 2010. Estimation of unit root spatial dynamic panel data models. *Econometric Theory* 26, 1332-1362.

APPENDIX A

Some Useful Lemmas

The following lemmas are extended versions of selected lemmas from Kelejian and Prucha (2010), Lee (2004), and Lin and Lee (2010), which are required in the proofs of the main results. Note that the following results are give in the most general form as required by this dissertation, however, for certain simpler models special cases of these results are applicable.

Lemma A.1: *Under Assumptions 6.7, 6.9 and 6.10, the projection matrices,* $\mathbf{M}_N(\rho) = \mathbf{I}_N - \mathbf{P}_N(\rho)$ *and* $\mathbf{P}_N(\rho) = \mathbf{A}_{2N}(\rho)\mathbf{X}_N[\mathbf{X}'_N\mathbf{A}'_{2N}(\rho)\mathbf{A}_{2N}(\rho)\mathbf{X}_N]^{-1}\mathbf{X}'_N\mathbf{A}'_{2N}(\rho)$ *are uniformly bounded in both row and column sums, where* $\mathbf{A}_{2N}(\rho)$ *and* $\mathbf{X}_N$ *are as defined in Section 6.2.*

Lemma A.2: *Let* $\mathbf{A}_N$ *be an* $N \times N$ *matrix, uniformly bounded in both row and column sums. Then for* $\mathbf{M}_N$ *defined in Lemma A.1,*

(i) $\operatorname{tr}(\mathbf{A}_N^m) = O(N)$ *for* $m \geq 1$*,*

(ii) $\operatorname{tr}(\mathbf{A}'_N\mathbf{A}_N) = O(N)$*,*

(iii) $\operatorname{tr}((\mathbf{M}_N\mathbf{A}_N)^m) = \operatorname{tr}(\mathbf{A}_N^m) + O(1)$ *for* $m \geq 1$ *and*

(iv) $\operatorname{tr}((\mathbf{A}'_N\mathbf{M}_N\mathbf{A}_N)^m) = \operatorname{tr}((\mathbf{A}'_N\mathbf{A}_N)^m) + O(1)$ *for* $m \geq 1$*.*

Let $\mathbf{B}_N$ *be another* $N \times N$ *matrix, uniformly bounded in both row and column sums. Then,*

(*iv*) $\mathbf{A}_N\mathbf{B}_N$ *is uniformly bounded in both row and column sums,*

(*v*) $\mathrm{tr}(\mathbf{A}_N\mathbf{B}_N) = \mathrm{tr}(\mathbf{B}_N\mathbf{A}_N) = O(N)$ *uniformly.*

Lemma A.3 (Moments and Limiting Distribution for Linear Quadratic forms): *For a given process of innovations* $\{v_{it}\}$*, let* $v_{it} \sim inid(0, \sigma_0^2 h_i)$*, where* $h_i > 0$ *for* $i = 1, \ldots, n$ *such that* $\frac{1}{n}\sum_{i=1}^{n} h_i = 1$*. Let* $H_n = diag(h_1, \ldots, h_n)$*,* $N = n \times T$*,* $\mathbf{H}_N = I_T \otimes H_n$*,* $\mathbf{B}_N$ *be an* $N \times N$ *matrix with diagonal elements* b_{it} *and* $\mathbf{c}_N$ *an* $N \times 1$ *vector with elements* c_{it} *for* $i = 1, \ldots, n$ *and* $t = 1, \ldots, T$*. For* $\mathbf{Q}_{rN} = \mathbf{V}_N'\mathbf{B}_{rN}\mathbf{V}_N + \mathbf{c}_{rN}'\mathbf{V}_N$*,* $r = 1, 2$*, where* $\mathbf{V}_N = (V_{n1}', \ldots, V_{nT}')'$*, then,*

(*i*) $\mathrm{E}(\mathbf{Q}_{rN}) = \sigma_0^2\mathrm{tr}(\mathbf{H}_N\mathbf{B}_{rN})$,

(*ii*) $\mathrm{Var}(\mathbf{Q}_{rN}) = \sigma_0^4\mathrm{tr}[\mathbf{H}_N\mathbf{B}_{rN}(\mathbf{H}_N\mathbf{B}_{rN} + \mathbf{B}_{rN}'\mathbf{H}_N)] + \sigma_0^2\mathbf{c}_{rN}'\mathbf{H}_N\mathbf{c}_{rN}$
$+ \sum_{i=1}^{n}\sum_{t=1}^{T}(\sigma_0^4 b_{r,it}^2 h_i^2\kappa_i + 2\sigma_0^3 b_{r,it}c_{r,it}h_i^{3/2}s_i)$ *and*

(*iii*) $\mathrm{Cov}(\mathbf{Q}_{1N}, \mathbf{Q}_{2N}) = 2\sigma_0^2\mathrm{tr}(\mathbf{B}_{1N}\mathbf{H}_N\mathbf{B}_{2N}\mathbf{H}_N) + \sigma_0^2\mathbf{c}_{1N}'\mathbf{H}_N\mathbf{c}_{2N}$
$+ \sum_{i=1}^{n}\sum_{t=1}^{T}\left[\sigma_0^4 b_{1,it}b_{2,it}h_i^2\kappa_i + \sigma_0^3(b_{1,it}c_{2,it} + b_{2,it}c_{1,it})h_i^{3/2}s_i\right]$,

where s_i *and* κ_i *are, respectively, the measures of skewness and excess kurtosis of* v_{it}*. Now, if* $\mathbf{B}_{rN}$ *is uniformly bounded in either row or column sums then,*

(*iv*) $\mathrm{E}(\mathbf{Q}_{rN}) = O(N)$,

(*v*) $\mathrm{Var}(\mathbf{Q}_{rN}) = O(N)$,

(*vi*) $\mathbf{Q}_{rN} = O_p(N)$,

(*vii*) $\frac{1}{N}\mathbf{Q}_{rN} - \frac{1}{N}\mathrm{E}(\mathbf{Q}_{rN}) = O_p\big(N^{-\frac{1}{2}}\big)$ *and*

(*viii*) $\mathrm{Var}(\frac{1}{N}\mathbf{Q}_{rN}) = O(N^{-1})$.

Further, if $\mathbf{B}_{rN}$ *is uniformly bounded in both row and column sums and Assumption 6.8 holds. Let* $\mathbf{Q}_N = (\mathbf{Q}_{1N}, \mathbf{Q}_{2N})'$ *and* $\Sigma_N = \Sigma_N^{1/2}\Sigma_N^{1/2'} = [\mathrm{Cov}(\mathbf{Q}_{rn}, \mathbf{Q}_{sn})]_{r,s=1,2}$*, then,*

(*ix*) $\frac{\mathbf{Q}_{rN} - \mathrm{E}(\mathbf{Q}_{rN})}{\sqrt{\mathrm{Var}(\mathbf{Q}_{rN})}} \xrightarrow{D} \mathcal{N}(0, 1)$ *and*

(*x*) $\Sigma_N^{-1/2}(\mathbf{Q}_N - \mathrm{E}(\mathbf{Q}_N)) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$.

APPENDIX B

Settings of Monte Carlo Experiments

Spatial Weight Matrix: Spatial weight matrices are generated according to: (i) `Rook contiguity`, (ii) `Queen contiguity`, (iii) `Circular neighbours` and (iv) `Group interaction`, (details given in Yang, 2015b). In (ii), neighbours could occur in the eight cardinal and ordinal positions of each unit while in (i) neighbours could occur only in the ordinal positions. In (iii), neighbours occur in the positions immediately ahead and behind a particular spatial unit. For example, for the ith spatial unit with 6 neighbours, the ith row of W_n matrix has non-zero elements in the positions: $i-3, i-2, i-1, i+1, i+2$, and $i+3$. We consider has 2, 4, 6, 8 and 10 neighbours with equal proportion. In (iv), neighbours occur in groups where each group member is spatially related to one another resulting in a symmetric W_n matrix. The degree of spatial dependence specified by layouts (i), (ii) and (iii) are fixed while in (iv) it grows with the increase in sample size. This is attained by allowing for the number of groups, k, to be related to n. Specifically in (iv), W_n is block-diagonal, with k blocks (groups) of sizes $n_1, \ldots, n_k$. The rth block is an $n_r \times n_r$ matrix with off-diagonal elements $\frac{1}{n_r-1}$ and diagonal elements zero. We have considered $k = n^{0.5}$ and $k = n^{0.65}$, where k is the number of groups for each n and hence the degree of spatial dependence indicated by the average group size is $m = n/k$. The actual sizes of the groups are generated from a discrete uniform distribution from $.5m$ to $1.5m$. Clearly in this case the degree of spatial dependence, indicated by the average group size m, increases with n, and it is stronger when $\delta = .5$ than when $\delta = .65$. In the case of experiments using heteroskedastic disturbances, to ensure the heteroskedasticity effect does not fade as n increases (so that the regular QML estimator is inconsistent), the degree of spatial dependence is fixed with respect to n. This is attained by fixing the group sizes in the Group Interaction scheme, and fixing the number of neighbours the Circular Neighbours scheme. The degree of spatial dependence is naturally bounded in the Queen Contiguity weight matrix. To analyse the performance of the original QML estimator when it is robust against heteroskedasticity, we use Queen Contiguity scheme and the `balanced Circular Neighbours` scheme where all spatial units have 6 peers each. All weights matrices are row normalised.

Regressors: Fixed regressors are generated by `REG1:` $\{x_{1i}, x_{2i}\} \overset{iid}{\sim} N(0,1)/\sqrt{2}$ when Rook contiguity, Queen contiguity or Circular neighbours is followed; and according to either `REG1` or `REG2:` $\{x_{1,ir}, x_{2,ir}\} \overset{iid}{\sim} (2z_r + z_{ir})/\sqrt{10}$, where, $(z_r, z_{ir}) \overset{iid}{\sim} N(0,1)$ when group interaction scheme is followed. The `REG2` scheme gives non-iid regressors where the group means of the regressors' values are different, see Lee (2004). Note that both schemes give a signal-to-noise ratio of 1 when $\beta_1 = \beta_2 = \sigma = 1$

Error Distribution: To generate disturbances, three DGPs are considered: `DGP1:` $\{e_{n,i}\}$ are iid standard normal, `DGP2:` $\{e_{n,i}\}$ are iid standardised normal mixture with 10% of values from $N(0,4)$ and the remaining from $N(0,1)$, and `DGP3:` $\{e_{n,i}\}$ iid standardised log-normal with parameters 0 and 1. Thus, the error distribution from `DGP2` is leptokurtic, and that of `DGP3` is both skewed and leptokurtic.

Heteroskedasticity: For the `unbalanced Circular Neighbour` scheme, $h_{n,i}$ is generated as the ratio of the total number of neighbours to the average number of neighbours for each i while for the Group Interaction scheme $h_{n,i}$ is generated as the ratio of the group size to mean group size. For the balanced Circular Neighbour and the Queen Contiguity schemes, we use $h_{n,i} = n[\sum_{i=1}^{n}(|X_{1n,i}| + |X_{2n,i}|)]^{-1}(|X_{1n,i}| + |X_{2n,i}|)$.

APPENDIX C

Additional Quantities for Bias Corrections

The expressions for $F_n^{(r)} = \frac{d^r}{d\rho_0^r} F_n(\rho_0), r = 1, 2$ needed in Chapter 3 are:

$F_n^{(1)} = F_n B_n^{-1} G_n^s B_n (X_n F_n - I_n)$, where $G_n^s = G_n + G_n'$

$F_n^{(2)} = F_n^{(1)} B_n^{-1} G_n^s B_n (X_n F_n - I_n) + F_n B_n^{-1} (G_n^{s2} - 2G_n' G_n) B_n (X_n F_n - I_n)$
$+ F_n B_n^{-1} G_n^s B_n X_n F_n^{(1)}$

For the SED model, the full expressions for $D_{jn}(\rho), j = 2, 3, 4$, required in the expressions of $R_{jn}(\rho)$ in (2.18), for up to third-order bias corrections are:

$$
\begin{aligned}
D_{2n}(\rho) &= 2G_n(\rho)P_n(\rho)G_n(\rho) + G_n(\rho)P_n(\rho)G_n'(\rho) - G_n'(\rho)M_n(\rho)G_n(\rho), \\
D_{3n}(\rho) &= \dot{D}_{2n}(\rho) + G_n(\rho)P_n(\rho)D_{2n}(\rho) + D_{2n}(\rho)P_n(\rho)G_n'(\rho) \\
&\quad - G_n'(\rho)M_n(\rho)D_{2n}(\rho) - D_{2n}(\rho)M_n(\rho)G_n(\rho), \\
D_{4n}(\rho) &= \dot{D}_{3n}(\rho) + G_n(\rho)P_n(\rho)D_{3n}(\rho) + D_{3n}(\rho)P_n(\rho)G_n'(\rho) \\
&\quad - G_n'(\rho)M_n(\rho)D_{3n}(\rho) - D_{3n}(\rho)M_n(\rho)G_n(\rho),
\end{aligned}
$$

where $P_n(\rho) = I_n - M_n(\rho)$ and $\dot{D}_{jn}(\rho) = \frac{d}{d\rho} D_{jn}(\rho), j = 2, 3$. Note that a predictable pattern emerges from $D_{3n}(\rho)$ onwards. Using the fact that $\frac{d}{d\rho} G_n^i = G_n^{i+1}$ for $i = 1, 2, \ldots$, we have,

$$
\begin{aligned}
\dot{D}_{2n}(\rho) &= 2G_n^2(\rho)P_n(\rho)G_n(\rho) - 2G_n(\rho)\dot{M}_n(\rho)G_n(\rho) + 2G_n(\rho)P_n(\rho)G_n^2(\rho) \\
&\quad + G_n^2(\rho)P_n(\rho)G_n'(\rho) - G_n(\rho)\dot{M}_n(\rho)G_n'(\rho) + G_n(\rho)P_n(\rho)G_n'^2(\rho) \\
&\quad - G_n'^2(\rho)M_n(\rho)G_n(\rho) - G_n'(\rho)\dot{M}_n(\rho)G_n(\rho) - G_n'(\rho)M_n(\rho)G_n^2(\rho), \\
\dot{D}_{3n}(\rho) &= G_n'^3(\rho)M_n(\rho)G_n(\rho) + 2G_n'^2(\rho)\dot{M}_n(\rho)G_n(\rho) + 2G_n'^2(\rho)M_n(\rho)G_n^2(\rho) \\
&\quad + G_n'(\rho)\ddot{M}_n(\rho)G_n(\rho) + 2G_n'(\rho)\dot{M}_n(\rho)G_n^2(\rho) + G_n'(\rho)M_n(\rho)G_n^3(\rho) \\
&\quad - 2G_n^3(\rho)P_n(\rho)G_n(\rho) + 4G_n^2(\rho)\dot{M}_n(\rho)G_n(\rho) - 4G_n^2(\rho)P_n(\rho)G_n^2(\rho) \\
&\quad + 2G_n(\rho)\ddot{M}_n(\rho)G_n(\rho) + 4G_n(\rho)\dot{M}_n(\rho)G_n^2(\rho) - 2G_n(\rho)P_n(\rho)G_n^3(\rho) \\
&\quad - G_n^3(\rho)P_n(\rho)G_n'(\rho) + 2G^2(\rho)\dot{M}_n(\rho)G_n'(\rho) - 2G_n^2(\rho)P_n(\rho)G_n'^2(\rho) \\
&\quad + G_n(\rho)\ddot{M}_n(\rho)G_n'(\rho) + 2G_n(\rho)\dot{M}_n(\rho)G_n'^2(\rho) - G_n(\rho)P_n(\rho)G_n'^3(\rho), \\
\dot{M}_n(\rho) &= P_n(\rho)G_n'(\rho)M_n(\rho) + M_n(\rho)G_n(\rho)P_n(\rho), \\
\ddot{M}_n(\rho) &= 2P_n(\rho)G_n'(\rho)P_n(\rho)G_n'(\rho)M_n(\rho) + 2P_n(\rho)G_n'(\rho)M_n(\rho)G_n(\rho)P_n(\rho) \\
&\quad + 2M_n(\rho)G_n(\rho)P_n(\rho)G_n(\rho)P_n(\rho) - 2M_n(\rho)G_n(\rho)P_n(\rho)G_n'(\rho)M_n(\rho).
\end{aligned}
$$

For the SED model with SMA errors, the additional quantities required by (2.28) are,

$$
\begin{aligned}
D_{2n}(\rho) &= G_n'(\rho)M_n(\rho)G_n(\rho) + 2G_n(\rho)M_n(\rho)G_n(\rho) - G_n(\rho)P_n(\rho)G_n'(\rho), \\
D_{3n}(\rho) &= \dot{D}_{2n}(\rho) - G_n(\rho)P_n(\rho)D_{2n}(\rho) - D_{2n}(\rho)P_n(\rho)G_n'(\rho) \\
&\quad + G_n'(\rho)M_n(\rho)D_{2n}(\rho) + D_{2n}(\rho)M_n(\rho)G_n(\rho), \\
\dot{D}_{2n}(\rho) &= G_n'^2(\rho)M_n(\rho)G_n(\rho) + G_n'(\rho)\dot{M}_n(\rho)G_n(\rho) + G_n'(\rho)M_n(\rho)G_n^2(\rho) \\
&\quad + 2G_n^2(\rho)M_n(\rho)G_n(\rho) + 2G_n(\rho)\dot{M}_n(\rho)G_n(\rho) + 2G_n(\rho)M_n(\rho)G_n^2(\rho) \\
&\quad - G_n^2(\rho)P_n(\rho)G_n'(\rho) + G_n(\rho)\dot{M}_n(\rho)G_n'(\rho) - G_n(\rho)P_n(\rho)G_n'^2(\rho), \\
\dot{M}_n(\rho) &= -P_n(\rho)G_n'(\rho)M_n(\rho) - M_n(\rho)G_n(\rho)P_n(\rho), \text{ and } P_n = I_n - M_n.
\end{aligned}
$$

For the SARAR model, the full expressions for $D_{jn}(\rho), j = 2, 3, 4$ and $F_n^{(r)}, r = 1, 2$ follow a similar pattern as in the quantities for the SED model with the exception that $G_n(\rho)$ in the SED must now be replaced with $G_{2n}(\rho)$.

APPENDIX D

Proofs of Results in Chapter 2

Proofs of Asymptotic Results

Proof of Theorem 2.1: Following Theorem 3.4 of White (1994), it is sufficient to show, (i) the identification uniqueness condition: $\limsup_{n\to\infty} \max_{\rho\in\mathcal{N}_\epsilon^c(\rho_0)} \frac{h_n}{n}[\bar{\ell}_n^c(\rho) - \bar{\ell}_n^c(\rho_0)] < 0$ for any $\epsilon > 0$, where $\mathcal{N}_\epsilon^c(\rho_0)$ is the compliment of an open neighbourhood of ρ_0 on $\mathcal{P}$ of radius ϵ, and (ii) the uniform convergence in probability: $\frac{h_n}{n}[\ell_n^c(\rho) - \bar{\ell}_n^c(\rho)] \xrightarrow{p} 0$ uniformly in $\rho \in \mathcal{P}$.

To show (i), first observing from (2.9) that $\sigma_n^2(\rho_0) = \sigma_0^2$, we have,

$$
\begin{aligned}
& \lim_{n\to\infty} \tfrac{h_n}{n}\left[\bar{\ell}_n^c(\rho) - \bar{\ell}_n^c(\rho_0)\right] \\
= \;& \lim_{n\to\infty} \left[\tfrac{h_n}{n}(\log|A_n(\rho)| - \log|A_n|) - \tfrac{h_n}{2}(\log\sigma_n^2(\rho) - \log\sigma_0^2)\right] \\
= \;& \lim_{n\to\infty} \left[\tfrac{h_n}{2n}(\log|A_n'(\rho)A_n(\rho)| - \log|A_n'A_n|) + \tfrac{h_n}{2n}(\log|\sigma_n^{-2}(\rho)I_n| - \log|\sigma_0^{-2}I_n|)\right] \\
\neq \;& 0 \text{ for } \rho \neq \rho_0, \text{ by Assumption 2.6.}
\end{aligned}
$$

Next, let $p_n(\theta) = \exp[\ell_n(\theta)]$ be the *quasi* joint pdf of $u_n(= Y_n - X_n\beta_0)$, and $p_n^0(\theta)$ the *true* joint pdf of u_n. Let E^q denote the expectation with respect to p_n, to differentiate from the usual notation E that corresponds to p_n^0. By Jensen's inequality (see Rao, 1973, p. 58), we have, $0 = \log \mathrm{E}^q\Big(\frac{p_n(\theta)}{p_n(\theta_0)}\Big) \geq \mathrm{E}^q\Big[\log\Big(\frac{p_n(\theta)}{p_n(\theta_0)}\Big)\Big] = \mathrm{E}\Big[\log\Big(\frac{p_n(\theta)}{p_n(\theta_0)}\Big)\Big]$, where, the last equation follows from the fact that $\log p_n(\theta_0)$ and $\log p_n(\theta)$ are either a quadratic form or a linear-quadratic form of u_n, and hence their expectations w.r.t $p_n(\theta_0)$ are the same as those w.r.t. $p_n^0(\theta_0)$. It follows that $\mathrm{E}[\log p_n(\theta)] \leq \mathrm{E}[\log p_n(\theta_0)]$, and that, $\bar{\ell}_n(\rho) = \max_{\beta,\sigma^2} \mathrm{E}[\log p_n(\theta)] \leq \mathrm{E}[\log p_n(\theta_0)] = \bar{\ell}_n(\rho_0)$ for $\rho \neq \rho_0$.

To show (ii), note that $\frac{h_n}{n}[\ell_n^c(\rho) - \bar{\ell}_n^c(\rho)] = -\frac{h_n}{2}[\log(\hat{\sigma}_n^2(\rho)) - \log(\sigma_n^2(\rho))]$. By the mean value theorem, $h_n[\log(\hat{\sigma}_n^2(\rho)) - \log(\sigma_n^2(\rho))] = \frac{h_n}{\tilde{\sigma}_n^2(\rho)}[\hat{\sigma}_n^2(\rho) - \sigma_n^2(\rho)]$ where $\tilde{\sigma}_n^2(\rho)$ lies between $\hat{\sigma}_n^2(\rho)$ and $\sigma_n^2(\rho)$. Note that, $\hat{\sigma}_n^2(\rho) = \frac{1}{n}Y_n' A_n'(\rho) M_n(\rho) A_n(\rho) Y_n = \frac{1}{n}\epsilon_n' A_n'^{-1} A_n'(\rho) M_n(\rho) A_n(\rho) A_n^{-1}\epsilon_n = \frac{1}{n}\epsilon_n' A_n'^{-1} A_n'(\rho) A_n(\rho) A_n^{-1}\epsilon_n - \mathit{\Delta}_n(\rho)$ where, $\mathit{\Delta}_n(\rho) \equiv \frac{1}{n}\epsilon_n' A_n'^{-1} A_n'(\rho) P_n(\rho) A_n(\rho) A_n^{-1}\epsilon_n$. By Assumption 2.3, $V_{1n}(\rho) \equiv \frac{1}{n}X_n' A_n'(\rho) A_n(\rho) X_n = O(1)$. By Lemma A.2, $\mathrm{tr}(G_n) = O(\frac{n}{h_n})$ and using $A_n(\rho) = A_n + (\rho_0 - \rho)W_n$, we have, $\mathit{\Delta}_n^*(\rho) = \frac{1}{\sqrt{n}}X_n' A_n'(\rho) A_n(\rho) A_n^{-1}\epsilon_n = \frac{1}{\sqrt{n}}\big[X_n' A_n'\epsilon_n + (\rho_0 - \rho)X_n'(W_n' + A_n' G_n)\epsilon_n + (\rho_0 - \rho)^2 X_n' W_n' G_n \epsilon_n\big] = O_p(\frac{1}{h_n})$. Hence, $\mathit{\Delta}_n(\rho) = \frac{1}{n}\mathit{\Delta}_n^{*'}(\rho) V_{1n}^{-1}(\rho) \mathit{\Delta}_n^*(\rho) = o_p(1)$, uniformly in $\rho \in \mathcal{P}$. Then by Lemma A.3(vi), $h_n[\hat{\sigma}_n^2(\rho) - \sigma_n^2(\rho)] = \frac{h_n}{n}[\epsilon_n' A_n'^{-1} A_n'(\rho) A_n(\rho) A_n^{-1}\epsilon_n - \sigma_0^2 \mathrm{tr}[A_n'^{-1} A_n'(\rho) A_n(\rho) A_n^{-1}] + o_p(1) = o_p(1)$, uniformly in $\rho \in \mathcal{P}$.

To show $\sigma_n^2(\rho)$ is uniformly bounded away from zero, we usea counter argument. Suppose $\sigma_n^2(\rho)$ is not uniformly bounded away from zero in $\mathcal{P}$. Then there exists a sequence $\rho_n \in \mathcal{P}$ s.t. $\sigma_n^2(\rho_n) \to 0$ as $n \to \infty$. Consider the truncated model by setting $\beta = 0$. The Gaussian log-likelihood is $\ell_{t,n}(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\rho)| - \frac{1}{2\sigma^2}Y_n' A_n'(\rho) A_n(\rho) Y_n$. Then $\bar{\ell}_{t,n}(\rho) = \max_{\sigma^2} \mathrm{E}[\ell_{t,n}(\theta)] = -\frac{n}{2}[\log(2\pi)+1] - \frac{n}{2}\log(\sigma_n^2(\rho)) + \log|A_n(\rho)|$. By Jensen's inequality, $\bar{\ell}_{t,n}(\theta) \leq \mathrm{E}[\ell_{t,n}(\theta_0)] = \bar{\ell}_{t,n}(\rho_0), \forall\rho$. This implies $\frac{1}{n}[\bar{\ell}_{t,n}(\theta) - \bar{\ell}_{t,n}(\theta_0)] \leq 0$ and $-\frac{1}{2}\log(\sigma_n^2(\rho)) \leq -\frac{1}{2}\log(\sigma_0^2) + \frac{1}{n}(\log|A_n(\rho_0)| - \log|A_n(\rho)|) = O(1)$ by Lemma A.2, i.e., $-\log(\sigma_n^2(\rho))$ is bounded from above which is a contradiction. Hence, $\log(\sigma_n^2(\rho))$ is well defined $\forall\rho \in \mathcal{P}$. Since $\sigma_n^2(\rho)$ is bounded away from zero and $h_n[\hat{\sigma}_n^2(\rho) - \sigma_n^2(\rho)] = o_p(1)$, $\hat{\sigma}_n^2(\rho)$ is bounded away from zero uniformly in $\mathcal{P}$. Collecting all these results together along with the mean value theorem, we have $h_n|\log(\hat{\sigma}_n^2(\rho)) - \log(\sigma_n^2(\rho))| = o_p(1)$ uniformly in $\rho \in \mathcal{P}$. Hence $\sup_{\rho\in\mathcal{P}} \frac{h_n}{n}|[\ell_n^c(\rho) - \bar{\ell}_n^c(\rho)]| = o_p(1)$.

Proof of Theorem 2.2: By applying the mean value theorem on the adjusted first order condition, we have,

$$
\begin{aligned}
0 = \tfrac{1}{\sqrt{n}} S_n^*(\hat{\theta}_n) &= \tfrac{1}{\sqrt{n}} S_n^*(\theta_0) + \tfrac{1}{\sqrt{n}} \tfrac{\partial}{\partial \theta'} S_n^*(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0) \\
&= \tfrac{1}{\sqrt{n}} S_n^*(\theta_0) - \tfrac{1}{n} K_n H_n(\tilde{\theta}_n) K_n . \sqrt{n} K_n^{-1}(\hat{\theta}_n - \theta_0) \qquad \text{(D-1)}
\end{aligned}
$$

where $\tilde{\theta}_n$ lies between the line segment joining θ_0 and $\hat{\theta}_n$, thus $\tilde{\theta} \xrightarrow{p} \theta_0$. Here $H_n(\theta)$ is the negative Hessian matrix and K_n is as defined in section 2.2.

Under Assumptions 2.1-2.5, the central limit theorem for linear-quadratic forms of Kelejian and Prucha (2001) is applicable, which gives $\frac{1}{\sqrt{n}} S_n^*(\theta_0) = \frac{K_n}{\sqrt{n}} \frac{\partial}{\partial \theta} \ell(\theta_0) \xrightarrow{D} N(0, \Gamma^*)$, where, $\Gamma^* = \lim_{n\to\infty} \frac{1}{n}\Gamma_n^*$ and $\Gamma_n^* = \mathrm{Var}[S_n^*(\theta_0)]$. The asymptotic normality of $\hat{\theta}_n$ thus follows from: (i) $\frac{1}{n} K_n H_n(\tilde{\theta}_n) K_n - \frac{1}{n} K_n H_n(\theta_0) K_n = o_p(1)$ and (ii) $\frac{1}{n} K_n H_n(\theta_0) K_n - \frac{1}{n} K_n \Sigma_n K_n = o_p(1)$, where, $\Sigma_n = \mathrm{E}[H_n(\theta_0)]$ is the information matrix given in section 2.2. To show (i), note that $H_n(\theta) =$

$$
\begin{pmatrix}
\frac{1}{\sigma^2} X_n' A_n'(\rho) A_n(\rho) X_n & \frac{1}{\sigma^4} X_n' A_n'(\rho) \epsilon_n(\delta) & \frac{2}{\sigma^2} X_n' A_n'(\rho) G_n'(\rho) \epsilon_n(\delta) \\
\frac{1}{\sigma^4} \epsilon_n'(\delta) A_n(\rho) X_n & \frac{1}{2\sigma^6}(2\epsilon_n'(\delta)\epsilon_n(\delta) - n\sigma^2) & \frac{1}{\sigma^4} \epsilon_n'(\delta) G_n'(\rho) \epsilon_n(\delta) \\
\frac{2}{\sigma^2} \epsilon_n'(\delta) G_n(\rho) A_n(\rho) X_n & \frac{1}{\sigma^4} \epsilon_n'(\delta) G_n(\rho) \epsilon_n(\delta) & \frac{1}{\sigma^2}[\epsilon_n'(\delta) G_n'(\rho) G_n(\rho) \epsilon_n(\delta) + \sigma^2 \mathrm{tr}(G_n^2(\rho))]
\end{pmatrix}
$$

where $\delta = (\beta', \rho)'$. Let $\tilde{A}_n = A_n(\tilde{\rho}_n)$. Under Assumption 2.3 and using $\tilde{\theta}_n \xrightarrow{p} \theta_0$,

$$
\begin{aligned}
&\tfrac{1}{n}\big(\tfrac{\partial^2}{\partial\beta\partial\beta'} \ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\beta\partial\beta'} \ell_n(\theta_0)\big) = \tfrac{1}{n}\big(\tfrac{1}{\sigma_0^2} X_n' A_n' A_n X_n - \tfrac{1}{\tilde{\sigma}_n^2} X_n' \tilde{A}_n' \tilde{A}_n X_n\big) \\
&\quad = \big(\tfrac{1}{\sigma_0^2} - \tfrac{1}{\tilde{\sigma}_n^2}\big) \tfrac{1}{n} X_n' A_n' A_n X_n + o_p(1) = o_p(1),
\end{aligned}
$$

noticing that $A_n' A_n - \tilde{A}_n' \tilde{A}_n = (\tilde{\rho}_n - \rho_0)(W_n + W_n') - (\tilde{\rho}_n^2 - \rho_0^2) W_n' W_n$. Similarly, it can be shown that, letting $\tilde{\epsilon}_n = \epsilon_n(\tilde{\rho}_n)$,

$$
\begin{aligned}
&\tfrac{1}{n}\big(\tfrac{\partial^2}{\partial(\sigma^2)^2} \ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial(\sigma^2)^2} \ell_n(\theta_0)\big) = \tfrac{1}{n\sigma_0^6} \epsilon_n' \epsilon_n - \tfrac{1}{n\tilde{\sigma}_n^6} \tilde{\epsilon}_n' \tilde{\epsilon}_n - \tfrac{1}{2}\big(\tfrac{1}{\sigma_0^4} - \tfrac{1}{\tilde{\sigma}_n^4}\big) \\
&\quad = \tfrac{1}{n\sigma_0^6}(\epsilon_n'\epsilon_n - \tilde{\epsilon}_n'\tilde{\epsilon}_n) + o_p(1) = o_p(1),
\end{aligned}
$$

since $\tilde{\epsilon}_n'\tilde{\epsilon}_n - \epsilon_n'\epsilon_n = 2(\rho_0 - \tilde{\rho}_n)\epsilon_n' G_n \epsilon_n + 2\epsilon_n' A_n X_n(\beta_0 - \tilde{\beta}_0) + (\rho_0 - \tilde{\rho}_0)^2 \epsilon_n' G_n' G_n \epsilon_n + 2(\rho_0 - \tilde{\rho}_n)\epsilon_n' W_n X_n(\beta_0 - \tilde{\beta}_n) + 2(\rho_0 - \tilde{\rho}_n)\epsilon_n' G_n' A_n X_n(\beta_0 - \tilde{\beta}_n) + (\beta_0 - \tilde{\beta}_n)' X_n' A_n' A_n X_n(\beta_0 - \tilde{\beta}_n) + 2(\rho_0 - \tilde{\rho}_n)^2 \epsilon_n' G_n' W_n X_n(\beta_0 - \tilde{\beta}_n) + 2(\rho_0 - \tilde{\rho}_n)(\beta - \tilde{\beta}_n)' X_n' A_n' W_n X_n(\beta_0 - \tilde{\beta}_n) + (\rho_0 - \tilde{\rho}_n)^2(\beta_0 - \tilde{\beta}_n)' X_n' W_n' W_n X_n(\beta_0 - \tilde{\beta}_n) = o_p(1)$.

By the mean value theorem, $\mathrm{tr}(G_n^2(\tilde{\rho}_n)) = \mathrm{tr}(G_n^2) + 2\mathrm{tr}[G_n^3(\bar{\rho}_n)](\tilde{\rho}_n - \rho_0)$, where $\bar{\rho}_n$ lies between ρ_0 and $\tilde{\rho}_n$. By Lemma A.2, Assumptions 2.4 and 2.5, $\mathrm{tr}[G_n^3(\bar{\rho}_n)] = O\big(\frac{n}{h_n}\big)$. Hence, $\frac{h_n}{n}[\mathrm{tr}(G_n^2(\tilde{\rho}_n)) - \mathrm{tr}(G_n^2)] = o_p(1)$ since $\tilde{\rho}_n \xrightarrow{p} \rho_0$. Further, $\epsilon_n' G_n' G_n \epsilon_n = Y_n' W_n' W_n Y_n - 2Y_n' W_n' W_n X_n \beta_0 + \beta_0' X_n' W_n' W_n X_n \beta_0 = O_p\big(\frac{n}{h_n}\big)$ by Lemmas A.2(i) and A.3(v). Hence, $\frac{h_n}{n}[\tilde{\epsilon}_n' \tilde{G}_n' \tilde{G}_n \tilde{\epsilon}_n - \epsilon_n' G_n' G_n \epsilon_n] = \frac{h_n}{n}[(\beta_0 - \tilde{\beta}_n)' X_n' W_n' W_n X_n (\beta_0 - \tilde{\beta}_n) - 2\epsilon_n' G_n' W_n X_n (\beta_0 - \tilde{\beta}_n)] = o_p(1)$, hence,

$$\begin{aligned}\tfrac{h_n}{n}\big(\tfrac{\partial^2}{\partial\rho^2}\ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\rho^2}\ell_n(\theta_0)\big) &= \tfrac{h_n}{n}\big(\tfrac{1}{\sigma_0^2}\epsilon_n' G_n' G_n \epsilon_n - \tfrac{1}{\tilde{\sigma}_n^2}\tilde{\epsilon}_n' \tilde{G}_n' \tilde{G}_n \tilde{\epsilon}_n + \mathrm{tr}(G_n^2) - \mathrm{tr}(\tilde{G}_n^2)\big)\\ &= \tfrac{h_n}{n}\big(\tfrac{1}{\sigma_0^2} - \tfrac{1}{\tilde{\sigma}_n^2}\big)\epsilon_n' G_n' G_n \epsilon_n + o_p(1) = o_p(1).\end{aligned}$$

Using similar arguments on the rest of the quantities, we have,

$$\begin{aligned}
\tfrac{\sqrt{h_n}}{n}\big(\tfrac{\partial^2}{\partial\beta\partial\rho}\ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\beta\partial\rho}\ell_n(\theta_0)\big) &= \tfrac{2\sqrt{h_n}}{n\sigma_0^2}(X_n' W_n' \epsilon_n - X_n' W_n' \tilde{\epsilon}_n) + o_p(1) = o_p(1),\\
\tfrac{1}{n}\big(\tfrac{\partial^2}{\partial\beta\partial\sigma^2}\ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\beta\partial\sigma^2}\ell_n(\theta_0)\big) &= \tfrac{1}{n\sigma_0^4}[(X_n' A_n' \epsilon_n) - (X_n' \tilde{A}_n' \tilde{\epsilon}_n)] + o_p(1) = o_p(1),\\
\tfrac{\sqrt{h_n}}{n}\big(\tfrac{\partial^2}{\partial\sigma^2\partial\rho}\ell_n(\tilde{\theta}_n) - \tfrac{\partial^2}{\partial\sigma^2\partial\rho}\ell_n(\theta_0)\big) &= \tfrac{\sqrt{h_n}}{n\sigma^4}(\epsilon_n' G_n' \epsilon_n - \tilde{\epsilon}_n' \tilde{G}_n' \tilde{\epsilon}_n) + o_p(1) = o_p(1).
\end{aligned}$$

Proof of (ii) is more straightforward, as the differences of the corresponding elements of $\frac{1}{n}K_n H_n(\theta_0) K_n$ and $\frac{1}{n}K_n \Sigma_n K_n$ are, 0, $\frac{1}{n\sigma^4}(X_n' A_n' \epsilon_n) = o_p(1)$, $\frac{1}{2n\sigma^6}(2\epsilon_n' \epsilon_n - n\sigma^2) - \frac{1}{2\sigma_0^4} = \frac{1}{n\sigma^6}\epsilon_n' \epsilon_n = o_p(1)$, $\frac{2\sqrt{h_n}}{n\sigma_0^2} X_n' A_n' G_n' = o_p(1)$, $\frac{\sqrt{h_n}}{n\sigma^4}\epsilon_n' G_n \epsilon_n - \frac{\sqrt{h_n}}{n\sigma_0^2}\mathrm{tr}(G_n) = o_p(1)$, and $\frac{h_n}{n\sigma_0^2}(\epsilon_n' G_n' G_n \epsilon_n + \sigma^2 \mathrm{tr}(G_n^2)) - \frac{h_n}{n}\mathrm{tr}(G_n^s G_n) = \frac{h_n}{n\sigma_0^2}\epsilon_n' G_n' G_n \epsilon_n = o_p(1)$.

The results (i) and (ii) give $0 = \frac{1}{\sqrt{n}} S_n^* - \frac{1}{n}\Sigma_n^* . \sqrt{n} K_n^{-1}(\hat{\theta}_n - \theta_0) + o_p(1)$.

Proof of Corollary 2.1: By using the block diagonal nature of Σ_n,

$$\Sigma_n^{-1} = \begin{pmatrix} \sigma_0^2 (X_n' A_n' A_n X_n)^{-1} & 0 & 0 \\ 0 & \frac{2\sigma_0^4}{n} T_{1n} & -\frac{2\sigma_0^2}{n} T_{2n} \\ 0 & -\frac{2\sigma_0^2}{n} T_{2n} & \frac{h_n}{n} T_{4n} \end{pmatrix}$$

where, $T_{1n} = \mathrm{tr}(G_n^s G_n)/\mathrm{tr}(C_n^s C_n)$, $T_{2n} = \mathrm{tr}(G_n)/\mathrm{tr}(C_n^s C_n)$, $T_{4n} = \frac{n}{h_n}\mathrm{tr}^{-1}(C_n^s C_n)$. Then deriving $\Sigma_n^{*-1} \Gamma_n^* \Sigma_n^{*-1} = K_n^{-1} \Sigma_n^{-1} \Gamma_n \Sigma_n^{-1} K_n^{-1}$ is just a matter of matrix multiplication.

Proofs of Higher Order Results

Proof of Lemma 2.1: Note, $\hat{\sigma}_n^2(\rho_0) \equiv \hat{\sigma}_{n0}^2 = \frac{1}{n}Y_n'A_n'M_nA_nY_n = \frac{1}{n}\epsilon_n'M_n\epsilon_n$. By the moments for quadratic forms, $\text{Var}(\hat{\sigma}_{n0}^2) = \frac{1}{n^2}O(n) = O(\frac{1}{n})$. Now by the generalised Chebyshev's inequality, $\text{P}(\sqrt{n}|\hat{\sigma}_{n0}^2 - \sigma_0^2| \geq \delta) \leq \frac{1}{\delta^2}n\text{Var}(\hat{\sigma}_{n0}^2) = O(1)$. Hence, by the definition of order of magnitudes[1] for stochastic components, $\hat{\sigma}_{n0}^2 = \sigma_0^2 + O_p(\frac{1}{\sqrt{n}})$.

In order to prove that $\hat{\sigma}_{n0}^{-2}$ is $\sqrt{n}$-consistent, by the Mean Value Theorem, we have, $\frac{1}{\hat{\sigma}_{n0}^2} - \frac{1}{\sigma_0^2} = -\frac{1}{\bar{\sigma}_{n0}^4}(\hat{\sigma}_{n0}^2 - \sigma_0^2)$, which can be written as, $\frac{1}{\hat{\sigma}_{n0}^2} = \frac{1}{\sigma_0^2} - \frac{1}{\sigma_0^4}(\hat{\sigma}_{n0}^2 - \sigma_0^2) - \big(\frac{1}{\bar{\sigma}_{n0}^4} - \frac{1}{\sigma_0^4}\big)(\hat{\sigma}_{n0}^2 - \sigma_0^2)$, where $\bar{\sigma}^2$ lies between $\hat{\sigma}_{n0}^2$ and σ_0^2. Hence, $\bar{\sigma}_{n0}^2 = \sigma_0^2 + O_p(\frac{1}{\sqrt{n}})$, $\bar{\sigma}_{n0}^4 = \big(\sigma_0^2 + O_p(\frac{1}{\sqrt{n}})\big)^2 = \sigma_0^4 + O_p(\frac{1}{\sqrt{n}})$, and $\bar{\sigma}_{n0}^{-4} = \big(\sigma_0^4 + O_p(\frac{1}{\sqrt{n}})\big)^{-1} = \sigma_0^{-4} + O_p(\frac{1}{\sqrt{n}})$. Therefore, we conclude that $\hat{\sigma}_{n0}^{-2} = \sigma_0^{-2} + O_p(\frac{1}{\sqrt{n}})$.

Consider, $h_nR_{1n} = \frac{h_n}{n\hat{\sigma}_{n0}^2}\epsilon_n'M_nG_nM_n\epsilon_n$. By Lemma A.3(v), $\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n = O_p(1)$. Hence,

$$h_nR_{1n} = \frac{1}{\sigma_0^2}\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n + O_p(\frac{1}{\sqrt{n}}) = O_p(1). \tag{D-2}$$

Using the expression for $\hat{\sigma}_{n0}^{-2}$,

$$\text{E}(h_nR_{1n}) = frac1\sigma_0^2\text{E}\big(\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n\big) - \frac{1}{\sigma_0^4}\text{E}\big(\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n(\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)$$
$$-\text{E}\big(\frac{h_n}{n}\epsilon_n'M_nG_nM_n\epsilon_n\big(\frac{1}{\bar{\sigma}_{n0}^4} - \frac{1}{\sigma_0^4}\big)(\hat{\sigma}_{n0}^2 - \sigma_0^2)\big).$$

The 1st term is, $\frac{h_n}{\sigma_0^2 n}\text{E}(\epsilon_n'\epsilon_n)\text{tr}(M_nG_nM_n) = O(1)$. The 3rd term is, $O((\frac{h_n}{n})^{\frac{1}{2}})$ by Assumption 2.7. For the 2nd term note that, $\text{E}(\hat{\sigma}_{n0}^2) = \sigma_0^2 + O(\frac{1}{n})$ and $\text{E}(\epsilon_n'M_nG_nM_n\epsilon_n) = \sigma_0^2\text{tr}(M_nG_nM_n) = O(\frac{n}{h_n})$. Then by Cauchy-Schwartz inequality,

$$\begin{aligned}
&|\text{E}\big(\epsilon_n'M_nG_nM_n\epsilon_n(\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)| \\
= \ &|\text{E}\big([\epsilon_n'M_nG_nM_n\epsilon_n - \text{E}(\epsilon_n'M_nG_nM_n\epsilon_n) + \text{E}(\epsilon_n'M_nG_nM_n\epsilon_n)](\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)| \\
\leq \ &|\text{E}\big([\epsilon_n'M_nG_nM_n\epsilon_n - \sigma_0^2\text{tr}(M_nG_nM_n)](\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)| + \sigma_0^2|\text{tr}(M_nG_nM_n)\text{E}\big(\hat{\sigma}_{n0}^2 - \sigma_0^2\big)| \\
= \ &|\text{Cov}\big([\epsilon_n'M_nG_nM_n\epsilon_n - \sigma_0^2\text{tr}(M_nG_nM_n)], (\hat{\sigma}_{n0}^2 - \sigma_0^2)\big)| + O(\tfrac{1}{h_n}) \\
\leq \ &\tfrac{1}{n}\big(\text{Var}(\epsilon_n'M_nG_nM_n\epsilon_n)\text{Var}(\epsilon_n'M_n\epsilon)\big)^{\frac{1}{2}} + O(\tfrac{1}{h_n}) \\
&= \tfrac{1}{n}\big(O(\tfrac{n}{h_n})O(n)\big)^{\frac{1}{2}} + O(\tfrac{1}{h_n}) = O(\tfrac{1}{\sqrt{h_n}}),
\end{aligned}$$

[1] If $\forall\epsilon > 0, \exists c \geq 0, n_0 > 0$ s.t. $\text{P}(|x_n| > cf_n) < \epsilon, \forall n \geq n_0$ then $x_n = O_p(f_n)$

where we have used the results for quadratic forms. Then, $\frac{1}{\sigma_0^4}\mathrm{E}\big[\frac{h_n}{n}\epsilon_n' M_n G_n M_n \epsilon_n(\hat{\sigma}_{n0}^2 - \sigma_0^2)\big] = O(\frac{\sqrt{h_n}}{n})$, which implies,

$$\mathrm{E}(h_n R_{1n}) = \mathrm{Max}\big\{O(1), O\big(\frac{\sqrt{h_n}}{n}\big), O\big((\frac{h_n}{n})^{\frac{1}{2}}\big)\big\} = O(1). \tag{D-3}$$

By (D-2) and (D-3), $h_n R_{1n} - \mathrm{E}(h_n R_{1n}) = \frac{h_n}{\sigma_0^2 n}\epsilon_n' M_n G_n M_n \epsilon_n - \frac{h_n}{\sigma_0^2 n}\mathrm{E}(\epsilon_n'\epsilon_n)\mathrm{tr}(M_n G_n M_n) + O_p(\frac{1}{\sqrt{n}}) - O(\frac{\sqrt{h_n}}{n}) - O((\frac{h_n}{n})^{\frac{1}{2}}) = O((\frac{h_n}{n})^{\frac{1}{2}})$.

By Lemma A.2 the remaining parts can be proved in a similar fashion noting that, $D_{jn} = O(\frac{n}{h_n})$, of the sandwich forms of R_{jn} for $j = 2, 3, 4$..

Proof of Proposition 2.1: We go on to prove the proposition using Lemma 2.1. To that effect consider the Taylor series expansion of $\tilde{\psi}_n(\rho)$ around ρ_0,

$$\begin{aligned} 0 = \tilde{\psi}_n(\hat{\rho}_n) \;&=\; \tilde{\psi}_n + H_{1n}(\hat{\rho}_n - \rho_0) + \tfrac{1}{2}H_{2n}(\hat{\rho}_n - \rho_0)^2 + \tfrac{1}{6}H_{3n}(\hat{\rho}_n - \rho_0)^3 \\ &\quad + \tfrac{1}{6}[H_{3n}(\bar{\rho}) - H_{3n}](\hat{\rho}_n - \rho_0)^3, \end{aligned}$$

where the last two terms sums up the mean value form of the remainder term with $\bar{\rho}$ lying between ρ_0 and $\hat{\rho}_n$. We have shown that $\hat{\rho}_n - \rho_0 \to_p \big(\frac{h_n}{n}\big)^{\frac{1}{2}}$. Next, note that $h_n T_{rn} = O(1)$ for $r = 0, 1, 2, 3$ by Assumptions 2.4 and 2.5. Now, in order to prove the result of the proposition, we need to establish the following conditions: (i) $\tilde{\psi}_n = O_p\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$ and $\mathrm{E}(\tilde{\psi}_n) = O(\frac{h_n}{n})$, (ii) $\mathrm{E}(H_{rn}) = O(1)$ and $H_{rn} - \mathrm{E}(H_{rn}) = O_p\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$ for $r = 1, 2, 3$, (iii) $H_{1n}^{-1} = O_{pu}(1)$ and $\mathrm{E}(H_{1n})^{-1} = O(1)$ and (iv) $H_{3n}(\bar{\rho}) - H_{3n} = O_p\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$.

For (i), note $\epsilon_n' M_n G_n M_n \epsilon_n - \sigma_0^2\mathrm{tr}(M_n G_n M_n) = O_p\big((\frac{n}{h_n})^{\frac{1}{2}}\big)$ and $\mathrm{tr}(M_n G_n M_n) = \mathrm{tr}(G_n) + O(1) = nT_{0n} + O(1)$. by Lemma A.2 Therefore, $\tilde{\psi}_n = -h_n T_{0n} + h_n R_{1n} = -h_n T_{0n} + \frac{h_n}{\sigma_0^2 n}\epsilon_n' M_n G_n M_n \epsilon_n + O_p(\frac{1}{\sqrt{n}}) = -h_n T_{0n} + \frac{h_n}{\sigma_0^2 n}\big[\sigma_0^2\mathrm{tr}(G_n) + O_p((\frac{n}{h_n})^{\frac{1}{2}})\big] + O_p(\frac{1}{\sqrt{n}})$ and $\mathrm{E}(\tilde{\psi}_n) = -h_n T_{0n} + \frac{h_n}{n}(\mathrm{tr}(G_n) + O(1)) + O\big((\frac{h_n}{n})^{\frac{1}{2}}\big) = O(\frac{h_n}{n})$.

For (ii), Lemma 2.1 implies, $(h_n R_{1n})^s = \mathrm{E}(h_n R_{1n})^s + O_p((\frac{h_n}{n})^{\frac{1}{2}})$ for $s = 2, 3, 4$, $(h_n R_{2n})^2 = \mathrm{E}(h_n R_{2n})^2 + O_p((\frac{h_n}{n})^{\frac{1}{2}})$, $(h_n R_{1n})^s h_n R_{2n} = \mathrm{E}(h_n R_{1n})^s\mathrm{E}(h_n R_{2n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$ for $s = 1, 2$, and $h_n R_{1n} h_n R_{3n} = \mathrm{E}(h_n R_{1n})\mathrm{E}(h_n R_{3n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$. Hence, Assumption 2.8 implies, $\mathrm{E}[(h_n R_{1n})^s] = \mathrm{E}(h_n R_{1n})^s + O((\frac{h_n}{n})^{\frac{1}{2}})$ for $s = 2, 3, 4$, $\mathrm{E}[(h_n R_{2n})^2] =$

$\mathrm{E}(h_nR_{2n})^2 + O((\frac{h_n}{n})^{\frac{1}{2}})$, $\mathrm{E}[(h_nR_{1n})^s h_nR_{2n}] = \mathrm{E}(h_nR_{1n})^s\mathrm{E}(h_nR_{2n}) + O((\frac{h_n}{n})^{\frac{1}{2}})$ for $s = 1, 2$, and $\mathrm{E}[h_nR_{1n}h_nR_{3n}] = \mathrm{E}(h_nR_{1n})\mathrm{E}(h_nR_{3n}) + O((\frac{h_n}{n})^{\frac{1}{2}})$. Combining these with (D) and Lemma 2.1, we have, $H_{rn} - \mathrm{E}(H_{rn}) = O_p((\frac{h_n}{n})^{\frac{1}{2}})$ and $\mathrm{E}(H_{rn}) = O(1)$ for $r = 1, 2, 3$.

For (iii), by Lemma 2.1 and $\mathrm{E}[(h_nR_{1n})^2] = \mathrm{E}(h_nR_{1n})^2 + O\big((\frac{h_n}{n})^{\frac{1}{2}}\big)$,

$$
\begin{aligned}
\mathrm{E}(H_{1n}) &= \tfrac{2}{h_n}\mathrm{E}[(h_nR_{1n})^2] - h_nT_{1n} - \mathrm{E}(h_nR_{2n}) \\
&= \tfrac{2}{h_n}\big(\tfrac{h_n}{n}\mathrm{tr}(M_nG_nM_n) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big)\big)^2 - h_nT_{1n} \\
&\quad -\big(\tfrac{h_n}{n}\mathrm{tr}(M_nD_{2n}M_n) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big)\big) \\
&= \tfrac{2}{h_n}\big(\big(\tfrac{h_n}{n}\mathrm{tr}(M_nG_nM_n)\big)^2\big) - h_nT_{1n} - \tfrac{h_n}{n}\mathrm{tr}(M_nD_{2n}M_n) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big) \\
&= \tfrac{2}{h_n}\big(\tfrac{h_n}{n}\mathrm{tr}(G_n)\big)^2 - \tfrac{h_n}{n}\mathrm{tr}(G_n^2) - \tfrac{h_n}{n}\mathrm{tr}(G_n'G_n) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big) \\
&= -\tfrac{h_n}{n}\big(\mathrm{tr}(G_n^2) + \mathrm{tr}(G_n'G_n) - 2T_{0n}^2\mathrm{tr}(I_n)\big) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big) \\
&= -\tfrac{h_n}{n}\big(\mathrm{tr}(G_n - T_{0n}I_n)^2 + \mathrm{tr}(G_n - T_{0n}I_n)'(G_n - T_{0n}I_n)\big) + O\big((\tfrac{h_n}{n})^{\frac{1}{2}}\big).
\end{aligned}
$$

That is, $\mathrm{E}(H_{1n})$ is negative for sufficiently large n and it is finite. Therefore, $\mathrm{E}(H_{1n})^{-1} = O(1)$. Also by, $H_{1n} = \mathrm{E}(H_{1n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$, we have, $H_{1n}^{-1} = O_p(1)$.

For (iv), consider (2.17) evaluated at $\bar{\rho}_n$. By the mean value theorem, $h_nT_{3n}(\bar{\rho}) = \frac{h_n}{n}\mathrm{tr}(G_n^4(\bar{\rho})) = \frac{h_n}{n}\mathrm{tr}(G_n^4) + 4\frac{h_n}{n}\mathrm{tr}(G_n^5(\tilde{\rho}))(\bar{\rho} - \rho_0)$, where, $\tilde{\rho}$ lies between $\bar{\rho}$ and ρ_0. By repeatedly applying the mean value theorem we can find a $\tilde{\rho}$ which is much closer to the true ρ_0. For such $\tilde{\rho}$, $\frac{h_n}{n}\mathrm{tr}(G_n^5(\tilde{\rho})) = O(1)$ by Assumptions 2.4 and 2.5. Combining with the $(\frac{n}{h_n})^{1/2}$-convergence of $\bar{\rho}$ to the true value we have, $h_nT_{3n}(\bar{\rho}) = O(1)$. Now consider $\hat{\sigma}_n^2(\bar{\rho}) = \frac{1}{n}Y_n'A_n'(\bar{\rho})M_n(\bar{\rho})A_n(\bar{\rho})Y_n$ and $\hat{\sigma}_{n0}^2 = \frac{1}{n}Y_n'A_n'M_nA_nY_n$. Similarly, by the mean value theorem we have, $\hat{\sigma}_n^2(\bar{\rho}) = \hat{\sigma}_{n0}^2 - \frac{2}{n}(\bar{\rho} - \rho_0)Y_n'A_n'(\tilde{\rho})M_n(\tilde{\rho})G_n(\tilde{\rho})M_n(\tilde{\rho})A_n(\tilde{\rho})Y_n = \hat{\sigma}_{n0}^2 - 2(\bar{\rho} - \rho_0)O_p(h_n^{-1}) = \hat{\sigma}_{n0}^2 + O_p((nh_n)^{-1/2})$. By continuity of $\hat{\sigma}_{n0}^{-2}$, it can be deduced that, $\hat\sigma_n^{-2}(\bar\rho) = \big(\hat\sigma_{n0}^2 + O_p((nh_n)^{-1/2})\big)^{-1} = \hat\sigma_{n0}^{-2} + O_p((nh_n)^{-1/2})$. Now,

$$\begin{aligned} h_n R_{1n}(\bar\rho) &= \hat\sigma_n^{-2}(\bar\rho)\tfrac{h_n}{n} Y_n' A_n'(\bar\rho) M_n(\bar\rho) G_n(\bar\rho) M_n(\bar\rho) A_n(\bar\rho) Y_n \\ &= \hat\sigma_n^{-2}(\bar\rho)\tfrac{h_n}{n}\big[Y_n' A_n' M_n G_n M_n A_n Y_n \\ &\quad -(\bar\rho-\rho_0) Y_n' A_n'(\tilde\rho) M_n(\tilde\rho) D_{2n}(\tilde\rho) M_n(\tilde\rho) A_n(\tilde\rho) Y_n\big] \\ &= \big(h_n R_{1n} + O_p\big((\tfrac{1}{nh_n})^{\frac12}\big)\big) - O_p((\tfrac{h_n}{n})^{\frac12}) = h_n R_{1n} + O_p(\tfrac{h_n}{n})^{\frac12} \quad \text{(D-4)} \end{aligned}$$

Using a similar set of arguments it can be shown that, $h_n R_{kn}(\bar\rho) = h_n R_{kn} + O_p((\frac{h_n}{n})^{\frac12})$ for $k = 2, 3, 4$. Then it follows that, $H_{3n}(\bar\rho) - H_{3n} = O_p((\frac{h_n}{n})^{\frac12})$.

Proof of Proposition 2.2: Arguments are similar to that of Proposition 2.1.

Proof of Proposition 2.3: Note that $b_2(\rho_0, \gamma_0) = O((\frac{n}{h_n})^{-1})$ and that it is differentiable. It follows that $\frac{\partial}{\partial(\rho_0,\gamma_0)} b_2(\rho_0, \gamma_0) = O((\frac{n}{h_n})^{-1})$. As $\hat\rho_n$, the QML estimator of ρ defined at the beginning of Section 5.2, is $\sqrt{n/h_n}$-consistent, it can be shown that $\hat\gamma_n = \gamma(\hat{\mathcal{F}}_n)$ is also $\sqrt{n/h_n}$-consistent. We have, under the additional assumptions in Proposition 2.3, $b_2(\hat\rho_n, \hat\gamma_n) = b_2(\rho_0, \gamma_0) + \frac{\partial}{\partial\rho_0} b_2(\rho_0, \gamma_0)(\hat\rho_n - \rho_0) + \frac{\partial}{\partial\gamma_0} b_2(\rho_0, \gamma_0)(\hat\gamma_n - \gamma_0) + O_p((\frac{n}{h_n})^{-2})$. Thus, $\mathrm{E}[b_2(\hat\rho_n, \hat\gamma_n)] = b_2(\rho_0, \gamma_0) + \frac{\partial}{\partial\rho_0} b_2(\rho_0, \gamma_0)\mathrm{E}(\hat\rho_n - \rho_0) + \frac{\partial}{\partial\gamma_0} b_2(\rho_0, \gamma_0)\mathrm{E}(\hat\gamma_n - \gamma_0) + O((\frac{n}{h_n})^{-2})]$. As $\mathrm{E}(\hat\rho_n - \rho_0) = O(\frac{h_n}{n})$, it can be shown that $\mathrm{E}(\hat\gamma_n - \gamma_0) = O(\frac{h_n}{n})$. These lead to $\mathrm{E}[b_2(\hat\rho_n, \hat\gamma_n)] = b_2(\rho_0, \gamma_0) + O((\frac{n}{h_n})^{-2})$. Similarly, we show that $\mathrm{E}[b_3(\hat\rho_n, \hat\gamma_n)] = b_3(\rho_0, \gamma_0) + o((\frac{n}{h_n})^{-2})$, noting that $b_3(\rho_0, \gamma_0) = O((\frac{n}{h_n})^{-3/2})$. Clearly, our bootstrap estimate has two step approximations, one is that described above, and the other is the bootstrap approximations to the various expectations in (2.23) given $\hat\rho_n$, e.g.,$\hat{\mathrm{E}}(H_{1n}\tilde\psi_n) = \frac{1}{B}\sum_{b=1}^{B} H_{1n}(e_{n,b}^*, \hat\rho_n)\tilde\psi_n(e_{n,b}^*, \hat\rho_n)$. However, these approximations can be made arbitrarily accurate, for a given $\hat\rho_n$ and $\mathcal{F}_n$, by choosing an arbitrarily large B.

APPENDIX E

Proofs of Results in Chapter 4

Some First-Order Results

The following list summarises some frequently used notations in the paper:

- $\delta = (\lambda, \rho)'$, and δ_0 is its true value.
- $J_m = I_m - \frac{1}{m} l_m l_m'$ where l_m is an $m \times 1$ vector of ones. $[F_{m,m-1}, \frac{1}{\sqrt{m}} l_m]$ is the eigenvector matrix of J_m, where $F_{m,m-1}$ corresponds to eigenvalue of ones.
- $W_{hn}^* = F_{n,n-1}' W_{hn} F_{n,n-1},\ h = 1, 2.$
- $A_n(\lambda) = I_n - \lambda W_{1n}$ and $B_n(\rho) = I_n - \rho W_{2n}$.
- $[Z_{n1}^*, \ldots, Z_{n,T-1}^*] = F_{n,n-1}'[Z_{n1}, \ldots, Z_{nT}]F_{T,T-1}$ for any $n \times T$ matrix $[Z_{n1}, \cdots, Z_{nT}]$.
- $\mathbf{Y}_N = (Y_{n1}^{*\prime}, \ldots, Y_{n,T-1}^{*\prime})'$, $\mathbf{X}_N = (X_{n1}^{*\prime}, \ldots, X_{n,T-1}^{*\prime})'$, and $\mathbf{W}_{hN} = I_{T-1} \otimes W_{hn}^*, h = 1, 2.$
- $\mathbf{A}_N(\lambda) = I_N - \lambda \mathbf{W}_{1N}$, and $\mathbf{B}_N(\rho) = I_N - \rho \mathbf{W}_{2N}$.

- $\mathbf{M}_N(\rho) = \mathbf{B}'_N(\rho)\{I_N - \mathbf{X}_N(\rho)[\mathbf{X}'_N(\rho)\mathbf{X}_N(\rho)]^{-1}\mathbf{X}'_N(\rho)\}\mathbf{B}_N(\rho)$.

The following set of regularity conditions from Lee and Yu (2010b) are sufficient for the $\sqrt{N}$-consistency of the QML estimator $\hat{\delta}_{nT}$ defined by maximizing (4.8), and hence the $\sqrt{N}$-consistency of the QML estimators $\hat{\beta}_N$ and $\hat{\sigma}^2_N$ of β and σ^2, which are clearly essential for the development of the higher-order results for the QML estimators.

Assumption E1. *W_{1n} and W_{2n} are row-normalised non-stochastic spatial weights matrices with zero diagonals.*

Assumption E2. *The disturbances $\{v_{it}\}$, $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$, are iid across i and t with zero mean, variance σ_0^2 and $\mathrm{E}|v_{it}|^{4+\eta} < \infty$ for some $\eta > 0$.*

Assumption E3. *$A_n(\lambda)$ and $B_n(\rho)$ are invertible for all $\lambda \in \Lambda$ and $\rho \in \mathbb{P}$, where Λ and $\mathbb{P}$ are compact intervals. Furthermore, λ_0 is in the interior of Λ, and ρ_0 is in the interior of $\mathbb{P}$.*[1]

Assumption E4. *The elements of X_{nt} are non-stochastic, and are bounded uniformly in n and t. Under the setting in Assumption E6, the limit of $\frac{1}{N}\mathbf{X}'_N\mathbf{X}_N$ exists and is non-singular.*

Assumption E5. *W_{1n} and W_{2n} are uniformly bounded in both row and column sums in absolute value (for short, UB).*[2] *Also $A_n^{-1}(\lambda)$ and $B_n^{-1}(\rho)$ are UB, uniformly in $\lambda \in \Lambda$ and $\rho \in \mathbb{P}$.*

Assumption E6. *n is large, where T can be finite or large.*[3]

Assumption E7. *Either*

(a): $\lim_{n\to\infty} \mathcal{H}_N(\rho)$ is *non-singular* $\forall \rho \in \mathbb{P}$ *and* $\lim_{n\to\infty} \mathcal{Q}_{1n}(\rho) \neq 0$ *for* $\rho \neq \rho_0$ *or*

[1] Due to the non-linearity of λ and ρ in the model, compactness of Λ and $\mathbb{P}$ is needed. However, the compactness of the space of β and σ^2 is not necessary because the β and σ^2 estimates given λ and ρ are least squares type estimates.

[2] A (sequence of $n \times n$) matrix P_n is said to be uniformly bounded in row and column sums in absolute value if $\sup_{n\geq 1} \|P_n\|_\infty < \infty$ and $\sup_{n\geq 1} \|P_n\|_1 < \infty$, where $\|P_n\|_\infty = \sup_{1\leq i\leq n} \sum_{j=1}^n |p_{ij,n}|$ and $\|P_n\|_1 = \sup_{1\leq j\leq n} \sum_{i=1}^n |p_{ij,n}|$ are, respectively, the row sum and column sum norms.

[3] The consistency and asymptotic normality of QML estimators still hold under a finite n and a large T, but this case is of less interest as the incidental parameter problem does not occur in this model.

(*b*): $\lim_{n\to\infty} \mathcal{Q}_{2n}(\delta) \neq 0$ *for* $\delta \neq \delta_0$, *where*

$$
\begin{aligned}
\mathcal{H}_N(\rho) &= \tfrac{1}{N}(\mathbf{X}_N, \mathbf{W}_{1N}\mathbf{A}_N^{-1}\mathbf{X}_N\beta_0)'\mathbf{B}_N'(\rho)\mathbf{B}_N(\rho)(\mathbf{X}_N, \mathbf{W}_{1N}\mathbf{A}_N^{-1}\mathbf{X}_N\beta_0),\\
\mathcal{Q}_{1n}(\rho) &= \tfrac{1}{n-1}\left(\ln\left|\sigma_0^2 B_n^{-1\prime}J_nB_n^{-1}\right| - \ln\left|\sigma_n^2(\rho)B_n^{-1}(\rho)'J_nB_n^{-1}(\rho)\right|\right),\\
\mathcal{Q}_{2n}(\delta) &= \tfrac{1}{n-1}\left(\ln\left|\sigma_0^2 B_n^{-1\prime}A_n^{-1\prime}J_nA_n^{-1}B_n^{-1}\right| - \ln\left|\sigma_n^2(\delta)B_n^{-1}(\rho)'A_n^{-1}(\lambda)'J_nA_n^{-1}(\lambda)B_n^{-1}(\rho)\right|\right),\\
\sigma_n^2(\delta) &= \tfrac{\sigma_0^2}{n-1}\mathrm{tr}[(B_n(\rho)A_n(\lambda)A_n^{-1}B_n^{-1})'J_n(B_n(\rho)A_n(\lambda)A_n^{-1}B_n^{-1})], \text{ and}\\
\sigma_n^2(\rho) &= \sigma_n^2(\delta)|_{\lambda=\lambda_0}.
\end{aligned}
$$

Assumption E8. *The limit of* $\frac{1}{(n-1)^2}\left[\mathrm{tr}(C_n^sC_n^s)\mathrm{tr}(D_n^sD_n^s) - \mathrm{tr}^2(C_n^sD_n^s)\right]$ *is strictly positive, where* $C_n = J_n\ddot{G}_n - \frac{\mathrm{tr}J_n\ddot{G}_n}{n-1}J_n$ *and* $D_n = J_nH_n - \frac{\mathrm{tr}J_nH_n}{n-1}J_n$*, with* $H_n = W_{2n}B_n^{-1}$ *and* $\ddot{G}_n = B_n(W_{1n}A_n^{-1})B_n^{-1}$.

Theorem E.1: *(Lee and Yu, 2010) Under Assumptions A1-A8, we have* $\hat{\theta}_N \xrightarrow{p} \theta_0$*, and*

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{D} N\big[0,\ \lim_{N\to\infty}\Sigma_N^{-1}(\theta_0)\Gamma_N(\theta_0)\Sigma_N^{-1}(\theta_0)\big], \tag{E-1}$$

where $\Sigma_N(\theta_0) = \frac{1}{N}\mathrm{E}[\frac{\partial^2}{\partial\theta_0\partial\theta_0'}\ell_N(\theta_0)]$ *assumed to be positive definite for large enough* N*, and* $\Gamma_N(\theta_0) = \frac{1}{N}\mathrm{E}[(\frac{\partial}{\partial\theta_0}\ell_N(\theta_0))(\frac{\partial}{\partial\theta_0}\ell_N(\theta_0))']$ *assumed to exist.*

Theorem E.1 serve two purposes: (*i*) the $\sqrt{N}$-consistency of $\hat{\theta}_N$, which is crucial for the higher-order results, and (*ii*) the asymptotic VC matrix of $\hat{\theta}_N$, which is needed in the third-order variance correction. The VC matrix takes the following form:

$$
\Sigma_N(\theta_0) = \begin{pmatrix}
\frac{1}{N\sigma_0^2}\mathbf{X}_N'\mathbf{B}_N'\mathbf{B}_N\mathbf{X}_N, & 0, & \frac{1}{N\sigma_0^2}\mathbf{X}_N'\mathbf{B}_N'\boldsymbol{\eta}_N, & 0\\
\sim, & \frac{1}{2\sigma_0^4}, & \frac{1}{N\sigma_0^2}\mathrm{tr}(\mathbf{B}_N'^{-1}\mathbf{G}_N\mathbf{B}_N), & \frac{1}{N\sigma_0^2}\mathrm{tr}(\mathbf{W}_{2N}\mathbf{B}_N^{-1})\\
\sim, & \sim, & T_{1N} + T_{1N}^* + \frac{1}{N\sigma_0^2}\boldsymbol{\eta}_N'\boldsymbol{\eta}_N, & T_{2N}^*\\
\sim, & \sim, & \sim, & K_{1N} + K_{1N}^*
\end{pmatrix},
$$

where $\boldsymbol{\eta}_N = \mathbf{G}_N\mathbf{X}_N\beta_0$,

$T_{1N}^* = \frac{1}{N}\mathrm{tr}(\mathbf{B}_N'^{-1}\mathbf{G}_N'\mathbf{B}_N'\mathbf{B}_N\mathbf{G}_N\mathbf{B}_N^{-1})$,

$K_{1N}^* = \frac{1}{N}\mathrm{tr}(\mathbf{B}_N'^{-1}\mathbf{W}_{2N}'^{-1}\mathbf{W}_{2N}^{-1}\mathbf{B}_N^{-1})$, and

$T_{2N}^* = \frac{1}{N}\mathrm{tr}(\mathbf{B}_N'^{-1}\mathbf{G}_N'\mathbf{W}_{2N} + \mathbf{B}_N'^{-1}\mathbf{G}_N'\mathbf{B}_N'\mathbf{W}_{2N}\mathbf{B}_N^{-1})$.

To obtain the other component $\Gamma_N(\theta_0)$ of the VC matrix, it is helpful to express the score vector in terms of the original errors using (4.29):

$$\frac{1}{N}\frac{\partial \ell_N(\theta_0)}{\partial \theta_0} = \begin{cases} \frac{1}{N\sigma_0^2}\mathbb{A}'_{1nT}\mathbb{V}_{nT} \\ -\frac{1}{2\sigma_0^2} + \frac{1}{2N\sigma_0^4}\mathbb{V}'_{nT}\mathbb{A}'_{2nT}\mathbb{V}_{nT} \\ -T_{0N} + \frac{1}{N\sigma_0^2}\mathbb{V}'_{nT}\mathbb{A}'_{3nT}\mathbb{V}_{nT} + \frac{1}{N\sigma_0^2}\mathbf{b}'_{nT}\mathbb{V}_{nT} \\ -K_{0N} + \frac{1}{N\sigma_0^2}\mathbb{V}'_{nT}\mathbb{A}'_{4nT}\mathbb{V}_{nT} \end{cases}$$

where $\mathbf{b}_{nT} = \mathbb{F}_{nT,N}\mathbf{B}_N\boldsymbol{\eta}_N$, $\mathbb{A}_{1nT} = \mathbb{F}_{nT,N}\mathbf{B}_N\mathbf{X}_N$, $\mathbb{A}_{2nT} = \mathbb{F}_{nT,N}\mathbb{F}'_{nT,N}$, $\mathbb{A}_{3nT} = \mathbb{F}_{nT,N}\mathbf{B}_N \cdot\mathbf{G}_N\mathbf{B}_N^{-1}\mathbb{F}'_{nT,N}$, and $\mathbb{A}_{4nT} = \mathbb{F}_{nT,N}\mathbf{W}_{2N}\mathbf{B}_N^{-1}\mathbb{F}'_{nT,N}$. Letting $\mathbf{a}_{inT}$ be the diagonal vector of $\mathbb{A}_{inT}$, and denoting $\Pi_{ij} = \frac{1}{N}\mathrm{tr}[\mathbb{A}_{inT}(\mathbb{A}_{jnT} + \mathbb{A}'_{jnT})] + \frac{1}{N}k_4\mathbf{a}'_{inT}\mathbf{a}_{jnT}$, we obtain, referring to Lemma A.4 of Lee and Yu (2010b) and its proof,

$$\Gamma_N(\theta_0) = \begin{pmatrix} \frac{1}{N\sigma_0^2}\mathbf{X}'_N\mathbf{B}'_N\mathbf{B}_N\mathbf{X}_N, & 0, & \frac{1}{N\sigma_0^2}\mathbf{A}'_{1nT}\mathbf{b}_{nT}, & 0 \\ \sim, & \frac{1}{4\sigma_0^4}\Pi_{22}, & \frac{1}{2\sigma_0^2}\Pi_{23}, & \frac{1}{2\sigma_0^2}\Pi_{24} \\ \sim, & \sim, & \Pi_{33} + \frac{1}{N\sigma_0^2}\mathbf{b}'_{nT}\mathbf{b}_{nT}, & \Pi_{34} \\ \sim, & \sim, & \sim, & \Pi_{44} \end{pmatrix}.$$

Some Higher-Order Results

Derivatives of $\mathbf{M}_N(\rho)$ defined below (4.7). Let, $C_N(\rho) = \mathbf{B}'_N(\rho)\mathbf{B}_N(\rho)$ and $D_N(\rho) = [\mathbf{X}'_N C_N(\rho)\mathbf{X}_N]^{-1}$, then write $\mathbf{M}_N(\rho) = C_N(\rho) - C_N(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}'_N C_N(\rho)$. Let $C_N^{(k)}(\rho)$ and $D_N^{(k)}(\rho)$ be, respectively, the kth order partial derivatives of $C_N(\rho)$ and $D_N(\rho)$ w.r.t. ρ. The derivatives of $\mathbf{M}_N(\rho)$ are:

$$\begin{aligned}
\mathbf{M}_N^{(1)}(\rho) &= C_N^{(1)}(\rho) - C_N^{(1)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}'_N C_N(\rho) - C_N(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}'_N C_N(\rho) \\
&\quad - C_N(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}'_N C_N^{(1)}(\rho), \\
\mathbf{M}_N^{(2)}(\rho) &= C_N^{(2)}(\rho) - C_N^{(2)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}'_N C_N(\rho) - 2C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}'_N C_N(\rho) \\
&\quad - 2C_N^{(1)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}'_N C_N^{(1)}(\rho) - 2C_N(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}'_N C_N^{(1)}(\rho) \\
&\quad - C_N(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}'_N C_N(\rho) - C_N(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}'_N C_N^{(2)}(\rho) \\
\mathbf{M}_N^{(3)}(\rho) &= -3C_N^{(2)}(\rho)\mathbf{X}_N D_N^{(1)}\mathbf{X}'_N C_N(\rho) - 3C_N^{(2)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}'_N C_N^{(1)}(\rho) \\
&\quad - 3C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}'_N C_N(\rho) - 6C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}'_N C_N^{(1)}(\rho)
\end{aligned}$$

$$-3C_N^{(1)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) - 3C_N(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho)$$
$$-3C_N(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) - C_N(\rho)\mathbf{X}_N D_N^{(3)}(\rho)\mathbf{X}_N' C_N(\rho)$$
$$\mathbf{M}_N^{(4)}(\rho) = -6C_N^{(2)}(\rho)\mathbf{X}_N D_N^{(2)}\mathbf{X}_N' C_N(\rho) - 12C_N^{(2)}(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho)$$
$$-6C_N^{(2)}(\rho)\mathbf{X}_N D_N(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) - 4C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(3)}(\rho)\mathbf{X}_N' C_N(\rho)$$
$$-4C_N(\rho)\mathbf{X}_N D_N^{(3)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho) - 12C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}_N' C_N^{(1)}(\rho)$$
$$-12C_N^{(1)}(\rho)\mathbf{X}_N D_N^{(1)}(\rho)\mathbf{X}_N' C_N^{(2)}(\rho) - 6C_N(\rho)\mathbf{X}_N D_N^{(2)}(\rho)\mathbf{X}_N' C_N^{(2)}(\rho)$$
$$-C_N(\rho)\mathbf{X}_N D_N^{(4)}(\rho)\mathbf{X}_N' C_N(\rho).$$

For the derivatives of $C_N(\rho)$, we have $C_N^{(1)}(\rho) = -\mathbf{W}_{2N}'\mathbf{B}_N(\rho) - \mathbf{B}_N'(\rho)\mathbf{W}_{2N}$, $C_N^{(2)}(\rho) = 2\mathbf{W}_{2N}'\mathbf{W}_{2N}$, and $C_N^{(k)}(\rho) = 0, k \geq 3$. For the derivatives of $D_N(\rho)$, denoting $P_N(\rho) = \mathbf{X}_N' C_N(\rho)\mathbf{X}_N$ and its kth derivative $P_N^{(k)}(\rho)$, we have,

$$D_N^{(1)}(\rho) = -D_N(\rho)P_N^{(1)}(\rho)D_N(\rho),$$
$$D_N^{(2)}(\rho) = -D_N^{(1)}(\rho)P_N^{(1)}(\rho)D_N(\rho) - D_N(\rho)P_N^{(2)}(\rho)D_N(\rho) - D_N(\rho)P_N^{(1)}(\rho)D_N^{(1)}(\rho),$$
$$D_N^{(3)}(\rho) = -D_N^{(2)}(\rho)P_N^{(1)}(\rho)D_N(\rho) - D_N(\rho)P_N^{(1)}(\rho)D_N^{(2)}(\rho) - 2D_N^{(1)}(\rho)P_N^{(2)}(\rho)D_N(\rho)$$
$$-2D_N^{(1)}(\rho)P_N^{(1)}(\rho)D_N^{(1)}(\rho) - 2D_N(\rho)P_N^{(2)}(\rho)D_N^{(1)}(\rho),$$
$$D_N^{(4)}(\rho) = -D_N^{(3)}(\rho)P_N^{(1)}(\rho)D_N(\rho) - D_N(\rho)P_N^{(1)}(\rho)D_N^{(3)}(\rho) - 3D_N^{(2)}(\rho)P_N^{(2)}(\rho)D_N(\rho)$$
$$-3D_N^{(2)}(\rho)P_N^{(1)}(\rho)D_N^{(1)}(\rho) - 3D_N^{(1)}(\rho)P_N^{(1)}(\rho)D_N^{(2)}(\rho) - 3D_N(\rho)P_N^{(2)}(\rho)D_N^{(2)}(\rho)$$
$$-6D_N^{(1)}(\rho)P_N^{(2)}(\rho)D_N^{(1)}(\rho).$$

Clearly, $P_N^{(k)}(\rho)$ can be obtained from $C_N^{(k)}(\rho)$, and both are zero when $k \geq 3$.

Additional quantities required in (4.17)**:** Letting $\mathrm{E}(Q_N^{(1)}) = (s_1, s_2)$, $q_N = (s_3, s_4)$ and $\mathrm{E}[Q_N^{(2)}(\delta_0) = (s_5, s_6, s_7, s_8)$, we have

$s_1 = -2\beta_0'\mathbf{X}_N'\mathbf{G}_{1N}'\mathbf{M}_N\mathbf{X}_N\beta_0 - 2\sigma_0^2\mathrm{tr}[\mathbf{G}_N\mathbf{M}_N(\mathbf{B}_N'\mathbf{B}_N)^{-1}]$,

$s_2 = 2\beta_0'\mathbf{X}_N'\mathbf{M}_N^{(1)}\mathbf{X}_N\beta_0 + \sigma_0^2\mathrm{tr}[\mathbf{M}_N^{(1)}(\mathbf{B}_N'\mathbf{B}_N)^{-1}]$,

$s_3 = -4\beta_0'\mathbf{X}_N'\mathbf{G}_{1N}'\mathbf{M}_N\mathbf{B}_N^{-1}\mathbf{V}_N - 2\mathbf{V}_N'\mathbf{B}_N'\mathbf{G}_N\mathbf{M}_N\mathbf{B}_N^{-1}\mathbf{V}_N + 2\sigma_0^2\mathrm{tr}[\mathbf{G}_N\mathbf{M}_N(\mathbf{B}_N'\mathbf{B}_N)^{-1}]$,

$s_4 = 2\beta_0'\mathbf{X}_N'\mathbf{M}_N^{(1)}\mathbf{B}_N^{-1}\mathbf{V}_N + \mathbf{V}_N'\mathbf{B}_N'^{-1}\mathbf{M}_N^{(1)}\mathbf{B}_N^{-1}\mathbf{V}_N - \sigma_0^2\mathrm{tr}[\mathbf{M}_N^{(1)}(\mathbf{B}_N'\mathbf{B}_N)^{-1}]$,

$s_5 = 2\beta_0'\mathbf{X}_N'\mathbf{G}_{1N}'\mathbf{M}_N\mathbf{G}_N\mathbf{X}_N\beta_0 + 2\sigma_0^2\mathrm{tr}[\mathbf{G}_{1N}'\mathbf{M}_N\mathbf{G}_N(\mathbf{B}_N'\mathbf{B}_N)^{-1}]$,

$s_6 = q_7 = -2\beta_0'\mathbf{X}_N'\mathbf{G}_{1N}'\mathbf{M}_N^{(1)}\mathbf{X}_N\beta_0 - 2\sigma_0^2\mathrm{tr}[\mathbf{G}_N\mathbf{M}_N^{(1)}(\mathbf{B}_N'\mathbf{B}_N)^{-1}]$,

$s_8 = \beta_0'\mathbf{X}_N'\mathbf{M}_N^{(2)}\mathbf{X}_N\beta_0 + \sigma_0^2\mathrm{tr}[\mathbf{M}_N^{(2)}(\mathbf{B}_N'\mathbf{B}_N)^{-1}]$,

where $\mathbf{M}_N \equiv \mathbf{M}_N(\rho_0)$ and $\mathbf{M}_N^{(k)} \equiv \mathbf{M}_N^{(k)}(\rho_0)$.

Proofs for Section 4.4

Proof of Lemma 4.1: The results of (a) follows from the following properties of cumulants: for two independent random variables X and Y and a constant c, (i) $\kappa_1(X+c) = \kappa_1(X)+c$, (ii) $\kappa_r(X+c) = \kappa_r(X), r \geq 2$, (iii) $\kappa_r(cX) = c^r\kappa_r(X)$, and (iv) $\kappa_r(X+Y) = \kappa_r(X)+\kappa_r(Y)$. See, e.g., Kendall and Stuart (1969, Sec. 3.12). The results of (b) follows from the definition of the joint cumulants, and some tedious derivations.

Proof of Lemma 4.2: Note that the rth cumulant w.r.t. the EDF $\mathcal{G}_N$ of $\{\mathbf{v}_i, i = 1,\ldots,N\}$ is just the rth sample cumulant of $\{\mathbf{v}_i, i = 1,\ldots,N\}$. This immediately gives $\kappa_1^\star(\mathbf{v}^\star) = \frac{1}{N}\sum_{i=1}^N \mathbf{v}_i = 0$. To show $\kappa_2^\star(\mathbf{v}^\star) = \sigma_0^2 + O_p(N^{-1/2})$, note that $\mathrm{E}(\kappa_2^\star(\mathbf{v}^\star)) = \frac{1}{N}\mathrm{E}(\mathbf{V}'_N\mathbf{V}_N) = \sigma_0^2$. From Lemma 4.1, we have $\mathrm{Var}(\mathbf{v}_i^2) = k_4 a_{4,i} + 2\sigma_0^4$, $\mathrm{Cov}(\mathbf{v}_i^2, \mathbf{v}_j^2) = k_4 a_{i,i,j,j} = k_4\sum_{m=1}^N f_{mi}^2 f_{mj}^2$, and thus

$$\begin{aligned}
\mathrm{Var}(\tfrac{1}{N}\mathbf{V}'_N\mathbf{V}_N) &= \tfrac{1}{N^2}\textstyle\sum_{i=1}^N \mathrm{Var}(\mathbf{v}_i^2) + \tfrac{2}{N^2}\sum_{i=1}^N\sum_{j\neq i}^N \mathrm{Cov}(\mathbf{v}_\mathrm{i}^2, \mathbf{v}_\mathrm{j}^2) \\
&= \tfrac{1}{N}(k_4\bar{a}_4 + 2\sigma_0^4) + \tfrac{2}{N^2}k_4\textstyle\sum_{i=1}^N\sum_{j\neq i}^N\sum_{m=1}^N f_{mi}^2 f_{mj}^2 \\
&= \tfrac{1}{N}(k_4\bar{a}_4 + 2\sigma_0^4) + \tfrac{2}{N^2}k_4\textstyle\sum_{i=1}^N\sum_{j=1}^N\sum_{m=1}^N f_{mi}^2 f_{mj}^2 - \tfrac{2}{N}k_4\bar{a}_4 \\
&= \tfrac{1}{N}(k_4\bar{a}_4 + 2\sigma_0^4) + \tfrac{2}{N^2}k_4\textstyle\sum_{m=1}^N(\sum_{i=1}^N f_{mi}^2)(\sum_{j=1}^N f_{mj}^2) - \tfrac{2}{N}k_4\bar{a}_4 \\
&= O(N^{-1}),
\end{aligned}$$

due to the fact that $\sum_{i=1}^N f_{mi}^2$ is bounded, uniformly in $m = 1, 2, \ldots, nT$. It follows by the generalised Chebyshev's inequality that $\kappa_2^\star(\mathbf{v}^\star) = \sigma_0^2 + O_p(N^{-1/2})$.

For the general results with $r \geq 3$, it is easy to verify that $\mathrm{E}(\kappa_r^\star(\mathbf{v}^\star)) = k_r\bar{a}_r + O(N^{-1/2})$. By the results of Lemma 4.1 and the fact that $\sum_{i=1}^N |f_{mi}|^r$ is bounded, uniformly in $m = 1, 2, \ldots, nT$, it is straightforward, though tedious, to show that $\mathrm{Var}(\kappa_r^\star(\mathbf{v}^\star)) = O(N^{-1})$. The result thus follows.

Proof of Lemma 4.3: As $\hat{\mathbf{V}}_N$ is defined by replacing θ_0 in $\mathbf{V}_N$ by $\hat{\theta}_N$, the result follows directly from the $\sqrt{N}$-consistency of $\hat{\theta}_N$.

Proof of Lemma 4.4: The proof is trivial.

APPENDIX F

Proofs of Results in Chapter 5

Proof of Theorem 5.1: We only prove the consistency of $\hat{\lambda}_n$ as the consistency of $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ immediately follows from identities similar to (5.13) and (5.14). Define $\bar{\ell}_n^c(\lambda) = \max_{\beta,\sigma^2} \mathrm{E}[\ell_n(\theta)]$. By Theorem 5.7 of van der Vaart (1998), it amounts to show, (a) identification uniqueness condition: $\sup_{\lambda:d(\lambda,\lambda_0)\geq\epsilon} \frac{1}{n}[\bar{\ell}_n^c(\lambda) - \bar{\ell}_n^c(\lambda_0)] < 0$ for any $\epsilon > 0$ and a distance measure $d(\lambda, \lambda_0)$ and (b) uniform convergence: $\frac{1}{n}[\ell_n^c(\lambda) - \bar{\ell}_n^c(\lambda)] \xrightarrow{p} 0$ uniformly in $\lambda \in \Lambda$.

Now $\bar{\ell}_n^c(\lambda) = -\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln(\bar{\sigma}_n^2(\lambda)) + \ln|A_n(\lambda)|$, where $\bar{\sigma}_n^2(\lambda) = \frac{1}{n}\big[(\lambda_0 - \lambda_n)^2\eta_n' M_n \eta_n + \sigma_0^2 \mathrm{tr}[H_n A_n'^{-1} A_n'(\lambda) A_n(\lambda) A_n^{-1}]\big]$. Recall $\ell_n^c(\lambda)$ defined in (5.3).

Condition (a)**:** Observe that $\bar{\sigma}_n^2(\lambda_0) = \sigma_0^2$, then,

$$
\begin{aligned}
& \lim_{n\to\infty} \tfrac{1}{n}\big[\bar{\ell}_n^c(\lambda) - \bar{\ell}_n^c(\lambda_0)\big] \\
= \; & \lim_{n\to\infty} \big[\tfrac{1}{2n}(\log|A_n'(\lambda)A_n(\lambda)| - \log|A_n' A_n|) + \tfrac{1}{2n}(\log|\sigma_n^{-2}(\lambda) I_n| - \log|\sigma_0^{-2} I_n|)\big] \\
\neq \; & 0 \text{ for } \lambda \neq \lambda_0, \text{ by Assumption 5.6.}
\end{aligned}
$$

Next, note that $p_n(\theta_0) = \exp[\ell_n(\theta_0)]$ is the *quasi* joint pdf of ϵ_n, which is $N(0, \sigma^2 I_n)$.

Let $p_n^0(\theta_0)$ be the *true* joint pdf of $\epsilon_n \sim (0, \sigma^2 H_n)$. Let E^q denote the expectation with respect to $p_n(\theta_0)$, to differentiate from the usual notation E that corresponds to $p_n^0(\theta_0)$.

Now consider $\epsilon_n(\beta, \lambda) = A_n(\lambda)Y_n - X_n\beta = B_n(\lambda)\epsilon_n + b_n(\beta, \lambda)$, where $B_n(\lambda) = A_n(\lambda)A_n^{-1}$ and $b_n(\beta, \lambda) = A_n(\lambda)A_n^{-1}X_n\beta_0 - X_n\beta$. Then, with $\ell_n(\theta)$ given in (5.2), we have

$$\begin{aligned}
\mathrm{E}^q[\ell_n(\theta_0)] &= -\tfrac{n}{2}\ln(2\pi\sigma^2) + \ln|A_n| - \tfrac{n}{2}, \\
\mathrm{E}[\ell_n(\theta_0)] &= -\tfrac{n}{2}\ln(2\pi\sigma^2) + \ln|A_n| - \tfrac{n}{2}, \text{ as } \tfrac{1}{n}\textstyle\sum_{i=1}^n h_{n,i} = 1 \\
\mathrm{E}^q[\ell_n(\theta)] &= -\tfrac{n}{2}\ln(2\pi\sigma^2) + \ln|A_n(\lambda)| - \tfrac{1}{2\sigma^2}[\sigma_0^2 \mathrm{tr}(B_n'(\lambda)B_n(\lambda)) + b_n'(\beta,\lambda)b_n(\beta,\lambda)], \\
\mathrm{E}[\ell_n(\theta)] &= -\tfrac{n}{2}\ln(2\pi\sigma^2) + \ln|A_n(\lambda)| - \tfrac{1}{2\sigma^2}[\sigma_0^2 \mathrm{tr}(H_n B_n'(\lambda)B_n(\lambda)) + b_n'(\beta,\lambda)b_n(\beta,\lambda)],
\end{aligned}$$

where we have used the identities, $B_n(\lambda_0) = I_n$ and $b_n(\beta_0, \lambda_0) = 0$. Now using the identities $A_n(\lambda) = A_n + (\lambda_0 - \lambda)W_n$ and $B_n(\lambda) = I_n + (\lambda_0 - \lambda)G_n$, we have,

$$\begin{aligned}
&\mathrm{E}[\ell_n(\theta)] - \mathrm{E}^q[\ell_n(\theta)] \\
&= 2(\lambda_0 - \lambda)[\mathrm{tr}(H_n G_n) - tr(G_n)] + (\lambda_0 - \lambda)^2[\mathrm{tr}(H_n G_n' G_n) - \mathrm{tr}(G_n' G_n)] = o(1),
\end{aligned}$$

where the last equality holds by assumptions $\mathrm{Cov}(g_n, h_n) = o(1)$ and $\mathrm{Cov}(q_n, h_n) = o(1)$.

By Jensen's inequality, $0 = \log \mathrm{E}^q\big(\frac{p_n(\theta)}{p_n(\theta_0)}\big) \geq \mathrm{E}^q\big[\log\big(\frac{p_n(\theta)}{p_n(\theta_0)}\big)\big]$, and above results, we conclude $\mathrm{E}\big[\log\big(\frac{p_n(\theta)}{p_n(\theta_0)}\big)\big] \leq 0$ or $\mathrm{E}[\log p_n(\theta)] \leq \mathrm{E}[\log p_n(\theta_0)]$, for large enough n. Thus, $\bar{\ell}_n(\lambda) = \max_{\beta,\sigma^2} \mathrm{E}[\log p_n(\theta)] \leq \max_{\beta,\sigma^2} \mathrm{E}[\log p_n(\theta_0)] = \mathrm{E}[\log p_n(\theta_0)] = \bar{\ell}_n(\lambda_0)$, for $\lambda \neq \lambda_0$ and n large enough.

Condition (*b*)**:** Note $\frac{1}{n}[\ell_n^c(\lambda) - \bar{\ell}_n^c(\lambda)] = -\frac{1}{2}[\log(\hat{\sigma}_n^2(\lambda)) - \log(\bar{\sigma}_n^2(\lambda))]$. By the mean value theorem, $\log(\hat{\sigma}_n^2(\lambda)) - \log(\bar{\sigma}_n^2(\lambda)) = \frac{1}{\dot{\sigma}_n^2(\lambda)}[\hat{\sigma}_n^2(\lambda) - \bar{\sigma}_n^2(\lambda)]$, where $\dot{\sigma}_n^2(\lambda)$ lies between $\hat{\sigma}_n^2(\lambda)$ and $\bar{\sigma}_n^2(\lambda)$. Using $M_n A_n(\lambda) Y_n = (\lambda_0 - \lambda)M_n\eta_n + M_n A_n(\lambda)A_n^{-1}\epsilon_n$ we have,

$$\hat{\sigma}_n^2(\lambda) = (\lambda_0 - \lambda)^2 \tfrac{1}{n}\eta_n' M_n \eta_n + 2(\lambda_0 - \lambda)T_{1n}(\lambda) + T_{2n}(\lambda), \tag{F-1}$$

where $T_{1n}(\lambda) = \frac{1}{n}\eta_n' M_n A_n(\lambda)A_n^{-1}\epsilon$ and $T_{2n}(\lambda) = \frac{1}{n}\epsilon_n' A_n^{-1} A_n'(\lambda) M_n A_n(\lambda) A_n^{-1}\epsilon_n$. Using $A_n(\lambda) = A_n + (\lambda_0 - \lambda)W_n$, we have, $T_{1n}(\lambda) = o_p(1)$ uniformly. Further, $T_{2n}(\lambda) =$

$\frac{1}{n}\epsilon_n' A_n^{-1} A_n'(\lambda) A_n(\lambda) A_n^{-1}\epsilon_n + o_p(1)$, since, $\frac{1}{n}\epsilon_n' A_n^{-1} A_n'(\lambda) P_n A_n(\lambda) A_n^{-1}\epsilon_n = \frac{1}{n}[\epsilon_n' P_n \epsilon + 2\epsilon_n' G_n' P_n \epsilon_n + \epsilon_n' G_n' P_n G_n \epsilon_n] = o_p(1)$ uniformly, using $\text{Cov}(h_n, g_n) = o(1)$. Lemmas A.1 - A.3 imply, $\frac{1}{n^2}\text{Var}(\epsilon_n' A_n^{-1} A_n'(\lambda) A_n(\lambda) A_n^{-1}\epsilon_n) = o(1)$. Then, together with Chebyshev inequality, $T_{2n}(\lambda) - \sigma_0^2 \frac{1}{n}\text{tr}[H_n A_n'^{-1} A_n'(\lambda) A_n(\lambda) A_n^{-1}] = o_p(1)$, uniformly for $\lambda \in \Lambda$.

It left to show $\sigma_n^2(\lambda)$ is uniformly bounded away from zero. Suppose $\sigma_n^2(\lambda)$ is not uniformly bounded away from zero. Then $\exists\{\lambda_n\} \subset \Lambda$ such that $\sigma_n^2(\lambda_n) \to 0$. Consider the model with $\beta_0 = 0$. The Gaussian log-likelihood is $\ell_{t,n}(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\lambda)| - \frac{1}{2\sigma^2} Y_n' A_n'(\lambda) A_n(\lambda) Y_n$ and $\bar{\ell}_{t,n}(\lambda) = \max_{\sigma^2} \text{E}[\ell_{t,n}(\theta)]$. By Jensen's inequality, $\bar{\ell}_{t,n}(\lambda) \leq \max_{\sigma^2} \text{E}[\ell_{t,n}(\theta_0)] = \bar{\ell}_{t,n}(\lambda_0)$. Then together with Lemma A.2, we have $\frac{1}{n}[\bar{\ell}_{t,n}(\lambda) - \bar{\ell}_{t,n}(\lambda_0)] \leq 0$, and $-\frac{n}{2}\log(\sigma_n^2(\lambda)) \leq -\frac{n}{2}\log(\sigma_0^2) + \frac{1}{n}(\log|A_n(\lambda_0)| - \log|A_n(\lambda)|) = O(1)$. That is, $-\frac{n}{2}\log(\sigma_n^2(\lambda))$ is bounded from above which is a contradiction. Hence, $\sigma_n^2(\lambda)$ is bounded away from zero uniformly, and $\frac{n}{2}\log(\sigma_n^2(\lambda))$ is well defined $\forall\lambda \in \Lambda$.

Collecting all these results we have, $\sup_{\lambda\in\Lambda}\frac{1}{n}|[\ell_n^c(\lambda) - \bar{\ell}_n^c(\lambda)]| = o_p(1)$, completing the proof of consistency part.

To prove the asymptotic normality, first note that $\text{tr}(H_n) = n$. By the mean value theorem, $\sqrt{n}(\hat{\theta}_n - \theta_0) = -\left[\frac{1}{n}\frac{\partial^2}{\partial\theta\partial\theta'}\ell_n(\tilde{\theta})\right]^{-1}\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\ell_n(\theta_0)$, where $\tilde{\theta}_n$ lies elementwise between $\hat{\theta}_n$ and θ_0. By Assumptions 5.1-5.6, the condition $\text{Cov}(g_n, h_n) = o(n^{-1/2})$, and the CLT for vector linear-quadratic forms of Kelejian and Prucha (2010, p. 63), we have $\frac{1}{\sqrt{n}}\frac{\partial}{\partial\theta}\ell_n(\theta_0) \xrightarrow{D} N(0, \Sigma)$, where Σ is defined in the theorem.

Let $\mathcal{H}_n(\theta) = \frac{\partial^2}{\partial\theta\partial\theta'}\ell_n(\theta)$. It left to show (i) $\frac{1}{n}\mathcal{H}_n(\tilde{\theta}_n) - \mathcal{H}_n = o_p(1)$ and (ii) $\mathcal{H}_n - \mathbb{I}_n = o_p(1)$.

Condition (i)**:** By Assumptions 5.3-5.5 and the assumption that $\text{Cov}(h_n, g_n) = o(1)$ stated in the theorem, Lemma A.2-A.3, $\tilde{\theta}_n - \theta_0 = o_p(1)$, $\epsilon_n(\tilde{\beta}_n, \tilde{\lambda}_n) = X_n(\beta_0 - \tilde{\beta}_n) + (\lambda_0 - \tilde{\lambda}_n) W_n Y_n + \epsilon_n$ and $\frac{1}{n}\epsilon_n'(\tilde{\beta}_n, \tilde{\lambda}_n)\epsilon_n(\tilde{\beta}_n, \tilde{\lambda}_n) = \frac{1}{n}\epsilon_n'\epsilon_n + o_p(1)$, we have,

$\mathcal{H}_{n,\beta\beta}(\tilde{\theta}_n) - \mathcal{H}_{n,\beta\beta} = \left(\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}_n^2}\right)\frac{1}{n} X_n' X_n = o_p(1)$,

$\mathcal{H}_{n,\sigma^2\beta}(\tilde{\theta}_n) - \mathcal{H}_{n,\sigma^2\beta} = \frac{1}{\sigma_0^4 n}\epsilon_n' X_n - \frac{1}{\tilde{\sigma}_0^4 n}(X_n(\beta_0 - \tilde{\beta}_n) + (\lambda_0 - \tilde{\lambda}_n) W_n Y_n + \epsilon_n)' X_n = o_p(1)$,

$\mathcal{H}_{n,\sigma^2\sigma^2}(\tilde{\theta}_n) - \mathcal{H}_{n,\sigma^2\sigma^2} = \frac{1}{n}\left(\frac{1}{\sigma_0^6}\epsilon_n'\epsilon_n - \frac{1}{\tilde{\sigma}_n^6}\epsilon_n'(\tilde{\delta}_n)\epsilon_n(\tilde{\delta}_n)\right) - \frac{1}{2}\left(\frac{1}{\sigma_0^4} - \frac{1}{\tilde{\sigma}_n^4}\right) = o_p(1)$,

$\mathcal{H}_{n,\lambda\beta}(\tilde{\theta}_n) - \mathcal{H}_{n,\lambda\beta} = \left(\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}_n^2}\right)\frac{1}{n} Y_n' W_n' X_n = o_p(1)$,

$\mathcal{H}_{n,\lambda\sigma^2}(\tilde{\theta}_n) - \mathcal{H}_{n,\lambda\sigma^2} = \frac{1}{\sigma_0^4 n} Y_n' W_n' \epsilon_n - \frac{1}{\tilde{\sigma}_0^4 n} Y_n' W_n'(X_n(\beta_0 - \tilde{\beta}_n) + (\lambda_0 - \tilde{\lambda}_n) W_n Y_n + \epsilon_n) = o_p(1)$,

$\mathcal{H}_{n,\lambda\lambda}(\tilde{\theta}_n) - \mathcal{H}_{n,\lambda\lambda} = \big(\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}_n^2}\big)\frac{1}{n}Y_n'W_n'W_nY_n + \frac{1}{n}\text{tr}(G_n^2) - \text{tr}(G_n^2(\tilde{\lambda}_n)) = o_p(1)$,

where the last equality holds since $\text{tr}(G_n^2) - \text{tr}(G_n^2(\tilde{\lambda}_n)) = 2\text{tr}(G_n^2(\bar{\lambda}_n))(\lambda_0 - \tilde{\lambda}_n)$ by the mean value theorem for some $\bar{\lambda}_n$ between λ_0 and $\tilde{\lambda}_n$.

Condition (ii)**:** Using the results $\text{E}(\epsilon_n'\epsilon_n) = \sigma_0^2\text{tr}(H_n)$, $\text{E}(\epsilon_n'G_n\epsilon_n) = \sigma_0^2\text{tr}(H_nG_n)$ and $\text{E}(\epsilon_n'G_n'G_n\epsilon_n) = \sigma_0^2\text{tr}(H_nG_n'G_n)$ with Lemmas A.1-A.3, $\text{Var}(\frac{1}{n}\epsilon_n'\epsilon_n) = \frac{1}{n^2}\big(\text{E}(\epsilon_{n,i}^4) - \sigma_0^4\text{tr}(H_n^2)\big) = o(1)$, $\text{Var}(\frac{1}{n}\epsilon_n'G_n\epsilon_n) = \frac{1}{n^2}\sum_{i=1}^n g_{n,ii}^2[\text{E}(\epsilon_{n,i}^4) - 3\sigma_0^4h_i^2] + \frac{1}{n^2}\sigma_0^4\text{tr}[H_nG_n(G_n'H_n + H_nG_n)] = o(1)$ and similarly $\text{Var}(\frac{1}{n}\epsilon_n'G_n'G_n\epsilon_n) = o_p(1)$. Then by Chebyshev inequality,

$$\begin{aligned}
&\mathcal{H}_{n,\beta\beta} - \mathbb{I}_{n,\beta\beta} = 0,\\
&\mathcal{H}_{n,\sigma^2\beta} - \mathbb{I}_{n,\sigma^2\beta} \;= O_p(\tfrac{1}{\sqrt{n}}) = o_p(1),\\
&\mathcal{H}_{n,\sigma^2\sigma^2} - \mathbb{I}_{n,\sigma^2\sigma^2} = \tfrac{1}{\sigma_0^6}\big(\tfrac{\epsilon_n'\epsilon_n}{n} - \sigma_0^2\big) = o_p(1),\\
&\mathcal{H}_{n,\lambda\beta} - \mathbb{I}_{n,\lambda\beta} = \tfrac{1}{n}X_n'G_n\epsilon_n = O_p(\tfrac{1}{\sqrt{n}}) = o_p(1),\\
&\mathcal{H}_{n,\lambda\sigma^2} - \mathbb{I}_{n,\lambda\sigma^2} = \tfrac{1}{\sigma_0^4 n}\epsilon_n'G_n\epsilon_n - \tfrac{1}{\sigma_0^2 n}\text{tr}(H_nG_n) + O_p(\tfrac{1}{\sqrt{n}}) = o_p(1) \text{ and}\\
&\mathcal{H}_{n,\lambda\lambda} - \mathbb{I}_{n,\lambda\lambda} = \tfrac{1}{n}\epsilon_n'G_n'G_n\epsilon_n - \tfrac{1}{n}\text{tr}(H_nG_n'G_n) + O_p(\tfrac{1}{\sqrt{n}}) = o_p(1).
\end{aligned}$$

Proof of Theorem 5.2: Let $\text{E}(\tilde{\psi}_n^*(\lambda)) = \bar{\psi}^*(\lambda)$. By Theorem 5.9 of van der Vaart (1998), the proof of consistency of $\tilde{\lambda}_n$ requires (a) Convergence: $\sup_{\lambda\in\Lambda}|\tilde{\psi}_n^*(\lambda) - \bar{\psi}^*(\lambda)| = o_p(1)$ and (b) Identification uniqueness: for $\epsilon > 0$, $\inf_{\lambda: d(\lambda,\lambda_0)\geq\epsilon}|\bar{\psi}^*(\lambda)| > 0 = |\bar{\psi}^*(\lambda_0)|$.

The proof of Theorem 5.1 implies that $\hat{\sigma}_n^2(\lambda)$ is uniformly bounded away from 0. Thus, the ACQS estimator $\tilde{\lambda}_n = \arg\{\tilde{\psi}_n^*(\lambda) = 0\}$ is equivalently defined as $\tilde{\lambda}_n = \arg\{Y_n'A_n'(\lambda)M_nG_n^\circ(\lambda)A_n(\lambda)Y_n = 0\}$, suggesting that we can work purely with the numerator $T_n(\lambda) = Y_n'A_n'(\lambda)M_nG_n^\circ(\lambda)A_n(\lambda)Y_n$ of $\tilde{\psi}_n^*(\lambda)$ to establish consistency. Note $T_n(\lambda) = Y_n'A_n'(\lambda)M_nG_n(\lambda)A_n(\lambda)Y_n - Y_n'A_n'(\lambda)M_n\text{diag}(M_n)^{-1}\text{diag}(M_nG_n(\lambda))A_n(\lambda)Y_n \equiv T_{1n}(\lambda) - T_{2n}(\lambda)$.

Condition (a): By $M_nX_n = 0$, $A_n(\lambda) = A_n + (\lambda_0 - \lambda)W_n$ and $G_nA_n = W_n = G_n(\lambda)A_n(\lambda)$,

$$\begin{aligned}
T_{1n}(\lambda) &= Y_n'A_n'(\lambda)M_nG_n(\lambda)A_n(\lambda)Y_n\\
&= Y_n'A_n'M_nG_nA_nY_n + (\lambda_0 - \lambda)Y_n'A_n'G_n'M_nG_nA_nY_n\\
&= \epsilon_n'M_nG_n(X_n\beta_0 + \epsilon_n) + (\lambda_0 - \lambda)(X_n\beta_0 + \epsilon_n)'G_n'M_nG_n(X_n\beta_0 + \epsilon_n), \text{ and}
\end{aligned}$$

$\mathrm{E}(T_{1n}(\lambda)) = (\lambda_0-\lambda)\beta_0' X_n G_n' M_n G_n X_n \beta_0 + \sigma_0^2 \mathrm{tr}(H_n M_n G_n) + \sigma_0^2(\lambda_0-\lambda)\mathrm{tr}(H_n G_n' M_n G_n)$. By Lemma A.3 and Assumptions 5.5 and 5.6, we have $\frac{1}{n}[T_{1n}(\lambda) - \mathrm{E}(T_{1n}(\lambda))] = o_p(1)$. Now, as M_n appeared in T_{2n} is a projection matrix, by Lemma A.2, similar arguments as for $T_{1n}(\lambda)$ lead to $\frac{1}{n}[T_{2n}(\lambda) - \mathrm{E}(T_{2n}(\lambda))] = o_p(1)$. Thus, $\frac{1}{n}\{T_n(\lambda) - \mathrm{E}[T_n(\lambda)]\} = o_p(1)$.

Condition (b): First, we have $\mathrm{E}[T_n(\lambda_0)] = 0$, as $\mathrm{tr}[H_n M_n \mathrm{diag}(M)^{-1}\mathrm{diag}(M_n G_n)] = \mathrm{tr}[\mathrm{diag}(H_n M_n \mathrm{diag}(M)^{-1})\mathrm{diag}(M_n G_n)] = \mathrm{tr}(H_n M_n G_n)$. Now,

$$
\begin{aligned}
\mathrm{E}[T_n(\lambda)] = {} & \beta_0' X_n' A_n'^{-1} A_n'(\lambda) M_n G_n^{\circ}(\lambda) A_n(\lambda) A_n^{-1} X_n \beta_0 \\
& + \sigma_0^2 \mathrm{tr}(H_n A_n'^{-1} A_n'(\lambda) M_n G_n^{\circ}(\lambda) A_n(\lambda) A_n^{-1}).
\end{aligned}
$$

By Assumption 5.6* and Lemma A.2, $\mathrm{E}[T_n(\lambda)] \neq 0$, for any $\lambda \neq \lambda_0$. Then the conditions of Theorem 5.9 of van der Vaart (1998) hold, and thus the consistency of $\tilde{\lambda}_n$.

To prove asymptotic normality, we have, by the mean value theorem,

$$
0 = \sqrt{n}\tilde{\psi}_n^*(\tilde{\lambda}_n) = \sqrt{n}\tilde{\psi}_n^*(\lambda_0) + \tfrac{d}{d\lambda}\tilde{\psi}_n^*(\bar{\lambda}_n)\sqrt{n}(\tilde{\lambda}_n - \lambda_0), \tag{F-2}
$$

where $\bar{\lambda}_n$ lies between $\tilde{\lambda}_n$ and λ_0. We have to show, (i) $\frac{d}{d\lambda}\tilde{\psi}_n^*(\bar{\lambda}_n) - \frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0) = o_p(1)$, (ii) $\frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0) - \mathrm{E}\big(\frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0)\big) = o_p(1)$, and (iii) $\mathrm{E}\big(\frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0)\big) \neq 0$ for large enough n.

Note,
$$
\begin{aligned}
\tfrac{d}{d\lambda}\tilde{\psi}_n^*(\lambda) = {} & \tfrac{2}{n^2\hat{\sigma}_n^4(\lambda)} Y_n' A_n'(\lambda) G_n^{\circ\prime}(\lambda) M_n A_n(\lambda) Y_n \cdot Y_n' W_n' M_n A_n(\lambda) Y_n \\
& - \tfrac{1}{n\hat{\sigma}_n^2(\lambda)} Y_n' W_n' G_n^{\circ\prime}(\lambda) M_n A_n(\lambda) Y_n - \tfrac{1}{n\hat{\sigma}_n^2(\lambda)} Y_n' A_n'(\lambda) G_n^{\circ\prime}(\lambda) M_n W_n Y_n \\
& + \tfrac{1}{n\hat{\sigma}_n^2(\lambda)} Y_n' A_n'(\lambda) \dot{G}_n^{\circ\prime}(\lambda) M_n A_n(\lambda) Y_n,
\end{aligned}
$$
where $\dot{G}_n^{\circ}(\lambda) = \frac{d}{d\lambda}G_n^{\circ}(\lambda) = G_n^2(\lambda) - \mathrm{diag}(M_n)^{-1}\mathrm{diag}(M_n G_n^2(\lambda))$.

Condition (i): $\frac{1}{n}Y_n' W_n' M_n A_n(\bar{\lambda}_n) Y_n = \frac{1}{n}Y_n' W_n' M_n A_n Y_n + \frac{1}{n}(\lambda_0 - \bar{\lambda}_n) Y_n' W_n' M_n W_n Y_n = \frac{1}{n}Y_n' W_n' M_n A_n Y_n + o_p(1)$. By Assumptions 5.4, 5.5 and continuous mapping theorem, $G_n^{\circ}(\bar{\lambda}_n) = G_n^{\circ} + o_p(1)$ and $\dot{G}_n^{\circ}(\bar{\lambda}_n) = \dot{G}_n^{\circ} + o_p(1)$. Thus $\frac{1}{n}Y_n' A_n'(\bar{\lambda}_n) G_n^{\circ\prime}(\bar{\lambda}_n) M_n A_n(\bar{\lambda}_n) Y_n = \frac{1}{n}Y_n' A_n' G_n^{\circ\prime} M_n A_n Y_n + o_p(1)$ and $\frac{1}{n}Y_n' A_n'(\bar{\lambda}_n) \dot{G}_n^{\circ\prime}(\bar{\lambda}_n) M_n A_n(\bar{\lambda}_n) Y_n = \frac{1}{n}Y_n' A_n' \dot{G}_n^{\circ\prime} M_n A_n Y_n + o_p(1)$. In a similar manner, $\frac{1}{n}Y_n' W_n' G_n^{\circ\prime}(\bar{\lambda}_n) M_n A_n(\bar{\lambda}_n) Y_n = \frac{1}{n}Y_n' W_n' G_n^{\circ\prime} M_n A_n Y_n + o_p(1)$, and $\frac{1}{n}Y_n' A_n'(\bar{\lambda}_n) G_n^{\circ\prime}(\bar{\lambda}_n) M_n W_n Y_n = \frac{1}{n}Y_n' A_n' G_n^{\circ\prime} M_n W_n Y_n + o_p(1)$. Collecting these results and observing $\hat{\sigma}_n^2(\bar{\lambda}_n) = \hat{\sigma}_n^2(\lambda_0) + o_p(1)$, we have $\frac{d}{d\lambda}\tilde{\psi}_n^*(\bar{\lambda}_n) - \frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0) = o_p(1)$.

Condition (ii)**:** Note that,

$$\begin{aligned}\tfrac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0) &= \tfrac{1}{n\sigma_0^2}Y_n'A_n'\dot{G}_n^{\circ\prime}M_nA_nY_n - \tfrac{1}{n\sigma_0^2}Y_nW_n'G_n^{\circ\prime}M_nA_nY_n - \tfrac{1}{n\sigma_0^2}Y_nA_n'G_n^{\circ\prime}M_nW_nY_n \\ &\quad + \tfrac{2}{n^2\sigma_0^4}(Y_n'A_n'G_n^{\circ\prime}M_nA_nY_n)\cdot(Y_n'W_n'M_nA_nY_n) + o_p(1) \equiv \textstyle\sum_{i=1}^4 T_{in} + o_p(1).\end{aligned}$$

Using $M_nA_nY_n = M_n\epsilon_n$ and $\frac{1}{n}a_n'\epsilon_n = o_p(1)$ for a vector a_n of uniformly bounded elements, we can readily verify that $T_{1n} = \frac{1}{n\sigma_0^2}\epsilon_n'\dot{G}_n^{\circ\prime}\epsilon_n + o_p(1)$, $T_{2n} = -\frac{1}{n\sigma_0^2}\epsilon_n'G_n^{\circ}G_n\epsilon_n + o_p(1)$, $T_{3n} = -\frac{1}{n\sigma_0^2}(c_n'\eta_n + \epsilon_n'G_n^{\circ\prime}G_n\epsilon_n) + o_p(1)$, and $T_{4n} = o_p(1)$, by Lemma A.2. It follows that $-\mathrm{E}\big[\frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0)\big] = \frac{1}{n}\mathrm{tr}[H_n(G_n^{\circ}G_n + G_n^{\circ\prime}G_n - \dot{G}_n^{\circ})] + \frac{1}{n\sigma_0^2}c_n'\eta_n + o(1) = \Phi_n + o(1)$, and that $\frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0) - \mathrm{E}\big[\frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0)\big] = o_p(1)$.

Condition (iii)**:** By Assumptions 5.3-5.6 and Lemmas A.2 and A.3, it is easy to see that $\Phi_n \neq 0$ for large enough n, and thus $\mathrm{E}\big(\frac{d}{d\lambda}\tilde{\psi}_n^*(\lambda_0)\big) \neq 0$ for large enough n.

Proof of Theorem 5.3: Recall $\tilde{\beta}_n = (X_n'X_n)^{-1}X_n'A_n(\tilde{\lambda}_n)Y_n$. We have,

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) = (\tfrac{1}{n}X_n'X_n)^{-1}\tfrac{1}{\sqrt{n}}X_n'\epsilon_n - \sqrt{n}(\tilde{\lambda}_n - \lambda_0)(\tfrac{1}{n}X_n'X_n)^{-1}\tfrac{1}{n}X_n'\eta_n + O_p\big(\tfrac{1}{\sqrt{n}}\big). \quad \text{(F-3)}$$

The proof of the asymptotic normality of $\tilde{\lambda}_n$ in Theorem 5.2 and the asymptotic representation for $\sqrt{n}\tilde{\psi}_n^*$ given in (5.10) imply that

$$\sqrt{n}(\tilde{\lambda}_n - \lambda_0) = \Phi_n^{-1}\sqrt{n}\tilde{\psi}_n^* + o_p(1) = (\sqrt{n}\sigma_0^2\Phi_n)^{-1}(\epsilon_n'B_n\epsilon_n + c_n'\epsilon_n) + o_p(1). \quad \text{(F-4)}$$

This shows that each component of $\sqrt{n}(\tilde{\beta}_n - \beta_0)$ is a linear-quadratic form in ϵ_n. Thus, Cramèr-Wold device and the CLT for linear-quadratic form of Kelejian and Prucha (2001) lead to the asymptotic normality of $\sqrt{n}(\tilde{\beta}_n - \beta_0)$. Clearly, the asymptotic mean of $\sqrt{n}(\tilde{\beta}_n - \beta_0)$ is zero and the first-order variance of it can be easily found using (F-3) and (F-4):

$$\begin{aligned}\tau^2(\tilde{\beta}_n) &= (X_n'X_n)^{-1}X_n'\mathrm{Var}(\epsilon_n)X_n(X_n'X_n)^{-1} + \tau^2(\tilde{\lambda}_n)(X_n'X_n)^{-1}X_n'\eta_n\eta_n'X_n(X_n'X_n)^{-1} \\ &\quad -2(\sigma_0^2\Phi_n)^{-1}(X_n'X_n)^{-1}X_n'\mathrm{Cov}(\epsilon_n, \epsilon_n'B_n\epsilon_n + c_n'\epsilon_n)\eta_n'X_n(X_n'X_n)^{-1} \\ &= (X_n'X_n)^{-1}X_n'\mathbb{A}_nX_n(X_n'X_n)^{-1},\end{aligned}$$

where $\mathbb{A}_n = n\sigma_0^2 H_n + \tau_n^2(\tilde{\lambda}_n)\eta_n\eta_n' - 2\Phi_n^{-1}(\sigma_0^{-2}\text{diag}(B_n)s_n + H_n c_n)\eta_n'$, and $s_n = \text{E}(\epsilon_n^3)$.

The limiting distribution of $\sqrt{n}(\tilde{\sigma}_n^2 - \sigma_0^2)$ can be found in a similar manner from

$$\begin{aligned}\sqrt{n}(\tilde{\sigma}_n^2 - \sigma_0^2) &= \sqrt{n}[\tfrac{1}{n}Y_n' A_n'(\tilde{\lambda}_n)M_n A_n(\tilde{\lambda}_n)Y_n - \sigma_0^2] \\ &= \tfrac{1}{\sqrt{n}}(\epsilon_n'\epsilon_n - n\sigma_0^2) + 2\sqrt{n}(\tilde{\lambda}_n - \lambda_0)\tfrac{1}{n}\sigma_0^2\text{tr}(H_n G_n) + o_p(1),\end{aligned}$$

which has a limiting mean zero and first-order variance: $\tau_n^2(\tilde{\sigma}_n^2) = \frac{1}{n}\sum_{i=1}^n \text{Var}(\epsilon_{n,i}^2) + \frac{4}{n^2}\sigma_0^4\tau_n^2(\tilde{\lambda}_n)\text{tr}^2(H_n G_n) + \frac{4}{n^2}\text{Cov}(\epsilon_n'\epsilon, \epsilon_n' B_n\epsilon_n + c_n'\epsilon_n)\text{tr}(H_n G_n)\Phi_n^{-1}$. $\text{Cov}(\epsilon_n'\epsilon, \epsilon_n' B_n\epsilon_n + c_n'\epsilon_n)$ can be easily derived but not needed in light of the discussions provided.

Proof of Theorem 5.4: To prove the consistency of $\tilde{\tau}_n^2(\tilde{\lambda}_n)$ as an estimator of $\tau_n^2(\tilde{\lambda}_n)$, we need to prove (a) $\tilde{\Phi}_n - \Phi_n = o_p(1)$, and (b) $\tilde{\tau}_n^2(\tilde{\psi}_n^*) - \tau_n^2(\tilde{\psi}_n^*) = o_p(1)$. First, (a) follows from the proof of Theorem 5.2 (the asymptotic normality part). To prove (b), as $\tilde{\sigma}_n^2 = \sigma_0^2 + o_p(1)$ by Theorem 5.3, it suffices to show that, by the consistency of $\tilde{\theta}_n$ and referring to (5.17) and (5.18),

$$\tfrac{1}{n}\sum_{i=1}^n \left(\epsilon_{n,i}^2\xi_{n,i}^2 - \text{Var}(\epsilon_{n,i}\xi_{n,i})\right) = o_p(1),$$

where $\xi_{n,i} = \zeta_{n,i} + b_{n,ii}\epsilon_{n,i} + c_{n,i}$. This follows immediately by Theorem A.1 and the poof of Theorem 5.1 of Baltagi and Yang (2013b).

The consistency of $\tilde{\tau}_n^2(\tilde{\beta}_n)$ follows that of $\tilde{\tau}_n^2(\tilde{\lambda}_n)$ and the consistency of $\tilde{\theta}_n$.

Finally, the same procedure proves the same set of the results for the regular QML estimators $\hat{\beta}_n$ and $\hat{\sigma}_n^2$.

APPENDIX G

Proofs of Results in Chapter 6

G.1 Proofs of Results for SARAR(p, q) Model

Proof of Theorem 6.1: Details of the proof of Theorem 6.1, that shows the consistency and the asymptotic normality of the QML estimators of θ under homoskedastic assumption is similar to the proofs of Theorem 3.1 and 3.2 of Lee (2004) after making adjustments for the autoregressive disturbances. As such we omit the proof of Theorem 6.1 and focus on the case where the errors are heteroskedastic.

Proof of Theorem 6.2: Let $\mathrm{E}(\tilde{\psi}_n^*(\delta)) = \bar{\psi}^*(\delta)$. By Theorem 5.9 of van der Vaart (1998), the proof of consistency of $\tilde{\delta}_n$ requires, (a) Uniform convergence: $\sup_{\delta \in \Delta} ||\tilde{\psi}_n^*(\delta) - \bar{\psi}^*(\delta)|| = o_p(1)$ and (b) Identification uniqueness: for $\epsilon > 0$, $\inf_{\delta : d(\delta, \delta_0) \geq \epsilon} ||\bar{\psi}^*(\delta)| > 0 = |\bar{\psi}^*(\delta_0)|$.

The proof of Theorem 6.1 implies that $\hat{\sigma}_n^2(\beta, \lambda)$ is bounded away from 0 with probability one for large enough n. Thus, the ACQS estimator $\tilde{\delta}_n = \arg\{\tilde{\psi}_n^*(\delta) = 0\}$ is equivalently defined as,

$$\tilde{\delta}_n = \arg \left\{ \begin{array}{l} Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}^{\circ}_{jn}(\delta)Y_n(\delta) = 0, \; j = 1, \ldots, p \\ Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}^{\circ}_{kn}(\rho)Y_n(\delta) = 0, \; k = 1, ..., q \end{array} \right\},$$

suggesting that we can work purely with the numerators of $\tilde{\psi}^*_n(\delta)$ in order to establish consistency of $\tilde{\delta}_n$. Let $R_{jn}(\delta) = T_{jn}(\delta) - S_{jn}(\delta)$ where $T_{jn}(\delta) = Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)Y_n(\delta)$ and $S_{jn}(\delta) = Y_n'(\delta)\mathcal{M}_n(\rho)\text{diag}(\mathcal{M}_n(\rho))^{-1}\text{diag}[\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)]Y_n(\delta)$ for $j = 1, \ldots, p$.

Condition (a): By a Taylor expansion around δ_0 we have,

$$\begin{aligned} T_{jn}(\delta) &= Y_n'(\delta_0)\mathcal{M}_n\bar{B}_{jn}Y_n(\delta_0) + (\delta - \delta_0)'\tfrac{\partial}{\partial\delta}T_{jn}(\delta) + o_p(1) \\ &= \epsilon_n'\mathcal{M}_n\bar{B}_{jn}[X_n(\rho_0)\beta_0 + \epsilon_n] + (\delta - \delta_0)'\tfrac{\partial}{\partial\delta}T_{jn}(\delta)|_{\delta=\delta_0} + o_p(1), \end{aligned} \quad \text{(G-1)}$$

where $\frac{\partial}{\partial\delta}T_{jn}(\delta) = \begin{pmatrix} \frac{\partial}{\partial\lambda}T_{jn}(\delta) \\ \frac{\partial}{\partial\rho}T_{jn}(\delta) \end{pmatrix} = \begin{pmatrix} \dot{T}_{jj',n}(\delta) \\ \dot{T}_{jk,n}(\delta) \end{pmatrix}$,

$\dot{T}_{jj',n}(\delta) = -Y_n'(\delta)\bar{B}'_{j'n}(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)Y_n(\delta)$ and

$\dot{T}_{jk,n}(\delta) = -Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}(\rho)\bar{B}_{jn}(\delta)Y_n(\delta) - Y_n'(\delta)\bar{G}'_{kn}(\rho)\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)Y_n(\delta)$ for $j, j' = 1, \ldots, p$ and $k = 1, \ldots, q$. Then,

$\text{E}[T_{rn}(\delta)] = \sigma_0^2\text{tr}(H_n\mathcal{M}_n\bar{B}_{jn}) + (\delta - \delta_0)'\text{E}(\frac{\partial}{\partial\delta}T_{jn}(\delta))$, where

$\text{E}[\dot{T}_{jj',n}(\delta)] = -\sigma_0^2\text{tr}(H_n\bar{B}'_{j'n}\mathcal{M}_n\bar{B}_{jn}) - \eta'_{j'n}\mathcal{M}_n\eta_{jn}$,

$\text{E}[\dot{T}_{jk,n}(\delta)] = -\sigma_0^2\text{tr}(H_n\mathcal{M}_n\bar{G}_{kn}(\rho)\bar{B}_{jn}) - \sigma_0^2\text{tr}(H_n\bar{G}'_{kn}(\rho)\mathcal{M}_n\bar{B}_{jn}) - \eta'_{kn}\mathcal{M}_n\eta_{jn}$,

$\eta_{jn} = B_nF_{jn}X_n\beta_0$ and $\eta_{kn} = \bar{G}_{kn}B_nX_n\beta_0$.

By Lemma A.3 and Assumptions 6.3-6.6, we have $\frac{1}{n}[T_{jn}(\delta) - \text{E}(T_{jn}(\delta))] = o_p(1)$. Now, as $\mathcal{M}_n$ in S_{jn} is a projection matrix, by Lemma A.2 and similar arguments as for $T_{jn}(\delta)$ lead to $\frac{1}{n}[S_{jn}(\delta) - \text{E}(S_{jn}(\delta))] = o_p(1)$. Thus, $\frac{1}{n}\{R_{jn}(\delta) - \text{E}[R_{jn}(\delta)]\} = o_p(1)$. We can reach a similar conclusion by following the same line of arguments for the second set of equations of the system of estimating equations.

Condition (b): Once again we prove the condition for the component of the adjusted score with respect to λ noting that the proof for the component with respect to

ρ follows in a similar manner. First, by design, we have $\mathrm{E}[R_{jn}(\delta_0)] = 0$ and

$$\begin{aligned}\mathrm{E}[R_{jn}(\delta)] &= \beta_0' X_n' A_n'^{-1} A_n'(\lambda) B_n'(\rho) \mathcal{M}_n(\rho) \bar{B}_n^{\circ}(\delta) B_n(\rho) A_n(\lambda) A_n^{-1} X_n \beta_0 \\ &\quad + \sigma_0^2 \mathrm{tr}(H_n B_n'^{-1} A_n'^{-1} A_n'(\lambda) B_n'(\rho) \mathcal{M}_n(\rho) \bar{B}_n^{\circ}(\delta) B_n(\rho) A_n(\lambda) A_n^{-1} B_n^{-1}).\end{aligned}$$

By Assumption 6.6* and Lemma A.2, $\mathrm{E}[R_{jn}(\lambda)] \neq 0$, for any $\lambda \neq \lambda_0$. Hence, the conditions of Theorem 5.9 of van der Vaart (1998) hold, and thus the consistency of $\tilde{\lambda}_n$.

To prove asymptotic normality, we have, by the mean value theorem,

$$\mathbf{0} = \sqrt{n}\tilde{\psi}_n^*(\tilde{\delta}_n) = \sqrt{n}\tilde{\psi}_n^*(\delta_0) + \tfrac{\partial}{\partial \delta'}\tilde{\psi}_n^*(\bar{\delta}_n)\sqrt{n}(\tilde{\delta}_n - \delta_0), \tag{G-2}$$

where $\bar{\delta}_n$ lies in the segment formed by $\tilde{\delta}_n$ and δ_0. It suffices to show that (i) $\frac{\partial}{\partial \delta'}\tilde{\psi}_n^*(\bar{\delta}_n) - \frac{\partial}{\partial \delta'}\tilde{\psi}_n^*(\delta_0) = o_p(1)$, (ii) $\frac{\partial}{\partial \delta'}\tilde{\psi}_n^*(\delta_0) - \mathrm{E}\big(\frac{\partial}{\partial \delta'}\tilde{\psi}_n^*(\delta_0)\big) = o_p(1)$, and (iii) $\mathrm{E}\big(\frac{\partial}{\partial \delta'}\tilde{\psi}_n^*(\delta_0)\big) \neq 0$ for large enough n. Let,

$$\frac{\partial}{\partial \delta'}\tilde{\psi}_n^*(\delta) = \begin{pmatrix} \tilde{\psi}_{jj',n}^*(\delta) & \tilde{\psi}_{jk,n}^*(\delta) \\ \tilde{\psi}_{kj,n}^*(\delta) & \tilde{\psi}_{kk',n}^*(\delta) \end{pmatrix}, \tag{G-3}$$

where $j, j' = 1, \ldots, p$ and $k, k' = 1, \ldots, q$. Then,

$$\begin{aligned}\tilde{\psi}_{jj',n}^*(\delta) &= \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\dot{\bar{B}}_{jj',n}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}^{\circ}(\delta)\bar{B}_{j'n}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\bar{B}_{j'n}'(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} \\ &\quad + 2\tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}_{j'n}(\delta)Y_n(\delta)\cdot Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}^{\circ}(\delta)Y_n(\delta)}{[Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)]^2},\end{aligned}$$

$$\begin{aligned}\tilde{\psi}_{jk,n}^*(\delta) &= \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\dot{\bar{B}}_{jk,n}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}^{\circ}(\delta)G_{kn}(\rho)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\bar{G}_{kn}'(\rho)\mathcal{M}_n(\rho)\bar{B}_{jn}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} \\ &\quad + \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)G_{kn}(\rho)\bar{B}_{jn}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}(\rho)\bar{B}_{jn}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} \\ &\quad + 2\tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}(\delta)Y_n(\delta)\cdot Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}^{\circ}(\delta)Y_n(\delta)}{[Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)]^2},\end{aligned}$$

$$\begin{aligned}\tilde{\psi}_{kj,n}^*(\delta) &= -\tfrac{Y_n'(\delta)\bar{B}_{jn}'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}^{\circ}(\rho)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}^{\circ}(\rho)\bar{B}_{jn}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} \\ &\quad + 2\tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)Y_n(\delta)\cdot Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}^{\circ}(\rho)Y_n(\delta)}{[Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)]^2} \text{ and}\end{aligned}$$

$$\begin{aligned}\tilde{\psi}_{kk',n}^*(\delta) &= \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\dot{\bar{G}}_{kk',n}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}^{\circ}(\delta)G_{k'n}(\rho)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\bar{G}_{k'n}'(\rho)\mathcal{M}_N(\rho)\bar{G}_{kn}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} \\ &\quad + \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)G_{k'n}(\rho)\bar{G}_{kn}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} - \tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{k'n}(\rho)\bar{G}_{kn}^{\circ}(\delta)Y_n(\delta)}{Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)} \\ &\quad + 2\tfrac{Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{k'n}(\delta)Y_n(\delta)\cdot Y_n'(\delta)\mathcal{M}_n(\rho)\bar{G}_{kn}^{\circ}(\delta)Y_n(\delta)}{[Y_n'(\delta)\mathcal{M}_n(\rho)Y_n(\delta)]^2}, \text{ where}\end{aligned}$$

$\dot{\bar{B}}_{jj',n}^{\circ}(\delta) = \frac{\partial}{\partial \lambda'}\bar{B}_{jn}^{\circ}(\delta)$

$$= B_n(\rho)F_{jn}(\lambda)F_{j'n}(\lambda)B_n^{-1}(\rho) - \text{diag}[\mathcal{M}_n(\rho)]^{-1}\text{diag}[\mathcal{M}_n(\rho)B_n(\rho)F_{jn}(\lambda)F_{j'n}(\lambda)B_n^{-1}(\rho)],$$

$$\dot{\bar{B}}^{\circ}_{jk,n}(\delta) = \tfrac{\partial}{\partial\rho'}\bar{B}^{\circ}_{j,n}(\delta) = \bar{B}_{jn}(\delta)G_{kn}(\rho) - G_{kn}(\rho)\bar{B}_{jn}(\delta)$$
$$+\text{diag}[\mathcal{M}_n^{-1}(\rho)\dot{\mathcal{M}}_{kn}(\rho)\mathcal{M}_n^{-1}(\rho)]\text{diag}[\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)] - \text{diag}[\mathcal{M}_n(\rho)]^{-1}\text{diag}[\dot{\mathcal{M}}_{kn}(\rho)\bar{B}_{jn}(\delta)]$$
$$-\text{diag}[\mathcal{M}_n(\rho)]^{-1}\text{diag}[\mathcal{M}_n(\rho)\bar{B}_{jn}(\delta)G_{kn}(\rho)] + \text{diag}[\mathcal{M}_n(\rho)]^{-1}\text{diag}[\mathcal{M}_n(\rho)G_{kn}(\rho)\bar{B}_{jn}(\delta)],$$

$$\dot{\bar{G}}^{\circ}_{kk',n}(\rho) = \tfrac{\partial}{\partial\rho'}\bar{G}^{\circ}_{k,n}(\rho) = G_{kn}(\rho)\bar{G}_{k'n}(\rho) + G_{kn}(\rho)\dot{\mathcal{M}}_{k'n}(\rho)$$
$$+\text{diag}[\mathcal{M}_n^{-1}(\rho)\dot{\mathcal{M}}_{k'n}(\rho)\mathcal{M}_n^{-1}(\rho)]\text{diag}[\mathcal{M}_n(\rho)\bar{G}_{kn}(\rho)] - \text{diag}[\mathcal{M}_n(\rho)]^{-1}\text{diag}[\dot{\mathcal{M}}_{k'n}(\rho)\bar{G}_{kn}(\rho)]$$
$$-\text{diag}[\mathcal{M}_n(\rho)]^{-1}\text{diag}[\mathcal{M}_n(\rho)G_{kn}(\rho)\bar{G}_{k'n}(\rho)] - \text{diag}[\mathcal{M}_n(\rho)]^{-1}\text{diag}[\mathcal{M}_n(\rho)G_{kn}(\rho)\dot{\mathcal{M}}_{k'n}(\rho)]$$

and $\dot{\mathcal{M}}_{kn}(\rho) = \mathcal{M}_n(\rho)G_{kn}(\rho)\mathcal{P}_n(\rho) + \mathcal{P}_n(\rho)G'_{kn}(\rho)\mathcal{M}_n(\rho)$.

Condition (i)**:** First note that the common term in the denominator of the components in $\frac{\partial}{\partial\delta'}\tilde{\psi}^*_n(\bar{\delta}_n)$ can be written as $Y'_n(\delta)\mathcal{M}_n(\rho)Y_n(\delta) = n\hat{\sigma}^2_n(\bar{\delta}_n)$, where $\hat{\sigma}^2_n(\bar{\delta}_n) = \hat{\sigma}^2_n(\delta_0) + o_p(1)$. By Assumptions 6.4 and 6.5 and continuous mapping theorem, we have, $\bar{B}^{\circ}_{jn}(\bar{\delta}_n) = \bar{B}^{\circ}_{jn}(\delta_0) + o_p(1)$, $\dot{\bar{B}}^{\circ}_n(\bar{\delta}_{jn}) = \dot{\bar{B}}^{\circ}_{jn}(\delta_0) + o_p(1)$, $\bar{G}^{\circ}_{kn}(\bar{\delta}_n) = \bar{G}^{\circ}_{kn}(\delta_0) + o_p(1)$ and $\dot{\bar{G}}^{\circ}_{kn}(\bar{\delta}_n) = \dot{\bar{G}}^{\circ}_{kn}(\delta_0) + o_p(1)$. Then, using a Taylor expansion, terms of the sort $T_{1n}(\bar{\delta}) = \frac{1}{n}Y'_n(\bar{\delta})\bar{B}'_{j'n}(\bar{\delta})\mathcal{M}_n(\bar{\rho})\bar{B}^{\circ}_{jn}(\bar{\delta})Y_n(\bar{\delta})$ can be written as, $T_{1n}(\delta_0) + (\bar{\delta} - \delta_0)'\frac{\partial}{\partial\delta}T_{1n}(\delta_0)$, where together with the continuous mapping theorem, Lemma A.2 and Assumptions 6.3-6.6 and some tedious algebra, we have $T_{1n}(\bar{\delta}) = T_{1n}(\delta_0) + o_p(1)$.

Condition (ii)**:** The result follows from a direct application of Lemmas A.2 and A.3 using Assumptions 6.3-6.6. See proof of Theorem 5.2 for further details.

Condition (iii)**:** By Assumptions 6.3-6.6 and Lemmas A.2 and A.3, it is easy to see that Φ_n is non-singular for large enough n, and thus $\text{E}\big(\frac{\partial}{\partial\delta'}\tilde{\psi}^*_n(\delta_0)\big)$ is non-singular for large enough n.

Proof of Theorem 6.3: Recall $\tilde{\beta}_n = (X'_n\tilde{B}'_n\tilde{B}_nX_n)^{-1}X'_n\tilde{B}'_n\tilde{B}_n\tilde{A}_nY_n$. Then by a Taylor expansion, we have,

$$\sqrt{n}(\tilde{\beta}_n - \beta_0) = (\tfrac{1}{n}X'_nB'_nB_nX_n)^{-1}\tfrac{1}{\sqrt{n}}X'_nB'_n\epsilon_n - \tfrac{\partial}{\partial\delta}\hat{\beta}_n\Big|_{\delta=\delta_0}\sqrt{n}(\tilde{\delta}_n - \delta_0) + O_p\big(\tfrac{1}{\sqrt{n}}\big), \quad \text{(G-4)}$$

where $\frac{\partial}{\partial\delta}\hat{\beta}_n\Big|_{\delta=\delta_0} = \begin{cases} \mathcal{Q}_n\eta_{jn} + o_p(1), \\ \mathcal{Q}_nG^s_{kn}B_n(I_n - X_n\mathcal{Q}_nB_n)X_n\beta_0 + o_p(1), \end{cases}$

for $\mathcal{Q}_n = (X'_nB'_nB_nX_n)^{-1}X'_nB'_n$, $\eta_{jn} = B_nF_{jn}X_n\beta_0$, and $G^s_{kn} = G'_{kn} + G_{kn}$. By Assumptions 6.3-6.5, the asymptotic order of the second component of $\frac{\partial}{\partial\delta}\hat{\beta}_n\Big|_{\delta=\delta_0}$ is $o_p(1)$. This shows that each component of $\sqrt{n}(\tilde{\beta}_n - \beta_0)$ is a linear-quadratic form in ϵ_n. Thus, Cramèr-Wold device and the CLT for linear-quadratic form of Kelejian and Prucha (2001) lead to the asymptotic normality of $\sqrt{n}(\tilde{\beta}_n - \beta_0)$. The asymptotic mean of $\sqrt{n}(\tilde{\beta}_n - \beta_0)$ is zero and the first-order variance of it can be found using (G-4):

$$\begin{aligned}\tau^2(\tilde{\beta}_n) &= \mathcal{Q}_n \mathrm{Var}(\epsilon_n)\mathcal{Q}_n' + \mathcal{Q}_n\eta_{jn}\tau^2(\tilde{\delta}_n)\eta_{jn}'\mathcal{Q}_n' - 2\sigma_0^{-2}\mathcal{Q}_n\mathrm{Cov}(\epsilon_n, R_n(\epsilon_n))\Phi_n^{-1}\eta_{jn}'\mathcal{Q}_n' \\ &= \mathcal{Q}_n\mathbb{A}_n\mathcal{Q}_n',\end{aligned}$$

where $\mathbb{A}_n = \eta_{jn}\tau_{n,j}^2(\tilde{\delta}_n)\eta_{jn}' + 2\sqrt{n}(\sigma_0^{-2}P_{jn}^d s_n + H_n c_{jn},\ \sigma_0^{-2}Q_{kn}^d s_n + H_n c_{kn})\Phi^{-1}(\eta_{jn}, 0_n)' + n\sigma_0^2 H_n$, $s_n = \mathrm{E}(\epsilon_n^3)$ and $R_n(\epsilon_n)$ is as defined in (6.18).

The limiting distribution of $\sqrt{n}(\tilde{\sigma}_n^2 - \sigma_0^2)$ can be found in a similar manner from

$$\begin{aligned}\sqrt{n}(\tilde{\sigma}_n^2 - \sigma_0^2) &= \sqrt{n}[\tfrac{1}{n}Y_n'(\tilde{\delta}_n)\mathcal{M}_n(\tilde{\rho}_n)Y_n(\tilde{\delta}_n) - \sigma_0^2] \\ &= \tfrac{1}{\sqrt{n}}(\epsilon_n'\epsilon_n - n\sigma_0^2) \\ &\quad + 2\sqrt{n}(\tilde{\delta}_n - \delta_0)'\tfrac{1}{n}\tfrac{\partial}{\partial\delta}Y_n'(\tilde{\delta}_n)\mathcal{M}_n(\tilde{\rho}_n)Y_n(\tilde{\delta}_n)\Big|_{\delta=\delta_0} + o_p(1),\end{aligned}$$

which has a limiting mean of zero and first-order variance that can be easily derived but not needed in light of Footnote 4. Thus by the consistency of $\tilde{\delta}_n$ from Theorem 6.2, we have the consistency of $\tilde{\sigma}_n^2$, in particular, it is $\sqrt{n}$-consistent.

Proof of Theorem 6.4: This follows immediately by the Central Limit Theorem for quadratic forms by Kelejian and Prucha (2001) and the poof of Theorem 1 of Baltagi and Yang (2013b). The details of the proof are similar to the proof of Theorem 6.8 given below but extended to include higher order spatial dependence.

The consistency of $\tilde{\tau}_n^2(\tilde{\beta}_n)$ follows that of $\tilde{\tau}_n^2(\tilde{\lambda}_n)$ and the consistency of $\tilde{\theta}_n$.

G.2 Proofs of Results for Fixed Effects SPD Model

Proof of Theorem 6.5: The concentrated expected quasi log-likelihood function is, $\bar{\ell}_N^c(\delta) = \max_{\beta,\sigma^2}\mathrm{E}[\ell_N(\theta)] = -\frac{N}{2}\ln(2\pi+1) + \ln|\mathbf{A}_{1N}(\lambda)| + \ln|\mathbf{A}_{2N}(\rho)| - \frac{N}{2}\ln(\bar{\sigma}_N^2(\delta))$, where the $\bar{\sigma}^2_N(\delta) = \frac{1}{N}(\mathbf{X}_N\beta_0)'\mathbf{A}^{-1'}_{1N}\mathbf{A}'_{1N}(\lambda)\mathbf{A}'_{2N}(\rho)\mathbf{M}_N(\rho)\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{X}_N\beta_0 + \frac{\sigma_0^2}{N}\text{tr}[\mathbf{H}_N\mathbf{A}^{-1'}_{2N}\mathbf{A}^{-1'}_{1N}\mathbf{A}'_{1N}(\lambda)\mathbf{A}'_{2N}(\rho)\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N}]$. To apply Theorem 5.9 of van der Vaart (1998) for consistency, we need (a) the identification uniqueness condition: $\inf_{\delta:d(\delta,\delta_0)\geq\epsilon}\frac{1}{N}||\bar{\ell}^c_N(\delta)|| > 0 = ||\bar{\ell}^c_N(\delta_0)||$ for any $\epsilon > 0$ and a distance measure $d(\delta,\delta_0)$, and (b) the uniform convergence: $\sup\frac{1}{N}||\ell^c_N(\delta) - \bar{\ell}^c_N(\delta)|| \xrightarrow{p} 0$ uniformly in $\delta \in \Delta$.

Condition (a)**:** Using the fact that $\bar{\sigma}^2_N(\delta_0) = \sigma_0^2$, we have,

$$\begin{aligned}
& \lim_{N\to\infty}\frac{1}{N}\left[\bar{\ell}_N(\theta) - \bar{\ell}_N(\theta_0)\right] \\
= & \lim_{N\to\infty}\left[\tfrac{1}{N}[(\log|\mathbf{A}_{1N}(\lambda)| - \log|\mathbf{A}_{1N}|) + (\log|\mathbf{A}_{2N}(\lambda)| - \log|\mathbf{A}_{2N}|)]\right. \\
& +\frac{1}{2N}(\log|\bar{\sigma}^{-2}I_N| - \log|\sigma_0^{-2}|I_N) \\
\neq & \ 0 \text{ for } \theta \neq \theta_0, \text{ by Assumption 6.12.}
\end{aligned}$$

Next, note $p_N(\delta_0) = \exp[\ell_N(\delta_0)]$ is the *quasi* joint pdf of $\mathbf{V}_N$, which is $\mathcal{N}(0,\sigma^2 I_N)$. Let $p^0_N(\delta_0)$ be the *true* joint pdf of $\mathbf{V}_N \sim (0,\sigma^2\mathbf{H}_N)$ and E^q denote the expectation w.r.t. $p_N(\delta_0)$, to differentiate from the usual notation E that corresponds to $p^0_N(\delta_0)$.

Now consider $\mathbf{V}_N(\zeta) = \mathbf{A}_{2N}(\rho)[\mathbf{A}_{1N}(\lambda)\mathbf{Y}_N - \mathbf{X}_N\beta] = \mathbf{B}_N(\delta)\mathbf{V}_N + \mathbf{b}_N(\zeta)$, where $\mathbf{B}_N(\delta) = \mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N}$ and $\mathbf{b}_N(\zeta) = \mathbf{A}_{2N}(\rho)[\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{X}_N\beta_0 - \mathbf{X}_N\beta]$. Then, for $\ell_N(\theta)$ given in (6.23), we have

$$\begin{aligned}
\text{E}^q[\ell_N(\theta_0)] &= -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{A}_{1N}| + \ln|\mathbf{A}_{2N}| - \tfrac{N}{2}, \\
\text{E}[\ell_N(\theta_0)] &= -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{A}_{1N}| + \ln|\mathbf{A}_{2N}| - \tfrac{N}{2}, \text{ as } \tfrac{1}{N}\text{tr}(\mathbf{H}_N) = 1 \\
\text{E}^q[\ell_N(\theta)] &= -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{A}_{1N}(\lambda)| + \ln|\mathbf{A}_{2N}(\rho)| \\
&\quad -\tfrac{1}{2\sigma^2}[\sigma_0^2\text{tr}(\mathbf{B}'_N(\delta)\mathbf{B}_N(\delta)) + \mathbf{b}'_N(\zeta)\mathbf{b}_N(\zeta)], \\
\text{E}[\ell_N(\theta)] &= -\tfrac{N}{2}\ln(2\pi\sigma^2) + \ln|\mathbf{A}_{1N}(\lambda)| + \ln|\mathbf{A}_{2N}(\rho)| \\
&\quad -\tfrac{1}{2\sigma^2}[\sigma_0^2\text{tr}(\mathbf{H}_N\mathbf{B}'_N(\delta)\mathbf{B}_N(\delta)) + \mathbf{b}'_N(\zeta)\mathbf{b}_N(\zeta)],
\end{aligned}$$

where we have used the identities, $\mathbf{B}_N(\delta_0) = I_N$ and $\mathbf{b}_N(\zeta_0) = 0$. Then, $\text{E}[\ell_N(\theta)] - \text{E}^q[\ell_N(\theta)] = o(1)$, where the last equality holds by assumptions $\text{Cov}(g_{rn},h_n) = o(1)$ and $\text{Cov}(q_{rn},h_n) = o(1)$ for $r = 1,2$. By Jensen's inequality, $0 = \log\text{E}^q\left(\frac{p_N(\theta)}{p_N(\theta_0)}\right) \geq$

$\mathrm{E}^q\big[\log\big(\frac{p_N(\theta)}{p_N(\theta_0)}\big)\big]$, and combining with the above results, we have $\mathrm{E}\big[\log\big(\frac{p_N(\theta)}{p_N(\theta_0)}\big)\big] \le 0$ or $\mathrm{E}[\log p_N(\theta)] \le \mathrm{E}[\log p_N(\theta_0)]$, for large enough N. Thus, $\bar{\ell}_N(\delta) = \max_{\beta,\sigma^2} \mathrm{E}[\log p_N(\theta)] \le \max_{\beta,\sigma^2} \mathrm{E}[\log p_N(\theta_0)] = \bar{\ell}_N(\delta_0)$, for $\theta \neq \theta_0$, and N large enough.

Condition (b)**:** Note, $\frac{1}{N}[\ell^c_N(\delta) - \bar{\ell}^c_N(\delta)] = -\frac{1}{2}[\log(\hat{\sigma}^2_N(\delta)) - \log(\bar{\sigma}^2_N(\delta))]$. By the mean value theorem, $\log(\hat{\sigma}^2_N(\delta)) - \log(\bar{\sigma}^2_N(\delta)) = \frac{1}{\dot{\sigma}^2_N(\delta)}[\hat{\sigma}^2_N(\delta) - \bar{\sigma}^2_N(\delta)]$, where $\dot{\sigma}^2_N(\delta)$ lies between $\hat{\sigma}^2_N(\delta)$ and $\bar{\sigma}^2_N(\delta)$. By $\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{Y}_N = \mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N}(\mathbf{A}_{2N}\mathbf{X}_N\beta + \mathbf{V}_N)$,

$$\hat{\sigma}^2_n(\delta) = \bar{\sigma}^2_N(\delta) + 2T_{1N}(\delta) + T_{2N}(\delta) - T_{3N}(\delta), \text{ where} \tag{G-5}$$

$T_{1N}(\delta) = \frac{1}{N}\mathbf{V}'_N\mathbf{A}^{-1'}_{2N}\mathbf{A}^{-1'}_{1N}\mathbf{A}'_{1N}(\lambda)\mathbf{A}'_{2N}(\rho)\mathbf{M}_N(\rho)\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{X}_N\beta_0$,
$T_{2N}(\delta) = \frac{1}{N}\mathbf{V}'_N\mathbf{A}^{-1'}_{2N}\mathbf{A}^{-1'}_{1N}\mathbf{A}'_{1N}(\lambda)\mathbf{A}'_{2N}(\rho)\mathbf{M}_N(\rho)\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N}\mathbf{V}_N$ and
$T_{3N}(\delta) = \frac{\sigma^2_0}{N}\mathrm{tr}[\mathbf{H}_N\mathbf{A}^{-1'}_{2N}\mathbf{A}^{-1'}_{1N}\mathbf{A}'_{1N}(\lambda)\mathbf{A}'_{2N}(\rho)\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N}]$.

Using $\mathbf{A}_{1N}(\lambda) = \mathbf{A}_{1N} + (\lambda_0 - \lambda)\mathbf{W}_{1N}$ and $\mathbf{A}_{2N}(\rho) = \mathbf{A}_{2N} + (\rho_0 - \rho)\mathbf{W}_{2N}$, we have, $\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N} = I_N + (\rho_0 - \rho)\mathbf{G}_{2N} + (\lambda_0 - \lambda)\bar{\mathbf{G}}_{1N} + (\lambda_0 - \lambda)(\rho_0 - \rho)\mathbf{G}_{2N}\bar{\mathbf{G}}_{1N}$. Combined with Assumptions 6.9-6.12, we have, $T_{1N}(\delta) = o_p(1)$ uniformly. Further, by Assumption 6.14, $T_{2N}(\delta) = \frac{1}{N}\mathbf{V}'_N\mathbf{A}^{-1'}_{2N}\mathbf{A}^{-1'}_{1N}\mathbf{A}'_{1N}(\lambda)\mathbf{A}'_{2N}(\rho)\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}^{-1}_{1N}\mathbf{A}^{-1}_{2N}\mathbf{V}_N + o_p(1)$. Now, Lemmas A.1 - A.3 imply, $\frac{1}{N^2}\mathrm{Var}(T_{2N}(\delta)) = o(1)$. Then, together with Chebyshev inequality, $T_{2N}(\delta) - T_{3N}(\delta) = o_p(1)$, uniformly for $\delta \in \Delta$.

It left to show that, $\sigma^2_N(\delta)$ (defined in Assumption 6.12 and the main part of $\bar{\sigma}^2_N(\delta)$) is uniformly bounded away from zero. Suppose $\sigma^2_N(\delta)$ is not uniformly bounded away from zero. Then $\exists\{\delta_n\} \subset \Delta$ such that $\sigma^2_N(\delta) \to 0$. Consider the model with $\beta_0 = 0$, which has the quasi log-likelihood, $\ell^*_N(\theta) = -\frac{N}{2}\log(2\pi\sigma^2) + \log|\mathbf{A}_{1N}(\lambda)| + \log|\mathbf{A}_{2N}(\rho)| - \frac{1}{2\sigma^2}\mathbf{Y}'_N(\delta)\mathbf{Y}_N(\delta)$ and $\bar{\ell}^*_N(\delta) = \max_{\sigma^2}\mathrm{E}[\ell^*_N(\theta)]$. By Jensen's inequality, we have $\bar{\ell}^*_N(\delta) \le \max_{\sigma^2}\mathrm{E}[\ell^*_N(\theta_0)] = \bar{\ell}^*_N(\delta_0)$. Then together with Lemma A.2, we have $\frac{1}{N}[\bar{\ell}^*_N(\delta) - \bar{\ell}^*_N(\delta_0)] \le 0$, and $-\frac{N}{2}\log(\sigma^2_N(\delta)) \le -\frac{N}{2}\log(\sigma^2_0) + \frac{1}{N}(\log|\mathbf{A}_{1N}(\lambda_0)| - \log|\mathbf{A}_{1N}(\lambda)|) + \frac{1}{N}(\log|\mathbf{A}_{2N}(\rho_0)| - \log|\mathbf{A}_{2N}(\rho)|) = O(1)$. That is, $-\frac{N}{2}\log(\sigma^2_N(\delta))$ is bounded from above which is a contradiction. Since $\sigma^2_N(\delta)$ is bounded away from zero uniformly, we have that $\bar{\sigma}^2_N(\delta)$ is also bounded away from zero since it is the sum of a quadratic term and $\sigma^2_N(\delta)$. Further by (G-5), $\hat{\sigma}^2_n(\delta)$ is also bounded away from zero.

Collecting these we have, $\sup_{\delta\in\Delta}\frac{1}{N}|[\ell_N^c(\delta)-\bar{\ell}_N^c(\delta)]|=o_p(1)$, completing the proof.

Proof of Theorem 6.6: Let $\mathrm{E}(\tilde{\psi}_N^*(\delta))=\bar{\psi}^*(\delta)$. By Theorem 5.9 of van der Vaart (1998), the proof of consistency of $\tilde{\delta}_N$ requires (a) Convergence: $\sup_{\delta\in\Delta}||\tilde{\psi}_N^*(\delta)-\bar{\psi}^*(\delta)||=o_p(1)$ and (b) Identification uniqueness: for $\epsilon>0$, $\inf_{\delta:d(\delta,\delta_0)\geq\epsilon}||\bar{\psi}^*(\delta)||>0=||\bar{\psi}^*(\delta_0)||$. The proof of Theorem 6.1 implies that $\hat{\sigma}_N^2(\delta)$ is bounded away from 0 with probability one for large enough N. Thus, the ACQS estimator $\tilde{\delta}_N=\arg\{\tilde{\psi}_N^*(\delta)=0\}$ is equivalently defined as $\tilde{\delta}_N=\arg\{\mathbf{Y}_N'(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{rN}^{\circ}(\delta)\mathbf{Y}_N(\delta)=0,\ r=1,2\}$, suggesting that we can work purely with the numerator of $\tilde{\psi}_N^*(\delta)$ to establish consistency. For $r=1,2$, write the numerators of the components of the adjusted score as,

$$
\begin{aligned}
R_{rN}(\delta) &= \mathbf{Y}_N'(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{rN}(\delta)\mathbf{Y}_N(\delta)\\
&\quad -\mathbf{Y}_N'(\delta)\mathbf{M}_N(\rho)\mathrm{diag}[\mathbf{M}_N(\rho)]^{-1}\mathrm{diag}[\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{rN}(\delta)]\mathbf{Y}_N(\delta)\\
&\equiv T_{rN}(\delta)-S_{rN}(\delta).
\end{aligned}
$$

Condition (a): Using the relations $\mathbf{Y}_N(\delta)=\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}_{1N}^{-1}\mathbf{A}_{2N}^{-1}(\mathbf{A}_{2N}\mathbf{X}_N\beta+\mathbf{V}_N)$, $\mathbb{B}_{1N}(\delta)=\mathbf{A}_{2N}(\rho)\mathbf{A}_{1N}(\lambda)\mathbf{A}_{1N}^{-1}\mathbf{A}_{2N}^{-1}=I_N+(\rho_0-\rho)\mathbf{G}_{2N}+(\lambda_0-\lambda)\bar{\mathbf{G}}_{1N}+(\lambda_0-\lambda)(\rho_0-\rho)\mathbf{G}_{2N}\bar{\mathbf{G}}_{1N}$, $\mathbb{B}_{2N}(\rho)=I_N+(\rho_0-\rho)\mathbf{G}_{2N}$ and the fact that the projection matrix $\mathbf{M}_N(\rho)$ is uniformly bounded, we have,

$$
\begin{aligned}
T_{1N}(\delta) &= \mathbf{Y}_N'(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{1N}(\delta)\mathbf{Y}_N(\delta)\\
&= (\mathbf{A}_{2N}\mathbf{X}_N\beta_0+\mathbf{V}_N)'\mathbb{B}_{1N}'(\delta)\mathbf{M}_N\mathbb{B}_{2N}(\rho)\bar{\mathbf{G}}_{1N}\bar{\mathbf{G}}_{1N}(\mathbf{A}_{2N}\mathbf{X}_N\beta_0+\mathbf{V}_N)+o_p(1)\\
T_{2N}(\delta) &= \mathbf{Y}_N'(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{2N}(\delta)\mathbf{Y}_N(\delta)\\
&= (\mathbf{A}_{2N}\mathbf{X}_N\beta+\mathbf{V}_N)'\mathbb{B}_{1N}'(\delta)\mathbf{M}_N\mathbf{G}_{2N}\mathbf{M}_N\mathbb{B}_{1N}(\delta)(\mathbf{A}_{2N}\mathbf{X}_N\beta+\mathbf{V}_N)+o_p(1).
\end{aligned}
$$

Then by Lemma A.3 and Assumptions 6.11 and 6.12, $\frac{1}{N}[T_{rN}(\delta)-\mathrm{E}(T_{rN}(\delta))]=o_p(1)$ for $r=1,2$. Now, as $\mathbf{M}_N(\rho)$ in $S_{rN}(\delta)$ is a projection matrix, by Lemma A.2 and similar arguments as for $T_{rN}(\delta)$ lead to $\frac{1}{N}[S_{rN}(\delta)-\mathrm{E}(S_{rN}(\delta))]=o_p(1)$. Thus, $\frac{1}{N}[R_{rN}(\delta)-\mathrm{E}[R_{rN}(\delta)]]=o_p(1)$.

Condition (b): First, we have $\mathrm{E}[R_{rN}(\delta_0)]=0$. By Assumption 6.12* and Lemma

A.2, $\mathrm{E}[R_{rN}(\delta)] \neq 0$, for any $\delta \neq \delta_0$. It follows that the conditions of Theorem 5.9 of van der Vaart (1998) hold, and thus the consistency of $\tilde{\delta}_N$ follows.

To prove asymptotic normality, we have, by the mean value theorem,

$$0 = \sqrt{N}\tilde{\psi}_N^*(\tilde{\delta}_N) = \sqrt{N}\tilde{\psi}_N^*(\delta_0) + \tfrac{d}{d\delta'}\tilde{\psi}_N^*(\bar{\delta}_N)\sqrt{N}(\tilde{\delta}_N - \delta_0), \tag{G-6}$$

where $\bar{\delta}_N$ lies between $\tilde{\delta}_N$ and δ_0. It suffices to show that (i) $\frac{d}{d\delta'}\tilde{\psi}_N^*(\bar{\delta}_N) - \frac{d}{d\delta'}\tilde{\psi}_N^*(\delta_0) = o_p(1)$, (ii) $\frac{d}{d\delta'}\tilde{\psi}_N^*(\delta_0) - \mathrm{E}\big(\frac{d}{d\delta'}\tilde{\psi}_N^*(\delta_0)\big) = o_p(1)$, and (iii) $\mathrm{E}\big(\frac{d}{d\delta'}\tilde{\psi}_N^*(\delta_0)\big)$ is non-singular for large enough N. Let,

$$\frac{d}{d\delta'}\tilde{\psi}_N^*(\delta) = \begin{pmatrix} \tilde{\psi}_{11,N}^*(\delta) & \tilde{\psi}_{12,N}^*(\delta) \\ \tilde{\psi}_{21,N}^*(\delta) & \tilde{\psi}_{22,N}^*(\delta) \end{pmatrix}, \text{ where} \tag{G-7}$$

$\tilde{\psi}_{11,N}^*(\delta) = \frac{1}{\hat{\sigma}_N^2(\delta)}\mathbf{Y}_N'(\delta)[\mathbf{M}_N(\rho)\dot{\bar{\mathbf{G}}}_{11,N}^\circ(\delta) - \mathbf{B}_{1N}(\delta)\bar{\mathbf{G}}_{1N}(\delta) - \bar{\mathbf{G}}_{1N}'(\delta)\mathbf{B}_{1N}(\delta)$
$\quad +2\tilde{\psi}_{1,N}^*(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{1N}(\delta)]\mathbf{Y}_N(\delta)$

$\tilde{\psi}_{12,N}^*(\delta) = \frac{1}{\hat{\sigma}_N^2(\delta)}\mathbf{Y}_N'(\delta)[\mathbf{M}_N(\rho)\dot{\bar{\mathbf{G}}}_{12,N}^\circ(\delta) + \mathbf{M}_N(\rho)\mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{1N}^\circ(\delta) - \mathbf{B}_{1N}(\delta)\mathbf{G}_{2N}(\rho)$
$\quad -\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{2N}(\rho)\bar{\mathbf{G}}_{1N}^\circ(\delta) - \bar{\mathbf{G}}_{2N}'(\rho)\mathbf{B}_{1N}(\delta) + 2\tilde{\psi}_{1,N}^*(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{2N}(\delta)]\mathbf{Y}_N(\delta)$

$\tilde{\psi}_{21,N}^*(\delta) = \frac{1}{\hat{\sigma}_N^2(\delta)}\mathbf{Y}_N'(\delta)[2\tilde{\psi}_{2,N}^*(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{1N}(\delta) - \bar{\mathbf{G}}_{1N}'(\delta)\mathbf{B}_{2N}(\delta)$
$\quad -\mathbf{B}_{2N}(\delta)\bar{\mathbf{G}}_{1N}(\delta)]\mathbf{Y}_N(\delta)$

$\tilde{\psi}_{22,N}^*(\delta) = \frac{1}{\hat{\sigma}_N^2(\delta)}\mathbf{Y}_N'(\delta)[\mathbf{M}_N(\rho)\dot{\bar{\mathbf{G}}}_{22,N}^\circ(\delta) + \mathbf{M}_N(\rho)\mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{2N}^\circ(\delta) - \mathbf{B}_{2N}(\delta)\mathbf{G}_{2N}(\rho)$
$\quad -\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{2N}(\rho)\bar{\mathbf{G}}_{2N}^\circ(\delta) - \bar{\mathbf{G}}_{2N}'(\rho)\mathbf{B}_{2N}(\delta) + 2\tilde{\psi}_{2,N}^*(\delta)\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{2N}(\delta)]\mathbf{Y}_N(\delta)$ where

$\mathbf{B}_{rN}(\delta) = \mathbf{M}_N(\rho)\bar{\mathbf{G}}_{rN}^\circ(\delta)$ for $r = 1, 2$,

$\dot{\bar{\mathbf{G}}}_{11,N}^\circ(\delta) = \frac{\partial}{\partial\lambda}\bar{\mathbf{G}}_{1N}^\circ(\delta) = \bar{\mathbf{G}}_{1N}^2(\delta) - \mathrm{diag}[\mathbf{M}_N(\rho)]^{-1}\mathrm{diag}[\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{1N}^2(\delta)]$,

$\dot{\bar{\mathbf{G}}}_{12,N}^\circ(\delta) = \frac{\partial}{\partial\rho}\bar{\mathbf{G}}_{1N}^\circ(\delta) = \bar{\mathbf{G}}_{1N}(\delta)\mathbf{G}_{2N}(\rho) - \mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{1N}(\delta)$
$\quad +\mathrm{diag}[\mathbf{M}_N(\rho)]^{-1}\mathrm{diag}[\mathbf{M}_N(\rho)\mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{1N}(\delta) - \mathbf{M}_N(\rho)\bar{\mathbf{G}}_{1N}(\delta)\mathbf{G}_{2N}(\rho) - \dot{\mathbf{M}}_N(\rho)\bar{\mathbf{G}}_{1N}(\delta)]$
$\quad +\mathrm{diag}[\mathbf{M}_N(\rho)]^{-2}\mathrm{diag}[\dot{\mathbf{M}}_N(\rho)]\mathrm{diag}[\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{1N}(\delta)]$,

$\dot{\bar{\mathbf{G}}}_{22,N}^\circ(\delta) = \frac{\partial}{\partial\rho}\bar{\mathbf{G}}_{2N}^\circ(\delta) = \mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{2N}(\rho) + \mathbf{G}_{2N}(\rho)\dot{\mathbf{M}}_N(\rho)$
$\quad -\mathrm{diag}[\mathbf{M}_N(\rho)]^{-1}\mathrm{diag}[\mathbf{M}_N(\rho)\mathbf{G}_{2N}(\rho)\bar{\mathbf{G}}_{2N}(\rho) + \mathbf{M}_N(\rho)\mathbf{G}_{2N}(\rho)\dot{\mathbf{M}}_N(\rho) + \dot{\mathbf{M}}_N(\rho)\bar{\mathbf{G}}_{2N}(\rho)]$
$\quad +\mathrm{diag}[\mathbf{M}_N(\rho)]^{-2}\mathrm{diag}[\dot{\mathbf{M}}_N(\rho)]\mathrm{diag}[\mathbf{M}_N(\rho)\bar{\mathbf{G}}_{2N}(\rho)]$ and

$\dot{\mathbf{M}}_N(\rho) = \mathbf{M}_N(\rho)\mathbf{G}_{2N}(\rho)\mathbf{P}_N(\rho) + \mathbf{P}_N(\rho)\mathbf{G}_{2N}'(\rho)\mathbf{M}_N(\rho)$.

Condition (i)**:** First note that the common term in the denominator of the components in $\frac{\partial}{\partial \delta'}\tilde{\psi}^*_N(\bar{\delta}_N)$ can be written as $\mathbf{Y}'_N(\delta)\mathbf{M}_N(\rho)\mathbf{Y}_N(\delta) = N\hat{\sigma}^2_N(\bar{\delta}_N)$, where $\hat{\sigma}^2_N(\bar{\delta}_N) = \hat{\sigma}^2_N(\delta_0) + o_p(1)$. By Assumptions 6.10 and 6.11 and continuous mapping theorem, $\bar{\mathbf{G}}^{\circ}_{rN}(\bar{\delta}_N) = \bar{\mathbf{G}}^{\circ}_{rN}(\delta_0) + o_p(1)$ and $\dot{\bar{\mathbf{G}}}^{\circ}_{rN}(\bar{\delta}_N) = \dot{\bar{\mathbf{G}}}^{\circ}_{rN}(\delta_0) + o_p(1)$ for $r = 1, 2$. Then, using a Taylor expansion, terms of the sort $T_{1N}(\bar{\delta}) = \frac{1}{N}\mathbf{Y}'_n(\bar{\delta})\bar{\mathbf{G}}'_{2N}(\bar{\delta})\mathbf{M}_N(\bar{\rho})\bar{\mathbf{G}}^{\circ}_{1N}(\bar{\delta})\mathbf{Y}_N(\bar{\delta})$ can be written as, $T_{1N}(\delta_0) + (\bar{\delta} - \delta_0)'\frac{\partial}{\partial \delta}T_{1N}(\delta_0)$, where together with the continuous mapping theorem, Lemma A.2 and Assumptions 6.9-6.12 and some tedious algebra, we have $T_{1N}(\bar{\delta}) = T_{1N}(\delta_0) + o_p(1)$. These results give, $\frac{d}{d\delta'}\tilde{\psi}^*_N(\bar{\delta}_N) - \frac{d}{d\delta'}\tilde{\psi}^*_N(\delta_0) = o_p(1)$.

Condition (ii)**:** The result follows from a direct application of Lemmas A.2 and A.3 using Assumptions 6.9-6.12. See proof of Theorem 5.2 for further details.

Condition (iii)**:** By Assumptions 6.9-6.12 Lemmas A.2 and A.3, Φ_N is non-singular for large enough N, and thus $\mathrm{E}\big(\frac{d}{d\delta'}\tilde{\psi}^*_N(\delta_0)\big)$ is non-singular for large enough N.

Proof of Theorem 6.7: By a Taylor expansion, we have,

$$\begin{aligned}\sqrt{N}(\tilde{\beta}_N - \beta_0) \;&=\; (\tfrac{1}{N}\mathbf{X}'_N\mathbf{A}'_{2N}\mathbf{A}_{2N}\mathbf{X}_N)^{-1}\tfrac{1}{\sqrt{N}}\mathbf{X}'_N\mathbf{A}'_{2N}\mathbf{V}_N \\ &\quad - \tfrac{\partial}{\partial\delta}\hat{\beta}_N\Big|_{\delta=\delta_0}\sqrt{N}(\tilde{\delta}_N - \delta_0) + O_p\big(\tfrac{1}{\sqrt{N}}\big),\end{aligned} \tag{G-8}$$

where $\frac{\partial}{\partial\delta}\hat{\beta}_N\Big|_{\delta=\delta_0} = \begin{cases}\mathbf{R}_N\eta_{1N} + o_p(1), \\ \mathbf{R}_N\mathbf{G}^s_{2N}\mathbf{A}_{2N}(I_N - \mathbf{X}_N\mathbf{R}_N\mathbf{A}_{2N})\mathbf{X}_N\beta_0 + o_p(1),\end{cases}$
for $\mathbf{R}_N = (\mathbf{X}'_N\mathbf{A}'_{2N}\mathbf{A}_{2N}\mathbf{X}_N)^{-1}\mathbf{X}'_N\mathbf{A}'_{2N}$, $\eta_{1N} = \mathbf{A}_{2N}\mathbf{G}_{1N}\mathbf{X}_N\beta_0$, and $\mathbf{G}^s_{2N} = \mathbf{G}'_{2N} + \mathbf{G}_{2N}$. By Assumptions 6.9-6.11, the asymptotic order of the second component of $\frac{\partial}{\partial\delta}\hat{\beta}_N\Big|_{\delta=\delta_0}$ is $o_p(1)$. This shows that each component of $\sqrt{N}(\tilde{\beta}_N - \beta_0)$ is a linear-quadratic form in $\mathbf{V}_N$. Thus, Cramèr-Wold device and the CLT for linear quadratic forms given in Lemma A.3 lead to the asymptotic normality of $\sqrt{N}(\tilde{\beta}_N - \beta_0)$. The asymptotic mean of $\sqrt{N}(\tilde{\beta}_N - \beta_0)$ is zero and the first-order variance of it can be found using (G-8):

$$\begin{aligned}\Sigma(\tilde{\beta}_N) \;&=\; \mathbf{R}_N\mathrm{Var}(\mathbf{V}_N)\mathbf{R}'_N + \mathbf{R}_N\eta_{1N}\Sigma(\tilde{\delta}_N)\eta'_{1N}\mathbf{R}'_N \\ &\quad -2\sigma_0^{-2}\mathbf{R}_N\mathrm{Cov}(\mathbf{V}_N, \mathbf{Q}_N(\mathbf{V}_N))\Phi_N^{-1}\eta'_{jn}\mathbf{R}'_N \\ &=\; \mathbf{R}_N\mathbb{A}_N\mathbf{R}'_N, \text{ where}\end{aligned}$$

$\mathbb{A}_N = 2\sqrt{N}(\frac{1}{\sigma_0^2}\mathbf{B}_{1N}^d s_N + H_N c_{1N}, \frac{1}{\sigma_0^2}\mathbf{B}_{2N}^d s_N + H_N c_{2N})\Phi^{-1}(\eta_{1N}, 0_N)' + \eta_{1N}\Sigma_{N,11}(\tilde{\delta}_N)\eta_{1N}' + N\sigma_0^2 H_N$, $s_N = \mathrm{E}(\mathbf{V}_N^3)$ and $Q_N(\mathbf{V}_N)$ is as defined in (6.33).

The limiting distribution of $\sqrt{N}(\tilde{\sigma}_N^2 - \sigma_0^2)$ can be found in a similar manner from

$$\begin{aligned}
\sqrt{N}(\tilde{\sigma}_N^2 - \sigma_0^2) &= \sqrt{N}[\tfrac{1}{N}\mathbf{Y}_N'(\tilde{\delta}_N)\mathbf{M}_N(\tilde{\rho}_N)\mathbf{Y}_N(\tilde{\delta}_N) - \sigma_0^2] \\
&= \tfrac{1}{\sqrt{N}}(\epsilon_N'\epsilon_N - N\sigma_0^2) \\
&\quad + 2\sqrt{N}(\tilde{\delta}_N - \delta_0)'\tfrac{1}{N}\tfrac{\partial}{\partial\delta}\mathbf{Y}_N'(\tilde{\delta}_N)\mathbf{M}_N(\tilde{\rho}_N)\mathbf{Y}_N(\tilde{\delta}_N)\Big|_{\delta=\delta_0} + o_p(1),
\end{aligned}$$

which has a limiting mean of zero and first-order variance that can be easily derived but not needed. Thus by the consistency of $\tilde{\delta}_N$ from Theorem 6.6, we have the consistency of $\tilde{\sigma}_N^2$, in particular, it is $\sqrt{N}$-consistent.

Proof of Theorem 6.8: To prove the consistency of $\tilde{\Sigma}_N(\tilde{\delta}_N)$ as an estimator of $\Sigma_N(\tilde{\delta}_N)$, we need to prove (a) $\tilde{\Phi}_N - \Phi_N = o_p(1)$, and (b) $\tilde{\Sigma}_N(\tilde{\psi}_N^*) - \Sigma_N(\tilde{\psi}_N^*) = o_p(1)$.

For, (a) note $\mathrm{E}(\mathbf{V}_N\mathbf{V}_N') = \sigma^2\mathbf{H}_N$. By Theorem 6.6, $\tilde{\mathbf{H}}_N = \frac{1}{\tilde{\sigma}^2}\mathrm{diag}(\tilde{V}_1^2, \ldots, \tilde{V}_N^2) = \mathbf{H}_N + o_p(1)$. Then following the proof of Theorem 6.6 (the asymptotic normality part) we have $\tilde{\Phi}_N = \Phi_N + o_p(1)$. To prove (b), as $\tilde{\sigma}_N^2 = \sigma_0^2 + o_p(1)$ by Theorem 6.7, it suffices to show that, by the consistency of $\tilde{\theta}_N$ and referring to (6.36),

$$\tfrac{1}{N}\textstyle\sum_{g=1}^N \left(V_g^{*2}\Upsilon_g\Upsilon_g' - \mathrm{Var}(V_g^*\Upsilon_g)\right) = (\Delta_{r,s})_{r,s=1,2} = o_p(1),$$

where $\Upsilon_g = (\zeta_{1N,g} + b_{1N,gg}V_g^* + c_{1N,g},\ \zeta_{2N,g} + b_{2N,gg}V_g^* + c_{2N,g})'$. To do so we use Theorem 19.7 in Davidson (1994) and the weak law for large numbers (WLLN) for martingale difference sequences. First note that when the errors are independent and normal, $\{V_g^*\Upsilon_g\}_{g=1,\ldots,N}$ form a martingale difference sequence since (i) $\mathrm{E}|V_g^*\Upsilon_g| < \infty$ and (ii) $\mathrm{E}(V_g^*\Upsilon_g|\mathcal{F}_{g-1}) = 0$, a.s. where $\mathcal{F}_{g-1}$ is the increasing σ-field generated by the transformed errors $\{V_1^*, \ldots, V_{g-1}^*\}$. For the second condition note, $b_{rN,gg}\mathrm{E}(V_g^{*2}|\mathcal{F}_{g-1}) = 0$ since $b_{rN,gg} = 0$ by design and $\mathrm{E}(V_g^*\zeta_{rN,g}|\mathcal{F}_{g-1}) = 0$ since $\zeta_{rN,g}$ is a triangular array measurable up to $\mathcal{F}_{g-1}$.

For $r, s = 1, 2$ let, $\Delta_{r,s} = \frac{1}{N}\sum_{g=1}^N b_{rN,gg}b_{sN,gg}[V_g^{*4} - \mathrm{E}(V_g^{*4})] + \frac{1}{N}\sum_{g=1}^N(b_{rN,gg}c_{sN,g} + b_{sN,gg}c_{rN,g})[V_g^{*3} - \mathrm{E}(V_g^{*3})] + \frac{1}{N}\sum_{g=1}^N c_{rN,g}c_{sN,g}[V_g^{*2} - \mathrm{E}(V_g^{*2})] + \frac{1}{N}\sum_{g=1}^N[\zeta_{rN,g}\zeta_{sN,g}V_g^{*2} -$

$d_{rsN,g}\mathrm{E}(V_j^{*2})]+\frac{1}{N}\sum_{g=1}^{N}(b_{rN,gg}\zeta_{sN,g}+b_{sN,gg}\zeta_{rN,g})V_g^{*3}+\frac{1}{N}\sum_{g=1}^{N}(c_{rN,g}\zeta_{sN,g}+c_{sN,g}\zeta_{rN,g})V_g^{*2}$ $\equiv \sum_{k=1}^{6} T_{kN}$, where, $d_{rsN,j} = 4\sum_{k=1}^{g-1} b_{rN,gk}b_{sN,gk}\mathrm{E}(V_k^{*2})$.

For T_{1N}, under Assumption 6.8 and the fact that the original disturbances are independently and Normally distributed, we have, $\{V_g^{*4} - \mathrm{E}(V_g^{*4})\}$ are independent with mean zero and it is also a martingale difference sequence. Further, $\max_{g=1,\ldots,N} \mathrm{E}|V_g^{*4} - \mathrm{E}(V_g^{*4})|^{1+\eta} < \infty$ for $\eta > 0$. Thus, $\{V_g^{*4} - \mathrm{E}(V_g^{*4})\}$ is a uniformly integrable sequence. Under Assumptions 6.9-6.11, we have, $\limsup_{N\to\infty} \frac{1}{N}\sum_{g=1}^{N} b_{rN,gg}b_{sN,gg} < \infty$ and $\limsup_{N\to\infty} \frac{1}{N^2}\sum_{g=1}^{N} b_{rN,gg}^2 b_{sN,gg}^2 \to 0$. Then using the WLLN for martingale difference arrays, $T_{1N} \xrightarrow{p} 0$. By similar arguments $T_{2N} \xrightarrow{p} 0$ and $T_{3N} \xrightarrow{p} 0$.

For T_{4N}, write $T_{4N} = T_{4N}^a + T_{4N}^b$ where $T_{4N}^a = \frac{1}{N}\sum_{g=1}^{N} \zeta_{rN,g}\zeta_{sN,g}[V_g^{*2} - \mathrm{E}(V_g^{*2})]$ and $T_{4N}^b = \frac{1}{N}\sum_{g=1}^{N} \mathrm{E}(V_g^{*2})[\zeta_{rN,g}\zeta_{sN,g} - d_{rsN,g}]$. For T_{4N}^a, note that $\frac{1}{N}\sum_{g=1}^{N} \zeta_{rN,g}\zeta_{sN,g}[V_g^{*2} - \mathrm{E}(V_g^{*2})]$ is $\mathcal{F}_g$-measurable and $\mathrm{E}[(V_g^{*2} - \mathrm{E}(V_g^{*2}))|\mathcal{F}_{g-1}] = 0$. Hence, $\zeta_{rN,g}\zeta_{sN,g}[V_g^{*2} - \mathrm{E}(V_g^{*2})]$ is a martingale difference array, thus using the WLLN for martingale difference arrays we have, $T_{4N} \xrightarrow{p} 0$. For T_{4N}^b, note that, $\zeta_{rN,g} = 2\sum_{k=1}^{g-1} b_{rN,gk}V_k^*$, hence, $\mathrm{E}(\zeta_{rN,g}\zeta_{sN,g}) = 4\sum_{k=1}^{g-1} b_{rN,gk}b_{sN,gk}\mathrm{E}(V_k^{*2}) = d_{rsN,g}$ and

$$
\begin{aligned}
T_{4N}^b &= \tfrac{1}{N}\textstyle\sum_{g=1}^{N} \mathrm{E}(V_g^{*2})[\zeta_{rN,g}\zeta_{sN,g} - d_{rsN,g}] \\
&= \tfrac{4}{N}\textstyle\sum_{g=1}^{N} \mathrm{E}(V_g^{*2})\sum_{k=1}^{g-1} b_{rN,gk}b_{sN,gk}[V_k^{*2} - \mathrm{E}(V_k^{*2})] \\
&\quad + \tfrac{8}{N}\textstyle\sum_{g=1}^{N} \mathrm{E}(V_g^{*2})\sum_{k=1}^{g-1} b_{rN,gk}V_k^*\sum_{l=1}^{k-1} b_{sN,gl}V_l^* \\
&= \tfrac{1}{N}\textstyle\sum_{g=1}^{N-1} \phi_{rsN,g}[V_k^{*2} - \mathrm{E}(V_k^{*2})] + \tfrac{1}{N}\sum_{g=1}^{N-1} \varphi_{rsN,g}V_g^*.
\end{aligned}
$$

For the last equality we use the re-arrangement, $\phi_{rsN,g} = \frac{4}{N}\sum_{k=g+1}^{N} b_{rN,kg}b_{sN,kg}\mathrm{E}(V_k^{*2})$, $\varphi_{rsN,g} = \sum_{k=1}^{g-1}\xi_{rsN,gk}V_k^*$ and $\xi_{rsN,gk} = 8\sum_{l=g+1}^{N} b_{rN,lg}b_{sN,lk}\mathrm{E}(V_l^{*2})$. Thus T_{4N}^b is the sum of two martingale difference arrays and the WLLN applies so that $T_{4N} \xrightarrow{p} 0$. Similar arguments can be given to show $T_{5N} \xrightarrow{p} 0$ and $T_{6N} \xrightarrow{p} 0$.

Proof of Theorem 6.9: To prove the consistency of $\tilde{\Sigma}_N(\tilde{\delta}_N)$ as an estimator of $\Sigma_N(\tilde{\delta}_N)$, we need to prove (a) $\tilde{\Phi}_N - \Phi_N = o_p(1)$, and (b) $\tilde{\Sigma}_N(\tilde{\psi}_N^*) - \Sigma_N(\tilde{\psi}_N^*) = o_p(1)$.

First, (a) follows an argument similar to that of the proof of Theorem 6.8. For

(b), as $\tilde{\sigma}_N^2 = \sigma_0^2 + o_p(1)$ by Theorem 6.7, it suffices to show that, by the consistency of $\tilde{\theta}_N$ and referring to (6.36), $\frac{1}{N}\sum_{g=1}^{N}\left(V_g^{*2}\Upsilon_g\Upsilon_g' - \text{Var}(V_g^*\Upsilon_g)\right) = o_p(1)$, where $\Upsilon_g = (\zeta_{1N,g} + b_{1N,gg}V_g^* + c_{1N,g},\ \zeta_{2N,g} + b_{2N,gg}V_g^* + c_{2N,g})' = (\Upsilon_{g,1}, \Upsilon_{g,2})'$. To do so consider the score function,

$$\mathbf{Q}_N(\mathbf{V}_N) = \begin{cases} \mathbf{V}_N'\mathbf{B}_{1N}\mathbf{V}_N + \mathbf{c}_{1N}'\mathbf{V}_N \\ \mathbf{V}_N'\mathbf{B}_{2N}\mathbf{V}_N + \mathbf{c}_{2N}'\mathbf{V}_N \end{cases} = \sum_{g=1}^{N} V_g^*\Upsilon_g$$

Define index, $\{g\} = \{1,\ldots,N\}$, $\{h\} = \{1,\ldots,N\} \Longleftrightarrow \{(i,t)\} = \{i = 1,\ldots,n$, for each $t = 1,\ldots,T\}$. Hence, $V_j^* = v_{it}^*$, for some i and t. As T is fixed, without loss of generality let $T = 3$. Then we have, (i) v_{it}^* are independent across i, $\forall t$; (v_{i1}^*, v_{i2}^*) uncorrelated $\forall t$; (ii) $\text{diag}(B_{rN}) = 0_{N\times N}$ for $r = 1,2$ by construction.

Write $B_{rN} = \begin{pmatrix} B_{rN,11} & B_{rN,12} \\ B_{rN,21} & B_{rN,22} \end{pmatrix}$ for $r = 1,2$, $V_N^* = \begin{pmatrix} V_{n1}^* \\ V_{n2}^* \end{pmatrix}$. Then,

$$\zeta_{rN} = (B_{rN}^{u'} + B_{rN}^{l})V_N^* = \begin{pmatrix} \zeta_{n1} \\ \zeta_{n2} \end{pmatrix} = \begin{pmatrix} (B_{rN,11}^{u'} + B_{rN,11}^{l})V_{n1}^* \\ (B_{rN,12}^{u'} + B_{rN,21}^{l})V_{n1}^* + (B_{rN,22}^{u'} + B_{rN,22}^{l})V_{n2}^* \end{pmatrix}$$

We have, $E(V_g^*\Upsilon_g) = 0$, $g = 1,\ldots,N$. Thus, $\text{Var}[\mathbf{Q}_N(\mathbf{V}_N)] = \sum_{g=1}^{N}\text{E}(V_g^{*2}\Upsilon_g\Upsilon_g') + 2\sum_{g=2}^{N}\sum_{h=1}^{g-1}\text{E}(V_g^*\Upsilon_g V_h^*\Upsilon_h')$. We need to show,

$$\sum_{g=2}^{N}\sum_{h=1}^{g-1}\text{E}(V_g^*\Upsilon_g V_h^*\Upsilon_h') = \sum_{g=2}^{N}\sum_{h=1}^{g-1}\text{E}\begin{pmatrix} V_g^*\Upsilon_{g,1}V_h^*\Upsilon_{h,1} & V_g^*\Upsilon_{g,1}V_h^*\Upsilon_{h,2} \\ V_g^*\Upsilon_{g,2}V_h^*\Upsilon_{h,1} & V_g^*\Upsilon_{g,2}V_h^*\Upsilon_{h,2} \end{pmatrix} = o(n) \text{ or } o(N). \tag{G-9}$$

Using (i,t) index, (Note: $t = 1,2$), consider the $(1,1)$ component of the covariance matrix given in (G-9),

$$\sum_{g=2}^{N}\sum_{h=1}^{g-1}\mathrm{E}(V_g^*\Upsilon_{g,1}V_h^*\Upsilon_{h,1}) = \underset{(t=1)}{\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}(v_{i1}^*\Upsilon_{i1,1}\cdot v_{j1}^*\Upsilon_{j1,1})} + \underset{(t=2)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}(v_{i2}^*\Upsilon_{i2,1}\cdot v_{j1}^*\Upsilon_{j1,1})}$$

$$+\underset{(t=2)}{\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}(v_{i2}^*\Upsilon_{i2,1}\cdot v_{j2}^*\Upsilon_{j2,1})}$$

$$\equiv Q_1+Q_2+Q_3$$

To keep the notation simple for the moment the $r=1$ index is suppressed. Note $\{\Upsilon_{i1}\}=\zeta_{n1}+c_{1n}, \mathbf{c}_N=(c_{1n},c_{2n})' \Longrightarrow \{\Upsilon_{i1}\}$ is $\mathcal{F}_{n,i-1}-$ measurable.

Let $\{\Upsilon_{i2}\}=\{\Upsilon_{i2}^a\}+\{\Upsilon_{i2}^b\}$ where $\{\Upsilon_{i2}^a\}=(B_{N,12}^{u'}+B_{N,21}^{l})V_{n1}^*+c_{2n}=b_{21,i}^{'}V_{n1}^*+c_{2n}$ and $\{\Upsilon_{i2}^b\}=(B_{N,22}^{u'}+B_{N,22}^{l})V_{n2}^*+c_{2n}=b_{22,i}^{'}V_{n2}^*+c_{2n}$.

From these we can see that $Q_1=0$. Now,

$$\begin{aligned}
Q_3 &= \textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}(v_{i2}^*\Upsilon_{i2}\cdot v_{j2}^*\Upsilon_{j2})\\
&= \textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}(v_{i2}^*\Upsilon_{i2}^a\cdot v_{j2}^*\Upsilon_{j2}^a+v_{i2}^*\Upsilon_{i2}^a\cdot v_{j2}^*\Upsilon_{j2}^b+v_{i2}^*\Upsilon_{i2}^b\cdot v_{j2}^*\Upsilon_{j2}^a+v_{i2}^*\Upsilon_{i2}^b\cdot v_{j2}^*\Upsilon_{j2}^b)\\
&\quad \Upsilon_{i2}^b \text{ is } \mathcal{F}_{n,i-1}-\text{measurable and the last term vanishes}\\
&= \textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}\left(v_{i2}^*(b_{21,i}^{'}V_{n1}^*)\cdot v_{j2}^*(b_{21,j}^{'}V_{n1}^*)+v_{i2}^*(b_{21,i}^{'}V_{n1}^*)\cdot v_{j2}^*(b_{22,j}^{'}V_{n2}^*)\right)\\
&\quad +\textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}\left(v_{i2}^*(b_{22,i}^{'}V_{n2}^*)\cdot v_{j2}^*(b_{21,j}^{'}V_{n1}^*)\right)\\
&= \textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}\left(v_{i2}^*b_{21,ii}v_{i1}^*\cdot v_{j2}^*b_{21,jj}v_{j1}^*+v_{i2}^*b_{21,ij}v_{j1}^*\cdot v_{j2}^*b_{21,ji}v_{i1}^*\right)\\
&\quad +\textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}\left(v_{i2}^*b_{21,ii}v_{i1}^*\cdot v_{j2}^*b_{22,jj}v_{j2}^*+v_{i2}^*b_{21,ij}v_{j1}^*\cdot v_{j2}^*b_{22,ji}v_{i2}^*\right)\\
&\quad +\textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}\left(v_{i2}^*b_{22,ii}v_{i2}^*\cdot v_{j2}^*b_{21,jj}v_{j1}^*+v_{i2}^*b_{22,ij}v_{j2}^*\cdot v_{j2}^*b_{21,ji}v_{i1}^*\right)\\
&= 0 \text{ since } (i,j) \text{ are independent and } (t,s) \text{ are uncorrelated and } b_{22,ii}=0.
\end{aligned}$$

Now $Q_2=\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{E}(v_{i2}^*\Upsilon_{i2}\cdot v_{j1}^*\Upsilon_{j1})=\sum_{i=1}^{n}\mathrm{E}(v_{i1}^*v_{i2}^*\Upsilon_{i1}\Upsilon_{i2})+\sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathrm{E}(v_{i2}^*\Upsilon_{i2}\cdot$

$v^*_{j1}\Upsilon_{j1})$, where,

$$
\begin{aligned}
& \textstyle\sum_{i=2}^{n}\sum_{j=1}^{i-1}\mathrm{E}(v^*_{i2}\Upsilon_{i2}\cdot v^*_{j1}\Upsilon_{j1}) \\
= & \textstyle\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Big(\mathrm{E}(v^*_{i2}\Upsilon^a_{i2}\cdot v^*_{j1}\Upsilon_{j1})+\mathrm{E}(v^*_{i2}\Upsilon^b_{i2}\cdot v^*_{j1}\Upsilon_{j1})\Big) \\
= & \textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\Big(\mathrm{E}(v^*_{i2}b'_{21,i}V^*_{n1}\cdot v^*_{j1}b'_{11,j}V^*_{n1})+\mathrm{E}(v^*_{i2}b'_{22,i}V^*_{n2}\cdot v^*_{j1}b'_{11,j}V^*_{n1})\Big) \\
= & \textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}(v^*_{i2}b'_{21,ii}v^*_{i1}\cdot v^*_{j1}b_{11,jj}v^*_{j1}+v^*_{i2}b'_{21,ij}v^*_{j1}\cdot v^*_{j1}b_{11,ji}v^*_{i1}) \\
& +\textstyle\sum_{i=1}^{n}\sum_{j=1}^{i-1}\mathrm{E}(v^*_{i2}b'_{22,ii}v^*_{i2}\cdot v^*_{j1}b_{11,jj}v^*_{j1}+v^*_{i2}b'_{22,ij}v^*_{j2}\cdot v^*_{j1}b_{11,ji}v^*_{i1}) \\
= & \ 0 \text{ since } (i,j) \text{ are independent and } (t,s) \text{ are uncorrelated and } b_{11,ii}=b_{22,ii}=0.
\end{aligned}
$$

Finally for $\Upsilon_{i1}=\zeta_{i1}+c_{1i}$ and $\Upsilon_{i2}=\zeta_{i2}+c_{2i}$,

$$
\begin{aligned}
\textstyle\sum_{i=1}^{n}\mathrm{E}(v^*_{i1}v^*_{i2}\Upsilon_{i1}\Upsilon_{i2}) &= \textstyle\sum_{i=1}^{n}\mathrm{E}\left(v^*_{i1}v^*_{i2}(\zeta_{i1}\zeta_{i2}+\zeta_{i1}c_{21}+\zeta_{i2}c_{1i})\right) \\
&= \textstyle\sum_{i=1}^{n}\mathrm{E}\left(v^*_{i1}v^*_{i2}[\zeta_{i1}(\zeta^a_{i2}+\zeta^b_{i2})+\zeta^a_{i2}c_{1i}]\right) \text{ since } b_{11,ii}=b_{22,ii}=0 \\
&= \textstyle\sum_{i=1}^{n}\mathrm{E}\left(v^*_{i1}v^*_{i2}b_{21,ii}v^*_{i2}c_{1i}\right)=\sum_{i=1}^{n}\mathrm{E}\left(v^*_{i1}v^{*2}_{i2}b_{21,ii}c_{1i}\right) \\
&= O(1),
\end{aligned}
$$

as off diagonal elements of B_N are $O(\frac{1}{N})$ by Assumptions 6.9-6.11. Similar workings on other components of (G-9), lead to $\sum_{g=2}^{N}\sum_{h=1}^{g-1}\mathrm{E}(V^*_g\Upsilon_g V^*_h\Upsilon'_h)=o(1)$.

This implies that $\mathrm{Var}[Q_N(V_N)]=\sum_{g=1}^{N}\mathrm{E}(V^{*2}_g\Upsilon_g\Upsilon'_g)+o(1)$.

Proof of Corollary 6.2: The consistency of $\tilde{\Sigma}_N(\tilde{\beta}_N)$ follows that of $\tilde{\Sigma}_N(\tilde{\lambda}_N)$ and the consistency of $\tilde{\theta}_N$.