HYPOTHESIS TESTING FOR SPATIAL SARAR TOBIT MODELS AND HIGH DIMENSIONAL DATA AND THEIR APPLICATIONS

BY

CHANG LU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Professor Anil K. Bera, Chair
Professor Xiaofeng Shao
Professor Geoffrey J.D. Hewings
Associate Professor JiHyung Lee

Abstract

The spatial autoregressive (SAR) model is one of the most important models in spatial econometrics to describe the connection between dependent and independent variables by taking account of spatial dependence. The emphasis on SAR models has become more prominent in not only practical econometrics, but also theoretical econometrics. While the traditional SAR model has been extensively researched in the literature, researchers in the field of spatial analysis are becoming increasingly interested in spatial models with limited dependent variables (LDV). The model specification problem and applications for both the SAR model and the SAR model with LDV are investigated in this dissertation.

Chapter 1 proposes the modified Lagrange multiplier (LM) tests for one of the SAR models with LDV, SAR Tobit models with SAR disturbances (SARAR Tobit model). There are two types of models considered in this chapter. One is the latent SARAR Tobit model, and the other one is the simultaneous SARAR Tobit model. The difference of these two types of SARAR Tobit model is whether a certain area's data is affected by the actual data of its neighbors or by the latent data of its neighbors. This chapter includes score functions, two information matrices, and proofs of asymptotic distributions for the suggested test statistics. The finite sample size and power are also investigated through a Monte Carlo simulation study. Our simulated results show that these proposed tests have good finite sample properties both in terms of size and power, even in the presence of misspecification.

Chapter 2 examines the geographical distribution of crime in the city of Chicago census tracts and its effects on house values using the SAR model. The results show the spillover effects of crime among census tracts in Chicago and the effect of crime on the local real estate market. We consider two kinds of spatial models: one is the SAR model, and the other one is the spatial Durbin model (SDM). We demonstrate the impact of crime on the local region as well as the large impact of neighborhood area crime on home values using data from 801 census tracts in the city of Chicago. Besides, in our results, theft and burglary have positive but insignificant effects on housing prices, which means they are associated with higher-income neighborhoods, while violent crimes tend to be found in lower-income neighborhood areas. This means that, contrary to the conventional thought that crime always has a negative effect on property prices, our analysis shows that the effects of different types of crimes can have various and complex effects on housing prices.

In addition to the SAR model and its application, this dissertation also investigates the hypothesis testing problem in a high-dimensional data setting. High-dimensional data means the dimension of the data can grow with sample size, or even larger than the sample size. When the dimension of data is larger than the sample size, the traditional Hotelling's T^2 test cannot be used since the inverse of the sample covariance does not exist. Chapter 3 studies the consistency of the wild bootstrap for the U-statistic in the inference of high-dimensional means. Both unstudentized and studentized U-statistics are investigated and the consistency of wild bootstrap is obtained under mild conditions, allowing for the dimension d growing at the faster rate than sample size n. The results are applied to three problems in statistics and econometrics: overidentification test with growing number of instrumental variables, spatial sign-based mean test and sphericity hypothesis test for the covariance matrix.

Overall, this dissertation discusses hypothesis testing for SAR models with LDV and high-dimensional data, as well as certain applications. The findings of this dissertation contribute to a better understanding of the use of robust LM tests in the SARAR Tobit model and the application of unstudentized and studentized U-statistics in the mean zero testing under high-dimensional settings, as well as the possibilities of future research.

To Father, Mother, Shijia and Luna.

Acknowledgments

I would like to offer my heartfelt gratitude to my advisor, Anil K. Bera, whose enlightening lectures helped me first grow interest in econometrics and then finally choose spatial econometrics as my field of study. I've known Professor Bera for more than seven years and have worked with him for six. Throughout the six years I worked with him, he was a constant source of encouragement and was always eager to help in any way he could. He offered not only academic advice, but also life advice.

Thank you also to my committee members, Xiaofeng Shao, Geoffrey J.D. Hewings, and JiHyung Lee, for their advice and support. Many thanks to Professor Xiaofeng Shao, who helped me develop critical thinking skills and taught me how to conduct research. Also I'm grateful to Professor Geoffrey J.D. Hewings. He has provided me with a lot of guidance and support, as well as working space in REAL. He and his REAL family gave me a warm welcome and embraced me as a member of their family. Professors Shao and Hewings both provided me with summer funding assistantship, which means a lot to a student who relies on that money to make ends meet. I also want to thank Professor Lee for his advice and suggestion for the prelim and final defense. Professor Lee's course, Financial Econometrics, was crucial in laying the groundwork for my dissertation. This dissertation would not have been possible without my committee members' tremendous effort and assistance.

I would also want to thank my parents, Yanqing Lu and Xiaojie Liu. They are the ones who have inspired me to persist in pursuing my academic goals and never give up on my goal of earning a Ph.D. degree. I would never have had the confidence to repeat the college admission exam in China and have a decent platform to pursue my career and goal if it hadn't been for their encouragement and support. They are my reliable back-up plan for whatever decision I make.

During my graduate studies, I was fortunate to have wonderful contemporaries and colleagues who assisted and supported me. Among them, I'd like to express my gratitude to my lifelong friend Jiarui Li in particular. With her by side, days and years become so much more joyful in Champaign. I would also want to thank my mother-in-law Professor Shi and father-in-law Professor Zhao. Without their assistance, I would be unable to concentrate on writing this dissertation thesis while still caring for my infant.

Finally, there are no words to adequately express my appreciation to my husband Dr. Shijia Zhao and my daughter Luna Zhao. My husband has been my solid support in my ups and downs. He encouraged me to pursue my dreams and assured me that he would be my staunchest ally, while Luna, my darling daughter, has become everything I want to fight for. I'm overjoyed to have them here with me.

When I was finishing this acknowledgement, my beloved grandfather, Doctor Baoquan Liu passed away on June 20, 2022. Throughout my childhood, my grandfather has showed me a lot of love and care. He was the first generation in our family and a famous oncologist who who helped many patients by saving or extending their lives. My grandfather has always served as my role model, and his achievements have inspired me to set lofty goals and work hard in my studies. May he rest in peace.

Table of Contents

Chapter 1

Robust LM Tests for Latent and Simultaneous Spatial Autoregressive Tobit Models with Spatial Autoregressive Disturbances

1.1 Introduction

The conventional Tobit model (Tobin [1]) can be expressed as

$$y_i = \max(\beta' x_i + \mu_i, 0)$$

where $u_i \sim i.i.d.N(0, \sigma_0^2)$, β and σ_0 are parameters and x_i is $k \times 1$ vector of exogenous variables. The Tobit model, also known as a censored regression model, is frequently used to estimate the linear relationship between variables when the dependent variable is censored. The conventional Tobit model has been well studied by many researchers (see, among others, Tobin [1]; Amemiya [2]; Robinson [3]; Robinson et al. [4]). When spatial settings are introduced to the Tobit model, for example, one certain area's data is not only left- or right-censoring, but also affected by the neighborhood locations. That's the basic concept of spatial models with LDV.

Spatial models with LDV have been an emerging topic with rising interests among spatial analysis researchers. The related publications can be generally classified into three categories including the model estimation, empirical applications and new test statistics for the detection of spatial correlations. The first category focuses on the estimation of the model. McMillen [5] proposed two categories of estimators for probit model with spatial autocorrelation. One was based on the EM algorithm and the other one could be applied to models derived using the spatial expansion method. Pinkse and Slade [6] developed tests for spatial-error correlation and proposed a generalized method of moments (GMM) estimation for the model. They also applied the method to evaluate spatial patterns in retail-gasoline contracts. Lesage [7] used the Gibbs sampling (Markov chain Monte Carlo simulation) method for the bayesian estimation of the LDV spatial autoregressive models. Wang et al. [8] demonstrated the consistency of a spatial Probit model for cross-sectional data in a binary choice form using partial maximum likelihood. Besides the estimation of the model, the second category focuses on the empirical applications. Pinkse et al. [9] used the one-step GMM estimator to study mine operating decions in a real-options context. Bhat et al. [10] used a composite marginal likelihood approach with a spatial dependent discrete choice model to investigate the daily episode frequency of teenagers' physically active and inactive reactional activity participation. In a recent survey paper, Bille and Arbia [11] discussed different methods used in spatial discrete choice models and LDV models, as well as their applications in health economics. Yang et al. [12] proposed a non-nested Cox-type test for the specification test in the Tobit model with social interactions with both complete and incomplete information. They used the property tax rates data in North Carolina to show that this Cox test has larger income marginal effects and smaller population marginal effects for complete information model than the incomplete ones. The final group is concerned with the novel test statistics for the discovery of spatial dependence. Testing spatial models has received a lot of attention in the literature. Moran [13] introduced the most commonly used test statistic for spatial correlation, which is referred as Moran's I test. In their survey paper, Anselin and Bera [14] documented various tests for spatial dependence. Anselin et al. [15] applied the modified LM test developed by Bera and Yoon [16] to the SAR model in the presence of spatial error auto-correlation, the spatial error model (SEM) part. Robust tests developed in Anselin et al. [15] have become standard diagnostic tools for spatial econometricians. In this chapter, we generalize Anselin et al. [15] from the standard SAR-SEM model to the Tobit framework. This generalization is importnat from the practical point of view, as evidence from the the papers cited above. And also theoretically, as we will note subsequently, derivations and proofs are relatively much more complex than in Anselin et al. [15].

There are two types of Tobit models that are frequently employed in the spatial setting. One is the SAR Tobit model and the other is the spatial error (SE) Tobit model. These two structures have different spatial correlations and have been proposed and studied by a series of research papers. Qu and Lee [17] proposed LM statistics for testing the spatial correlation and derived the asymptotic distributions. Xu and Lee [18] came up with the maximum likelihood estimation for the SAR Tobit model and showed the consistency for the estimator. Xu and Lee [19] investigated a distinct environment for disturbances that were i.i.d. but had an unidentified distribution. They used the sieve maximum likelihood estimation and showed the consistency. However, none of the papers combined the two spatial correlations together. Therefore, in this chapter, we proposed the spatial auto-regressive Tobit model with spatial auto-regressive disturbances (SARAR) and applied Bera and Yoon's [16] modified LM test technique.

For developing robust LM test statistics, we consider two types of SARAR Tobit models. One is the latent SARAR Tobit model as follows.

$$Y_n^* = \lambda W_n Y_n^* + X_n\beta + u_n, \quad u_n = \rho M_n u_n + \varepsilon_n$$

$$y_{i,n} = \max(0, y_{i,n}^*),$$

where $Y_n = (y_{1,n}, y_{2,n}, ..., y_{n,n})'$ are observable variables, $Y_n^* = (y_{1,n}^*, y_{2,n}^*, ..., y_{n,n}^*)'$ is the vector of latent variables, $\varepsilon_{i,n} \sim i.i.d.N(0,\sigma^2)$, $E|\varepsilon_{i,n}|^{4+v} < \infty$, for $v > 0$ and $\varepsilon_n \sim N(0_n, \Sigma_\varepsilon)$, where $\Sigma_\varepsilon = \sigma^2 I_n$.

Another version is the simultaneous SARAR Tobit model, which has the following structure

$$Y_n^* = \lambda W_n Y_n + X_n\beta + u_n, \quad u_n = \rho M_n u_n + \varepsilon_n, \qquad y_{i,n} = \max(0, y_{i,n}^*)$$

$$w_{ii,n} = 0, m_{ii,n} = 0, \qquad \varepsilon_n \sim N(0_n, \sigma^2 I_n).$$

The difference of these two types of SARAR Tobit model is whether a certain location's observation is affected by the latent data of its neighbors or by the actual data of its neighbors.

The rest of the chapter is organized as follows. In Section 1.2, we introduce the latent SARAR Tobit model and present the details of its score functions and information matrix. In Section 1.3 we provide the model structure of simultaneous SARAR Tobit model and the corresponding score functions and information matrix. Section 1.4 is about the asymptotic distribution of test statistics. Section 1.5 shows the Monte Carlo simulation results under different data generating processes. Section 1.6 provides the conclusions and some ideas for future research, and the technical proofs are in the Appendix A.1.

1.2 Latent SARAR Tobit Model

1.2.1 Model Construction

Assume we have the latent SARAR Tobit model and the model can be written in the following matrix form:

$$Y_n^* = \lambda W_n Y_n^* + X_n\beta + u_n, \quad u_n = \rho M_n u_n + \varepsilon_n$$

$$y_{i,n} = \max(0, y_{i,n}^*),$$

where $Y_n = (y_{1,n}, y_{2,n}, ..., y_{n,n})'$ are observable, $Y_n^* = (y_{1,n}^*, y_{2,n}^*, ..., y_{n,n}^*)'$, $\varepsilon_{i,n} \sim i.i.d.N(0, \sigma^2)$. Using the same settings in Qu and Lee [17], we let $E|\varepsilon_{i,n}|^{4+v} < \infty$, for $v > 0$ and $\varepsilon_n \sim N(0_n, \Sigma_\varepsilon)$, where $\Sigma_\varepsilon = \sigma^2 I_n$, $|\lambda| < \frac{1}{\max_i \sum_{j=1}^n W_{ij}}$ and W_n is bounded in row sums and $|\rho| < \frac{1}{\max_i \sum_{j=1}^n M_{ij}}$ and M_n is bounded in row sums. W_n and M_n are both non-negative and they are not necessarily the same. It should be noted that the spatial correlations are between the latent variables, as well as the disturbances.

Then the model can be rewritten as, when $y_{i,n}^* > 0$ for every i,

$$Y_n = (I_n - \lambda W_n)^{-1} X_n\beta + (I_n - \lambda W_n)^{-1}(I_n - \rho M_n)^{-1}\varepsilon_n.$$

Since $\varepsilon_n \sim N(0_n, \sigma^2 I_n)$, let $S_n(\lambda) = I_n - \lambda W_n$, $R_n(\rho) = I_n - \rho M_n$, then $(I_n - \lambda W_n)^{-1}(I_n - \rho M_n)^{-1}\varepsilon_n \sim N(0_n, \sigma^2 S_n^{-1}(\lambda) R_n^{-1}(\rho) S_n^{-1'}(\lambda) R_n^{-1'}(\rho)$.

Denote $\theta = (\beta, \sigma^2, \lambda, \rho)'$ and the density function takes the following form

$$f(Y_n^*|X_n, \theta) = (\frac{1}{\sqrt{2\pi}\sigma})^n |S_n(\lambda)||R_n(\rho)| \exp\left[-\frac{[S_n(\lambda)Y_n^* - X_n\beta]' R_n(\rho)' R_n(\rho)[S_n(\lambda)Y_n^* - X_n\beta]}{2\sigma^2}\right].$$

Following the results in Qu and Lee [17], we denote $I(Y_n, Y_n^*)$ which takes value 1 if $Y_n^* > 0$, and 0 if $Y_n^* \leq 0$. Then the joint function of (Y_n, Y_n^*) will be

$$f(Y_n, Y_n^*|X_n, \theta) = I(Y_n, Y_n^*)f(Y_n^*|X_n, \theta),$$

and the density function of Y_n will be

$$f(Y_n|X_n, \theta) = \int_{\Omega_y^*} f(Y_n, Y_n^*|X_n, \theta)dY_n^* = \int_{\Omega_y^*} I(Y_n, Y_n^*)f(Y_n^*|X_n, \theta)dY_n^*.$$

Considering the log-likelihood function for Y_n might be difficult and our primary focus is the score function and the information matrix, we can utilize the property to interchange differentiation w.r.t. θ and integration w.r.t. y_n^*.

$$\begin{aligned}\frac{\partial \ln f(Y_n|X_n, \theta)}{\partial \theta} &= \int_{\Omega_{y_n^*}} \frac{\partial \ln f(Y_n^*|X_n, \theta)}{\partial \theta} \frac{I(Y_n, Y_n^*)f(Y_n^*|X_n, \theta)}{f(Y_n|X_n, \theta)} dY_n^* \\ &= \int_{\Omega_{y_n^*}} \frac{\partial \ln f(Y_n^*|X_n, \theta)}{\partial \theta} f(Y_n^*|Y_n, X_n, \theta)dY_n^* \\ &= E_\theta\left[\frac{\partial \ln f(Y_n^*|X_n, \theta)}{\partial \theta}|Y_n, X_n\right].\end{aligned} \tag{1.2.1}$$

Since the density function $f(Y_n^*|X_n, \theta)$ belongs to the exponential family of distributions, we can express it in the following form.

$$f(Y_n^*|X_n, \theta) = \exp\left[\eta(X_n, \theta)'m(Y_n^*) + a(X_n, Y_n^*)\right]/K(X_n, \eta(X_n, \theta)).$$

where

$$K(X_n, \eta(X_n, \theta)) = \int_{\Omega_{y_n^*}} \exp\left[\eta(X_n, \theta)'m(Y_n^*) + a(X_n, Y_n^*)\right] dY_n^*$$

thus

$$\begin{aligned}\frac{\partial \ln f(Y_n^*|X_n, \theta)}{\partial \theta} &= \frac{\eta(X_n, \theta)'}{\partial \theta}\left[m(Y_n^*) - \frac{\partial \ln K(X_n, \eta(X_n, \theta))}{\partial \eta}\right] \\ &= \frac{\partial \eta(X_n, \theta)'}{\partial \theta}\left[m(Y_n^*) - E_\theta(m(Y_n^*)|X_n)\right].\end{aligned}$$

Then Equation (1.2.1) becomes

$$\frac{\partial \ln f(Y_n|X_n, \theta)}{\partial \theta} = \frac{\partial \eta(X_n, \theta)'}{\partial \theta}\left[E_\theta(m(Y_n^*)|X_n, Y_n) - E_\theta(m(Y_n^*)|X_n)\right].$$

In order to get the explicit form for Equation (1.2.1), $f(Y_n^*|X_n, \theta)$ can be decomposed as

$$\begin{aligned}f(Y_n^*|X_n, \theta) &= (\frac{1}{\sqrt{2\pi}\sigma})^n|S_n(\lambda)||R_n(\rho)|\exp\left[-\frac{[S_n(\lambda)Y_n^* - X_n\beta]'R_n(\rho)'R_n(\rho)[S_n(\lambda)Y_n^* - X_n\beta]}{2\sigma^2}\right] \\ &= (\frac{1}{\sqrt{2\pi}\sigma})^n|S_n(\lambda)||R_n(\rho)|\exp[-\frac{1}{2\sigma^2}(Y_n^{*'}S_n(\lambda)'R_n(\rho)'R_n(\rho)S_n(\lambda)Y_n^* - \beta'X_n'R_n(\rho)'R_n(\rho)S_n(\lambda)Y_n^* \\ &\quad - Y_n^{*'}S_n(\lambda)'R_n(\rho)'R_n(\rho)X_n\beta + \beta'X_n'R_n(\rho)'R_n(\rho)X_n\beta)].\end{aligned} \tag{1.2.2}$$

A part of Equation (1.2.2) can be further decomposed as

$$\begin{aligned}
S_n(\lambda)'R_n(\rho)'R_n(\rho)S_n(\lambda) &= (I_n - \lambda W_n)'(I_n - \rho M_n)'(I_n - \rho M_n)(I_n - \lambda W_n) \\
&= (I_n - \lambda W_n' - \rho M_n' + \lambda\rho W_n' M_n')(I_n - \lambda W_n - \rho M_n + \lambda\rho M_n W_n) \\
&= I_n - \lambda W_n - \rho M_n + \lambda\rho M_n W_n - \lambda W_n' + \lambda^2 W_n' W_n + \lambda\rho W_n' M_n - \lambda^2\rho W_n' M_n W_n \\
&- \rho M_n' + \rho\lambda M_n' W_n + \rho^2 M_n' M_n - \lambda\rho^2 M_n' M_n W_n + \lambda\rho W_n' M_n' \\
&- \lambda^2\rho W_n' M_n' W_n - \lambda\rho^2 W_n' M_n' M_n + \lambda^2\rho^2 W_n' M_n' M_n W_n.
\end{aligned}$$

Then the density function becomes

$$\begin{aligned}
f(Y_n^*|X_n,\theta) &= \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n |S_n(\lambda)||R_n(\rho)| \exp[-\frac{1}{2}(Y_n^{*'} S_n(\lambda)'R_n(\rho)'R_n(\rho)S_n(\lambda)Y_n^* - \beta' X_n' R_n(\rho)'R_n(\rho)S_n(\lambda)Y_n^* \\
&- Y_n^{*'} S_n(\lambda)'R_n(\rho)'R_n(\rho)X_n\beta + \beta' X_n' R_n(\rho)'R_n(\rho)X_n\beta)] \\
&= \Big(\frac{1}{\sqrt{2\pi\sigma^2}}\Big)^n |S_n(\lambda)||R_n(\rho)| \exp[-\frac{1}{2\sigma^2}(Y_n^{*'} Y_n^* - 2\lambda Y_n^{*'} W_n Y_n^* - 2\rho Y_n^{*'} M_n Y_n^* + 2\lambda\rho Y_n^{*'} M_n W_n Y_n^* \\
&+ \lambda^2 Y_n^{*'} W_n' W_n Y_n^* - 2\lambda^2\rho Y_n^{*'} W_n' M_n W_n Y_n^* \\
&+ 2\rho\lambda Y_n^{*'} M_n' W_n Y_n^* + \rho^2 Y_n^{*'} M_n' M_n Y_n^* - 2\lambda\rho^2 Y_n^{*'} M_n' M_n W_n Y_n^* \\
&+ \lambda^2\rho^2 Y_n^{*'} W_n' M_n' M_n W_n Y_n^* - 2\beta' X_n' R_n(\rho)'R_n(\rho)S_n(\lambda)Y_n^*) \\
&- \frac{1}{2\sigma^2}\beta' X_n' R_n(\rho)'R_n(\rho)X_n\beta].
\end{aligned} \tag{1.2.3}$$

Equation (1.2.3) is used to determine the values of $\eta(X_n,\theta)$, $m(Y_n^*)$ and $a(X_n,Y_n^*)$. As a result, we have

$$\begin{aligned}
\eta(X_n,\theta) &= (-\frac{1}{2\sigma^2}, \frac{1}{\sigma^2}\lambda, \frac{1}{\sigma^2}\rho, -\frac{1}{\sigma^2}\lambda\rho, -\frac{1}{2\sigma^2}\lambda^2, \frac{1}{\sigma^2}\lambda^2\rho, \\
&- \frac{1}{\sigma^2}\lambda\rho, -\frac{1}{2\sigma^2}\rho^2, \frac{1}{\sigma^2}\lambda\rho^2, -\frac{1}{2\sigma^2}\lambda^2\rho^2, \frac{1}{\sigma^2}S_n(\lambda)'R_n(\rho)'R_n(\rho)X_n\beta, -\frac{1}{2\sigma^2}\beta' X_n' R_n(\rho)'R_n(\rho)X_n\beta)'
\end{aligned} \tag{1.2.4}$$

and

$$\begin{aligned}
m(Y_n^*) &= (Y_n^{*'} Y_n^*, Y_n^{*'} W_n Y_n^*, Y_n^{*'} M_n Y_n^*, Y_n^{*'} M_n W_n Y_n^*, Y_n^{*'} W_n' W_n Y_n^*, Y_n^{*'} W_n' M_n W_n Y_n^*, Y_n^{*'} M_n' W_n Y_n^*, \\
&Y_n^{*'} M_n' M_n Y_n^*, Y_n^{*'} M_n' M_n W_n Y_n^*, Y_n^{*'} W_n' M_n' M_n W_n Y_n^*, Y_n^*, 1)'
\end{aligned} \tag{1.2.5}$$

Though the forms of $\eta(X_n,\theta)$ and $m(Y_n^*)$ are rather complicated, it's possible to get a neat form for the score functions and the information matrix.

1.2.2 Score Functions

Now we denote $\hat{\theta} = (\beta, \sigma^2, 0, 0)'$. The score functions for latent SARAR Tobit model take the following forms:

$$d_\beta(\hat{\theta}) = \frac{1}{\sigma^2} X_n'[E(Y_n^*|X_n, Y_n) - E(Y_n^*|X_n)] \tag{1.2.6}$$

$$d_{\sigma^2}(\hat{\theta}) = \frac{1}{2\sigma^4}[E(Y_n^{*'} Y_n^*|X_n, Y_n) - E(Y_n^{*'} Y_n^*|X_n)] - \frac{1}{\sigma^4}\beta' X_n'[E(Y_n^*|X_n, Y_n) - E(Y_n^*|X_n)] \tag{1.2.7}$$

$$d_\lambda(\hat{\theta}) = \frac{1}{\sigma^2} E(\varepsilon_n|X_n, Y_n)' W_n E(\varepsilon_n|X_n, Y_n) + \frac{1}{\sigma^2} E(\varepsilon_n|X_n, Y_n)' W_n X_n \beta \tag{1.2.8}$$

$$d_\rho(\hat{\theta}) = \frac{1}{\sigma^2} E(\varepsilon_n|X_n, Y_n)' M_n E(\varepsilon_n|X_n, Y_n) \tag{1.2.9}$$

The derivation for Equations (1.2.8) and (1.2.9) is given in Appendix A.1.

1.2.3 Information Matrix

The information matrix for the SARAR Tobit model takes the following form:

$$\begin{aligned} J(\theta) &= -E\left[\frac{1}{n}\frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}\right] \\ &= \begin{bmatrix} J_\beta & J_{\beta\sigma^2} & J_{\beta\lambda} & J_{\beta\rho} \\ J_{\sigma^2\beta} & J_{\sigma^2} & J_{h\lambda} & J_{h\rho} \\ J_{\lambda\beta} & J_{\lambda\sigma^2} & J_\lambda & J_{\lambda\rho} \\ J_{\rho\beta} & J_{\rho\sigma^2} & J_{\rho\lambda} & J_\rho \end{bmatrix} \end{aligned}$$

Next we compute the components of the information matrix one by one. Since the second derivative might be very complicated, we can use the score function and compute the information matrix using the following information matrix equality:

$$E\left[\left(\frac{\partial L(\hat{\theta})}{\partial\theta}\right)\left(\frac{\partial L(\hat{\theta})}{\partial\theta}\right)'\right] = E\left[-\frac{\partial^2 L(\hat{\theta})}{\partial\theta\partial\theta'}\right].$$

We simplify the score functions as follows because the information matrix will be more complex than the scoring function.

Equation (1.2.6) can be further broken down into

$$d_\beta(\hat{\theta}) = \frac{1}{\sigma^2} X_n'[E(Y_n^*|X_n, Y_n) - E(Y_n^*|X_n)] = \frac{1}{\sigma^2} X_n' E(\varepsilon_n|X_n, Y_n) \tag{1.2.10}$$

because

$$E(Y_n^*|X_n, Y_n) = E(X_n\beta + \varepsilon_n|X_n, Y_n) = X_n\beta + E(\varepsilon_n|X_n, Y_n)$$

and

$$E(Y_n^*|X_n) = E(X_n\beta + \varepsilon_n|X_n) = X_n\beta.$$

It can be deduced from definition of $\hat{\theta}$ that

$$E(\varepsilon_{i,n}|X_n, Y_n) = (y_{i,n} - x'_{i,n}\beta)I(y_{i,n} > 0) - \frac{\phi(x'_{i,n}\beta, \sigma^2)\sigma^2}{(1 - \Phi(x'_{i,n}\beta, \sigma^2))}I(y_{i,n} = 0). \tag{1.2.11}$$

The derivation of Equation (1.2.11) is shown in Appendix A.1. Denoting $v_i(\hat{\theta}) = E(\varepsilon_{i,n}|X_n, Y_n)$, the score function $d_\beta(\hat{\theta})$ can be simplified as

$$d_\beta(\hat{\theta}) = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_{i,n}v_i(\hat{\theta}). \tag{1.2.12}$$

Similarly, for $d_{\sigma^2}(\hat{\theta})$ we have,

$$\begin{aligned} d_{\sigma^2}(\hat{\theta}) &= \frac{\partial \ln L_{ns}(\beta, \sigma^2)}{\partial\sigma^2} \\ &= \frac{\partial\left[\sum_{i=1}^{n} I(y_{i,n} = 0)\ln(1 - \Phi(x'_i\beta, \sigma^2)) - \frac{1}{2}\sum_{i=1}^{n} I(y_{i,n} > 0)(\ln(2\pi\sigma^2) + \frac{1}{\sigma^2}(y_{i,n} - x'_{i,n}\beta)^2)\right]}{\partial\sigma^2} \\ &= \frac{1}{2\sigma^4}\sum_{i=1}^{n}\left[I(y_{i,n} = 0)\frac{x'_{i,n}\beta\sigma^2\phi(x'_{i,n}\beta, \sigma^2)}{1 - \Phi(x'_{i,n}\beta, \sigma^2)} + I(y_{i,n} > 0)((y_{i,n} - x'_{i,n}\beta)^2 - \sigma^2)\right] \\ &= \frac{1}{2\sigma^4}\sum_{i=1}^{n} q_i(\hat{\theta}), \end{aligned} \tag{1.2.13}$$

where $q_i(\hat{\theta}) = I(y_{i,n} = 0)\frac{x'_{i,n}\beta\sigma^2\phi(x'_{i,n}\beta,\sigma^2)}{1-\Phi(x'_{i,n}\beta,\sigma^2)} + I(y_{i,n} > 0)((y_{i,n} - x'_{i,n}\beta)^2 - \sigma^2)$ and $E(q_i(\hat{\theta})) = 0$.

For $d_\lambda(\hat{\theta})$,

$$\begin{aligned} d_\lambda(\hat{\theta}) &= \frac{1}{\sigma^2}E(\varepsilon_n|X_n, Y_n)'W_nE(\varepsilon_n|X_n, Y_n) + \frac{1}{\sigma^2}E(\varepsilon_n|X_n, Y_n)'W_nX_n\beta \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_{ij}v_i(\hat{\theta})v_j(\hat{\theta}) + \frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_{ij}v_i(\hat{\theta})x'_j\beta. \end{aligned} \tag{1.2.14}$$

For $d_\rho(\hat{\theta})$,

$$d_\rho(\hat{\theta}) = \frac{1}{\sigma^2}E(\varepsilon_n|X_n, Y_n)'M_nE(\varepsilon_n|X_n, Y_n) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} m_{ij}v_i(\hat{\theta})v_j(\hat{\theta}). \tag{1.2.15}$$

With these simplified score functions, it will be much easier to get a simplified version of the information matrix.

1.2.3.1 J_β

According to the information matrix equality,

$$\begin{aligned}
J_\beta &= \frac{1}{n} E\left[d_\beta(\hat\theta) d_\beta(\hat\theta)'\right] \\
&= \frac{1}{n} E\left[\frac{1}{\sigma^4}\sum_{i=1}^n \sum_{j=1}^n x_{i,n} x'_{j,n} v_i(\hat\theta) v_j(\hat\theta)\right] \\
&= \frac{1}{\sigma^4 n}\sum_{i=1}^n x_{i,n} x'_{i,n} E(v_i(\hat\theta)^2).
\end{aligned}$$

Based on the property of $v_i(\theta)$, under $\hat\theta$, $E(v_i(\hat\theta)) = 0$. Since

$$\begin{aligned}
E\left[(y_{i,n} - x'_{i,n}\beta)^2 I(y_{i,n} > 0)\right] &= \int_{-x'_i\beta}^{+\infty} (x'_i\beta)^2 \phi(z,\sigma^2)dz + \int_{-x'_i\beta}^{+\infty} 2(x'_i\beta)\phi(z,\sigma^2)dz + \int_{-x'_i\beta}^{+\infty} (-\sigma^2 z)\phi(z,\sigma^2) d(\frac{-z^2}{2\sigma^2}) \\
&= (x'_{i,n}\beta)^2 \Phi(x'_{i,n}\beta.\sigma^2) + x'_{i,n}\beta\sigma^2\phi(x'_{i,n}\beta,\sigma^2) + \sigma^2\Phi(x'_{i,n}\beta,\sigma^2),
\end{aligned}$$

then

$$\begin{aligned}
E[v_i(\hat\theta)^2] &= E\left[[(y_{i,n} - x'_{i,n}\beta) I(y_{i,n} > 0) - \frac{\phi(x'_{i,n}\beta,\sigma^2)\sigma^2}{(1-\Phi(x'_{i,n}\beta,\sigma^2))} I(y_{i,n} = 0)]^2\right] \\
&= \sigma^2\left[\Phi(x'_{i,n}\beta,\sigma^2) - \frac{x'_{i,n}\beta\phi(x'_{i,n}\beta,\sigma^2)(1-\Phi(x'_{i,n}\beta,\sigma^2)) - \sigma^2\phi(x'_{i,n}\beta,\sigma^2)^2}{1-\Phi(x'_{i,n}\beta,\sigma^2)}\right].
\end{aligned}$$

Denoting $h_i(\hat\theta) = \frac{1}{\sigma^2}E[v_i(\hat\theta)^2]$, J_β can be simplified as

$$J_\beta = \frac{1}{\sigma^2 n}\sum_{i=1}^n x_{i,n} x'_{i,n} h_i(\hat\theta).$$

1.2.3.2 $J_{\beta\sigma^2}$

For $J_{\beta\sigma^2}$,

$$\begin{aligned}
J_{\beta\sigma^2}(\hat\theta) &= \frac{1}{n}E[d_\beta(\hat\theta) d_{\sigma^2}(\hat\theta)] \\
&= \frac{1}{n}E\left[\frac{1}{\sigma^2}\sum_{i=1}^n x_{i,n} v_i(\hat\theta)\frac{1}{2\sigma^4}\sum_{i=1}^n q_i(\hat\theta)\right] \\
&= \frac{1}{2\sigma^6 n}E\left[\sum_{i=1}^n\sum_{j=1}^n x_{i,n} v_i(\hat\theta) q_j(\hat\theta)\right] \\
&= \frac{1}{2\sigma^4 n}\sum_{i=1}^n x_{i,n}\left[(x'_i\beta)^2\phi_{i,n}(x'_{i,n}\beta,\sigma^2) + \sigma^2\phi_{i,n}(x'_{i,n}\beta,\sigma^2) - \frac{\sigma^2 x'_i\beta\phi^2_{i,n}(x'_{i,n}\beta,\sigma^2)}{1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2)}\right].
\end{aligned}$$

1.2.3.3 $J_{\beta\lambda}$

For $J_{\beta\lambda}$,

$$\begin{aligned} J_{\beta\lambda}(\hat{\theta}) &= \frac{1}{n}E[d_{\beta}(\hat{\theta})d_{\lambda}(\hat{\theta})] \\ &= \frac{1}{\sigma^2 n}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} w_{ij}x_{i,n}x'_{j,n}\beta h_i(\hat{\theta}). \end{aligned}$$

1.2.3.4 $J_{\beta\rho}$

For $J_{\beta\rho}$,

$$\begin{aligned} J_{\beta\rho}(\hat{\theta}) &= \frac{1}{n}E[d_{\beta}(\hat{\theta})d_{\rho}(\hat{\theta})] \\ &= \mathbf{0}_{k\times 1}. \end{aligned}$$

1.2.3.5 J_{σ^2}

Since

$$\begin{aligned} E\left[q_i(\hat{\theta})^2\right] &= E\left[((y_{i,n}-x'_{i,n}\beta)^2-\sigma^2)^2 I(y_{i,n}>0)+\left(\frac{\sigma^2 x'_{i,n}\beta\phi(x'_{i,n}\beta,\sigma^2)}{1-\Phi(x'_{i,n}\beta,\sigma^2)}\right)^2 I(y_{i,n}=0)\right] \\ &= \frac{\sigma^4(x'_{i,n}\beta)^2\phi^2(x'_{i,n}\beta,\sigma^2)}{1-\Phi(x'_{i,n}\beta,\sigma^2)}+\sigma^2[-(x'_{i,n}\beta)^3\phi(z,\sigma^2)-3\sigma^2 x'_{i,n}\beta\phi(x'_{i,n}\beta,\sigma^2)+3\sigma^2\Phi(x'_{i,n}\beta,\sigma^2)] \\ &\quad -2\sigma^2\sigma^2[-x'_{i,n}\beta\phi(x'_{i,n}\beta,\sigma^2)+\Phi(x'_{i,n}\beta,\sigma^2)]+\sigma^4\Phi(x'_{i,n}\beta,\sigma^2) \\ &= \sigma^2\left[-(x'_{i,n}\beta)^3\phi(x'_{i,n}\beta,\sigma^2)-\sigma^2(x'_{i,n}\beta)\phi(x'_{i,n}\beta,\sigma^2)+2\sigma^2\Phi(x'_{i,n}\beta,\sigma^2)+\frac{\sigma^2(x'_{i,n}\beta)^2\phi(x'_{i,n}\beta,\sigma^2)^2}{1-\Phi(x'_{i,n}\beta,\sigma^2)}\right]. \end{aligned}$$

Then

$$\begin{aligned} J_{\sigma^2}(\hat{\theta}) &= \frac{1}{n}E[d_{\sigma^2}(\hat{\theta})^2] \\ &= \frac{1}{n}E\left[\frac{1}{4\sigma^8}\sum_{i=1}^{n} q_i(\hat{\theta})^2\right] \\ &= \frac{1}{4\sigma^6 n}\sum_{i=1}^{n}\left[-(x'_{i,n}\beta)^3\phi(x'_{i,n}\beta,\sigma^2)-\sigma^2(x'_{i,n}\beta)\phi(x'_{i,n}\beta,\sigma^2)+2\sigma^2\Phi(x'_{i,n}\beta,\sigma^2)+\frac{\sigma^2(x'_{i,n}\beta)^2\phi^2(x'_{i,n}\beta,\sigma^2)}{1-\Phi(x'_{i,n}\beta,\sigma^2)}\right]. \end{aligned}$$

1.2.3.6 $J_{\sigma^2\lambda}$

For $J_{\sigma^2\lambda}$,

$$\begin{aligned}
\frac{1}{n}E[d_{\sigma^2}(\hat{\theta})d_{\lambda}(\hat{\theta})] &= \frac{1}{n}E[\frac{1}{2\sigma^4}\sum_{i=1}^{n} q_i(\hat{\theta})(\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta}) + \frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})x'_{i,n}\beta)] \\
&= \frac{1}{n}E[\frac{1}{2\sigma^4}\sum_{i=1}^{n} q_i(\hat{\theta})(\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})x'_{i,n}\beta)] = \frac{1}{2\sigma^6 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}x'_{i,n}\beta E[v_i(\hat{\theta})q_i(\hat{\theta})] \\
&= \frac{1}{2\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}x'_{j,n}\beta\left[\frac{-\sigma^2 x'_{i,n}\beta\phi^2(x'_{i,n}\beta,\sigma^2)}{1-\Phi(x'_{i,n}\beta,\sigma^2)} + (x'_{i,n}\beta)^2\phi(x'_{i,n}\beta,\sigma^2) + \sigma^2\phi(x'_{i,n}\beta,\sigma^2)\right].
\end{aligned}$$

1.2.3.7 $J_{\sigma^2\rho}$

For $J_{\sigma^2\rho}$,

$$\frac{1}{n}E[d_{\sigma^2}(\hat{\theta})d_{\rho}(\hat{\theta})] = \frac{1}{n}E[\frac{1}{2\sigma^4}\sum_{i=1}^{n} q_i(\hat{\theta})\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})] = 0.$$

1.2.3.8 J_{ρ}

$$\begin{aligned}
&\frac{1}{n}E[d_{\rho}(\hat{\theta})d_{\rho}(\hat{\theta})] = \frac{1}{n}E[\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})] \\
&= \frac{1}{n}E[\frac{1}{\sigma^4}\sum_{i,j=1}^{n(j\neq i)}\sum_{a,b=1}^{n(a\neq b)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})m_{ab,n}v_a(\hat{\theta})v_b(\hat{\theta})] \\
&= \frac{1}{\sigma^4 n}E[\sum_{i,j=1(j\neq i)}^{n(a=i,b=j)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})m_{ab,n}v_a(\hat{\theta})v_b(\hat{\theta}) + \sum_{i,j=1(j\neq i)}^{n(a=j,b=i)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})m_{ab,n}v_a(\hat{\theta})v_b(\hat{\theta}) \\
&\quad + \sum_{i,j=1}^{n(j\neq i)}\sum_{a,b=1(a\neq b)}^{n(other)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})m_{ab,n}v_a(\hat{\theta})v_b(\hat{\theta})] \\
&= \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n}(m_{ij,n}^2 + m_{ij,n}m_{ji,n})E[v_i(\hat{\theta})^2]E[v_j(\hat{\theta})^2].
\end{aligned}$$

1.2.3.9 $J_{\rho\lambda}$

For $J_{\rho\lambda}$, we have

$$\begin{aligned}
\frac{1}{n}E[d_{\rho}(\hat{\theta})d_{\lambda}(\hat{\theta})] &= \frac{1}{n}E[\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})(\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta}) + \frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})x'_{j,n}\beta)] \\
&= \frac{1}{\sigma^4 n}E[\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})\sum_{a,b=1}^{n(a\neq b)} w_{ab,n}v_a(\hat{\theta})v_b(\hat{\theta})] \\
&= \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}(w_{ij,n} + w_{ji,n})E[v_i(\hat{\theta})^2]E[v_j(\hat{\theta})^2].
\end{aligned}$$

1.2.3.10 J_λ

$$
\begin{aligned}
J_\lambda(\hat\theta) &= \frac{1}{n}E[d_\lambda(\hat\theta)^2] \\
&= \frac{1}{n}E\left[(\frac{1}{\sigma^2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_{ij}v_i(\hat\theta)v_j(\hat\theta) + \frac{1}{\sigma^2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_{ij}v_i(\hat\theta)x_j'\beta)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_{ij}(w_{ij,n}+w_{ji,n})h_i(\hat\theta)h_j(\hat\theta) + \frac{1}{n}\frac{1}{\sigma^2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\sum_{t=1,t\neq i,j}^n w_{ij}w_{it}x_j'\beta x_t'\beta h_i(\hat\theta) \\
&+ \frac{1}{n}\frac{1}{\sigma^2}\sum_{i=1}^n\sum_{j=1,j\neq i}^n w_{ij}^2(x_j'\beta)^2h_i(\hat\theta).
\end{aligned}
$$

1.3 Simultaneous SARAR Model

1.3.1 Model Construction

Similar to the settings in latent SARAR Tobit model, the simultaneous SARAR Tobit model can be expressed as

$$Y_n^* = \lambda W_nY_n + X_n\beta + u_n, \quad u_n = \rho M_nu_n + \varepsilon_n, \qquad y_{i,n} = \max(0, y_{i,n}^*).$$

Let $R_n(\rho) = I_n - \rho M_n$, then

$$Y_n^* = \lambda W_nY_n + X_n\beta + R_n^{-1}(\rho)\varepsilon_n, \quad \varepsilon_n = R_n(\rho)(Y_n^* - \lambda W_nY_n) - R_n(\rho)X_n\beta.$$

Denote $Y_n = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$, and correspondingly $Y_n^* = \begin{pmatrix} Y_1^* \\ Y_2^* \end{pmatrix}$, where all elements of $Y_1 = 0$, and all elements of Y_2 are positive. Then Y_1^* and Y_2^* can be written as

$$
\begin{aligned}
Y_1^* &= \lambda W_{12}Y_2 + X_1\beta + (R^{-1}(\rho)\varepsilon)_1 = \lambda 0_{11}Y_1^* + \lambda W_{12}Y_2^* + X_1\beta + (R^{-1}(\rho)\varepsilon)_1 \\
Y_2^* &= \lambda W_{22}Y_2 + X_2\beta + (R^{-1}(\rho)\varepsilon)_2 = \lambda 0_{21}Y_1^* + \lambda W_{22}Y_2^* + X_2\beta + (R^{-1}(\rho)\varepsilon)_2.
\end{aligned}
$$

Denote $W_{Y11} = 0, W_{Y12} = W_{12}, W_{Y21} = 0, W_{Y22} = W_{22}$, then

$$Y_n^* = \lambda W_{Yn}Y_n^* + X_n\beta + R_n^{-1}(\rho)\varepsilon_n.$$

Let $S_{Yn}(\lambda) = I_n - \lambda W_{Yn}$, then

$$Y_n^* = S_{Yn}^{-1}(\lambda)X_n\beta + S_{Yn}^{-1}(\lambda)R_n^{-1}(\rho)\varepsilon_n, \quad \varepsilon_n = R_n(\rho)S_{Yn}(\lambda)Y_n^* - R_n(\rho)X_n\beta.$$

This equation is the same as the latent SARAR Model, but $w_{ij,n} = 0$ for $y_{j,n} = 0$. When $\lambda = 0, \rho = 0$

$$S_n = I_n - \lambda W_n = I_n, \quad R_n = I_n - \rho M_n = I_n.$$

For $\hat{\theta} = (\beta, \sigma^2, \lambda = 0, \rho = 0)'$, similar to the latent model, it can be computed that

$$\begin{aligned} v_i(\hat{\theta}) =& E_\theta(y^*_{i,n} - x'_{i,n}\beta | X_n, Y_n) \\ =& \frac{-\sigma^2 \phi(x'_{i,n}\beta, \sigma^2)}{1 - \Phi(x'_{i,n}\beta, \sigma^2)} I(y_{i,n} = 0) + (y_{i,n} - x'_{i,n}\beta) I(y_{i,n} > 0) \end{aligned}$$

$$\begin{aligned} q_i(\hat{\theta}) &= E_\theta((y^*_{i,n} - x'_{i,n}\beta)^2 - \sigma^2 | X_n, Y_n) \\ &= \frac{\sigma^2 x'_{i,n}\beta \phi(x'_{i,n}\beta, \sigma^2)}{1 - \Phi(x'_{i,n}\beta, \sigma^2)} I(y_{i,n} = 0) + ((y_{i,n} - x'_{i,n}\beta)^2 - \sigma^2) I(y_{i,n} > 0). \end{aligned}$$

1.3.2 Score Functions

The score functions for simultaneous SAR Tobit model have the following forms:

$$d_\beta(\hat{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_{i,n} v_i(\hat{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_{i,n} \left[\frac{-\sigma^2 \phi(x'_{i,n}\beta, \sigma^2)}{1 - \Phi(x'_{i,n}\beta, \sigma^2)} I(y_{i,n} = 0) + (y_{i,n} - x'_{i,n}\beta) I(y_{i,n} > 0) \right] \tag{1.3.1}$$

$$d_{\sigma^2}(\hat{\theta}) = \frac{1}{2\sigma^4} \sum_{i=1}^{n} q_i(\hat{\theta}) = \frac{1}{2\sigma^4} \sum_{i=1}^{n} \left[\frac{\sigma^2 x'_{i,n}\beta \phi(x'_{i,n}\beta, \sigma^2)}{1 - \Phi(x'_{i,n}\beta, \sigma^2)} I(y_{i,n} = 0) + ((y_{i,n} - x'_{i,n}\beta)^2 - \sigma^2) I(y_{i,n} > 0) \right] \tag{1.3.2}$$

$$d_\lambda(\hat{\theta}) = \frac{1}{\sigma^2} \sum_{i,j=1}^{n(j \neq i)} w_{ij,n} v_i(\hat{\theta}) [(y_{j,n} - x'_{j,n}\beta) I(y_{j,n} > 0)] + \frac{1}{\sigma^2} \sum_{i,j=1}^{n(j \neq i)} w_{ij,n} v_i(\hat{\theta}) x'_{j,n}\beta \tag{1.3.3}$$

$$d_\rho(\hat{\theta}) = \frac{1}{\sigma^2} \sum_{i,j=1}^{n(j \neq i)} m_{ij,n} v_i(\hat{\theta}) v_j(\hat{\theta}). \tag{1.3.4}$$

The alteration of $d_\lambda(\hat{\theta})$ is the main distinction from the latent SARAR Tobit model.

1.3.3 The Information Matrix

Similar to latent SARAR Tobit model, considering the variation of score function d_λ, then there are 4 items containing $d_\lambda(\hat{\theta})$ that have changed correspondingly.

1.3.3.1 $J_{\beta\lambda}$

$$\begin{aligned} &\frac{1}{n} E[d_\beta(\hat{\theta}) d_\lambda(\hat{\theta})] \\ &\quad = \frac{1}{n} E[\frac{1}{\sigma^2} \sum_{i=1}^{n} x_{i,n} v_i(\hat{\theta}) (\frac{1}{\sigma^2} \sum_{i,j=1}^{n(j \neq i)} w_{ij,n} v_i(\hat{\theta}) [(y_{j,n} - x'_{i,n}\beta_{j,n}) I(y_{j,n} > 0)] + \frac{1}{\sigma^2} \sum_{i,j=1}^{n(j \neq i)} w_{ij,n} v_i(\hat{\theta}) x'_{j,n}\beta)] \\ &\quad = \frac{1}{\sigma^4 n} \sum_{i,j=1}^{n(j \neq i)} x_{i,n} w_{ij,n} E[v_i(\hat{\theta})^2] E[(y_{j,n} - x'_{j,n}\beta) I(y_{j,n} > 0)] + \frac{1}{\sigma^4 n} \sum_{i,j=1}^{n(j \neq i)} x_{i,n} w_{ij,n} x'_{j,n}\beta E[v_i(\hat{\theta})^2] \\ &\quad = \frac{1}{\sigma^2 n} \sum_{i,j=1}^{n(j \neq i)} x_{i,n} w_{ij,n} [\sigma^2 \phi(x'_{j,n}\beta, \sigma^2) + x'_{j,n}\beta] \left[\frac{\sigma^2 \phi^2(x'_{i,n}\beta, \sigma^2)}{1 - \Phi(x'_{i,n}\beta, \sigma^2)} - x'_{i,n}\beta \phi(x'_{i,n}\beta, \sigma^2) + \Phi(x'_{i,n}\beta, \sigma^2) \right] \\ &\quad = \frac{1}{\sigma^2 n} \sum_{i,j=1}^{n(j \neq i)} x_{i,n} w_{ij,n} [\sigma^2 \phi(x'_{j,n}\beta, \sigma^2) + x'_{j,n}\beta] h_i(\hat{\theta}). \end{aligned}$$

1.3.3.2 $J_{\sigma^2\lambda}$

$$
\begin{aligned}
&\frac{1}{n}E[d_{\sigma^2}(\hat{\theta})d_{\lambda}(\hat{\theta})] \\
&\quad= \frac{1}{n}E[\frac{1}{2\sigma^4}\sum_{i=1}^{n} q_i(\hat{\theta})(\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})[(y_{j,n}-x'_{i,n}\beta_{j,n})I(y_{j,n}>0)] + \frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})x'_{j,n}\beta)] \\
&\quad= \frac{1}{2\sigma^6 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}E[v_i(\hat{\theta})q_i(\hat{\theta})]E[(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)] + \frac{1}{2\sigma^6 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}x'_{j,n}\beta E[v_i(\hat{\theta})q_i(\hat{\theta})] \\
&\quad= \frac{1}{2\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}[\sigma^2\phi(x'_{j,n}\beta,\sigma^2)+x'_{j,n}\beta]\left[\frac{-\sigma^2 x'_{i,n}\beta\phi^2(x'_{i,n}\beta,\sigma^2)}{1-\Phi(x'_{i,n}\beta,\sigma^2)} + x'_{i,n}\beta^2\phi(x'_{i,n}\beta,\sigma^2) + \sigma^2\phi(x'_{i,n}\beta,\sigma^2)\right].
\end{aligned}
$$

1.3.3.3 $J_{\rho\lambda}$

$$
\begin{aligned}
&\frac{1}{n}E[d_{\rho}(\hat{\theta})d_{\lambda}(\hat{\theta})] \\
&\quad= \frac{1}{n}E\left[\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})(\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})[(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)] + \frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})x'_{j,n}\beta)\right] \\
&\quad= \frac{1}{\sigma^4 n}E\left[\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}v_i(\hat{\theta})v_j(\hat{\theta})\sum_{a,b=1}^{n(a\neq b)} w_{ab,n}v_a(\hat{\theta})[(y_{b,n}-\mu_{b,n})I(y_{b,n}>0)]\right] \\
&\quad= \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}w_{ij,n}E[v_i(\hat{\theta})^2]E[v_j(\hat{\theta})(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)] \\
&\quad+ \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} m_{ij,n}w_{ji,n}E[v_j(\hat{\theta})^2]E[v_i(\hat{\theta})(y_{i,n}-x'_{i,n}\beta)I(y_{i,n}>0)] \\
&\quad= \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} (m_{ij,n}w_{ij,n}+m_{ji,n}w_{ji,n})E[v_i(\hat{\theta})^2]E[v_j(\hat{\theta})(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)],
\end{aligned}
$$

where

$$
E[v_i(\hat{\theta})^2] = \sigma^2\left[\frac{\sigma^2\phi^2(x'_{i,n}\beta,\sigma^2)}{1-\Phi(x'_{i,n}\beta,\sigma^2)} - x'_{i,n}\beta\phi(x'_{i,n}\beta,\sigma^2) + \Phi(x'_{i,n}\beta,\sigma^2)\right] \tag{1.3.5}
$$

$$
E[v_j(\hat{\theta})(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)] = E[(y_{j,n}-x'_{j,n}\beta)^2 I(y_{j,n}>0)] = \sigma^2[-x'_{j,n}\beta\phi(x'_{j,n}\beta,\sigma^2)+\Phi(x'_{j,n}\beta,\sigma^2)]. \tag{1.3.6}
$$

1.3.3.4 J_λ

$$
\begin{aligned}
&\frac{1}{n}E[d_\lambda(\hat{\theta})d_\lambda(\hat{\theta})] \\
&\quad= \frac{1}{n}E\left[(\frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})[(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)] + \frac{1}{\sigma^2}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}v_i(\hat{\theta})x'_{j,n}\beta)^2\right] \\
&\quad= \frac{1}{n}E\left[\frac{1}{\sigma^4}\sum_{i,j=1;a,b=1}^{n(j\neq i,b\neq a)} w_{ij,n}v_i(\hat{\theta})[(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)]w_{ab,n}v_a(\hat{\theta})[(y_{b,n}-\mu_{b,n})I(y_{b,n}>0)]\right] \\
&\quad+ \frac{1}{n}E\left[\frac{2}{\sigma^4}\sum_{i,j=1;a,b=1}^{n(j\neq i,b\neq a)} w_{ij,n}v_i(\hat{\theta})[(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)]w_{ab,n}v_a(\hat{\theta})\mu_{b,n}\right] \\
&\quad+ \frac{1}{n}E\left[\frac{1}{\sigma^4}\sum_{i,j=1;a,b=1}^{n(j\neq i,b\neq a)} w_{ij,n}v_i(\hat{\theta})x'_{j,n}\beta w_{ab,n}v_a(\hat{\theta})\mu_{b,n}\right] \\
&\quad= \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}^2 E\left[v_i(\hat{\theta})^2\right]E[(y_{j,n}-x'_{j,n}\beta)^2 I(y_{j,n}>0)] \\
&\quad+ \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}w_{ji,n}E\left[v_i(\hat{\theta})[(y_{i,n}-x'_{i,n}\beta)I(y_{i,n}>0)]\right]E\left[v_j(\hat{\theta})[(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)]\right] \\
&\quad+ \frac{2}{\sigma^4 n}\sum_{i,j,k=1}^{n(j\neq i,k\neq i)} w_{ij,n}w_{ik,n}\mu_{k,n}E\left[v_i(\hat{\theta})^2\right]E[(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)] \\
&\quad+ \frac{1}{\sigma^4 n}\sum_{i,j,k=1}^{n(j\neq i,k\neq i)} w_{ij,n}w_{ik,n}x'_{j,n}\beta\mu_{k,n}E\left[v_i(\hat{\theta})^2\right] \\
&\quad= \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}^2 E\left[v_i(\hat{\theta})^2\right]E[(y_{j,n}-x'_{j,n}\beta)^2 I(y_{j,n}>0)] \\
&\quad+ \frac{1}{\sigma^4 n}\sum_{i,j=1}^{n(j\neq i)} w_{ij,n}w_{ji,n}E\left[(y_{i,n}-x'_{i,n}\beta)^2 I(y_{i,n}>0)\right]E\left[(y_{j,n}-x'_{j,n}\beta)^2 I(y_{j,n}>0)\right] \\
&\quad+ \frac{2}{\sigma^4 n}\sum_{i,j,k=1}^{n(j\neq i,k\neq i)} w_{ij,n}w_{ik,n}x'_{k,n}\beta E\left[v_i(\hat{\theta})^2\right]E[(y_{j,n}-x'_{j,n}\beta)I(y_{j,n}>0)] \\
&\quad+ \frac{1}{\sigma^4 n}\sum_{i,j,k=1}^{n(j\neq i,k\neq i)} w_{ij,n}w_{ik,n}x'_{j,n}\beta x'_{k,n}\beta E\left[v_i(\hat{\theta})^2\right].
\end{aligned}
$$

1.4 Asymptotic Distributions of Test Statistics

1.4.1 Asymptotic Distribution of Conventional LM Test Statistic

Now we denote $\tilde{\theta} = (\hat{\beta}, \hat{\sigma}^2, 0, 0)'$, where $\hat{\beta}$ and $\hat{\sigma}^2$ are the maximum likelihood estimators of β and σ^2, which lead to the discussion on the test statistics under different hypotheses.

(1) $H_0^a : \rho = 0$.

(2) $H_0^b : \lambda = 0$.

(3) $H_A^a : \rho = \rho_0 + \frac{\delta_\rho}{\sqrt{n}}$.

(4) $H_A^b : \lambda = \lambda_0 + \frac{\delta_\lambda}{\sqrt{n}}$.

Under H_0^a, the conventional LM test statistic is

$$LM_\rho = \frac{1}{n} d_\rho(\widetilde{\theta})' J_{\rho\cdot\gamma}^{-1}(\widetilde{\theta}) d_\rho(\widetilde{\theta})$$

where $J_{\rho\cdot\gamma}(\widetilde{\theta}) = J_\rho(\widetilde{\theta}) - J_{\rho\gamma}(\widetilde{\theta}) J_\gamma^{-1}(\widetilde{\theta}) J_{\gamma\rho}(\widetilde{\theta})$. Under H_0^a (or $\widetilde{\theta}$), we have $J_{\rho\beta}(\widetilde{\theta}) = \mathbf{0}_{1\times k}$, $J_{\rho\sigma^2}(\widetilde{\theta}) = 0$ and $J_\gamma(\widetilde{\theta})$ is nonsingular under the conditions of Amemiya (1973). Thus we have $J_{\rho\cdot\gamma}(\widetilde{\theta}) = J_\rho(\widetilde{\theta})$. Let the log-likelihood function to be $L(\gamma, \rho, \lambda)$. Suppose we set $\lambda = 0$ have the log-likelihood function as $L_1(\gamma, \rho, 0)$. Then we set $\rho = 0$ and have the log-likelihood function $L_2(\gamma, 0, \lambda)$. If $L_1(\gamma, \rho, 0)$ is the true model, then it is obvious that LM_ρ follows a chi-square distribution. In our case, we can simplify LM_ρ as

$$LM_\rho = \frac{1}{n} \frac{d_\rho(\widetilde{\theta})^2}{J_\rho(\widetilde{\theta})} = \frac{1}{n} \frac{d_\rho(\widetilde{\theta})^2}{\frac{1}{n} E[d_\rho(\widetilde{\theta})^2]} = \frac{d_\rho(\widetilde{\theta})^2}{Var[d_\rho(\widetilde{\theta})^2]} \to_d \chi_1^2(0).$$

And according to Davidson and MacKinnon (1987) and Saikkonen (1989), if true model is $L_2(\gamma, 0, \lambda)$, and under H_A^b,

$$LM_\rho \to_d \chi_1^2(\delta_\lambda^2 J_{\lambda\rho\cdot\gamma} J_{\rho\cdot\gamma}^{-1} J_{\rho\lambda\cdot\gamma})$$

where $J_{\rho\lambda\cdot\gamma}(\widetilde{\theta}) = J_{\rho\lambda}(\widetilde{\theta}) - J_{\rho\gamma}(\widetilde{\theta}) J_\gamma^{-1}(\widetilde{\theta}) J_{\gamma\lambda}(\widetilde{\theta})$.

Under H_0^b, the conventional test statistic is

$$LM_\lambda = \frac{1}{n} d_\lambda(\widetilde{\theta})' J_{\lambda\cdot\gamma}^{-1}(\widetilde{\theta}) d_\lambda(\widetilde{\theta})$$

where $J_{\lambda\cdot\gamma}(\widetilde{\theta}) = J_\lambda(\widetilde{\theta}) - J_{\lambda\gamma}(\widetilde{\theta}) J_\gamma^{-1}(\widetilde{\theta}) J_{\gamma\lambda}(\widetilde{\theta})$. Since γ is a $(k+1) \times 1$ vector, $J_\gamma(\widetilde{\theta})$ is a $(k+1) \times (k+1)$ matrix, $J_{\lambda\gamma}(\widetilde{\theta})$ is a $1 \times (k+1)$ vector and $J_{\gamma\lambda}(\widetilde{\theta})$ is a $(k+1) \times 1$ vector. Similarly, if $\rho = 0$ is the true model, under H_0^b, it is obvious that

$$LM_\lambda \to_d \chi_1^2(0)$$

Similarly, if $L_1(\gamma, \rho, 0)$ is the true model, under H_A^a,

$$LM_\lambda \to_d \chi_1^2(\delta_\rho^2 J_{\rho\lambda\cdot\gamma} J_{\lambda\cdot\gamma}^{-1} J_{\lambda\rho\cdot\gamma})$$

In Qu and Lee [17], there are five types of LM test statistics. The first one is the LM test for simultaneous SAR Tobit model. The second one is LM test for the latent SAR Tobit model. The third is for the latent SE Tobit model. The fourth is latent SAR Probit model, and the fifth is the latent SE Probit model. After comparing with our test statistics, our LM_ρ for the latent SARAR Tobit model is just the LM test statistic for latent SE model in Qu and Lee [17], which takes the following form:

$$\frac{g_{SE}(\hat{\alpha})}{\tilde{\sigma}_{Q_n}} \sim N(0,1),$$

where

$$g_{SE}(\hat{\alpha}) = \tilde{\varepsilon}_n' W_n \tilde{\varepsilon}_n,$$

which is just the d_ρ in this chapter; and

$$\sigma^2_{Q_n} = tr(W_n\Sigma_n W_n\Sigma_n + W_n'\Sigma_n W_n\Sigma_n),$$

where $\Sigma_n = diag(E(\varepsilon^2_{i,n}(\alpha))$ and

$$E(\varepsilon^2_{i,n}(\alpha)) = \frac{\phi(x_{i,n}'\beta,\sigma^2)^2\Phi(x_{i,n}'\beta,\sigma^2)}{1-\Phi(x_{i,n}'\beta,\sigma^2)} + \phi(x_{i,n}'\beta,\sigma^2)^2 + \Phi(x_{i,n}'\beta,\sigma^2)^2 - x_{i,n}'\beta\phi(x_{i,n}'\beta,\sigma^2),$$

which is just J_ρ in this chapter.

Similarly, the LM_λ under the latent SARAR Tobit model in our chapter is just the LM test statistic for latent SAR Tobit model in Qu and Lee [17], which has the following form:

$$\frac{g_{SAR}(\hat{\alpha})}{\tilde{\sigma}_{Q_n}} \sim N(0,1),$$

where

$$g_{SAR}(\hat{\alpha}) = \tilde{\varepsilon}_n' W_n\hat{\varepsilon}_n + \tilde{\varepsilon}_n' W_n X_n\tilde{\beta},$$

which is just the d_λ in this chapter; and

$$\begin{aligned}\sigma^2_{Q_n} &= tr(W_n\Sigma_n W_n\Sigma_n + W_n'\Sigma_n W_n\Sigma_n) + (W_nX_n\beta)'\Sigma_n W_nX_n\beta \\ &\quad + E(\frac{\partial g_{SAR}(\hat{\alpha})}{\partial\alpha'})[E\frac{\partial^2\ln L_n(\alpha)}{\partial\alpha\partial\alpha'}]^{-1}E(\frac{\partial g_{SAR}(\hat{\alpha})}{\partial\alpha}),\end{aligned}$$

where $\Sigma_n = diag(E(\varepsilon^2_{i,n}(\alpha))$, and

$$E(\varepsilon^2_{i,n}(\alpha)) = \frac{\phi(x_{i,n}'\beta,\sigma^2)^2\Phi(x_{i,n}'\beta,\sigma^2)}{1-\Phi(x_{i,n}'\beta,\sigma^2)} + \phi(x_{i,n}'\beta,\sigma^2)^2 + \Phi(x_{i,n}'\beta,\sigma^2)^2 - x_{i,n}'\beta\phi(x_{i,n}'\beta,\sigma^2).$$

The whole term of $\sigma^2_{Q_n}$ is just $J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})$ in this paper.

For the simultaneous SARAR Tobit model, the LM_λ under the simultaneous setting is just the LM test statistic for simultaneous SAR Tobit model in Qu and Lee [17], which has the following form:

$$\frac{g_n(\hat{\alpha})}{\tilde{\sigma}_{Q_n}} \sim N(0,1),$$

where

$$g_n(\hat{\alpha}) = \tilde{\varepsilon}_n' W_n\tilde{\varepsilon}_n + \tilde{\varepsilon}_n' W_n(X_n\tilde{\beta}),$$

which is just the d_λ in this chapter. And $\tilde{\sigma}_{Q_n}$ has the similar form as above, which is exactly $J_{\lambda\dot{\gamma}}$ in this chapter.

1.4.1.1 Robust LM Test Statistic

According to Bera and Yoon [16] and Anselin et al. [15], the modified robust LM test statistic is as follows

$$LM_\rho^* = \frac{1}{n}[d_\rho(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\widetilde{\theta})d_\lambda(\widetilde{\theta})]'[J_{\rho\cdot\gamma}(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\widetilde{\theta})J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})]^{-1}[d_\rho(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\widetilde{\theta})d_\lambda(\widetilde{\theta})]$$

$$LM_\lambda^* = \frac{1}{n}[d_\lambda(\widetilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})J_{\rho\cdot\gamma}^{-1}(\widetilde{\theta})d_\rho(\widetilde{\theta})]'[J_{\lambda\cdot\gamma}(\widetilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})J_{\rho\cdot\gamma}^{-1}(\widetilde{\theta})J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})]^{-1}[d_\lambda(\widetilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})J_{\rho\cdot\gamma}^{-1}(\widetilde{\theta})d_\rho(\widetilde{\theta})].$$

According to Anselin et al. [15], for LM_ρ^*, $d_\rho(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J_{\lambda\cdot\gamma}^{-1}(\widetilde{\theta})d_\lambda(\widetilde{\theta})$ is the part of $d_\rho(\widetilde{\theta})$ that remains after eliminating the effect of $d_\lambda(\widetilde{\theta})$. It's shown in Bera and Yoon [16] that under H_0^a, LM_ρ^* has a central chi-square distribution, similarly for LM_λ^*, which is the following theorem.

The asymptotic distributions of the LM test statistics under SAR Tobit model settings have been well-proved by Qu and Lee [17]. We can further illustrate a brief proof of the modified LM test statistics in our SARAR Tobit model setting. We denote $\gamma_0 = (\beta_0', \sigma_0^2)'$, which is the true parameter contained in the interior of the compact set Θ. And $\hat{\gamma} = (\hat{\beta}', \hat{\sigma}^2)'$ is the MLE estimator for γ. Now we define assumptions as follows.

Assumption 1.4.1. *W and M spatial weight matrix is bounded in row and column sums norms.*

Assumption 1.4.2. *$\varepsilon_{i,n} \sim$ i.i.d. $N(0, \sigma^2)$ for all i.*

To show the asymptotic distribution of the LM test statistics, we need the following Lemma.

Lemma 1.4.1. *Under the null and Assumption 1.4.1 and 1.4.2, $\frac{1}{n}\frac{\partial^2 L(\gamma^*)}{\partial\gamma\partial\gamma'} - E\left(\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'}\right) \rightarrow_p 0$ where γ^* lies between the MLE estimator $\hat{\gamma}$ and the true parameter γ_0.*

The proof of Lemma 1.4.1 is in Section 7.

Using Taylor series expansion:

$$0 = \frac{1}{n}\frac{\partial L(\gamma)}{\partial \gamma}|_{\hat{\gamma}} = \frac{1}{n}\frac{\partial L(\gamma_0)}{\partial \gamma} + \frac{1}{n}\frac{\partial^2 L(\gamma)}{\partial\gamma\partial\gamma'}|_{\gamma^*}(\hat{\gamma} - \gamma_0),$$

where γ^* lies between γ_0 and $\hat{\gamma}$. Since $\frac{\partial L(\gamma)}{\partial \gamma}|_{\hat{\gamma}} = 0$,

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = \left(-\frac{1}{n}\frac{\partial^2 L(\gamma^*)}{\partial\gamma\partial\gamma'}\right)^{-1} \frac{1}{\sqrt{n}}\frac{\partial L(\gamma_0)}{\partial \gamma}.$$

Based on Lemma 1.4.1, using the Central Limit Theorem, we have the following result:

$$\sqrt{n}(\hat{\gamma} - \gamma_0) \rightarrow_d N(0, J^{-1}), \tag{1.4.1}$$

where $J = E\left(-\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'}\right)$.

Theorem 1.4.1. *Under assumptions in Amemiya [2] and also Theorem 2 in Qu and Lee [17], as $n \rightarrow \infty$, the following result hold.*
Under H_0^a and H_0^b,

$$LM_\rho^* \rightarrow_d \chi_1^2 \quad \textit{and} \quad LM_\lambda^* \rightarrow_d \chi_1^2.$$

Proof of Theorem 1.4.1: According to (1.4.1) and the results from Bera and Yoon [16], we have

$$\sqrt{n}\begin{bmatrix} d_\rho(\hat{\gamma},0,0) \\ d_\lambda(\hat{\gamma},0,0) \end{bmatrix} \rightarrow_d N\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} J_{\rho\cdot\gamma} & J_{\rho\lambda\cdot\gamma} \\ J_{\lambda\rho\cdot\gamma} & J_{\lambda\cdot\gamma} \end{bmatrix}\right]$$

Since we have our robust LM test with the modified score $d^*_\rho(\widetilde{\theta})$ as

$$d^*_\rho(\widetilde{\theta}) = [d_\rho(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})d_\lambda(\widetilde{\theta})],$$

then we can get

$$\sqrt{n}d^*_\rho(\widetilde{\theta}) = (1, -J_{\rho\lambda\cdot\gamma}J^{-1}_{\lambda\cdot\gamma})\sqrt{n}\mathbf{d}(\widetilde{\theta}) + o_p(1) \rightarrow_d N(0, J_{\rho\cdot\gamma}(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})),$$

which shows that $LM^*_\rho \rightarrow_d \chi^2_1(0)$. The similar procedure holds for LM^*_λ

When $H^a_0 : \rho = 0$, the conventional one-direction test is

$$LM_\rho = \frac{1}{n}d_\rho(\widetilde{\theta})' J^{-1}_{\rho\cdot\gamma}(\widetilde{\theta})d_\rho(\widetilde{\theta}) = \frac{1}{n}\frac{d_\rho(\widetilde{\theta})^2}{J_\rho(\widetilde{\theta})},$$

and the robust test statistic is

$$LM^*_\rho = \frac{1}{n}[d_\rho(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})d_\lambda(\widetilde{\theta})]'[J_{\rho\cdot\gamma}(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})]^{-1}[d_\rho(\widetilde{\theta}) - J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})d_\lambda(\widetilde{\theta})],$$

where $J_{\rho\lambda\cdot\gamma}(\widetilde{\theta}) = J_{\rho\lambda}(\widetilde{\theta}) - J_{\rho\gamma}(\widetilde{\theta})J^{-1}_{\gamma}(\widetilde{\theta})J_{\gamma\lambda}(\widetilde{\theta})$, $J_{\rho\cdot\gamma}(\widetilde{\theta}) = J_\rho(\widetilde{\theta}) - J_{\rho\gamma}(\widetilde{\theta})J^{-1}_{\gamma}(\widetilde{\theta})J_{\gamma\rho}(\widetilde{\theta})$, $J_{\lambda\cdot\gamma}(\widetilde{\theta}) = J_\lambda(\widetilde{\theta}) - J_{\lambda\gamma}(\widetilde{\theta})J^{-1}_{\gamma}(\widetilde{\theta})J_{\gamma\lambda}(\widetilde{\theta})$. Since we have $J_{\beta\rho} = \mathbf{0}_{k\times 1}$ and $J_{\sigma^2\rho} = 0$, then $J_{\rho\lambda\cdot\gamma}(\widetilde{\theta}) = J_{\rho\lambda}(\widetilde{\theta})$, $J_{\rho\cdot\gamma}(\widetilde{\theta}) = J_\rho(\widetilde{\theta})$. We can simplify LM^*_ρ as

$$LM^*_\rho = \frac{1}{n}[d_\rho(\widetilde{\theta}) - J_{\rho\lambda}(\widetilde{\theta})J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})d_\lambda(\widetilde{\theta})]'[J_\rho(\widetilde{\theta}) - J_{\rho\lambda}(\widetilde{\theta})J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})J_{\lambda\rho}(\widetilde{\theta})]^{-1}[d_\rho(\widetilde{\theta}) - J_{\rho\lambda}(\widetilde{\theta})J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})d_\lambda(\widetilde{\theta})].$$

When $H^b_0 : \lambda = 0$, the conventional one-direction test is

$$LM_\lambda = \frac{1}{n}d_\lambda(\widetilde{\theta})' J^{-1}_{\lambda\cdot\gamma}(\widetilde{\theta})d_\lambda(\widetilde{\theta}),$$

and the robust test statistic is

$$LM^*_\lambda = \frac{1}{n}[d_\lambda(\widetilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})J^{-1}_{\rho\cdot\gamma}(\widetilde{\theta})d_\rho(\widetilde{\theta})]'[J_{\lambda\cdot\gamma}(\widetilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})J^{-1}_{\rho\cdot\gamma}(\widetilde{\theta})J_{\rho\lambda\cdot\gamma}(\widetilde{\theta})]^{-1}[d_\lambda(\widetilde{\theta}) - J_{\lambda\rho\cdot\gamma}(\widetilde{\theta})J^{-1}_{\rho\cdot\gamma}(\widetilde{\theta})d_\rho(\widetilde{\theta})],$$

where $J_{\lambda\rho\cdot\gamma}(\widetilde{\theta}) = J_{\lambda\rho}(\widetilde{\theta}) - J_{\lambda\gamma}(\widetilde{\theta})J^{-1}_{\gamma}(\widetilde{\theta})J_{\gamma\rho}(\widetilde{\theta}) = J_{\lambda\rho}(\widetilde{\theta})$. The simplified version is

$$LM^*_\lambda = \frac{1}{n}[d_\lambda(\widetilde{\theta}) - J_{\lambda\rho}(\widetilde{\theta})J^{-1}_{\rho}(\widetilde{\theta})d_\rho(\widetilde{\theta})]'[J_{\lambda\cdot\gamma}(\widetilde{\theta}) - J_{\lambda\rho}(\widetilde{\theta})J^{-1}_{\rho}(\widetilde{\theta})J_{\rho\lambda}(\widetilde{\theta})]^{-1}[d_\lambda(\widetilde{\theta}) - J_{\lambda\rho}(\widetilde{\theta})J^{-1}_{\rho}(\widetilde{\theta})d_\rho(\widetilde{\theta})].$$

When $H_0 : \rho = 0, \lambda = 0$, the joint test statistic is

$$LM_{\rho\lambda} = \frac{1}{n} \begin{bmatrix} d_\rho(\widetilde{\theta}) \\ d_\lambda(\widetilde{\theta}) \end{bmatrix}' \begin{bmatrix} J_{\rho\cdot\gamma}(\widetilde{\theta}) & J_{\rho\lambda\cdot\gamma}(\widetilde{\theta}) \\ J_{\lambda\rho\cdot\gamma}(\widetilde{\theta}) & J_{\lambda\cdot\gamma}(\widetilde{\theta}) \end{bmatrix}^{-1} \begin{bmatrix} d_\rho(\widetilde{\theta}) \\ d_\lambda(\widetilde{\theta}) \end{bmatrix}$$

which has the simplified version as

$$LM_{\rho\lambda} = \frac{1}{n} \begin{bmatrix} d_\rho(\widetilde{\theta}) \\ d_\lambda(\widetilde{\theta}) \end{bmatrix}' \begin{bmatrix} J_{\rho}(\widetilde{\theta}) & J_{\rho\lambda}(\widetilde{\theta}) \\ J_{\lambda\rho}(\widetilde{\theta}) & J_{\lambda\cdot\gamma}(\widetilde{\theta}) \end{bmatrix}^{-1} \begin{bmatrix} d_\rho(\widetilde{\theta}) \\ d_\lambda(\widetilde{\theta}) \end{bmatrix}$$

1.5 Monte Carlo Results

In this section, we use Monte Carlo simulations to demonstrate the limit sample performance of our test statistics. 1000 Monte Carlo replications are performed for each simulation. In each table, N is the sample size; ρ is the spatial error parameter; λ is the spatial lag parameter. For our Monte Carlo study, we consider five test statistics LM_ρ, LM_ρ^*, LM_λ, LM_λ^* and $LM_{\rho\lambda}$, which are given below.

Similar to Anselin et. al.[15], the joint test statistic is also the sum of uncorrected one-direction test for one type of alternative and the adjusted form for the other alternative, which is

$$LM_{\rho\lambda} = LM_\lambda + LM_\rho^* = LM_\rho + LM_\lambda^* \tag{1.5.1}$$

First we verify the identities given in Equation (1.5.1).From Table A.1, we notice that the identities are satisfied. This observation also verifies all our algebraic derivation. The test statistics have complex expression; thus verifying Equation (1.5.1) is very difficult. Thus Table A.1 results are very reassuring.

[Table A.1 is here]

In the following tables, the performance of five test statistics are studied. For the latent SARAR Tobit model, LM_ρ and LM_λ are exactly the sqaure form of the LM test statistics for latent SE and latent SAR Tobit models in Qu and Lee [17]. For the simultaneous SARAR Tobit model, LM_λ is the square form of the LM test statistic for the simultaneous SAR Tobit model in Qu and Xi [17]. There is no corresponding test statistic for LM_ρ in Qu and Lee [17] under the simultaneous SARAR Tobit model setting. Table A.2 is a brief description of the test statistics involved in the Monte Carlo simulation study.

[Table A.2 is here]

1.5.1 Data Generating Processes

The model under null hypothesis of no spatial dependence is a classical Tobit model:

$$Y_n^* = X_n\beta + u_n$$

$$y_{i,n} = \max(0, y_{i,n}^*)$$

The explanatory variable X_n is a $N \times 2$ matrix. There are two types of model considered in this chapter, latent SARAR Tobit model and simulaneous SARAR Tobit model.

For the latent SARAR Tobit model, we let $x_{1i,n}$ to be intercept and $X_{2i,n} \sim N(1,1)$ as in Qu and Lee [17]. $u_i's$ are from normal distribution with mean equal to 0 and σ 6 and the coefficients $\beta = (1,1)'$. For each combination of parameters, 1000 replications are performed and the proportions of rejections are reported.

The choice of weight matrix is the same with Qu and Lee [17]. We construct the weight matrix W_n and M_n using Lesage's econometrics toolbox (Lesage and Pace [20]). For the construction of W_n and M_n, we firstly generate two random normal vectors of coordinates as geographic location, and then we let k to be the number of nearest neighbors for each observation and denote weight to be 1, otherwise 0. Finally we row-normalize W_n and M_n, where N is chosen to be 50 and 100 and k is 3 and 5.

The empirical sizes of the test for latent SARAR Tobit model are reported in Table A.3.

[Table A.3 is here]

For the simultaneous SARAR Tobit model, we let $x_{1i,n}$ to be intercept and $X_{2i,n} \sim N(1,1)$ as in Qu and Lee [17]. $u_i's$ are from normal distribution with mean 0 and σ 6. The coefficients $\beta = (1,-1)'$. For each combination of parameters, 1000 replications are performed and the proportions of rejections are reported. The empirical sizes of the test for simultaneous SARAR Tobit model are in Table A.4.

[Table A.4 is here]

The nominal significant levels in Table A.3 and Table A.4 are $\alpha = 0.05$ and $\alpha = 0.1$. According to Table A.3 and Table A.4, the sizes for LM_λ and $LM_{\rho\lambda}$ are not so good for they reject the null too often. LM_ρ and LM_ρ^* behave the best among the test statistics. For all combinations of parameters, their rejection frequencies are all within range. When $N = 50$ and $K = 3$, LM_ρ and LM_ρ^* under-reject the null while others over-reject at the same time. If we keep the sample size fixed and increase the number of neighbors, they all have experienced the over-rejection problem. Therefore, with fewer neighbors, the outcomes are substantially better. If we keep the number of neighbors fixed and increase the sample size from 50 to 100, all the test statistics fairly reject the null compared with a smaller sample size. The robust test statistic LM_ρ^* has the good size and no penalty for using it. For the other three statistics, they have experienced the over-rejection problem. This may be caused by the settings of the data generation. The same results apply to the simultaneous SARAR Tobit model.

In this chapter, we consider three types of alternative hypothesis of spatial dependence.

1.5.1.1 Spatial AR Error

In this setting we have the spatial AR error as follows.

$$u_n = \rho M_n u_n + \varepsilon_n$$

where $\varepsilon_{i,n}$ follows the normal distribution with mean 0 and $\sigma = 6$. We let the parameter ρ take values from 0.1 to 0.5. The tables and figures are as follows:

[Table A.5 is here]

[Table A.6 is here]

[Figure A.1 is here]

[Figure A.2 is here]

Table A.5 and Table A.6, as well as Figure A.1 and Figure A.2 illustrate the estimated rejection rates for the five test statistics against the spatial AR errors for both latent and simultaneous SARAR Tobit models. The first and seventh rows of Table A.5 and Table A.6 are just the empirical sizes for the five test statistics when both ρ and λ equal to 0. Rejection probabilities are presented in the remaining rows of the table, some of which can be interpreted as powers. Here ρ and λ change from 0.1 to 0.5 and the results for ρ and λ bigger than 0.5 are not reported in the table since their estimated rejection rates are just 1 or close to 1. LM_ρ and LM_ρ^* have the highest power, which means that they both have the power against spatial errors. When $N = 50$, $\rho > 0.4$, their power is higher than 90% rejection rates, and when $N = 100$, their power is higher than 90% even when $\rho > 0.3$. From Figure A.1, when $\lambda = 0$, it is obvious that employing the adjusted test LM_ρ^* over LM_ρ results in a minor reduction of power, which is due to the difference in the non-centrality parameter of $\chi_1^2(\delta' J_{\lambda\rho\cdot\gamma} J_{\rho\cdot\gamma}^{-1} J_{\rho\lambda\cdot\gamma}\delta)$. The findings indicate that employing robust test statistics rather than a one-directional alternative does not result in a significant loss of power.

According to Figure A.1, when ρ changes from 0.1 to 0.4, the power function of LM_λ^* is flatter than other test statistics, which represents the meaning of robustness of this robust LM test statistic, especially for small values of ρ. The joint test $LM_{\rho\lambda}$ also has good power against the AR error dependence, much better than LM_λ and close to LM_ρ. Since LM_ρ is a χ_1^2 test while $LM_{\rho\lambda}$ is a χ_2^2, it should come as no surprise that the former has higher power than that of the latter. When ρ value is high, the two tests are indistinguishable. The advantage of using LM_ρ^* is that when compared to LM_λ^*, it can help us to point to the correct alternative because LM_ρ^* has higher power than LM_λ^* under the AR error dependence. On the contrary, $LM_{\rho\lambda}$ can not provide the indication which alternative cause the misspecification. In comparison to the power of a smaller sample size, all statistics tend to have more power as the sample size grows.

1.5.1.2 Spatial AR Lag

The spatial AR lag has the following form for the latent SARAR Tobit model:

$$Y_n^* = \lambda W_n Y_n^* + X_n\beta + u_n, \qquad y_{i,n} = \max(0, y_{i,n}^*)$$

where $x_{1i,n}$ to be intercept and $X_{2i,n} \sim N(1,1)$ as in Qu and Lee [17]. $u_i's$ are from normal distribution with mean equal to 0 and variance to be 6. The coefficients $\beta = (1,1)'$.

For the simultaneous SARAR Tobit model, the spatial lag has the following form:

$$Y_n^* = \lambda W_n Y_n + X_n\beta + u_n, \qquad y_{i,n} = \max(0, y_{i,n}^*)$$

where $x_{1i,n}$ to be intercept and $X_{2i,n} \sim N(1,1)$ as in Qu and Lee [17]. $u_i's$ are from normal distribution with mean equal to 0 and variance to be 6. The coefficients $\beta = (1,1)'$. We let the parameter λ take values from 0.1 to 0.5 and get the following empirical power for both of the models. The tables and figures are as follows:

[Table A.7 is here]

[Table A.8 is here]

[Figure A.3 is here]

[Figure A.4 is here]

Table A.7 and Table A.8, as well as Figure A.3 and Figure A.4 show the empirical power under the spatial AR lag for both latent and simultaneous SARAR Tobit models. Similar to Table A.5 and Table A.6, the first and seventh rows of Table A.7 and Table A.8 are just the empirical sizes for the five test statistics when both ρ and λ equal to 0. Different from Table A.5 and Table A.6, now LM_λ and LM_λ^* have the highest power, which means that they both have the power against spatial lags.

From Figure A.3 and Figure A.4, the power function of LM_ρ^* is flatter than rest of the statistics, especially for the small values of λ. The difference of power between LM_ρ and LM_ρ^* for simultaneous SARAR Tobit model is not as big as in the latent SARAR Tobit model which is due to the difference in the non-centrality parameters.

An interesting finding is that the power function of LM_ρ^* is flatter for simulatneous SARAR Tobit model when there is the presence of spatial AR lags; while the power function of LM_λ^* is much flatter for latent SARAR Tobit model when there is a presence of spatial AR errors. The reason behind this phenomenon is because the simultaneous and latent SARAR Tobit models' information matrices diverge, resulting in a variation in the non-centrality parameter.

1.5.1.3 Spatial AR Lag and AR Error

In this setting, both λ and ρ will not equal to 0. The latent SARAR Tobit model has the following

$$Y_n^* = \lambda W_n Y_n^* + X_n\beta + u_n, \quad u_n = \rho M_n u_n + \varepsilon_n$$

$$y_{i,n} = \max(0, y_{i,n}^*)$$

and the simultaneous SARAR Tobit model:

$$Y_n^* = \lambda W_n Y_n + X_n\beta + u_n, , \quad u_n = \rho M_n u_n + \varepsilon_n$$

$$y_{i,n} = \max(0, y_{i,n}^*)$$

where the β and X_n are just the same with above. For the selection of parameters, first we keep $\lambda = 0.05$ fixed and let ρ change from 0.1 to 0.5. Then we keep $\lambda = 0.1$ fixed and let ρ change from 0.1 to 0.5, vice versa. The tables and figure below present the estimated rejection probability for the latent SARAR Tobit model and the simultaneous SARAR Tobit model:

[Table A.9 is here]

[Table A.10 is here]

[Figure A.5 is here]

Table A.9 and Table A.10 show the estimated rejection probabilities for both latent and simultaneous SARAR Tobit models. The results are similar to the case when there is only spatial AR errors or only spatial AR lags. When ρ is not zero, but remains a small value, as λ gets bigger, the power function of LM_ρ^* is still flatter, compared to all the other test statistics. Similar situation holds for LM_λ^* as λ stays a small value and ρ increases.

Besides LM_ρ, LM_λ, LM_ρ^*, LM_λ^*, additionally, we are quite intrigued by the performance of $LM_{\rho\lambda}$ for the relevance of its size and power characteristics are so crucial. Figure A.5 show the power surface plots of latent SARAR Tobit model and simultaneous SARAR Tobit model for different combinations of (ρ, λ) varying from (0, 0) to (0.5, 0.5). From the power surface plots, it is obvious that the power of $LM_{\rho\lambda}$ climbs extremely quickly and approaches 1 very quickly when any one of ρ and λ increases.

1.6 Conclusions and Future Research

This chapter derives the robust LM tests to test the spatial correlations for SARAR Tobit model, which is a generalization of Anselin et al. [15] to Tobit model. Additionally, we derive the joint LM test for endogeneity and spatial dependence, which we recommend as the place to start when performing a specification search in the given situation. Two types of SARAR Tobit models are considered in this chapter. One is the latent SARAR Tobit model and the other is the simultaneous SARAR Tobit model. The Monte Carlo simulation results show that the robust test statistic for both models have good properties under finite samples for local misspecification. For the practical use, our recommendation would be to apply $LM_{\rho\lambda}$ for the joint test $H_0 : \rho = \lambda = 0$. If H_0 is accepted, then it would be totally fine to use the standard Tobit model. If rejected, then apply LM_ρ^* or LM_λ^* to find the case.

In this chapter, we do not address the problem of choosing between two types of SARAR Tobit models. It is possible to design Cox-type non-nested tests to determine which Tobit model, latent or simultaneous is more appropriate for the data in hand. For the current data generating process, we only consider the normal error terms. According to Anselin et al. [15], tests are still robust under non-normality. Thus for the future research, we can also consider the cases under non-normality. Besides, we also intend to do further research on the empirical case dedicated to demonstrating the effectiveness of test statistics in future works of this study.

Chapter 2

Modelling the Spatial Impact of Crime on Housing Prices: Evidence from Chicago City

2.1 Introduction

The City of Chicago (hereafter, Chicago) has 801 census tracts, 77 community areas that have been grouped into 9 districts; Chicago has one of the highest crime rates in the United States compared to all other cities, especially the violent crime rates. The crime rate in Chicago experienced a major rise in the late 1960s; however, there is a drop in recent years. Different census tracts in Chicago have different changing patterns in crime but somehow affect each other at the same time, which reveal significant signs of clustering.

In general, research tend to operate on the assumption that crime plays an important role in affecting housing prices, and the effect is negative since people are more willing to buy houses located in "safe" areas. The underlying relationship between property prices and crime, on the other hand, may be more intricate than a straightforward downward effect. Ceccato and Wilhelmsson [21] proved the fact that the overall influence of crime on property values (measured as crime rates) was negligible, but the effect was non-negligible when assessed by distance to a crime hot spot. When different types of crime were considered instead than just looking at overall crime instances, the impact of crime might vary. According to Gibbons [22], different types of crime might have different impacts on property prices. Violent crimes, such as sexual assaults or homicides, might have a significant negative influence on prices, although property crimes, such as burglaries, might have a minor impact. According to the findings of Pope [23], when a sex offender came into a neighborhood, house prices in the surrounding area dropped by 2.3 percent ($3500 on average). While Ihlanfeldt and Mayock [24] found that only robbery and severe assault crimes (per acre) had a significant impact on housing values in the neighborhood.

In this chapter, the goal is to use the spatial autoregressive (SAR) model to show the spatial impact of crime on housing prices using the data from Chicago. The rest of the chapter is organized as follows. In Section 2.2 a brief review of the relevant literature is provided. Section 2.3 describes the data sources and the methodologies used in data acquisition and organization; Section 2.4 provides an overview of crime analysis in Chicago. Section 2.5 presents the two models used in this chapter, a SAR model and spatial Durbin model (SDM) and their estimation methods. Section 2.6 presents the results and the final section 2.7 provides some concluding remarks.

2.2 Literature Review

Researchers have historically focused on within-neighborhood crime for offenders who commit their crimes mostly in the vicinity of their homes (Rhodes and Conly [25]) and rarely commit crimes outside of their familiar surroundings (Brantingham and Brantingham [26]). However, as people travel from one area to another and connect local crime between spaces, there is mounting evidence that crime in one area is also associated with their neighborhood areas. As a result, the use of spatial analytical methods in crime pattern analysis has become common (see, for example, Anselin [27] and Townsley [28]).

First introduced by Paelinck [29], spatial econometrics has recently become of considerable importance in estimating appropriate models that incorporate spatial auto-correlation or neighborhood spillovers. Using spatial econometric methods can help identify the difference of predictors in different areas and shrink the estimate bias by introducing and accounting for spatial dependence. Anselin and Bera [30] used the observations of crime rates from 49 contiguous planning neighborhoods in Columbus, Ohio to analyze the determinants (housing values and income) of neighborhood crime. Morenoff and Sampson [31] examined the role of violent crime in generating population decline in Chicago neighborhoods from 1970 to 1990. Ye and Wu [32] used exploratory spatial data analysis (ESDA) to examine the dynamic spatial patterns of the indicators of homicide rates of Chicago. Schnell et. al. [33] employed street segments instead of neighborhood units to analyze the spatial variability of violent crime in Chicago. According to Huang et al. [34], both the spatial distribution and the spatial clustering of crimes had impacts on house values. The volume and severity of crimes had an increasing negative impact on housing values. Regardless of crime category, the price discount effect of crimes reduced as distance from crime places increased.

When taking the spatial setting into consideration, the relationship between property price and crime will be more complicated. One not only needs to consider the effect that neighboring areas may have on a certain area's property price, but also need to consider the effect of crime from the neighbor areas, since crime could also have the spillover effect. Hence, two kinds of spatial model will be considered in this chapter; one is the SAR model with the following form:

$$y_i = \lambda \sum_{j=1}^{N} w_{ij} y_j + \beta' x_i + u_i, \quad i = 1, 2, ..., N.$$

The other one is the SDM with the following form:

$$y_i = \lambda \sum_{j=1}^{N} w_{ij} y_j + \beta' x_i + \theta \sum_{j=1}^{N} w_{ij} x_j + v_i, \quad i = 1, 2, ..., N.$$

With these two models and the annual Chicago crime data for 77 communities and 801 census tracts, we can observe the spatial trends of major crime rates in Chicago from 2012 to 2018 and analyze the spatial impact of crime activities on house prices in Chicago.

The innovative contribution of this chapter is to combine Hedonic price modeling with spatial models. Hedonic price modeling has been employed by many researchers in the real estate market; it's a useful tool to model the price of goods based on the characteristics of those goods. When estimating the house prices, those properties often include floor area, number of bedrooms and so forth. For example, according to Fotheringham et al. [35], house prices are positively related to property size and number of rooms and negatively to the distance from the Central Business Districts (CBDs). Crime has also been included as a negative factor in influencing the house prices. A price index based on this approach using housing characteristics has been used for a number of years in Chicago, as a complement to the more traditional median price indices (see Lopez and Hewings [36]).

2.3 Data Source

Data for this research are extracted from several public sources. Census tracts are chosen here as the spatial unit because of the homogeneity of socioeconomic and demographic characteristics. The boundaries of census tracts are from the Chicago data portal in a shapefile form. The house value data is the residential sales data from Illinois REALTORS. Each house price is the actual transaction price with characteristics provided like number of bathrooms, floor area and X, Y coordinate for its address. All these transactions are geocoded using ArCGIS pro and aggregated to census tract to compute a mean price for each community. According to the property sale data, we choose year 2017 since it has enough houses sold for each census tract in this year.

Several typical hedonic pricing variables are considered here, representing structural, locational and neighborhood attributes of the property. For the structural variables, the number of bathrooms and floor area are constructed as averages for each census tract. For the locational variable, we use the distance from each census tract to the Central Business District (CBD). The spatial distance of this location variable is calculated in ArcGIS by its Euclidean distance. The structural variables and locational variables are from REALTORS.

The crime data is from Chicago data portal that provides reported incidents of crime that occurred in Chicago every year. These data are from the Chicago Police Department's CLEAR crime victims with X, Y coordinates. Using definitions from the U.S. Bureau of Justice Statistics, violent crime includes homicide, assault, crime sexual assault, battery and robbery. Property crime includes burglary, motor vehicle theft, theft and arson. These cases are sorted into violent crime and property crime and aggregated using ArcGIS pro into different census tracts. The core of this chapter is to study the impacts of different crimes on property values in Chicago. Therefore, this chapter applies several kinds of crime variables, including three main categories: total crime, violent crime, property crime, and three specific crime variables, including criminal damage, theft, and burglary. The crime data are the total number of events happened in each census tract. Table B.1 presents the overall summary statistics for each of the variables with data from 2017 and the description of these variables are listed in Table B.2.

2.4 Crime Analysis in Chicago: Initial Explorations

In recent years, there has been an overall decline in the crime rate in the United States, as well as in Chicago. From Figure B.1a, we can see there is a sharp decline from 2012 to 2014 and a small increase in 2016. Figure B.3 and B.4 show the quantile plots of housing prices, total crime, violent crime and property crime in 2017. In the housing price quantile plot of census tracts, we can observe a clear clustering pattern. The darker purple color in the left sub-figure indicates higher prices for those areas, while those areas in the right sub-figure tend to have lighter green colors, which means a low number of crime cases. From these figures, we can derive a general pattern of the negative relationship between housing prices and total crime. However the plot for property crime has a different pattern compared with total crime and violent crime, leading to our conjecture that the relationship between housing price and different types of crime may be more complicated than simply negatively correlated.

Based on the figures, it is obvious that crimes, both violent and property crimes have spatial dependence; this can be tested using Moran's tests, a widely accepted test of spatial dependence. In Table B.3, we have two sets of weighting matrices, Queen criterion and Rook criterion matrix. It can be seen that the results are not sensitive to the choice of weight matrix and all crime types show a high degree of spatial dependence.

2.5 Formal Analyses: Spatial Autoregressive Model and Spatial Durbin Model

2.5.1 Spatial Autoregressive Model

As the house prices show a spatial dependence pattern according to the Moran's I test result, this chapter employs a spatial autoregressive (SAR) model.

$$Y = \lambda WY + X\beta + \varepsilon,$$

where $Y = (y_1, y_2, ..., y_n)'$ and ε to be error terms. W is a $n \times n$ nonstochastic spatial weights matrix. When the queen criteria is applied, w_{ij} is 1 when two neighbors share a side or vertex, and 0 otherwise. The explanatory variables X_n consist of different types of crime, number of bathrooms, floor area and distance to CBD. House prices, floor area and crime types, those continuous variables are measured in natural log terms. The model is estimated with two methods, one is maximum likelihood function in Anselin [37] and the other is a general spatially weighted two-stage least square (2SLS) method following Kelejian and Prucha [38].

2.5.1.1 Maximum Likelihood Estimation (MLE)

If we assume that ε is a multivariate white noise, then its density function follows the multivariate Gaussian density function. Thus we can write the log-likelihood function as follows

$$log(L(\lambda, \beta, \Sigma|X, Y, W)) = -\frac{1}{2}[(I - \lambda W)Y - X\beta]'\Sigma^{-1}[(I - \lambda W)Y - X\beta] - \frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| + \log|\frac{\partial\varepsilon}{\partial V}|,$$

where $V = (X, Y, W)$ and $\Sigma = \sigma^2 I$. Then by setting the first derivative of log-likelihood function to 0, we can obtain an estimate for β.

$$\frac{\partial l}{\partial \beta} = 0$$

$$\hat{\beta} = (X'X)^{-1}X'(I - \lambda W)Y.$$

The regression results are shown in Table B.4. Note that the p-values for LM test for the residual correlation for each model are all greater than 0.273, providing for acceptance of the null hypothesis of no residual correlation.

2.5.1.2 Spatial 2SLS Estimation

Consider the SAR model

$$Y = \lambda WY + X\beta + \varepsilon.$$

Now we can rewrite it in a compact form

$$Y = Z\gamma + \varepsilon,$$

where $Z = [X, WY]$ and $\gamma = [\beta', \lambda]'$. Since WY can be correlated with the innovations ε, then we can propose an instrumental variable. The matrix of instruments can be defined as

$$H = (X, WX, WWX).$$

where $q \leq 2$. Then the spatial 2SLS estimator for γ is:

$$\hat{\gamma} = [\hat{Z}'\hat{Z}]^{-1}\hat{Z}'y,$$

where $\hat{Z} = PZ$, $P = H(H'H)^{-1}H'$. The asymptotic variance covariance matrix is

$$Var(\hat{\gamma}) = \hat{\sigma}^2(\hat{Z}'\hat{Z})^{-1},$$

where $\hat{\sigma}^2 = e'e/n$ with $e = y - Z\hat{\gamma}$. The regression results are shown in Table B.5. The results will be discussed in more detail later in Section 2.6.

2.5.2 Spatial Durbin Model (SDM)

As an explanatory variable, crime also has the spillover effects to the neighboring areas. One area's property price may also be affected by the neighboring areas' crimes. Thus, when the spatial dependencies not only occur in the dependent variable, but also in the independent variables, a SDM can be employed. The SDM has the following structure:

$$Y = \lambda WY + \beta_0 + X\beta_1 + WX\beta_2 + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2 I_n)$.

The SDM can be viewed as a generalization of SAR model in the sense that the SDM adds spatial lags on the independent variables. To estimate the SDM model, here we employ the Maximum Likelihood estimation as in Bekti and Rahayu [39].

2.5.2.1 Maximum Likelihood Estimation

SDM can be formed into the following equation:

$$Y = (I - \lambda W)^{-1} Z\beta + \varepsilon,$$

where $\beta = (\beta_0, \beta_1, \beta_2)'$ and $Z = (I, X, WX)$.

Then the loglikelihood function is

$$L(\lambda, \beta, \sigma^2 | y) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} |I - \lambda W| \exp\left(-\frac{1}{2\sigma^2}((I - \lambda W)y - Z\beta)^T (I - \lambda W)y - Z\beta)\right).$$

Letting $\frac{\partial ln(L)}{\partial \beta} = 0$, the MLE $\hat{\beta}$ of β is as follows:

$$\hat{\beta} = (Z^T Z)^{-1} Z^T (I - \lambda W)y,$$

which is unbiased since:

$$E(\hat{\beta}) = E\left((Z^T Z)^{-1} Z^T (I - \lambda W)y\right) = (Z^T Z)^{-1} Z^T (I - \lambda W)(I - \lambda W)^{-1} Z\beta = \beta.$$

The results for the SDM are shown in Table B.6.

2.6 Results and Findings

Table B.4 table B.5 and table B.6 all show high levels of association between the dependent and independent variables. Instead of using the traditional R^2, a pseudo R^2 is used for the spatial model. According to table B.4, the pseudo R^2 values are all greater than 0.824, indicating that the spatial relationships provide a strong source of explanation for house prices in Chicago. The spatial lag coefficients are all significant and positive across the different models, suggesting that a neighborhood's house prices have positive effects.

The hedonic price model captures the effects of housing characteristics on house price change. The regression results show the expected signs and impacts for the housing characteristics. In Table B.4, when estimating the model using ML estimation and the total crime considered in the model, when number of bathroom increase 1 percent, the natural log of house prices will rise about 0.196%. Different types of crime will not change the behavior of those housing characteristics very much. Distance to CBD is not very significant with small coefficients, which may result in the fact that many of these census tracts are very close to CBD and so this factor will not matter so much. In addition, distance to CBD maybe sensitive to the certain types of crime. While all of the models have the same sign for this variable, some of them are not significant, showing that certain types of crime will influence the assumptions regarding the effect of distance on house prices.

Most models show a statistically significant association with house price and crime and most of the signs are negative, showing that crime activities have a negative impact on house prices in Chicago. However, there is a positive relationship between theft and house prices, with 1 percent increase in theft activities, there will be a 0.018% increase in house price, indicating that such crime is primarily connected with less dynamic locations. So the general idea of crime having a negative impact on house prices is actually more complex, depending on the type of crimes.

When the spatial dependence of the independent variables is introduced into the model, the results become more interesting and complex. It is still clear that crime, notwithstanding the types of crimes, plays a relatively small part in affecting the property price; all of them reveal small values that are not significant. However, what is interesting here is that though crime does not play a significant part in the current area, they can have bigger influence on the neighbor areas' property prices. Refer to Table B.6 and explore violent crime as an example; the effect on current area's housing price is not significant. However, the neighbor areas' violent crime cases can have significant effects. The same patterns hold for criminal damage. The effects of housing characteristics on housing price are still within expectations, all significant except for the distance to CBD.

When considering burglary and theft in the area, they can have positive effect on the housing price. The different part is that the neighborhood burglary cases still plays a negative effect on a certain area but the neighboring theft cases can also play a positive part in affecting a certain area's housing price. The explanation about why burglary and theft have positive effects not only on a certain area, but also on the neighborhood is that criminals usually choose the places for property crime based on perceived wealthiness degree of that area. If an area's housing price is rather high, it will have a higher probability to have property crimes such as burglary and theft. So we can't simply analyze the effects of theft and burglary on the housing price using a simple regression model. Clearly, there is an important endogeneity issue here that needs to be resolved.

2.7 Conclusion

In this chapter, we examine the spatial pattern of crime in 801 census tracts in Chicago and its effect on housing prices. For different types of crimes, the crime in the neighborhood areas all have significant and positive effects on the housing prices. We not only show the effect of current area crime but also show the significant effect of neighborhood area crime on housing prices. This result is different from Lynch and Rasmussen [40], where they showed that the cost of crime has almost little effect on overall housing values, but residences in high-crime regions were heavily discounted.

In contrast to the conventional thinking that crime always has a negative effect on housing prices, our results show that crime doesn't have a uniform effect on the housing prices, which is consistent to McHatton et. al. [41]. In our results, theft and burglary have positive but not significant effects on housing prices, which means they are associated with higher-income neighborhoods, while violent crimes tend to be found in lower-income neighborhood areas.

Our next step is to expand our cross-sectional study to panel data analysis. It would be interesting to see what is the pattern of crime within a period and whether the effects of theft will change for a long period of time. Besides, we consider the possibility of using this SAR model to make predictions about housing prices for the future.

Chapter 3

Bootstrap-Based Inference of High-Dimensional Means and Its Applications

3.1 Introduction

Suppose we have some independently and identically distributed (i.i.d.) d-dimensional random vectors $Z_1, ..., Z_n$ with $E(Z_t) = \mu$, $Cov(Z_t) = \Sigma$. Now consider the following high-dimensional hypothesis:

$$H_0 : \mu = 0 \text{ versus } H_1 : \mu \neq 0.$$

Here we allow the dimension d to increase with the sample size n and can be even larger than n. This is a typical "large d, small n" problem which arises with the analysis of the gene sets. In this case, the conventional Hotelling's T^2 test cannot work properly because the inverse of the sample covariance does not exist. There have been many papers about this high-dimensional problem and some new tests to deal with it. Bai and Saranadasa [42] compared the Hotelling's T^2 statistics and Dempster's Non-Exact statistics, and then came up with their own statistic when $d/n \to \lambda \in (0, 1)$. Chen and Qin [43] proposed a two-sample test for the high-dimensional data when dimension d was much larger than the sample size n and their proposed test didn't require explicit conditions on the relationship of dimension and sample size. Pouzo [44] studied the asymptotic behavior of quadratic forms with increasing dimension and proved the consistency of the wild bootstrap for the quadratic forms using the Lindeberg interpolation. Xu et al. [45] showed the asymptotic theory for the L^2 norm of the sample mean vector of the high-dimensional data. Instead of bootstrapping, they used the plug-in method to estimate the eigenvalues of the sample covariance matrix and an alternative sub-sampling approach to avoid estimating eigenvalues.

In this chapter we propose a new test statistic in the form of U-statistic as follows:

$$T_n = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} Z_t^T Z_s.$$

To approximate this statistic, we apply the wild bootstrap to estimate our U-statistics. The bootstrapped statistic takes the following form:

$$T_n^* = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} w_t Z_t^T Z_s w_s,$$

where $\{w_t\}_{t=1}^n$ are the wild bootstrap weights. Bootstrap is a very simple and efficient way to estimate statistics and is widely used ever since Efron [46] first introduces it. There are many papers about the application of bootstrap (see Mammen [47]; Chang [48] among others). In our simulation study, we compared both normal weights wild bootstrap and Mamen weights bootstrap. The results show that compared with Xu et al. [45], both of the bootstrap estimations have smaller approximation errors than the plug-in procedure and sub-sampling procedure in Section 3.6.

Our main goal in this chapter is to show that

$$sup_x \left| P\left(\frac{T_n}{\sqrt{Var(T_n)}} \leq x\right) - P^*\left(\frac{T_n^*}{\sqrt{Var^*(T_n^*)}} \leq x\right) \right| \xrightarrow{p} 0.$$

Since our statistics T_n and T_n^* turn out to be martingales, we can directly apply Heyde and Brown's [49] results and compute the exact bounds of the U-statistics departure from the bootstrapped version. Heyde and Brown [49] studied the departure from normality of a class of martingales and came up with the main theorem about the bound of this distance. Besides, we also apply the studentized statistic as the testing statistic, which shows a better result than the unstudentized one under some data generating processes in the simulation study.

In order to show the applicability of our statistic, we apply our results in three aspects. The first application is about the overidentification tests for models with many instrumental variables. This topic has been studied by many researchers in recent years (see, among others, Lee and Okui [50]; Anatolyev and Gospodinov [51]; Chao et al. [52]). Lee and Okui [50] proposed a modified version of Sargan test when the number of instruments increased with the sample size. Anatolyev and Gospodinov [51] carried modifications on AR tests and J tests for the models of many instrument variables in the homoskedasticity case. Chao et al. [53] used a jackknife version of the overidentifying test statistic under heteroskedasticity when the number of instrumental variables grew at a rate up to the sample size.

In our application part, we apply the U-statistic as the testing statistic for the overidentifying restrictions and use the bootstrapped version for approximation. The technical assumptions for our statistic are much weaker than the ones in Anatolyev and Gospodinov [51].

The second application is on the spatial sign statistics. Spatial sign is often used to construct robust test statistics, and its applications in high-dimensional data have been studied by many papers in recent years. Wang et al. [54] proposed a novel high-dimensional non-parametric test for the mean vector of the non-normal high-dimensional data. Feng and Sun [55] employed a scalar transform invariant test based on spatial sign in the high-dimensional settings. In our second application, we apply our results in the spatial sign form of data and compare the simulation results with Wang et al. [54]. The approximation errors are much smaller in our case.

The third application is about testing the high-dimensional covariance matrix. The null hypothesis is $H_0 : \Sigma$ is a diagonal matrix. We construct the test statistics with U-statistics and use the bootstrapped version to approximate it. Compared with the tests in Chen et al. [43], our approximate errors are smaller under the same settings, leading to smaller size distortion.

This chapter is organized as follows. In Section 3.2, we introduce some lemmas and technical assumptions. Section 3.3 presents the theorems and corollaries for unstudentized statistics. In Section 3.4 we provide the main theorem for the studentized statistics. Section 3.5 is about the applications of our results. Section 3.6 shows the Monte-Carlo simulation results about the approximation errors for our studentized and unstudentized statistics compared with other statistics. Section 3.7 concludes while Section C.1 collects the proofs of lemmas, propositions and theorems in this chapter.

3.2 Technical Assumptions and Auxiliary Lemmas

Consider an i.i.d sample $Z = (Z_1, ..., Z_n)$ with $E(Z_t) = 0$, $Var(Z_t) = \Sigma$ and the dimension of Z_i is d. We will stick to this assumption throughout the whole chapter. The U-statistic is constructed as follows:

$$T_n = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} Z_t^T Z_s = \sum_{t=2}^{n} Y_t,$$

where $Y_t = \sum_{s=1}^{t-1} \frac{1}{\binom{n}{2}} Z_t^T Z_s$. The filtration is set up in the following way: $\mathcal{F}_t = \sigma(Z_1, Z_2, ..., Z_t)$. Then it is obvious that

$$E(Y_t|\mathcal{F}_{t-1}) = \sum_{s=1}^{t-1} \frac{1}{\binom{n}{2}} E(Z_t^T|\mathcal{F}_{t-1}) Z_s = \sum_{s=1}^{t-1} \frac{1}{\binom{n}{2}} E(Z_t^T) Z_s = 0.$$

Thus $\{Y_t, \mathcal{F}_{t-1}\}$ is a martingale difference sequence. So T_n is a martingale. We have the following lemma.

Lemma 3.2.1. *For the random sample* $\{Z_t\}_{t=1}^n$*, we have*

$$Var(T_n) = \frac{1}{\binom{n}{2}} ||\Sigma||_F^2.$$

Then we construct the wild bootstrapped version of T_n, T_n^*.

$$T_n^* = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} w_t Z_t^T Z_s w_s = \sum_{t=2}^{n} Y_t^*,$$

where $Y_t^* = \sum_{s=1}^{t-1} \frac{1}{\binom{n}{2}} w_t Z_t^T Z_s w_s$. The filtration is set up in the following way: $\mathcal{F}_t^* = \sigma(w_1, w_2, ..., w_t)$, $\mathcal{F}_n = \sigma(Z_1, ..., Z_n)$. Throughout we assume the wild bootstrap weights $w_t's$ as iid scalar random variables and $w_t's$ are independent of $Z_t's$. $E(w_t) = 0$, $E(w_t^2) = 1$, $E[w_t^4] = \kappa < \infty$.

Assumption 3.2.1. *Define* $D_2(Z)^4 := E\frac{|Z_1^T Z_2|^4}{||\Sigma||_F^4}$*, and we have*

$$\frac{D_2(Z)^4}{n} \to 0.$$

Assumption 3.2.2. *As* $n \wedge d \to \infty$*,*

$$\frac{tr(\Sigma^4)}{||\Sigma||_F^4} \to 0.$$

Proposition 3.2.1. *Assumption 3.2.1 is implied by*

$$\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{j_3=1}^{d}\sum_{j_4=1}^{d} cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4})^2 = o(n||\Sigma||_F^4). \tag{3.2.1}$$

Remark 3.2.1. *To satisfy Assumption 3.2.1, we need to assume*

$$sup_{1\le j\le d} E(X_{tj}^4) \le C \le \infty.$$

If we further assume $Z_{ti} \perp Z_{tj}$ when $|i-j| > L$ (i.e. L-dependence), then we have

$$\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{j_3=1}^{d}\sum_{j_4=1}^{d} cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4})^2 = O(dL^3).$$

If $||\Sigma||_F^4 \asymp (dL)^2$, then $L = o(dn)$.

We use $E^*(T_n^*)$, $Var^*(T_n^*)$ to represent the expectation and variance of T_n^* conditional on $\mathcal{F}_n$, $E^*(T_n^*) = E(T_n^*|\mathcal{F}_n)$, $E^*(Y_t^*|\mathcal{F}_{t-1}^*) = E(Y_t^*|\mathcal{F}_{t-1}^*, \mathcal{F}_n)$ and $Var^*(T_n^*) = Var(T_n^*|\mathcal{F}_n)$. It is obvious that

$$E^*(T_n^*) = E(T_n^*|\mathcal{F}_n) = E(\sum_{t=2}^{n} Y_t^*|\mathcal{F}_n) = \sum_{t=2}^{n} E(Y_t^*|\mathcal{F}_n) = 0,$$

and

$$\begin{aligned} E^*(Y_t^*|\mathcal{F}_{t-1}^*) &= E(Y_t^*|\mathcal{F}_{t-1}^*, \mathcal{F}_n) \\ &= \sum_{s=1}^{t-1} \frac{1}{\binom{n}{2}} E(w_t|\mathcal{F}_{t-1}^*) Z_t^T Z_s w_s = 0. \end{aligned}$$

Thus $\{Y_t^*, \mathcal{F}_{t-1}^*\}$ is a martingale difference sequence conditional on $\mathcal{F}_n$.
Then we have the following result about the consistency of bootstrap variance estimator.

Proposition 3.2.2. *The variance of the bootstrapped statistics T_n^* is*

$$Var^*(T_n^*) = \frac{1}{\binom{n}{2}^2} \sum_{t=2}^{n}\sum_{s=1}^{t-1} Z_t^T Z_s Z_s^T Z_t.$$

Under Assumption 3.2.1, we have

$$\frac{Var^*(T_n^*)}{Var(T_n)} \xrightarrow{p} 1.$$

3.3 Unstudentized U-Statistic

In this section, we directly apply the results in Heyde and Brown [49] about the convergence rate for the martingale to get the convergence rate for the standardized U-statistic to the standard normal and the wild bootstrap U-statistic. Before presenting the following theorems, we first give a definition of convergence in distribution in probability introduced by Li et al. [56].

Definition 3.3.1. *Let S_n be a statistic that is a function of $\{Z_i, w_i\}_{i=1}^n$, we say that $(S_n|Z_1, Z_2, ...)$ converges to $(S|Z_1, Z_2, ...)$ in distribution in probability if for any subsequence $S_{n'}$, there exists a subsequence of $S_{n'}$, $S_{n''}$, such that $(S_{n''}|Z_1, Z_2, ...)$ converges to $(S|Z_1, Z_2, ...)$ for almost every sequence $(Z_1, Z_2, ...)$.*

Then we introduce the following theorems:

Theorem 3.3.1. *We have*

$$sup_x\left|P\left(\frac{T_n}{\sqrt{Var(T_n)}} \leq x\right) - \Phi(x)\right| \leq K\left(\frac{tr(\Sigma^4)}{||\Sigma||_F^4} + \frac{D_2(Z)^4 + 1}{n}\right)^{\frac{1}{5}},$$

where $\Phi(x) = \frac{1}{2\pi}\int_{\infty}^{x} e^{-\frac{1}{2}u^2}du$ and K is a generic positive constant.
Therefore under Assumptions 3.2.1 and 3.2.2, we have

$$\frac{T_n}{\sqrt{Var(T_n)}} \xrightarrow{D} N(0,1).$$

Similarly for the bootstrapped version, we have the following theorem.

Theorem 3.3.2. *Under Assumption 3.2.1, we have*

$$sup_x\left|P^*\left(\frac{T_n^*}{\sqrt{Var^*(T_n^*)}} \leq x\right) - \Phi(x)\right| = O_p\left[\left(\frac{tr(\Sigma^4)}{||\Sigma||_F^4} + \frac{D_2(Z)^4 + 1}{n}\right)^{\frac{1}{5}}\right],$$

where $P^*(\cdot)$ means the probability conditional on $\mathcal{F}_n$.
Thus under Assumptions 3.2.1 and 3.2.2, we have

$$\frac{T_n^*}{\sqrt{Var^*(T_n^*)}} \text{ converges to N(0,1) in distribution in probability.}$$

Finally we combine Theorems 3.3.1 and 3.3.2 and get

Corollary 3.3.1. *Under Assumption 3.2.1, we have*

$$sup_x\left|P\left(\frac{T_n}{\sqrt{Var(T_n)}} \leq x\right) - P^*\left(\frac{T_n^*}{\sqrt{Var^*(T_n^*)}} \leq x\right)\right| = O_p\left[\left(\frac{tr(\Sigma^4)}{||\Sigma||_F^4} + \frac{D_2(Z)^4 + 1}{n}\right)^{\frac{1}{5}}\right].$$

Corollary 3.3.1 gives the convergence rate for the bootstrap approximation.

3.4 Studentized U-Statistic

We can also prove the consistency under the studentized form. Similar to the way we construct the wild bootstrapped U-statistics T_n^*, we can construct the bootstrapped version of sample variance, $\widehat{Var^*(T_n^*)}$. Since

$$Var^*(T_n^*) = \frac{1}{\binom{n}{2}^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} Z_t^T Z_s Z_s^T Z_t,$$

thus we define

$$\widehat{Var^*(T_n^*)} = \frac{1}{\binom{n}{2}^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} w_t Z_t^T Z_s w_s w_s Z_s^T Z_t w_t$$
$$= \frac{1}{\binom{n}{2}^2} \sum_{t=2}^{n} \sum_{s=1}^{t-1} Z_t^T Z_s Z_s^T Z_t w_t^2 w_s^2.$$

Based on Definition 3.3.1, we have the following theorem.

Theorem 3.4.1. *Under Assumptions 3.2.1 and 3.2.2, we have*

$$\frac{T_n^*}{\sqrt{\widehat{Var^*(T_n^*)}}} \text{ converges to } N(0,1) \text{ in distribution in probability.}$$

Then we can follow the previous results in Section 3.4 and show the convergence.

3.5 Applications

In this section we apply the results in section 3.3 to three problems. The first application is the test for the overidentifying restrictions in linear model with many instruments.

3.5.1 Application 1: Overidentification Tests

Now consider the simple IV regression model

$$y = X\beta + e,$$

where

$$Y = (y_1, y_2, ..., y_n)'_{n\times 1}$$

$$X = (x_1, x_2, ..., x_n)'_{n\times k}$$

$$Z = (Z_1, Z_2, ..., Z_n)'_{n\times d}$$

$$e = (e_1, e_2, ..., e_n)'_{n\times 1}$$

$\{y_i, x_i, Z_i\}_{i=1}^n$ is a random sample.
We impose the following assumptions.

Assumption 3.5.1. *The errors* e_i *satisfies* $E(e|Z) = 0, E(ee'|Z) = \sigma^2 I_n$ *and* $E(|e_i|^4) < \infty$.

Assumption 3.5.2. *Suppose* $Z_t e_t = Z_t^*$ *with* $E(Z_t e_t) = 0$, $Var(Z_t e_t) = \Sigma^*$.

Assumption 3.5.3. *Suppose* $Z_t x_t^T = A_t$, $E(A_t) = A \neq 0$ $E(A_t A_t^T) = \Sigma_a$ *and as* $n \wedge d \to \infty$,

$$\frac{||A^T A||_F}{||\Sigma^*||_F} = o(1) \tag{3.5.1}$$

$$\frac{tr(E(A_1^T A_2 A_1^T) A))}{n||\Sigma^*||_F^2} = o(1) \tag{3.5.2}$$

$$\frac{||\Sigma_a||_F}{n||\Sigma^*||_F} = o(1) \tag{3.5.3}$$

$$\frac{tr(\Sigma^* AA^T)}{||\Sigma^*||_F^2} = o(1) \tag{3.5.4}$$

$$\frac{tr(\Sigma^*\Sigma_a)}{n||\Sigma^*||_F^2} = o(1). \tag{3.5.5}$$

Assumption 3.5.4. *As $n \to \infty$, $d/n = O(1)$.*

To test the overidentifying restrictions, the null hypothesis of correct moment restrictions is $H_0 : E(e_t Z_t) = 0$.

Then we apply our unfeasible U-statistic here and get

$$T_n = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} e_t Z_t^T Z_s e_s = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} Z_t^{*T} Z_s^*.$$

We can compute the conditional variance of T_n as

$$Var(T_n) = \frac{1}{\binom{n}{2}} ||\Sigma^*||_F^2.$$

Then the feasible test statistic is

$$\hat{T_n} = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \hat{e_t} Z_t^T Z_s \hat{e_s},$$

where $\hat{e_i}$ is the residual. $\hat{e} = y - X\hat{\beta}$. According to Bekker [57], since the 2SLS estimator is not consistent in the model with many instrument variables, in stead we use the LIML estimator to estimate β. According to Anatolyev et al. [51],

$$\hat{\beta}_{LIML} = \frac{X'(I_n - kM)Y}{X'(I_n - kM_z)X},$$

where k is the smallest characteristic root of $(\bar{Y}'\bar{Y})(\bar{Y}'M\bar{Y})^{-1}$, $\bar{Y} = (y, X)$ and $M = In - Z(Z'Z)^{-1}Z'$. Since

$$\hat{e_t} - e_t = x_t^T(\beta - \hat{\beta}) = (\beta - \hat{\beta})^T x_t,$$

then we can use e_t to replace $\hat{e_t}$ in $\hat{T_n}$ and have the following theorem.

Theorem 3.5.1. *Under Assumptions 3.5.1, 3.5.2 and 3.5.3, we have*

$$\frac{\hat{T_n}}{\sqrt{Var(T_n)}} = \frac{T_n}{\sqrt{Var(T_n)}} + o_p(1).$$

Further assume Assumptions 3.2.1 and 3.2.2 hold for $\{Z_t^*\}$, by Theorem 3.1 we have

$$\frac{\hat{T_n}}{\sqrt{Var(T_n)}} \xrightarrow{D} N(0,1).$$

For the bootstrapped version

$$\hat{T_n^*} = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} w_t \hat{e_t} Z_t^T Z_s \hat{e_s} w_s,$$

we have the following theorem.

Theorem 3.5.2. *Under Assumptions 3.5.1, 3.5.2 and 3.5.3, we have*

$$\frac{\hat{T_n^*}}{\sqrt{Var(T_n)}} = \frac{T_n^*}{\sqrt{Var(T_n)}} + o_{p^*}(1),$$

where $o_{p^*}(1)$ means converging to zero in probability conditional on the data. Further assume Assumptions 3.2.1 and 3.2.2 hold for $\{Z_t^*\}$, then by Theorem 3.2 and Proposition 2.2 we have

$$\frac{\hat{T_n^*}}{\sqrt{Var(T_n)}} \text{ converges to N(0,1) in distribution in probability.}$$

Thus we have

$$sup_x \left| P\left(\hat{T_n} \le x\right) - P^*\left(\hat{T_n^*} \le x\right) \right| \xrightarrow{p} 0.$$

Remark 3.5.1. *We can use an example to examine the validity of Assumption 3.5.3. Now we assume that*

$$x_t = \gamma Z_t + v_t,$$

where γ is a $k \times d$ matrix and $E(Z_t v_t^T) = 0, E(Z_t x_t^T) \neq 0,$ and $E(Z_t Z_t^T) = \Sigma_z$. Then we have

$$A_t = Z_t x_t^T = Z_t Z_t^T \gamma^T + Z_t v_t^T.$$

Then for Assumption 3.5.3, we apply the conditions in Anatolyev et al. [51] with

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$x_i = \sum_{j=1}^{d} \gamma_j Z_{ij} + v_i,$$

where $\begin{bmatrix} e_i \\ v_i \end{bmatrix} = chol(\Sigma)\varepsilon_i$, $\begin{bmatrix} \varepsilon_i \\ Z_i \end{bmatrix} = iidN(0, I_{d+1})$, $\Sigma = \begin{bmatrix} 0.25 & 0.20 \\ 0.20 & 0.25 \end{bmatrix}$, $\gamma = (\frac{1}{\sqrt{d}}, ..., \frac{1}{\sqrt{d}})$. *Thus* $\Sigma_z = I_d, ||\Sigma^*||_F^2 \asymp d^2, A = E(A_t) = \Sigma_z \gamma^T = \gamma^T, ||\Sigma_a||_F^2 \asymp d^2, tr(E(A_1^T A_2 A_1^T)\Sigma_z \gamma^T \asymp 1$. *Then we have* $\frac{||A^T A||_F^2}{||\Sigma^*||_F^2} = \frac{||\gamma \Sigma_z^2 \gamma^T||_F^2}{||\Sigma^*||_F^2} \asymp \frac{1}{d^2} \to 0; \frac{tr(E(A_1^T A_2 A_1^T)\Sigma_z \gamma^T)}{n||\Sigma^*||_F^2} \asymp \frac{1}{nd^2} \to 0; \frac{||\Sigma_a||_F^2}{n^2||\Sigma^*||_F^2} \asymp \frac{1}{n^2} \to 0; \frac{tr(\Sigma^* \Sigma_a)}{n||\Sigma^*||_F^2} \asymp \frac{1}{n} \to 0.$

3.5.2 Application 2: Spatial Sign Statistics

We follow the paper of Wang et al. [54] and use the spatial sign function of the data. We denote $\tilde{Z_t} = \frac{Z_t}{||Z_t||}$, where $||Z_t||$ is the L_2 norm of Z_t. Then our new test statistic $\tilde{T_n}$ is

$$\tilde{T_n} = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \frac{Z_t^T}{||Z_t||} \frac{Z_s}{||Z_s||} = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \tilde{Z_t}^T \tilde{Z_s},$$

and under the settings in Wang et al. [54], we have $E(\tilde{Z_t}) = 0$, and we let $B = Var(\tilde{Z_t}) = E(\frac{Z_t Z_t^T}{||Z_t||^2})$. Then we have

$$Var(\tilde{T_n}) = \frac{1}{\binom{n}{2}} ||B||_F^2.$$

The bootstrapped version is

$$\tilde{T_n}^* = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} w_t \tilde{Z_t}^T \tilde{Z}_s w_s.$$

Our assumption is as follows

Assumption 3.5.5.

$$\frac{tr^4(\Sigma)}{||\Sigma||_F^4} exp\{-\frac{tr^2(\Sigma)}{128d\lambda_{max}^2(\Sigma)}\} = o(1),$$

where $\lambda_{max}(\Sigma) = o\{\sqrt{Tr^2(\Sigma^2)}\}$ *and* $\lambda_{max}(\cdot)$ *means the largest eigenvalue of a matrix.*

Then we have the following theorem.

Theorem 3.5.3. *Under Assumptions 3.2.2 and 3.5.5, we have*

$$sup_x \left| P\left(\frac{\tilde{T_n}}{\sqrt{Var(\tilde{T_n})}} \leq x\right) - P^*\left(\frac{\tilde{T_n}^*}{\sqrt{Var^*(\tilde{T_n}^*)}} \leq x\right) \right| \xrightarrow{p} 0.$$

We compare this new spatial sign test statistic with our old one in our simulation study in Section 3.6.

3.5.3 Application 3: Tests for the Covariance Matrix

For this chapter, suppose we have the data $Z_1, Z_2, ..., Z_n$, which are i.i.d. with mean 0 and covariance matrix Σ. Now the null hypothesis $H_0 : \Sigma$ is a diagonal matrix, which contains the sphericity hypothesis $\Sigma = \sigma^2 I_d$, where σ^2 is unknown. We denote

$$W_t = \begin{bmatrix} Z_{t1}Z_{t2} \\ Z_{t1}Z_{t3} \\ ... \\ Z_{t(d-1)}Z_{td} \end{bmatrix}$$

Then our statistics can be rewritten as $T_n = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} W_t^T W_s$, $T_n^* = \frac{1}{\binom{n}{2}} \sum_{t=2}^{n} \sum_{s=1}^{t-1} w_t W_t^T W_s w_s$. Now it is easy to see that under the null, $E(W_t) = 0, Var(W_t) = \Sigma_w$ and $Var(T_n) = \frac{1}{\binom{n}{2}} ||\Sigma_w||_F^2$.
From Section 3.2, we have the following proposition for W_t:

Proposition 3.5.1. *If we assume that* $Z_{ti} \perp Z_{tj}$ *for* $i \neq j$

$$0 < c_1 \leq sup_i E(Z_{ti}^2) \leq C_1 < \infty$$

$$0 < c_2 \leq sup_i E(Z_{ti}^4) \leq C_2 < \infty,$$

then

$$\frac{D_2(W)^4}{n} \to 0$$

and

$$\frac{tr(\Sigma_w^4)}{||\Sigma_w||_F^4} \to 0.$$

Theorem 3.5.4. *Under the assumption of Proposition 3.5.1, we have*

$$sup_x \left| P\left(\frac{T_n}{\sqrt{Var(T_n)}} \leq x\right) - P^*\left(\frac{T_n^*}{\sqrt{Var^*(T_n^*)}} \leq x\right)\right| \to 0.$$

For the simulation part, we apply the numerical settings by Chen et al. [58].

Suppose $X_1, X_2, ..., X_n$ are IID d-dimensional random vectors such that

$$X_i = \Gamma Z_i,$$

where Γ is a $d \times d$ constant matrix so that $\Gamma\Gamma^T = \Sigma$, and $E(Z_1) = 0$, and $Var(Z_1) = I_d$.
Chen et al. [58] introduced the statistics for testing the high-dimensional covariance matrix. There are two testing structures for the covariance matrix. $H_0 : \Sigma = \sigma^2 I_d$ vs. $H_1 : \Sigma \neq \sigma^2 I_d$ and $H_0 : \Sigma = I_d$ vs. $H_1 : \Sigma \neq I_d$.

Let $Y_{1,n} = \frac{1}{n}\sum_{t=1}^n X_t^T X_t$, $Y_{2,n} = \frac{1}{P_n^2}\sum_{i\neq j}(X_i^T X_j)^2$, $Y_{3,n} = \frac{1}{P_n^2}\sum_{i\neq j} X_i^T X_j$, $Y_{4,n} = \frac{1}{P_n^3}\sum_{i,j,k}^* X_i^T X_j X_j^T X_k$, $Y_{5,n} = \frac{1}{P_n^4}\sum_{i,j,k,l}^* X_i^T X_j X_k^T X_l$. where $P_n^r = n!/(n-r)!$, $\sum_{i,j,k}^*$ and $\sum_{i,j,k,l}^*$ mean that the summations over mutually different indices. Then let $T_{1,n} = Y_{1,n} - Y_{3,n}$, $T_{2,n} = Y_{2,n} - 2Y_{4,n} + Y_{5,n}$. The test statistic for the first testing problem is

$$U_n = d(\frac{T_{2,n}}{T_{1,n}^2}) - 1.$$

The testing statistic for the second testing problem is

$$V_n = \frac{1}{d}T_{2,n} - \frac{2}{d}T_{1,n} + 1.$$

Under the first null hypothesis $H_0 : \Sigma = \sigma^2 I_d$ vs. $H_1 : \Sigma \neq \sigma^2 I_d$, we have

$$nU_n \xrightarrow{D} N(0, 4).$$

Under the second null hypothesis $H_0 : \Sigma = I_d$ vs. $H_1 : \Sigma \neq I_d$, we have

$$nV_n \xrightarrow{D} N(0, 4).$$

The simulation reults are listed in section 3.6.

3.6 Simulation Results

In this section we use Monte Carlo study to present the finite sample performance of our both studentized and unstudentized U-statistics compared with other statistics. 10000 Monte Carlo repetitions are performed here and for every repetition, we have 2000 bootstraps. In each table, n is the sample size, d is the dimension, a is the ath percentile of the distributions. In the first three subsections, we compare the approximation errors for different statistics under different data generating processes. In the following tables,
$\varepsilon_{T_{us,n}}$ represents approximation error of unstudentized statistics under normal weights.
$\varepsilon_{T_{s,n}}$ represents approximation error of studentized statistics under normal weights.
$\varepsilon_{T_{us,m}}$ represents approximation error of unstudentized statistics under Mammen weights.
$\varepsilon_{T_{s,m}}$ represents approximation error of studentized statistics under Mammen weights.

ε_{V_p} and ε_{F_s} represent approximation error of the plug-in method and sub-sampling method in Xu et al. [45]. ε_{Q_n} and ε_{Q_m} are the approximation errors of the quadratic forms under normal weights and Mammen weights in Pouzo [44].

3.6.1 $\chi^2 - 1$ case

In the first data generating process, we construct Z_i as $Z_{ij} = \chi^2 - 1, j \leq d$. In Table C.1, we set the sample size n to be 50 and 100 and the dimension d from 5 to 300.

[Table C.1 is here]

Then the following linear and factor model data generating processes are from Xu et al. [45].

3.6.2 Linear Process

For the second data generating process we let $\xi_{i,k}$ be i.i.d. Student t_5. Let

$$Z_{i,j} = \sum_{k=0}^{2000} (k+1)^{-\beta} \xi_{i,j-k}.$$

Then if $\beta < 1$, the process $(Z_{i,j})$ is long memory. Here we let $\beta = 0.6$ and $\beta = 2$, $n = 50$ and $n = 100$ in Table C.2.

[Table C.2 is here]

3.6.3 Factor Model

Then we let

$$Z_{i,j} = \sqrt{4 + U_i^2} \xi_{i,j} + b(2N_i + N_i^2 - 1), 1 \leq i \leq n, 1 \leq j \leq d,$$

where $U_i \sim Uniform[-1, 1], \xi_{i,j}, N_i \sim N(0, 1)$ and they are all independent. In the following Table C.3, we consider the case when $b = 0.05$ and $b = 0.5$.

[Table C.3 is here]

In the above six tables, the first four statistics in each table are our statistics under studentized forms or under unstudentized forms. It's clear that the approximate errors for our statistic are very low, all below 5%, which show the better consistency of the bootstrap under Mammen weights and wild weights. When the dimension of the sample becomes bigger than the sample size, the error becomes even smaller, which show the validity of our statistics under the high dimensional settings.

3.6.4 Spatial Sign

In our second application, to compare our spatial sign version statistics and the statistics in Wang et al. [54], we use the settings in their paper. There are three types of data generating processes (DGP).

1. DGP1: Z_i follows a $N_d(\mu, \Sigma)$ where $\mu = (0, ..., 0)^T$ and for Σ, $\sigma_{ij} = 0.8^{|i-j|}$.
2. DGP2: Z_i follows a $N_d(\mu, \Sigma)$ where $\mu = (0, ..., 0)^T$ and for $\Sigma = DRD$, $D = diag(d_1, d_2, ..., d_d)$ with $d_i = 2 + (d - i + 1)/d$, $R = (r_{ij})$ with $r_{ii} = 1$ and $r_{ij} = (-1)^{i+j} 0.2^{|i-j|}$ for $i \neq j$.

3. DGP3: Z_i follows a $N_d(\mu, \Sigma)$ where $\mu = (0, ..., 0)^T$ and for $\Sigma = \sigma_{ij}$, where $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.2(i \neq j)$.

4. DGP4: Z_i follows a d-variate t distribution with mean vector to be $\mu = (0, ..., 0)^T$ and $\Sigma = \sigma_{ij}$, where $\sigma_{ii} = 1$ and $\sigma_{ij} = 0.2(i \neq j)$, and 3 degrees of freedom.

5. DGP5: Z_i follows a d-variate t distribution with mean vector to be $\mu = (0, ..., 0)^T$ and Σ, $\sigma_{ij} = 0.8^{|i-j|}$, and 3 degrees of freedom.

Here are Table C.4, C.5, C.6, C.7, and C.8 generated from the above three DGPs. In each table, the sample size is set to 30, and dimensions are from 20 to 1000.
$\varepsilon_{T_{us,n}}$ represents approximation error of unstudentized statistics under normal weights.
$\varepsilon_{T_{s,n}}$ represents approximation error of studentized statistics under normal weights.
ε_{W_n} represents approximation error of Wang's statistics compared with normal distribution.

[Table C.4 is here]

[Table C.5 is here]

[Table C.6 is here]

[Table C.7 is here]

[Table C.8 is here]

It's obvious that bootstrapping shows a better consistency under the high dimensional settings.

3.6.5 Covariance Testing

In this part we apply the settings in Chen et al. [58] and compare the behaviors of our test statistic and their test statistics under the two hypotheses. In the following Table C.9 and C.10, we have two kinds of settings for Z_i.

1. DGP1: Z_i were IID d-dimensional normal random vector with mean 0_m and covariance I_d; $\Gamma = I_d$.

2. DGP2: $Z_i = (Z_{i1}, Z_{i2}, ..., Z_{id})$ consists of IID random variables Z_{ij} which follows a standardized Gamma(4,0.5) distribution with mean 0 and unit variance; $\Gamma = I_d$.

For each scenario, we let $n = 30$ and dimensions change from 20 to 1000.
Here $\varepsilon_{E_{us,n}}$ represents approximation error of our unstudentized statistics.
$\varepsilon_{E_{us,n}}$ represents approximation error of our studentized statistics.
ε_{U_n} represents approximation error of statistics nU_n in Chen et al. [58] under the first null hypothesis.
ε_{V_n} represents approximation error of statistics nV_n in Chen et al. [58] under the second null hypothesis.

[Table C.9 is here]

[Table C.10 is here]

Table C.9 and C.10 show the consistency between the U-statistics and Bootstrapped version. Compared with the two statistics under different hypotheses in Chen et al. [58], our statistic behaves much better even in the high dimensional case.

3.7 Conclusion and Discussion

This chapter shows the consistency of the bootstrap for U-statistic under high-dimensional settings and proposes the accurate convergence rate by directly applying the main theorem in Heyde and Brown [49]. Under some proper assumptions, we can use the results in testing for the overidentifying restrictions, covariance testing and spatial sign statistics. The simulation studies for our statistics compared with other statistics are pretty good under different data generating processes. For future study, we can derive some high order accuracy results when Z_t has special dependence structure, like m-dependence. Besides, we can use martingale expansion to find the exact convergence rate of studentized statistic.

References

[1] J. Tobin, "Estimation of relationships for limited dependent variables," *Econometrica: journal of the Econometric Society*, pp. 24–36, 1958.

[2] T. Amemiya, "Regression analysis when the dependent variable is truncated normal," *Econometrica: Journal of the Econometric Society*, pp. 997–1016, 1973.

[3] P. M. Robinson, "On the asymptotic properties of estimators of models containing limited dependent variables," *Econometrica: Journal of the Econometric Society*, pp. 27–41, 1982.

[4] P. M. Robinson, A. K. Bera, and C. M. Jarque, "Tests for serial defendence in limited dependent variable models," *International Economic Review*, pp. 629–638, 1985.

[5] D. P. McMillen, "Probit with spatial autocorrelation," *Journal of Regional Science*, vol. 32, no. 3, pp. 335–348, 1992.

[6] J. Pinkse and M. E. Slade, "Contracting in space: An application of spatial statistics to discrete-choice models," *Journal of Econometrics*, vol. 85, no. 1, pp. 125–154, 1998.

[7] J. P. LeSage, "Bayesian estimation of limited dependent variable spatial autoregressive models," *Geographical Analysis*, vol. 32, no. 1, pp. 19–35, 2000.

[8] H. Wang, E. M. Iglesias, and J. M. Wooldridge, "Partial maximum likelihood estimation of spatial probit models," *Journal of Econometrics*, vol. 172, no. 1, pp. 77–89, 2013.

[9] J. Pinkse, M. Slade, and L. Shen, "Dynamic spatial discrete choice using one-step gmm: An application to mine operating decisions," *Spatial Economic Analysis*, vol. 1, no. 1, pp. 53–99, 2006.

[10] C. R. Bhat, I. N. Sener, and N. Eluru, "A flexible spatially dependent discrete choice model: Formulation and application to teenagers' weekday recreational activity participation," *Transportation research part B: methodological*, vol. 44, no. 8-9, pp. 903–921, 2010.

[11] A. G. Billé and G. Arbia, "Spatial discrete choice and spatial limited dependent variable models: A review with an emphasis on the use in regional health economics," *arXiv preprint arXiv:1302.2267*, 2013.

[12] C. Yang, L.-f. Lee, and X. Qu, "Tobit models with social interactions: Complete vs incomplete information," *Regional Science and Urban Economics*, vol. 73, pp. 30–50, 2018.

[13] P. A. Moran, "A test for the serial independence of residuals," *Biometrika*, vol. 37, no. 1/2, pp. 178–181, 1950.

[14] L. Anselin and A. K. Bera, "Spatial dependence in linear regression models with an introduction to spatial econometrics," *Statistics textbooks and monographs*, vol. 155, pp. 237–290, 1998.

[15] L. Anselin, A. K. Bera, R. Florax, and M. J. Yoon, “Simple diagnostic tests for spatial dependence,” *Regional science and urban economics*, vol. 26, no. 1, pp. 77–104, 1996.

[16] A. K. Bera and M. J. Yoon, “Specification testing with locally misspecified alternatives,” *Econometric theory*, vol. 9, no. 4, pp. 649–658, 1993.

[17] X. Qu and L.-f. Lee, “Lm tests for spatial correlation in spatial models with limited dependent variables,” *Regional Science and Urban Economics*, vol. 42, no. 3, pp. 430–445, 2012.

[18] X. Xu and L.-f. Lee, “Maximum likelihood estimation of a spatial autoregressive tobit model,” *Journal of Econometrics*, vol. 188, no. 1, pp. 264–280, 2015.

[19] ——, “Sieve maximum likelihood estimation of the spatial autoregressive tobit model,” *Journal of Econometrics*, vol. 203, no. 1, pp. 96–112, 2018.

[20] J. LeSage and R. K. Pace, *Introduction to spatial econometrics.* Chapman and Hall/CRC, 2009.

[21] V. Ceccato and M. Wilhelmsson, “Do crime hot spots affect housing prices?” *Nordic journal of criminology*, vol. 21, no. 1, pp. 84–102, 2020.

[22] S. Gibbons, “The costs of urban property crime,” *The Economic Journal*, vol. 114, no. 499, F441–F463, 2004.

[23] J. C. Pope, “Fear of crime and housing prices: Household reactions to sex offender registries,” *Journal of Urban Economics*, vol. 64, no. 3, pp. 601–614, 2008.

[24] K. Ihlanfeldt and T. Mayock, “Panel data estimates of the effects of different types of crime on housing prices,” *Regional Science and Urban Economics*, vol. 40, no. 2-3, pp. 161–172, 2010.

[25] W. M. Rhodes and C. Conly, “Crime and mobility: An empirical study,” in *Principles of geographical offender profiling*, Routledge, 2017, pp. 143–164.

[26] P. L. Brantingham and P. J. Brantingham, “Notes on the geometry of crime,” *Environmental criminology*, 2017.

[27] L. Anselin, J. Cohen, D. Cook, W. Gorr, and G. Tita, “Spatial analyses of crime,” *Criminal justice*, vol. 4, no. 2, pp. 213–262, 2000.

[28] M. Townsley, “Spatial autocorrelation and impacts on criminology,” *Geographical Analysis*, vol. 41, no. 4, pp. 452–461, 2009.

[29] J. Paelinck, “Spatial econometrics,” *Economics Letters*, vol. 1, no. 1, pp. 59–63, 1978.

[30] L. Anselin and A. K. Bera, “Introduction to spatial econometrics,” *Handbook of applied economic statistics*, vol. 237, no. 5, 1998.

[31] J. D. Morenoff and R. J. Sampson, “Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970–1990,” *Social forces*, vol. 76, no. 1, pp. 31–64, 1997.

[32] X. Ye and L. Wu, “Analyzing the dynamics of homicide patterns in chicago: Esda and spatial panel approaches,” *Applied Geography*, vol. 31, no. 2, pp. 800–807, 2011.

[33] C. Schnell, A. A. Braga, and E. L. Piza, “The influence of community areas, neighborhood clusters, and street segments on the spatial variability of violent crime in chicago,” *Journal of quantitative criminology*, vol. 33, no. 3, pp. 469–496, 2017.

[34] Y. Huang, R. W. Spahr, M. A. Sunderman, and E. Ozdenerol, “The geospatial impact of crime on neighbourhood property values,” Paper presentation At the 2015 FMA Annual Meeting Orlando, Florida, 2015.

[35] A. S. Fotheringham, R. Crespo, and J. Yao, “Exploring, modelling and predicting spatiotemporal variations in house prices,” *The Annals of Regional Science*, vol. 54, no. 2, pp. 417–436, 2015.

[36] E. Lopez and G. J. Hewings, “Housing price indices for small spatial units,” *Regional Science and Urban Economics*, vol. 70, pp. 57–71, 2018.

[37] L. Anselin, *Spatial econometrics: methods and models*. Springer Science & Business Media, 1988, vol. 4.

[38] H. H. Kelejian and I. R. Prucha, “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances,” *Journal of econometrics*, vol. 157, no. 1, pp. 53–67, 2010.

[39] R. Bekti and A. Rahayu, “Maximum likelihood estimation for spatial durbin model,” *Journal of Mathematics and Statistics*, vol. 9, no. 3, p. 169, 2013.

[40] A. K. Lynch and D. W. Rasmussen, “Measuring the impact of crime on house prices,” *Applied Economics*, vol. 33, no. 15, pp. 1981–1989, 2001.

[41] D. McIlhatton, W. McGreal, P. T. de La Paz, and A. Adair, “Impact of crime on spatial analysis of house prices: Evidence from a uk city,” *International Journal of Housing Markets and Analysis*, 2016.

[42] Z. Bai and H. Saranadasa, “Effect of high dimension: By an example of a two sample problem,” *Statistica Sinica*, pp. 311–329, 1996.

[43] S. X. Chen and Y.-L. Qin, “A two-sample test for high-dimensional data with applications to gene-set testing,” *The Annals of Statistics*, vol. 38, no. 2, pp. 808–835, 2010.

[44] D. Pouzo, “Bootstrap consistency for quadratic forms of sample averages with increasing dimension,” *Electronic Journal of Statistics*, vol. 9, no. 2, pp. 3046–3097, 2015.

[45] M. Xu, D. Zhang, and W. B. Wu, “$L2$ Asymptotics for high-dimensional data,” *arXiv preprint arXiv:1405.7244*, 2014.

[46] B. Efron, “Bootstrap methods: Another look at the jackknife,” in *Breakthroughs in statistics*, Springer, 1992, pp. 569–593.

[47] E. Mammen, “Bootstrap and wild bootstrap for high dimensional linear models,” *The annals of statistics*, vol. 21, no. 1, pp. 255–285, 1993.

[48] Y. Chang, “Bootstrap unit root tests in panels with cross-sectional dependency,” *Journal of econometrics*, vol. 120, no. 2, pp. 263–293, 2004.

[49] C. Heyde and B. Brown, “On the departure from normality of a certain class of martingales,” in *Selected Works of CC Heyde*, Springer, 2010, pp. 129–133.

[50] Y. Lee and R. Okui, “A specification test for instrumental variables regression with many instruments,” 2009.

[51] S. Anatolyev and N. Gospodinov, “Specification testing in models with many instruments,” *Econometric Theory*, vol. 27, no. 2, pp. 427–441, 2011.

[52] J. C. Chao, J. A. Hausman, W. K. Newey, N. R. Swanson, and T. Woutersen, “Testing overidentifying restrictions with many instruments and heteroskedasticity,” *Journal of econometrics*, vol. 178, pp. 15–21, 2014.

[53] J. C. Chao, N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen, “Asymptotic distribution of jive in a heteroskedastic iv regression with many instruments,” *Econometric Theory*, vol. 28, no. 1, pp. 42–86, 2012.

[54] L. Wang, B. Peng, and R. Li, “A high-dimensional nonparametric multivariate test for mean vector,” *Journal of the American Statistical Association*, vol. 110, no. 512, pp. 1658–1669, 2015.

[55] L. Feng and F. Sun, “Spatial-sign based high-dimensional location test,” *Electronic Journal of Statistics*, vol. 10, no. 2, pp. 2420–2434, 2016.

[56] Q. Li, C. Hsiao, and J. Zinn, “Consistent specification tests for semiparametric/nonparametric models based on series estimation methods,” *Journal of Econometrics*, vol. 112, no. 2, pp. 295–325, 2003.

[57] P. A. Bekker, “Alternative approximations to the distributions of instrumental variable estimators,” *Econometrica: Journal of the Econometric Society*, pp. 657–681, 1994.

[58] S. X. Chen, L.-X. Zhang, and P.-S. Zhong, “Tests for high-dimensional covariance matrices,” *Journal of the American Statistical Association*, vol. 105, no. 490, pp. 810–819, 2010.

[59] S. Lahiri and S. Lahiri, *Resampling methods for dependent data*. Springer Science & Business Media, 2003.

Appendix A

Robust LM Tests for Latent and Simultaneous Spatial Autoregressive Tobit Models with Spatial Autoregressive Disturbances

A.1 Proofs of Lemmas and Theories

A.1.1 Derivation of Equation (1.2.8) and (1.2.9)

According to Equation (1.2.4) and Equation (1.2.5), the score function for λ has the following equation

$$\begin{aligned}
d_\lambda(\hat\theta)=&\frac{1}{\sigma^2}[E(Y_n^{*'}W_nY_n^*|X_n,Y_n)-E(Y_n^{*'}W_nY_n^*|X_n)]-\frac{1}{\sigma^2}\rho[E(Y_n^{*'}M_nW_nY_n^*|X_n,Y_n)-E(Y_n^{*'}M_nW_nY_n^*|X_n)]\\
&-\frac{1}{\sigma^2}\lambda[E(Y_n^{*'}W_n'W_nY_n^*|X_n,Y_n)-E(Y_n^{*'}W_n'W_nY_n^*|X_n)]\\
&+2\frac{1}{\sigma^2}\lambda\rho[E(Y_n^{*'}W_n'M_nW_nY_n^*|X_n,Y_n)-E(Y_n^{*'}W_n'M_nW_nY_n^*|X_n)]\\
&-\frac{1}{\sigma^2}\rho[E(Y_n^{*'}M_n'W_nY_n^*|X_n,Y_n)-E(Y_n^{*'}M_n'W_nY_n^*|X_n)]\\
&+\frac{1}{\sigma^2}\rho^2[E(Y_n^{*'}M_n'M_nW_nY_n^*|X_n,Y_n)-E(Y_n^{*'}M_n'M_nW_nY_n^*|X_n)]\\
&-\frac{1}{\sigma^2}\lambda\rho^2[E(Y_n^{*'}W_n'M_n'M_nW_nY_n^*|X_n,Y_n)-E(Y_n^{*'}W_n'M_n'M_nW_nY_n^*|X_n)]\\
&-\frac{1}{\sigma^2}(X_n\beta)'R_n(\rho)'R_n(\rho)W_n[E(Y_n^*|X_n,Y_n)-E(Y_n^*|X_n)]|_{\hat\theta}\\
=&\frac{1}{\sigma^2}[E(Y_n^{*'}W_nY_n^*|X_n,Y_n)-E(Y_n^{*'}W_nY_n^*|X_n)]-\frac{1}{\sigma^2}(X_n\beta)'W_n[E(Y_n^*|X_n,Y_n)-E(Y_n^*|X_n)]\\
=&\frac{1}{\sigma^2}[E(\varepsilon_n'W_nX_n\beta|X_n,Y_n)-E(\varepsilon_n'W_nX_n\beta|X_n)]+\frac{1}{\sigma^2}[E(\varepsilon_n'W_n\varepsilon_n|X_n,Y_n)-E(\varepsilon_n'W_n\varepsilon_n|X_n)]\\
=&\frac{1}{\sigma^2}[E(\varepsilon_n'W_n\varepsilon_n|X_n,Y_n)-E(\varepsilon_n'W_n\varepsilon_n|X_n)]+\frac{1}{\sigma^2}[E(\varepsilon_n'|X_n,Y_n)-E(\varepsilon_n'|X_n)]W_nX_n\beta
\end{aligned}$$

Since

$$E(\varepsilon_{i,n}|X_n) = 0, \quad E(\varepsilon_{i,n}^2|X_n) = \sigma^2, \quad E(\varepsilon_{i,n}\varepsilon_{j\neq i,n}|X_n) = 0,$$

then

$$E(\varepsilon_n|X_n) = \mathbf{0}, \quad E(\varepsilon_n'\varepsilon_n|X_n) = n\sigma^2$$

$$\begin{aligned}
E(\varepsilon_n' W_n \varepsilon_n|X_n) &= E(\sum_{i,j=1}^{n} \varepsilon_{i,n} W_{ij,n} \varepsilon_{j,n}|X_n) \\
&= \sum_{i=1,i=j}^{n} W_{ij,n} E(\varepsilon_{i,n}^2|X_n) + \sum_{i=1,i\neq j}^{n} W_{ij,n} E(\varepsilon_{i,n}\varepsilon_{j\neq i,n}|X_n) \\
&= \sum_{i=1}^{n} W_{ii,n}\sigma^2.
\end{aligned}$$

According to the properties of the spatial weight matrix, W_n have zero diagonals, which means $E(\varepsilon_n' W_n \varepsilon_n|X_n) = 0$. The final form of score function of λ is

$$d_\lambda(\hat{\theta}) = \frac{1}{\sigma^2} E(\varepsilon_n|X_n, Y_n)' W_n E(\varepsilon_n|X_n, Y_n) + \frac{1}{\sigma^2} E(\varepsilon_n|X_n, Y_n)' W_n X_n \beta.$$

The derivation for $d_\rho(\hat{\theta})$ is similar to $d_\lambda(\hat{\theta})$:

$$\begin{aligned}
d_\rho(\hat{\theta}) =& \frac{1}{\sigma^2}[E(Y_n^{*'} M_n Y_n^*|X_n, Y_n) - E(Y_n^{*'} M_n Y_n^*|X_n)] - \frac{1}{\sigma^2}\lambda[E(Y_n^{*'} M_n W_n Y_n^*|X_n, Y_n) - E(Y_n^{*'} M_n W_n Y_n^*|X_n)] \\
&+ \frac{1}{\sigma^2}\lambda^2[E(Y_n^{*'} W_n' M_n W_n Y_n^*|X_n, Y_n) - E(Y_n^{*'} W_n' M_n W_n Y_n^*|X_n)] \\
&- \frac{1}{\sigma^2}\lambda[E(Y_n^{*'} M_n' W_n Y_n^*|X_n, Y_n) - E(Y_n^{*'} M_n' W_n Y_n^*|X_n)] \\
&- \frac{1}{\sigma^2}\rho[E(Y_n^{*'} M_n' M_n Y_n^*|X_n, Y_n) - E(Y_n^{*'} M_n' M_n Y_n^*|X_n)] \\
&+ 2\frac{1}{\sigma^2}\lambda\rho[E(Y_n^{*'} M_n' M_n W_n Y_n^*|X_n, Y_n) - E(Y_n^{*'} M_n' M_n W_n Y_n^*|X_n)] \\
&- \frac{1}{\sigma^2}\lambda^2\rho[E(Y_n^{*'} W_n' M_n' M_n W_n Y_n^*|X_n, Y_n) - E(Y_n^{*'} W_n' M_n' M_n W_n Y_n^*|X_n)] \\
&- \frac{1}{\sigma^2}(X_n\beta)'(M_n' + M_n)S_n(\lambda)[E(Y_n^*|X_n, Y_n) - E(Y_n^*|X_n)]|_{\hat{\theta}} \\
=& \frac{1}{\sigma^2}[E(Y_n^{*'} M_n Y_n^*|X_n, Y_n) - E(Y_n^{*'} M_n Y_n^*|X_n)] - \frac{1}{\sigma^2}(X_n\beta)'(M_n' + M_n)[E(Y_n^*|X_n, Y_n) - E(Y_n^*|X_n)] \\
=& \frac{1}{\sigma^2}\{E[(Y_n^* - X_n\beta)' M_n (Y_n^* - X_n\beta)|X_n, Y_n] - E[(Y_n^* - X_n\beta)' M_n (Y_n^* - X_n\beta)|X_n]\} \\
=& \frac{1}{\sigma^2} E(\varepsilon_n|X_n, Y_n)' M_n E(\varepsilon_n|X_n, Y_n).
\end{aligned}$$

A.1.2 Derivation of Equation (1.2.11)

Let $\lambda = 0, \rho = 0$, $\varepsilon_{i,n} = y_{i,n}^* - x_{i,n}'\beta$, then

$$y_{i,n} = \max\{0, y_{i,n}^*\} = y_{i,n}^* I(y_{i,n} > 0), \quad I(y_{i,n} > 0) + I(y_{i,n} = 0) = 1.$$

When $y_{i,n} > 0$, $\varepsilon_{i,n} = y^*_{i,n} - x'_{i,n}\beta = y_{i,n} - x'_{i,n}\beta$;
When $y_{i,n} = 0$, $y^*_{i,n} \leq 0$, $\varepsilon_{i,n} = y^*_{i,n} - x'_{i,n}\beta = (-\infty, -x'_{i,n}\beta]$. Thus

$$\begin{aligned}\int_{-\infty}^{-x'_{i,n}\beta} z\phi(z,\sigma)dz &= -\sigma^2 \int_{-\infty}^{-x'_{i,n}\beta} \phi(z,\sigma)d(\frac{-z^2}{2\sigma^2}) \\ &= -\sigma^2 \int_{-\infty}^{-x'_{i,n}\beta} d\phi(z,\sigma) \\ &= -\sigma^2\phi(-x'_{i,n}\beta,\sigma) \\ &= -\sigma^2\phi(x'_{i,n}\beta,\sigma).\end{aligned} \tag{A.1.1}$$

Because for normal distribution, we have $\phi(x'_{i,n}\beta,\sigma) = \phi(-x'_{i,n}\beta,\sigma)$ and $\Phi(x'_{i,n}\beta,\sigma) = 1 - \Phi(-x'_{i,n}\beta,\sigma)$. Thus

$$\begin{aligned}E(\varepsilon_{i,n}|X_n, Y_n) &= E(\varepsilon_{i,n}|X_n, Y_n)I(y_{i,n} > 0) + E(\varepsilon_{i,n}|X_n, Y_n)I(y_{i,n} = 0) \\ &= (y_{i,n} - x'_{i,n}\beta)I(y_{i,n} > 0) + [\frac{1}{\Phi(-x'_{i,n}\beta,\sigma)} \int_{-\infty}^{-x'_{i,n}\beta} z\phi(z,\sigma)dz]I(y_{i,n} = 0) \\ &= (y_{i,n} - x'_{i,n}\beta)I(y_{i,n} > 0) - \frac{\sigma^2\phi(x'_{i,n}\beta,\sigma)}{1-\Phi(x'_{i,n}\beta,\sigma)}I(y_{i,n} = 0).\end{aligned} \tag{A.1.2}$$

A.1.3 Proof of Lemma 1.4.1

The proof procedure is similar to the proof of Theorem 1 in Qu and Lee (2012). By the triangle inequality,

$$\begin{aligned}|\frac{1}{n}\frac{\partial^2 L(\gamma^*)}{\partial\gamma\partial\gamma'} - E\left(\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'}\right)| &\leq |\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'} - E\left(\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'}\right)| + |\frac{1}{n}\frac{\partial^2 L(\gamma^*)}{\partial\gamma\partial\gamma'} - \frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'}| \\ &\leq |\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'} - E\left(\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'}\right)| + \sup_{\theta} \in \Theta\frac{1}{n}\left|\frac{\partial^3 \ln L(\beta,\sigma^2,0,0)}{\partial\gamma\partial\gamma\partial\gamma'}\right||\gamma^* - \gamma_0|.\end{aligned}$$

It's sufficient to show that

$$\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'} - E\left(\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'}\right) \to_p 0$$

and

$$\sup_{\theta} \in \Theta\frac{1}{n}\left|\frac{\partial^3 \ln L(\beta,\sigma^2,0,0)}{\partial\gamma\partial\gamma\partial\gamma'}\right| < \infty.$$

Under the null, the second order derivatives are

$$\begin{aligned}\frac{\partial^2 \ln L(\beta,\sigma^2,0,0)}{\partial\beta\partial\beta'} &= \frac{1}{\sigma^2}\sum_{i=1}^{n} x_{i,n}x'_{i,n}[-I(y_{i,n} > 0) \\ &+ I(y_{i,n} = 0)\frac{\phi_{i,n}(x'_{i,n}\beta,\sigma^2)x'_{i,n}\beta(1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2)) - \sigma^2\phi^2_{i,n}(x'_{i,n}\beta,\sigma^2)}{(1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2))^2}]\end{aligned}$$

$$
\begin{aligned}
&\frac{\partial^2 \ln L(\beta, \sigma^2, 0, 0)}{\partial\sigma^2\partial\sigma^2} \\
&= \frac{\partial\left[\sum_{i=1}^{n} I(y_{i,n}=0)\ln(1-\Phi(x_i'\beta,\sigma^2)) - \frac{1}{2}\sum_{i=1}^{n} I(y_{i,n}>0)(\ln(2\pi\sigma^2)+\frac{1}{\sigma^2}(y_{i,n}-x_{i,n}'\beta)^2)\right]}{\partial\sigma^2} \\
&= \frac{1}{2\sigma^4}\sum_{i=1}^{n}\left[I(y_{i,n}=0)\frac{x_{i,n}'\beta\sigma^2\phi(x_{i,n}'\beta,\sigma^2)}{1-\Phi(x_{i,n}'\beta,\sigma^2)} + I(y_{i,n}>0)((y_{i,n}-x_{i,n}'\beta)^2-\sigma^2)\right] \\
&= \frac{1}{2\sigma^4}\sum_{i=1}^{n} q_i(\hat{\theta})
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial^2 \ln L(\beta, \sigma^2, 0, 0)}{\partial\beta\partial\sigma^2} \\
&= \frac{1}{2\sigma^4}\sum_{i=1}^{n} x_{i,n} I(y_{i,n}=0)\left[\frac{(x_i'\beta)^2\phi_{i,n}(x_{i,n}'\beta,\sigma^2)}{(1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2))} + \frac{\sigma^2\phi_{i,n}(x_{i,n}'\beta,\sigma^2)}{1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2)} - \frac{\sigma^2 x_i'\beta\phi_{i,n}^2(x_{i,n}'\beta,\sigma^2)}{(1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2))^2}\right].
\end{aligned}
$$

Under H_0, according to Lemma 1 and Lemma 2 in Qu and Lee (2013), each elements are i.i.d. and have the same mean, so it follows the law of large numbers that

$$
\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'} - E\left(\frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'}\right) \to_p 0.
$$

To show $\frac{1}{n}\frac{\partial^2 L(\hat{\gamma})}{\partial\gamma\partial\gamma'} - \frac{1}{n}\frac{\partial^2 L(\gamma_0)}{\partial\gamma\partial\gamma'} \to_p 0$, we need to show

$$
\sup_{\theta} \in \Theta \frac{1}{n}\left|\frac{\partial^3 \ln L(\beta,\sigma^2,0,0)}{\partial\gamma\partial\gamma\partial\gamma'}\right| < \infty.
$$

The third order derivatives are

$$
\begin{aligned}
\frac{1}{n}\frac{\partial^3 \ln L(\beta,\sigma^2,0,0)}{\partial\beta\partial\beta'\partial\beta'} &= \frac{1}{\sigma^2 n}\sum_{i=1}^{n} x_{i,n}x_{i,n}'x_{i,n}' I(y_{i,n}=0)\frac{\phi_{i,n}(x_{i,n}'\beta,\sigma^2)}{1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2)}[1-\frac{2\sigma^2\phi_{i,n}(x_{i,n}'\beta,\sigma^2)^2}{(1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2))^2} \\
&+ \frac{3x_{i,n}'\beta\phi_{i,n}^2(x_{i,n}'\beta,\sigma^2)}{1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2)} - \frac{1}{\sigma^2}(x_{i,n}'\beta)^2]
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{n}\frac{\partial^3 \ln L(\beta,\sigma^2,0,0)}{\partial\beta\partial\beta'\partial\sigma^2} &= \frac{1}{\sigma^2 n}\sum_{i=1}^{n} x_{i,n}x_{i,n}' I(y_{i,n}=0)\frac{\phi_{i,n}(x_{i,n}'\beta,\sigma^2)}{1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2)}[-\frac{x_{i,n}'\beta}{2\sigma^2} \\
&+ \frac{(x_{i,n}'\beta)^2}{2\sigma^4} - \frac{3(x_{i,n}'\beta)^2\phi_{i,n}(x_{i,n}'\beta,\sigma^2)}{2\sigma^2(1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2))} + \frac{2x_{i,n}'\beta\phi_{i,n}^2(x_{i,n}'\beta,\sigma^2)}{(1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2))^2} + \frac{\phi_{i,n}(x_{i,n}'\beta,\sigma^2)}{(1-\Phi_{i,n}(x_{i,n}'\beta,\sigma^2))}]
\end{aligned}
$$

$$\begin{aligned}&\frac{1}{n}\frac{\partial^2 \ln L(\beta,\sigma^2,0,0)}{\partial\sigma^2\partial\sigma^2\partial\beta'}\\&=\frac{1}{2\sigma^4 n}\sum_{i=1}^{n}x'_{i,n}[I(y_{i,n}=0)\frac{\sigma^2\phi_{i,n}(x'_{i,n}\beta,\sigma^2)}{1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2)}-I(y_{i,n}=0)\frac{(x'_{i,n}\beta)^2\phi_{i,n}(x'_{i,n}\beta,\sigma^2)}{1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2)}\\&+I(y_{i,n}=0)\frac{x'_{i,n}\beta\sigma^2\phi^2_{i,n}(x'_{i,n}\beta,\sigma^2)}{(1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2))^2}-I(y_{i,n}>0)2(y_{i,n}-x'_{i,n}\beta)]\end{aligned}$$

$$\begin{aligned}&\frac{1}{n}\frac{\partial^2 \ln L(\beta,\sigma^2,0,0)}{\partial\sigma^2\partial\sigma^2\partial\sigma^2}\\&=\frac{1}{2\sigma^4 n}\sum_{i=1}^{n}x'_{i,n}[I(y_{i,n}=0)\frac{x'_{i,n}\beta\phi_{i,n}(x'_{i,n}\beta,\sigma^2)}{1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2)}-I(y_{i,n}=0)\frac{x'_{i,n}\beta\phi_{i,n}(x'_{i,n}\beta,\sigma^2)}{2(1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2))}\\&+I(y_{i,n}=0)\frac{(x'_{i,n}\beta)^3\phi_{i,n}(x'_{i,n}\beta,\sigma^2)}{2\sigma^2(1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2))}-I(y_{i,n}=0)\frac{(x'_{i,n}\beta)^2\phi^2_{i,n}(x'_{i,n}\beta,\sigma^2)}{2(1-\Phi_{i,n}(x'_{i,n}\beta,\sigma^2))^2}-I(y_{i,n}>0)].\end{aligned}$$

According to Qu and Lee (2012), under the null, $E(y_{i,n})$ is bounded and these third order derivatives are all bounded, which implies that

$$\sup_{\theta}\in\Theta\frac{1}{n}\left|\frac{\partial^3 \ln L(\beta,\sigma^2,0,0)}{\partial\gamma\partial\gamma\partial\gamma'}\right|<\infty.$$

Under the null, $\hat{\gamma}$ is a consistent estimator of γ_0. Thus $|\hat{\gamma}-\gamma_0|=o_p(1)$, which implies that $|\gamma^*-\gamma_0|=o_p(1)$. Thus the second argument is also proved.

A.2 Simulation Results

Table A.1: Verification of the identities in Equation (1.5.1)

	LM_ρ	LM^*_ρ	LM_λ	LM^*_λ	$LM_{\rho\lambda}$
$\rho=0,\lambda=0$					
Latent SARAR Tobit Model	0.902	0.777	0.373	0.248	1.150
Simultaneous SARAR Tobit Model	0.422	0.729	0.507	0.814	1.236
$\rho=0.1,\lambda=0.1$					
Latent SARAR Tobit Model	0.190	0.290	0.950	1.050	1.240
Simultaneous SARAR Tobit Model	3.843	4.150	0.657	0.964	4.807

Table A.2: Test statistics descriptions

Null Hypothesis	Parameter		Test statistics
	Spatial lag, λ	Spatial error, ρ	
$H_0 : \rho = 0$	$\lambda = 0$	-	LM_ρ
$H_0 : \rho = 0$	$\lambda \neq 0$	-	LM_ρ^*
$H_0 : \lambda = 0$	-	$\rho = 0$	LM_λ
$H_0 : \lambda = 0$	-	$\rho \neq 0$	LM_λ^*
$H_0 : \rho = 0, \lambda = 0$	-	-	$LM_{\rho\lambda}$

Table A.3: Empirical size of test for latent SARAR Tobit model ($\rho = 0$ and $\lambda = 0$)

Test	LM_ρ	LM_ρ^*	LM_λ	LM_λ^*	$LM_{\rho\lambda}$
$N = 50$, $\alpha = 0.05$					
$K = 3$	0.043	0.045	0.087	0.056	0.081
$K = 5$	0.060	0.054	0.119	0.067	0.094
$N = 50$, $\alpha = 0.10$					
$K = 3$	0.093	0.097	0.147	0.080	0.124
$K = 5$	0.099	0.100	0.175	0.091	0.180
$N = 100$, $\alpha = 0.05$					
$K = 3$	0.054	0.054	0.108	0.065	0.101
$K = 5$	0.057	0.060	0.169	0.097	0.131
$N = 100$, $\alpha = 0.10$					
$K = 3$	0.103	0.104	0.194	0.107	0.161
$K = 5$	0.106	0.113	0.236	0.124	0.200

Table A.4: Empirical size of test for simultaneous SARAR Tobit model ($\rho = 0$ and $\lambda = 0$)

Test	LM_ρ	LM_ρ^*	LM_λ	LM_λ^*	$LM_{\rho\lambda}$
$N = 50$, $\alpha = 0.05$					
$K = 3$	0.038	0.043	0.078	0.060	0.077
$K = 5$	0.051	0.049	0.126	0.095	0.119
$N = 50$, $\alpha = 0.10$					
$K = 3$	0.090	0.093	0.127	0.084	0.118
$K = 5$	0.098	0.099	0.209	0.119	0.178
$N = 100$, $\alpha = 0.05$					
$K = 3$	0.060	0.060	0.111	0.071	0.106
$K = 5$	0.053	0.052	0.175	0.104	0.142
$N = 100$, $\alpha = 0.10$					
$K = 3$	0.104	0.105	0.176	0.095	0.161
$K = 5$	0.102	0.106	0.263	0.139	0.219

Table A.5: Estimated rejection probabilities of test against spatial AR errors for latent SARAR Tobit model ($\lambda = 0$)

N	ρ	LM_ρ	LM_ρ^*	LM_λ	LM_λ^*	$LM_{\rho\lambda}$
50	0.0	0.043	0.045	0.087	0.056	0.081
	0.1	0.174	0.175	0.107	0.070	0.192
	0.2	0.487	0.479	0.143	0.100	0.458
	0.3	0.836	0.840	0.239	0.155	0.798
	0.4	0.987	0.986	0.406	0.281	0.979
	0.5	0.999	0.999	0.680	0.509	0.999
100	0.0	0.054	0.054	0.108	0.065	0.101
	0.1	0.300	0.300	0.132	0.084	0.302
	0.2	0.782	0.780	0.189	0.117	0.743
	0.3	0.993	0.993	0.300	0.194	0.987
	0.4	1.000	1.000	0.525	0.389	1.000
	0.5	1.000	1.000	0.832	0.668	1.000

Table A.6: Estimated rejection probabilities of test against spatial AR errors for simultaneous SARAR Tobit model ($\lambda = 0$)

N	ρ	LM_ρ	LM_ρ^*	LM_λ	LM_λ^*	$LM_{\rho\lambda}$
50	0.0	0.038	0.043	0.078	0.060	0.077
	0.1	0.159	0.156	0.100	0.074	0.162
	0.2	0.433	0.446	0.163	0.123	0.428
	0.3	0.783	0.782	0.269	0.210	0.749
	0.4	0.966	0.969	0.448	0.365	0.951
	0.5	1.000	1.000	0.671	0.550	0.998
100	0.0	0.060	0.060	0.111	0.071	0.106
	0.1	0.284	0.284	0.149	0.094	0.300
	0.2	0.736	0.737	0.246	0.170	0.718
	0.3	0.977	0.977	0.396	0.294	0.974
	0.4	1.000	1.000	0.645	0.523	1.000
	0.5	1.000	1.000	0.886	0.801	1.000

Figure A.1: Estimated rejection probabilities of test against spatial AR errors for latent SARAR Tobit model ($\lambda = 0$)

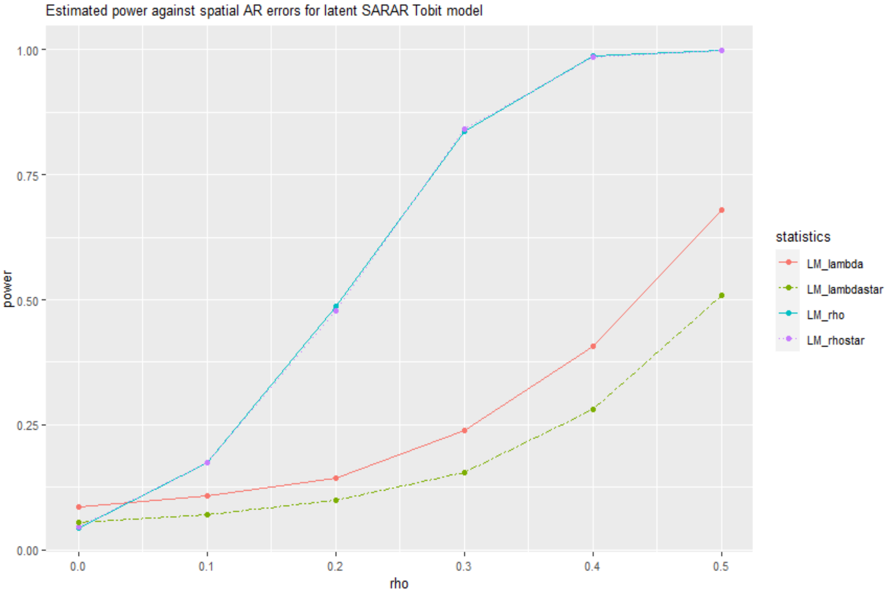


Figure A.2: Empirical rejection probabilities of test against spatial AR errors for simultaneous SARAR Tobit model ($\lambda = 0$)

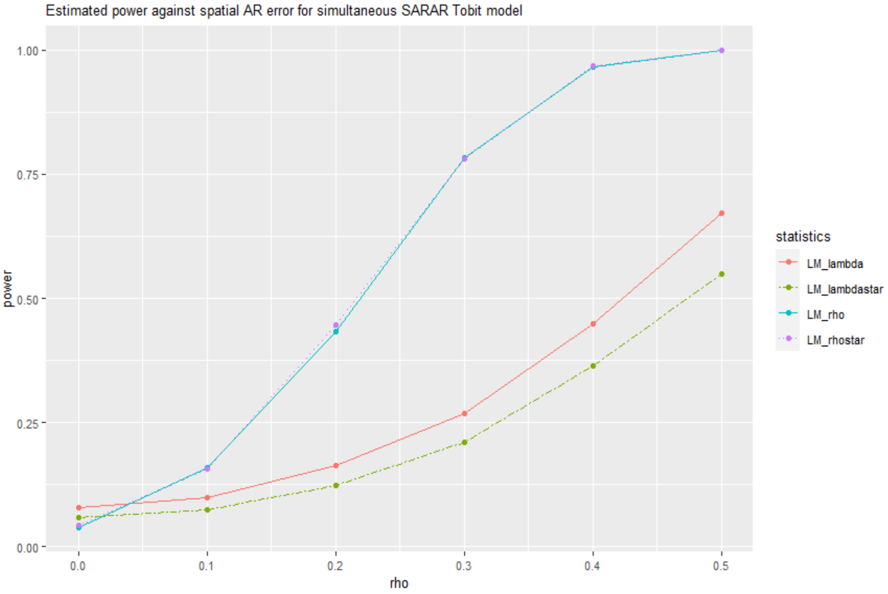


Table A.7: Estimated rejection probabilities of test against spatial AR lag for latent SARAR Tobit model ($\rho = 0$)

N	λ	LM_ρ	LM_ρ^*	LM_λ	LM_λ^*	$LM_{\rho\lambda}$
50	0.0	0.043	0.045	0.087	0.056	0.081
	0.1	0.061	0.061	0.277	0.269	0.253
	0.2	0.153	0.125	0.662	0.667	0.646
	0.3	0.411	0.320	0.937	0.934	0.916
	0.4	0.733	0.624	0.996	0.995	0.996
	0.5	0.956	0.920	1.000	1.000	1.000
100	0.0	0.054	0.054	0.108	0.065	0.101
	0.1	0.075	0.075	0.500	0.498	0.475
	0.2	0.185	0.152	0.947	0.946	0.934
	0.3	0.557	0.465	0.999	0.999	0.999
	0.4	0.929	0.867	1.000	1.000	1.000
	0.5	0.997	0.996	1.000	1.000	1.000

Table A.8: Estimated rejection probabilities of test against spatial AR lag for simultaneous SARAR Tobit model ($\rho = 0$)

N	λ	LM_ρ	LM_ρ^*	LM_λ	LM_λ^*	$LM_{\rho\lambda}$
50	0.0	0.038	0.043	0.078	0.060	0.077
	0.1	0.051	0.052	0.175	0.166	0.167
	0.2	0.072	0.070	0.386	0.387	0.360
	0.3	0.145	0.134	0.633	0.631	0.616
	0.4	0.354	0.311	0.812	0.810	0.838
	0.5	0.664	0.657	0.889	0.887	0.966
100	0.0	0.060	0.060	0.111	0.071	0.106
	0.1	0.061	0.064	0.324	0.317	0.293
	0.2	0.089	0.085	0.694	0.693	0.654
	0.3	0.161	0.149	0.930	0.929	0.910
	0.4	0.416	0.367	0.990	0.990	0.991
	0.5	0.863	0.788	0.999	0.998	1.000

Figure A.3: Estimated rejection probabilities of test against spatial AR lag for latent SARAR Tobit model ($\rho = 0$)

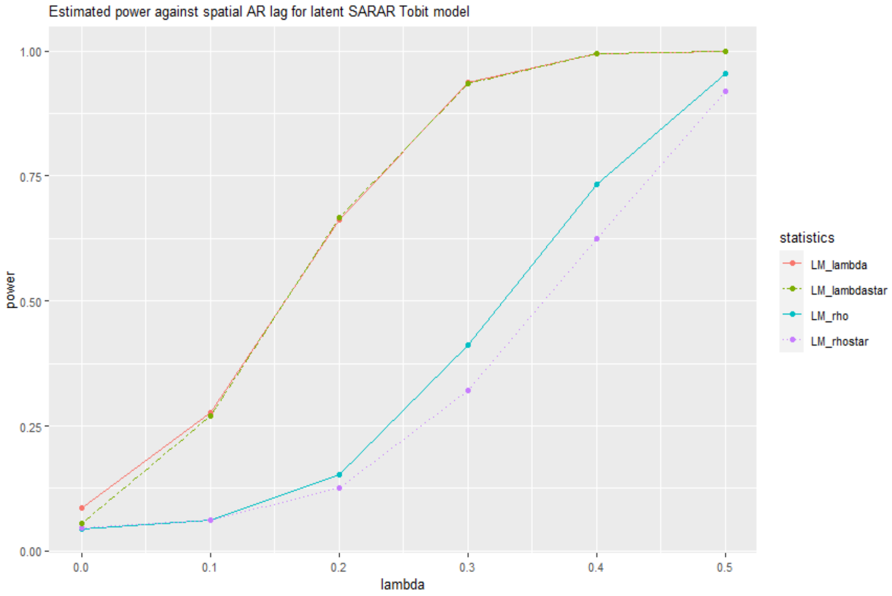


Figure A.4: Estimated rejection probabilities of test against spatial AR lag for simultaneous SARAR Tobit model ($\rho = 0$)

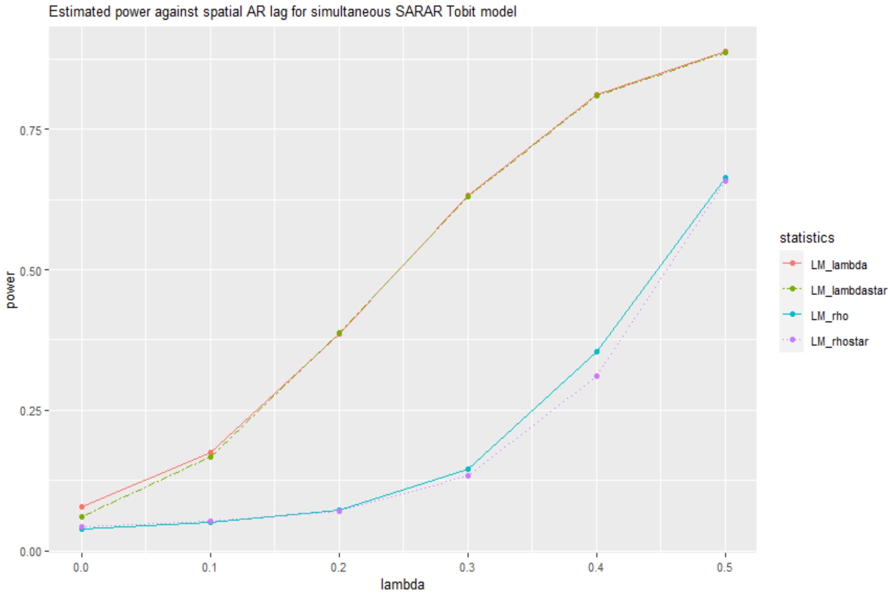


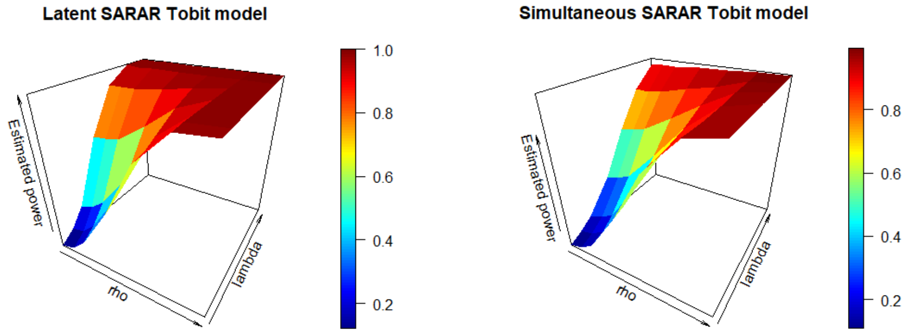
Table A.9: Estimated rejection probabilities for latent SARAR Tobit model

ρ	λ	LM_ρ	LM_ρ^*	LM_λ	LM_λ^*	$LM_{\rho\lambda}$
0.05	0.1	0.122	0.099	0.287	0.274	0.284
0.05	0.2	0.235	0.170	0.672	0.664	0.654
0.05	0.3	0.411	0.320	0.937	0.934	0.916
0.05	0.4	0.479	0.380	0.928	0.919	0.914
0.05	0.5	0.954	0.921	1.000	0.999	1.000
0.1	0.1	0.218	0.196	0.306	0.290	0.351
0.1	0.2	0.364	0.276	0.685	0.668	0.672
0.1	0.3	0.556	0.469	0.927	0.924	0.919
0.1	0.4	0.786	0.672	0.994	0.994	0.995
0.1	0.5	0.943	0.918	1.000	0.999	1.000
0.1	0.05	0.199	0.178	0.183	0.156	0.248
0.2	0.05	0.527	0.509	0.228	0.194	0.505
0.3	0.05	0.846	0.837	0.323	0.261	0.811
0.4	0.05	0.988	0.989	0.474	0.381	0.983
0.5	0.05	1.000	0.999	0.718	0.582	1.000
0.1	0.1	0.218	0.196	0.306	0.290	0.351
0.2	0.1	0.557	0.528	0.350	0.317	0.570
0.3	0.1	0.864	0.843	0.436	0.380	0.848
0.4	0.1	0.988	0.986	0.566	0.478	0.989
0.5	0.1	1.000	1.000	0.775	0.656	1.000

Table A.10: Estimated rejection probabilities for simultaneous SARAR Tobit model

ρ	λ	LM_ρ	LM^*_ρ	LM_λ	LM^*_λ	$LM_{\rho\lambda}$
0.05	0.1	0.082	0.078	0.191	0.179	0.199
0.05	0.2	0.125	0.107	0.393	0.390	0.377
0.05	0.3	0.232	0.183	0.637	0.630	0.636
0.05	0.4	0.424	0.362	0.802	0.802	0.836
0.05	0.5	0.699	0.674	0.879	0.878	0.971
0.1	0.1	0.174	0.158	0.214	0.200	0.250
0.1	0.2	0.236	0.198	0.410	0.407	0.423
0.1	0.3	0.340	0.280	0.645	0.641	0.665
0.1	0.4	0.501	0.413	0.792	0.788	0.845
0.1	0.5	0.713	0.696	0.875	0.871	0.969
0.1	0.05	0.162	0.154	0.148	0.128	0.196
0.2	0.05	0.454	0.452	0.211	0.182	0.452
0.3	0.05	0.782	0.781	0.327	0.266	0.761
0.4	0.05	0.965	0.963	0.488	0.418	0.948
0.5	0.05	0.998	0.999	0.666	0.569	0.996
0.1	0.1	0.174	0.158	0.214	0.200	0.250
0.2	0.1	0.474	0.457	0.281	0.261	0.493
0.3	0.1	0.792	0.788	0.386	0.348	0.767
0.4	0.1	0.962	0.967	0.523	0.474	0.948
0.5	0.1	1.000	1.000	0.664	0.591	0.996

Figure A.5: Power surface plots of $LM_{\rho\lambda}$ for latent and simultaneous SARAR Tobit models



(a) Power surface plots for $LM_{\rho\lambda}$ for latent SARAR Tobit model

(b) Power surface plots for $LM_{\rho\lambda}$ for simultaneous SARAR Tobit model

Appendix B

Modelling the Spatial Impact of Crime on Housing Prices: Evidence from Chicago City

B.1 Tables and Figures

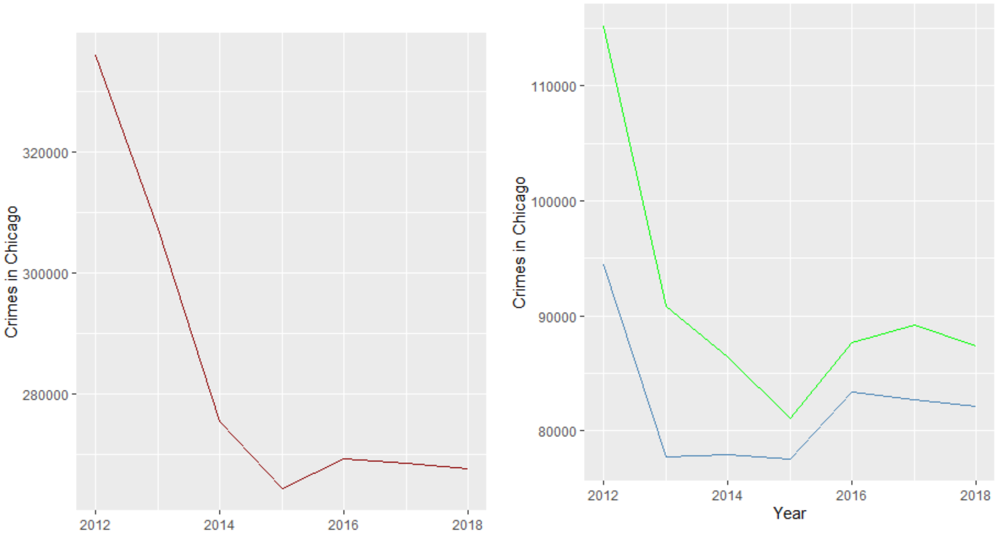
Table B.1: Summary of Variables

	Mean	Range	Std.deviation
House prices(in thousands)	280.4	5.0-1471.6	218.3
Total crime	329.3	3.0-4380.0	306.6
Violent crime	102.5	1.0-793.0	92.1
Property crime	30.9	1.0-110.0	20.6
Criminal Damage	36.2	1.0-168.0	26.4
Burglary	16.2	1.0-70.0	12.1
Theft	79.1	1-2329	133.5
Number of bathrooms	1.9	1.0-3.0	0.4
Floor area	1259	0.0-4000.0	469.9

Table B.2: Crime Variable Descriptions

Crime Variable	Description
Violent crime	Violent crime includes murder, rape and sexual assault, robbery, and assault. Information about murder is obtained on a yearly basis from the FBI's Uniform Crime Reports.
Property crime	Property crime includes burglary/trespassing, motor-vehicle theft, and property theft.
Criminal damage	Criminal damage is a crime against property and is a charge that can lead to prison or jail time, fines, and costs to repair the damaged property.
Burglary	Burglary is the unlawful entry of a building at night with the intent to commit a felony therein.
Theft	Theft is the generic term for all crimes in which a person intentionally takes personal property of another without permission or consent and with the intent to convert it to the taker's use (including potential sale).

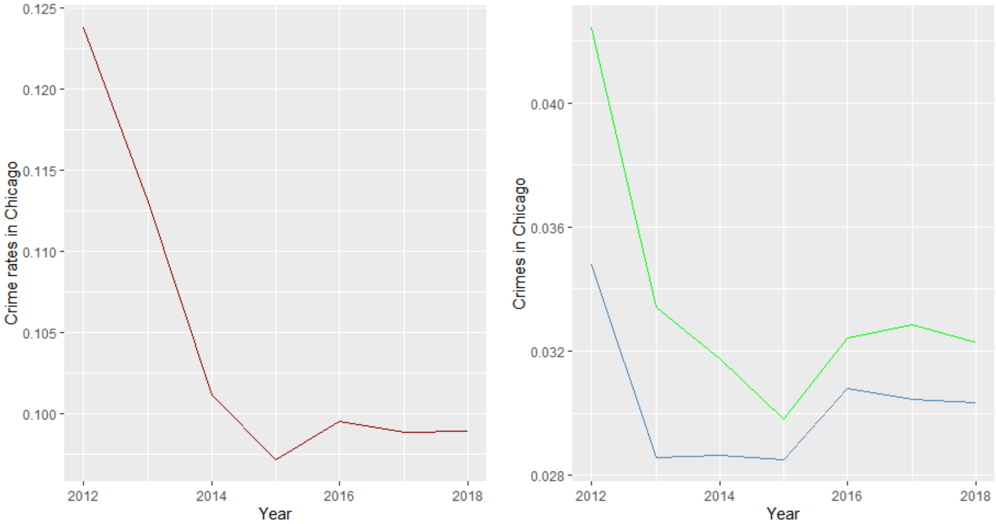
Figure B.1: Different Crime Changes in Chicago City



(a) Total Crimes in Chicago from 2012 to 2018

(b) Violent and Property Crimes in Chicago from 2012 to 2018

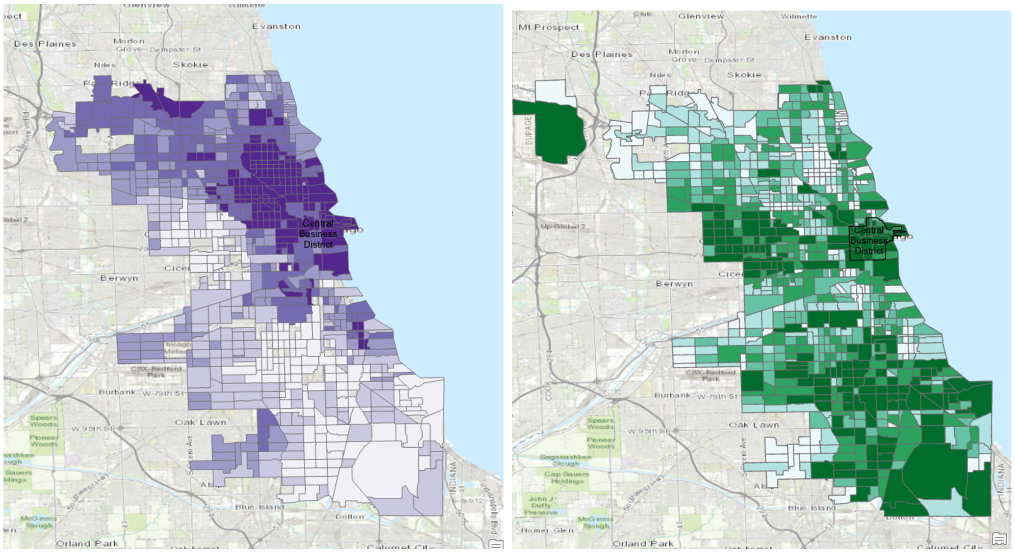
Figure B.2: Different Crime Rates Change in Chicago City



(a) Total Crimes Rates in Chicago from 2012 to 2018

(b) Violent and Property Crimes Rates in Chicago from 2012 to 2018

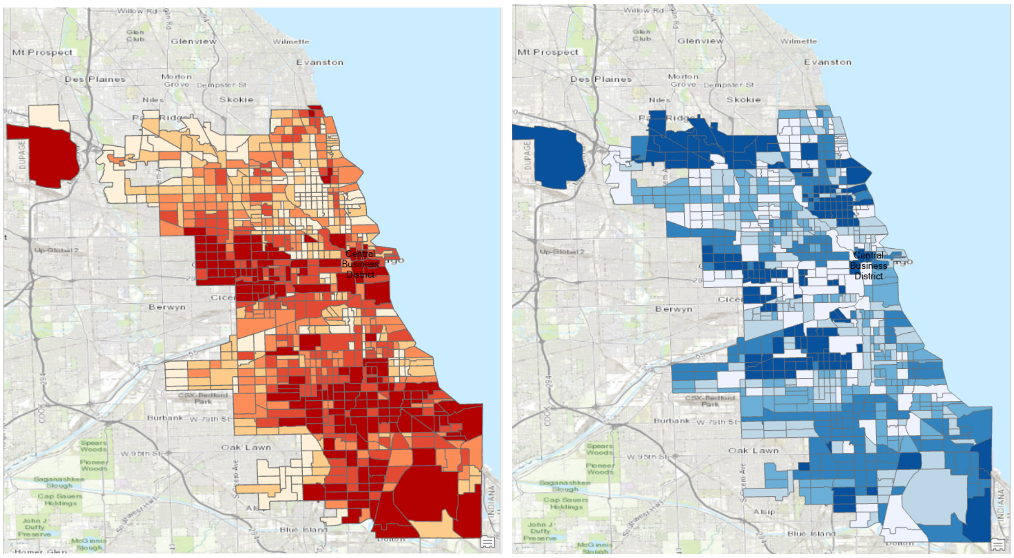
Figure B.3: House prices and total crime quantiles in Chicago city (darker color means higher prices or larger number of crime)



(a) Chicago City Housing Prices Quantile for Census Tracts in 2017

(b) Chicago City Total Crime Quantile for Census Tracts in 2017

Figure B.4: Violent crime and property crime quantiles in Chicago city (darker color means larger number of crime)



(a) Chicago City Violent Crime Quantile for Census Tracts in 2017

(b) Chicago City Property Crime Quantile for Census Tracts in 2017

Table B.3: Moran's I Test Results

Moran's I tests	
Weight matrix: Queen criterion	
	Moran's I statistics
Housing prices	0.758***
Total crimes	0.460***
Violent crimes	0.338***
Property crimes	0.474***
Weight matrix: Rook criterion	
	Moran's I statistics
Housing prices	0.765***
Total crimes	0.401***
Violent crimes	0.367***
Property crimes	0.494***

Table B.4: Spatial MLE results based on queen criterion

	Total Crime	Violent Crime	Property Crime	Criminal Damage	Burglary	Theft
Intercept	0.276.	0.443**	0.215	0.276.	0.124	0.003
	(0.150)	(0.154)	(0.143)	(0.147)	(0.139)	(0.211)
No. of Baths	0.196***	0.192***	0.205***	0.199***	0.207***	0.211***
	(0.017)	(0.017)	(0.016)	(0.017)	(0.016)	(0.017)
Floor Area	0.132***	0.139***	0.125**	0.125**	0.115**	0.113**
	(0.040)	(0.040)	(0.039)	(0.040)	(0.039)	(0.039)
Distance CBD	-0.004*	-0.005*	-0.003	-0.003	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Crime Types	-0.060***	-0.088***	-0.066***	-0.068***	-0.003.	0.018
	(0.018)	(0.015)	(0.018)	(0.018)	(0.016)	(0.017)
Spatial Lag (λ)	0.835***	0.805***	0.837***	0.828***	0.847***	0.856***
Pseudo R^2	0.826	0.831	0.827	0.827	0.825	0.824

Note: Standard error s are in the parentheses. Significance level are as follows: '***' for $p < 0.001$; '**' for $p < 0.01$; '*' for $p < 0.05$; '.' for $p < 0.1$.

Table B.5: Spatial 2SLS estimation results based on queen criterion

	Total Crime	Violent Crime	Property Crime	Criminal Damage	Burglary	Theft
Intercept	0.210	0.156	0.043	0.076	-0.137	-0.266
	(0.286)	(0.269)	(0.244)	(0.259)	(0.236)	(0.231)
No. of Baths	0.195***	0.186***	0.199***	0.193***	0.197***	0.200***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)
Floor Area	0.135***	0.150***	0.134**	0.134**	0.131**	0.132**
	(0.041)	(0.040)	(0.041)	(0.041)	(0.041)	(0.041)
Distance CBD	-0.004	-0.003	-0.002	-0.002	-0.001	0.000
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Crime Types	-0.057*	-0.071***	-0.057***	-0.056**	-0.022	0.019
	(0.022)	(0.021)	(0.020)	(0.022)	(0.018)	(0.017)
Spatial Lag (λ)	0.844***	0.847***	0.862***	0.857***	0.886***	0.987***
	(0.021)	(0.035)	(0.032)	(0.034)	(0.031)	(0.031)

Note: Standard error s are in the parentheses. Significance level are as follows: '***' for $p < 0.001$; '**' for $p < 0.01$; '*' for $p < 0.05$; '.' for $p < 0.1$.

Table B.6: Spatial MLE results for SDM based on queen criterion

	Total Crime	Violent Crime	Property Crime	Criminal Damage	Burglary	Theft
Intercept	1.267***	1.381***	1.170***	1.213***	1.144***	1.020***
	(0.255)	(0.265)	(0.249)	(0.255)	(0.246)	(0.243)
No. of Baths	0.215***	0.212***	0.223***	0.218***	0.225***	0.228***
	(0.017)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)
Floor Area	0.206**	0.211***	0.196***	0.202***	0.193***	0.187***
	(0.042)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Distance CBD	-0.012	-0.017	0.001	-0.010	0.002	-0.000
	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)	(0.024)
Crime Types	-0.030	-0.044*	-0.031	-0.017	0.007	0.019
	(0.021)	(0.019)	(0.021)	(0.021)	(0.019)	(0.018)
Lag of Baths	-0.115***	-0.130***	-0.078*	-0.108**	-0.075*	-0.051
	(0.034)	(0.033)	(0.032)	(0.033)	(0.031)	(0.034)
Lag of floor	-0.312***	-0.260**	-0.354***	-0.301***	-0.376**	-0.417***
	(0.079)	(0.079)	(0.076)	(0.079)	(0.075)	(0.076)
Lag of CBD	0.007	0.010	-0.004	0.007	-0.004	0.000
	(0.025)	(0.025)	(0.025)	(0.024)	(0.025)	(0.025)
Lag of Crime	-0.080*	-0.108***	-0.053	-0.113***	-0.064*	0.036
	(0.035)	(0.031)	(0.033)	(0.034)	(0.031)	(0.030)
Spatial Lag (λ)	0.844***	0.847***	0.846***	0.829***	0.854***	0.865***
Pseudo R^2	0.826	0.844	0.839	0.840	0.838	0.837

Note: Standard errors are in the parentheses. Significance level are as follows: '***' for $p < 0.001$; '**' for $p < 0.01$; '*' for $p < 0.05$; '.' for $p < 0.1$.

Appendix C

Bootstrap-Based Inference of High-Dimensional Means and Its Applications

C.1 Proofs of Lemmas and Theories

C.1.1 Proof of Lemma 3.2.1

$$
\begin{aligned}
Var(T_n) &= Var\left(\frac{2}{n(n-1)}\sum_{t=2}^{n}\sum_{s=1}^{t-1} Z_t^T Z_s\right) \\
&= \frac{4}{n^2(n-1)^2}\sum_{t=2}^{n} E\left(\left(\sum_{s=1}^{t-1} Z_t^T Z_s\right)^2\right) \\
&= \frac{4}{n^2(n-1)^2}\sum_{t=2}^{n} E\left(\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} Z_t^T Z_s Z_{s'}^T Z_t\right) \\
&= \frac{4}{n^2(n-1)^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} E(tr(Z_t^T Z_s Z_{s'}^T Z_t)) \\
&= \frac{4}{n^2(n-1)^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} tr(E(Z_s Z_{s'}^T))E(Z_t Z_t^T)) \\
&= \frac{4}{n^2(n-1)^2}\sum_{t=2}^{n}(t-1)tr(\Sigma^2) \\
&= \frac{2}{n(n-1)}tr(\Sigma^2) \\
&= \frac{1}{\binom{n}{2}}||\Sigma||_F^2
\end{aligned}
$$

where $||\Sigma||_F$ is the Frobenius norm.

C.1.2 Proof of Proposition 3.2.1

Note that

$$D_2(Z)^4 := \frac{E|Z_t^T Z_s|^4}{||\Sigma||_F^4},$$

We can write the numerator as

$$\begin{aligned}E[|Z_t^T Z_s|^4] &= E[Z_t^T Z_s Z_s^T Z_t Z_t^T Z_s Z_s^T Z_t] \\ &= \sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \sum_{j_3=1}^{d} \sum_{j_4=1}^{d} E[Z_{tj_1} Z_{tj_2} Z_{tj_3} Z_{tj_4}] E[Z_{sj_1} Z_{sj_2} Z_{sj_3} Z_{sj_4}],\end{aligned}$$

where

$$\begin{aligned}E[Z_{tj_1} Z_{tj_2} Z_{tj_3} Z_{tj_4}] &= cov(Z_{tj_1}, Z_{tj_2}) cov(Z_{tj_3}, Z_{tj_4}) + cov(Z_{tj_1}, Z_{tj_3}) cov(Z_{tj_2}, Z_{tj_4}) \\ &\quad + cov(Z_{tj_1}, Z_{tj_4}) cov(Z_{tj_2}, Z_{tj_3}) + cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4}),\end{aligned}$$

with $cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4})$ being the fourth order joint cumulant. Since $Var(Z_t) = \Sigma$, $cov(Z_{tj_i}, Z_{tj_k})$ equals to the (j_i, j_k) component of Σ, which can be represented as Σ_{j_i,j_k}. Then

$$\begin{aligned}&\sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \sum_{j_3=1}^{d} \sum_{j_4=1}^{d} E[Z_{tj_1} Z_{tj_2} Z_{tj_3} Z_{tj_4}] E[Z_{sj_1} Z_{sj_2} Z_{sj_3} Z_{sj_4}] \\ &= \sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \sum_{j_3=1}^{d} \sum_{j_4=1}^{d} [\Sigma_{j_1,j_2} \Sigma_{j_3,j_4} + \Sigma_{j_1,j_3} \Sigma_{j_2,j_4} + \Sigma_{j_1,j_4} \Sigma_{j_2,j_3} \\ &+ cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4})]^2,\end{aligned}$$

where the summand can be decomposed as

$$\begin{aligned}&[\Sigma_{j_1,j_2}^2 \Sigma_{j_3,j_4}^2 + \Sigma_{j_1,j_3}^2 \Sigma_{j_2,j_4}^2 + \Sigma_{j_1,j_4}^2 \Sigma_{j_2,j_3}^2 + 2\Sigma_{j_1,j_2} \Sigma_{j_3,j_4} \Sigma_{j_1,j_3} \Sigma_{j_2,j_4} \\ &+ 2\Sigma_{j_1,j_2} \Sigma_{j_3,j_4} \Sigma_{j_1,j_4} \Sigma_{j_2,j_3} + 2\Sigma_{j_1,j_3} \Sigma_{j_2,j_4} \Sigma_{j_1,j_4} \Sigma_{j_2,j_3} \\ &+ 2\Sigma_{j_1,j_2} \Sigma_{j_3,j_4} cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4}) + 2\Sigma_{j_1,j_3} \Sigma_{j_2,j_4} cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4}) \\ &+ 2\Sigma_{j_1,j_4} \Sigma_{j_2,j_3} cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4}) + cum(Z_{tj_1}, Z_{tj_2}, Z_{tj_3}, Z_{tj_4})^2].\end{aligned}$$

Since there are four types of items in the above equation, we can discuss them separately.
(1) For $\sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \sum_{j_3=1}^{d} \sum_{j_4=1}^{d} \Sigma_{j_1,j_2}^2 \Sigma_{j_3,j_4}^2$, it is just $||\Sigma||_F^4$.
(2) For $\sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \sum_{j_3=1}^{d} \sum_{j_4=1}^{d} \Sigma_{j_1,j_2} \Sigma_{j_3,j_4} \Sigma_{j_1,j_3} \Sigma_{j_2,j_4}$, we can use Cauchy-Schwarz inequality here. Thus

$$\begin{aligned}\sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \sum_{j_3=1}^{d} \sum_{j_4=1}^{d} \Sigma_{j_1,j_2} \Sigma_{j_3,j_4} \Sigma_{j_1,j_3} \Sigma_{j_2,j_4} &\leq \left(\sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \sum_{j_3=1}^{d} \sum_{j_4=1}^{d} \Sigma_{j_1,j_2}^2 \Sigma_{j_3,j_4}^2 \right)^{\frac{1}{2}} \\ &\times \left(\sum_{j_1=1}^{d} \sum_{j_2=1}^{d} \sum_{j_3=1}^{d} \sum_{j_4=1}^{d} \Sigma_{j_1,j_3}^2 \Sigma_{j_2,j_4}^2 \right)^{\frac{1}{2}} \\ &= ||\Sigma||_F^4.\end{aligned}$$

(3) For $\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{j_3=1}^{d}\sum_{j_4=1}^{d}\Sigma_{j_1,j_2}\Sigma_{j_3,j_4}cum(Z_{tj_1},Z_{tj_2},Z_{tj_3},Z_{tj_4})$, we can also apply Cauchy-Schwarz inequality. Under Equation (3.2.1), we have

$$\begin{aligned}
&\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{j_3=1}^{d}\sum_{j_4=1}^{d}\Sigma_{j_1,j_2}\Sigma_{j_3,j_4}cum(Z_{tj_1},Z_{tj_2},Z_{tj_3},Z_{tj_4})\\
&\leq\left(\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{j_3=1}^{d}\sum_{j_4=1}^{d}\Sigma_{j_1,j_2}^2\Sigma_{j_3,j_4}^2\right)^{\frac{1}{2}}\left(\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{j_3=1}^{d}\sum_{j_4=1}^{d}cum(Z_{tj_1},Z_{tj_2},Z_{tj_3},Z_{tj_4})^2\right)^{\frac{1}{2}}\\
&=o(n||\Sigma||_F^4).
\end{aligned}$$

(4) For $\sum_{j_1=1}^{d}\sum_{j_2=1}^{d}\sum_{j_3=1}^{d}\sum_{j_4=1}^{d}cum(Z_{tj_1},Z_{tj_2},Z_{tj_3},Z_{tj_4})^2$, it is just $o(n||\Sigma||_F^4)$ under Equation (3.2.1). Combining all the cases, the conclusion follows.

C.1.3 Proof of Proposition 3.2.2

$$\begin{aligned}
Var^*(T_n^*) &= Var(T_n^*|\mathcal{F}_n)\\
&=\frac{1}{\binom{n}{2}^2}\sum_{t=2}^{n}Var\left(\sum_{s=1}^{t-1}w_tZ_t^TZ_sw_s|\mathcal{F}_n\right)\\
&=\frac{1}{\binom{n}{2}^2}\sum_{t=2}^{n}E\left(\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}w_tZ_t^TZ_sw_sw_{s'}Z_{s'}^TZ_tw_t|\mathcal{F}_n\right)\\
&=\frac{1}{\binom{n}{2}^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}Z_t^TZ_sZ_{s'}^TZ_tE(w_tw_sw_{s'}w_t)\\
&=\frac{1}{\binom{n}{2}^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}Z_t^TZ_sZ_{s'}^TZ_tE(w_t^2)E(w_sw_{s'})\\
&=\frac{1}{\binom{n}{2}^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}Z_t^TZ_sZ_s^TZ_t
\end{aligned}\tag{C.1.1}$$

Since $Var(T_n)=\frac{1}{\binom{n}{2}}||\Sigma||_F^2$,

$$\frac{Var^*(T_n^*)}{Var(T_n)}=\frac{1}{\binom{n}{2}||\Sigma||_F^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}Z_t^TZ_sZ_s^TZ_t.$$

Then

$$\begin{aligned}
E\left[\frac{Var^*(T_n^*)}{Var(T_n)}\right]&=\frac{1}{\binom{n}{2}||\Sigma||_F^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}E[tr(Z_sZ_s^TZ_tZ_t^T)]\\
&=\frac{1}{\binom{n}{2}||\Sigma||_F^2}\sum_{t=2}^{n}(t-1)tr(\Sigma^2)=1.
\end{aligned}\tag{C.1.2}$$

We can compute the variance as

$$
\begin{aligned}
Var\left[\frac{Var^*(T_n^*)}{Var(T_n)}\right] &= E\left[\left(\frac{1}{\binom{n}{2}||\Sigma||_F^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} Z_t^T Z_s Z_s^T Z_t - 1\right)^2\right] \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4} E\left[\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1} Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}\right] - 1.
\end{aligned}
\tag{C.1.3}
$$

We have different cases for Equation (C.1.3) under different values of $\{t, t', s, s'\}$.
(1) When none of them are equal, then

$$
\begin{aligned}
&\frac{4}{n^2(n-1)^2||\Sigma||_F^4} E\left[\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2, t'\neq t, t'\neq s}^{n}\sum_{s'=1, s'\neq s, s'\neq t}^{t'-1} Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}\right] \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{s=1}^{t-1} E[Z_t^T Z_s Z_s^T Z_t] \sum_{t'=2, t'\neq t, t'\neq s}^{n}\sum_{s'=1, s'\neq s, s'\neq t}^{t'-1} E[Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}] \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{s=1}^{t-1} tr(\Sigma^2) \sum_{t'=2, t'\neq t, t'\neq s}^{n}\sum_{s'=1, s'\neq s, s'\neq t}^{t'-1} tr(\Sigma^2) \to 1.
\end{aligned}
$$

(2) When there is at least one equality among $\{t, t', s, s'\}$, then there are at most three summations. Since the absolute value of the summand $E(Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'})$ is bounded by $D_2(Z)^4||\Sigma||_F^4$ by Hölder's inequality, which is

$$
E|Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}| \leq \{E[(Z_t^T Z_s)^4)]\}^{\frac{1}{2}}\{E[(Z_{t'}^T Z_{s'})^4)]\}^{\frac{1}{2}} = D_2(Z)^4||\Sigma||_F^4. \tag{C.1.4}
$$

Then the summation is bounded by $\frac{D_2(Z)^4}{n}$, which converges to zero under Assumption 3.2.1.
Thus

$$
\frac{Var^*(T_n^*)}{Var(T_n)} \xrightarrow{p} 1,
$$

in view of Equations (C.1.2) and (C.1.3).

C.1.4 Proof of Theorem 3.3.1

According to the main theorem in Heyde and Brown [49], we have

$$
sup_x\left|P\left(\frac{T_n}{\sqrt{Var(T_n)}} \leq x\right) - \Phi(x)\right| \leq K_2(V_{n,1} + V_{n,2})^{\frac{1}{5}},
$$

where

$$
V_{n,1} = \frac{\sum_{t=2}^{n} E|Y_t|^4}{Var(T_n)^2},\ V_{n,2} = \frac{E|(\sum_{t=2}^{n} Y_t^2) - Var(T_n)|^2}{Var(T_n)^2},
$$

and K_2 is a generic positive constant.
Let $a(n) = \frac{16}{n^4(n-1)^4}$, then

$$\begin{aligned}
\sum_{t=2}^{n} E(Y_t^4) &= \sum_{t=2}^{n} E\Big[\Big(\sum_{s=1}^{t-1} \frac{2}{n(n-1)} Z_t^T Z_s\Big)^4\Big] \\
&= a(n) \sum_{t=2}^{n} E\Big[\Big(\sum_{s=1}^{t-1} Z_t^T Z_s\Big)^4\Big] \\
&= a(n) \sum_{t=2}^{n} E\Big(\sum_{s_1=1}^{t-1}\sum_{s_2=1}^{t-1}\sum_{s_3=1}^{t-1}\sum_{s_4=1}^{t-1} Z_t^T Z_{s_1} Z_{s_2}^T Z_t Z_t^T Z_{s_3} Z_{s_4}^T Z_t\Big) \\
&= a(n) \sum_{t=2}^{n}\sum_{s=1}^{t-1} E[(Z_t^T Z_s)^4] + 3a(n) \sum_{t=2}^{n}\sum_{s_1=1}^{t-1}\sum_{s_2\neq s_1}^{t-1} E[Z_t^T Z_{s_1} Z_{s_1}^T Z_t Z_t^T Z_{s_2} Z_{s_2}^T Z_t].
\end{aligned}$$

By Lemma 3.2.1 and Equation (C.1.4), we have

$$\begin{aligned}
V_{n,1} &= \frac{\sum_{t=2}^{n} E(Y_t^4)}{Var(T_n)^2} \\
&\leq \frac{4}{n^2(n-1)^2} \sum_{t=2}^{n}\sum_{s=1}^{t-1} D_2(Z)^4 + \frac{12}{n^2(n-1)^2} \sum_{t=2}^{n}\sum_{s_1=1}^{t-1}\sum_{s_2\neq s_1}^{t-1} D_2(Z)^4 \\
&\leq K \frac{D_2(Z)^4}{n}.
\end{aligned}$$

Then for $V_{n,2}$, we have

$$\begin{aligned}
V_{n,2} &= E\Big\{\frac{|(\sum_{t=2}^{n} Y_t^2) - Var(T_n)|^2}{Var(T_n)^2}\Big\} \\
&= E\Big(\frac{1}{\binom{n}{2}||\Sigma||_F^2} \sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} Z_t^T Z_s Z_{s'}^T Z_t - 1\Big)^2 \\
&= \frac{1}{\binom{n}{2}^2||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{s_1=1}^{t'-1}\sum_{s_1'=1}^{t'-1} E(Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'}) \\
&- \frac{2}{\binom{n}{2}||\Sigma||_F^2} \sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} E(Z_t^T Z_s Z_{s'}^T Z_t) + 1 \\
&= \frac{1}{\binom{n}{2}^2||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{s_1=1}^{t'-1}\sum_{s_1'=1}^{t'-1} E(Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'}) - 1.
\end{aligned}$$

Let Ω denote the set of indices with no particular equality constraints among $\{t, t', s, s', s_1, s_1'\}$, i.e.

$$\Omega = \{(t, t', s, s', s_1, s_1')|t, t' = 2, ..., n; s, s' = 1, ..., t-1; s_1, s_1' = 1, ..., t'-1\}.$$

Then we can write $\Omega = \Omega_0 \cup \Omega_1 \cup \Omega_2 \cup \Omega_3 \cup \Omega_4$, where

$$\Omega_j = \{(t, t', s, s', s_1, s_1')|\text{there are j equalities among } (t, t', s, s', s_1, s_1') \in \Omega\}$$

for $j = 0, 1, 2, 3, 4$.

(1) When $(t, t', s, s', s_1, s_1') \in \Omega_0$, we can always find one index in $\{s, s', s_1, s_1'\}$ that is distinct from all others, which leads to

$$\frac{1}{\binom{n}{2}^2 ||\Sigma||_F^4} \sum_{\Omega_0} E(Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'}) = 0.$$

(2) When $(t, t', s, s', s_1, s_1') \in \Omega_1$, this is similar to **(1)**. There is also at least one index in $\{s, s', s_1, s_1'\}$ that is distinct from all others, which leads to

$$\frac{1}{\binom{n}{2}^2 ||\Sigma||_F^4} \sum_{\Omega_1} E(Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'}) = 0.$$

(3) When $(t, t', s, s', s_1, s_1') \in \Omega_2$, we have two kinds of cases, and we use one example for each case.

(3a) When $s = s' \neq s_1 = s_1'$, $t \neq t', t' \neq s, t \neq s_1$, here we define $\Omega_3' = \{(t, t', s, s', s_1, s_1') | s = s' \neq s_1 = s_1', t = t', \text{or } t' = s, \text{or } t = s_1\}$.

$$\begin{aligned}
&\frac{4}{n^2(n-1)^2 ||\Sigma||_F^4} E\Big[\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2, t'\neq t, t'\neq s}^{n}\sum_{s_1=1, s_1\neq s, s_1\neq t}^{t'-1} Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s_1} Z_{s_1}^T Z_{t'}\Big] \\
&= \frac{4}{n^2(n-1)^2 ||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2, t'\neq t, t'\neq s}^{n}\sum_{s_1=1, s_1\neq s, s_1\neq t}^{t'-1} E(Z_t^T Z_s Z_s^T Z_t) E(Z_{t'}^T Z_{s_1} Z_{s_1}^T Z_{t'}) \\
&= \frac{4}{n^2(n-1)^2 ||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2, t'\neq t, t'\neq s}^{n}\sum_{s_1=1, s_1\neq s, s_1\neq t}^{t'-1} tr(\Sigma^2)^2 \\
&= \frac{4}{n^2(n-1)^2 ||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2}^{n}\sum_{s_1=1}^{t'-1} tr(\Sigma^2)^2 - \frac{4}{n^2(n-1)^2 ||\Sigma||_F^4} \sum_{\Omega_3'} tr(\Sigma^2)^2 \\
&= 1 - H_n,
\end{aligned}$$

where $|H_n| \leq \frac{K}{n}$, since there are at most three summations in $\Sigma_{\Omega_3'}$.

Thus we have

$$\left|\frac{4}{n^2(n-1)^2 ||\Sigma||_F^4} E\Big[\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2, t'\neq t, t'\neq s}^{n}\sum_{s_1=1, s_1\neq s, s_1\neq t}^{t'-1} Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s_1} Z_{s_1}^T Z_{t'}\Big] - 1\right| \leq \frac{K}{n}. \tag{C.1.5}$$

(3b) When $s = s_1 \neq s' = s_1'$, $t \neq t'$, we have

$$\begin{aligned}
&\frac{1}{\binom{n}{2}^2 ||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{t'=2, t'\neq t}^{n}\sum_{s=1}^{t\wedge t'-1}\sum_{s'=1, s'\neq s}^{t\wedge t'-1} E(Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s'} Z_s^T Z_{t'}) \\
&= \frac{1}{\binom{n}{2}^2 ||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{t'=2, t'\neq t}^{n}\sum_{s=1}^{t\wedge t'-1}\sum_{s'=1, s'\neq s}^{t\wedge t'-1} E[Z_s^T Z_t Z_t^T Z_{s'} Z_{s'}^T Z_{t'} Z_{t'}^T Z_s] \\
&= \frac{1}{\binom{n}{2}^2 ||\Sigma||_F^4} \sum_{t=2}^{n}\sum_{t'=2, t'\neq t}^{n}\sum_{s=1}^{t\wedge t'-1}\sum_{s'=1, s'\neq s}^{t\wedge t'-1} tr\{E[Z_t Z_t^T] E[Z_{s'} Z_{s'}^T] E[Z_{t'} Z_{t'}^T] E[Z_s Z_s^T]\} \\
&\leq K \frac{tr(\Sigma^4)}{||\Sigma||_F^4}.
\end{aligned} \tag{C.1.6}$$

Thus

$$\left|\frac{1}{\binom{n}{2}^2||\Sigma||_F^4}\sum_{\Omega_2}E(Z_t^TZ_sZ_{s'}^TZ_tZ_{t'}^TZ_{s_1'}Z_{s_1}^TZ_{t'})-1\right|\leq K\left[\frac{1}{n}+\frac{tr(\Sigma^4)}{||\Sigma||_F^4}\right].$$

(4) When $(t,t',s,s',s_1,s_1')\in\Omega_3$ or Ω_4, there are three or two summations for the term, and the summand is bounded by $D_2(Z)^4||\Sigma||_F^4$ by Equation (C.1.4). The term is bounded by $K\frac{D_2(Z)^4}{n}$.
We combine all the cases and get

$$\begin{aligned}V_{n,2}&=\frac{1}{\binom{n}{2}^2||\Sigma||_F^4}\sum_{\Omega}E(Z_t^TZ_sZ_{s'}^TZ_tZ_{t'}^TZ_{s_1'}Z_{s_1}^TZ_{t'})-1\\&\leq K\Big(\frac{D_2(Z)^4+1}{n}+\frac{tr(\Sigma^4)}{||\Sigma||_F^4}\Big).\end{aligned}$$

The result follows.

C.1.5 Proof of Theorem 3.3.2

According to the main theorem in Heyde and Brown [49], we have

$$sup_x\left|P^*\Bigg(\frac{T_n^*}{\sqrt{Var^*(T_n^*)}}\leq x\Bigg)-\Phi(x)\right|\leq K_2(V_{n,3}+V_{n,4})^{\frac{1}{5}},$$

where

$$V_{n,3}=\frac{\sum_{t=2}^nE^*|Y_t^*|^4}{Var^*(T_n^*)^2},\ \ V_{n,4}=\frac{E^*|(\sum_{t=2}^nY_t^{*2})-Var^*(T_n^*)|^2}{Var^*(T_n^*)^2}.$$

Since

$$\begin{aligned}&\sum_{t=2}^nE^*(Y_t^{*4})\\&=\sum_{t=2}^nE^*\Bigg[\Bigg(\sum_{s=1}^{t-1}\frac{2}{n(n-1)}w_tZ_t^TZ_sw_s\Bigg)^4\Bigg]\\&=a(n)\sum_{t=2}^n\sum_{s_1=1}^{t-1}\sum_{s_2=1}^{t-1}\sum_{s_3=1}^{t-1}\sum_{s_4=1}^{t-1}E(Z_t^TZ_{s_1}Z_{s_2}^TZ_tZ_t^TZ_{s_3}Z_{s_4}^TZ_tw_t^4w_{s_1}w_{s_2}w_{s_3}w_{s_4}|F_n)\\&=a(n)\kappa\sum_{t=2}^n\sum_{s_1=1}^{t-1}\sum_{s_2=1}^{t-1}\sum_{s_3=1}^{t-1}\sum_{s_4=1}^{t-1}Z_t^TZ_{s_1}Z_{s_2}^TZ_tZ_t^TZ_{s_3}Z_{s_4}^TZ_tE(w_{s_1}w_{s_2}w_{s_3}w_{s_4})\\&=a(n)\kappa^2\sum_{t=2}^n\sum_{s=1}^{t-1}(Z_t^TZ_s)^4+3a(n)\kappa\sum_{t=2}^n\sum_{s_1=1}^{t-1}\sum_{s_2\neq s_1}^{t-1}Z_t^TZ_{s_1}Z_{s_1}^TZ_tZ_t^TZ_{s_2}Z_{s_2}^TZ_t,\end{aligned}$$

under Assumption 3.2.1, it follows from Proposition 3.2.2 that

$$\frac{Var^*(T_n^*)}{Var(T_n)}\xrightarrow{p}1.$$

Similar to the argument used in bounding $V_{n,1}$ in the proof of Theorem 3.3.1, we have

$$E\Bigg(\frac{\sum_{t=2}^nE^*(Y_t^{*4})}{Var(T_n)^2}\Bigg)\leq K\frac{D_2(X)^4}{n},$$

which yields

$$V_{n,3} = O_p\Big(\frac{D_2(Z)^4}{n}\Big).$$

To analyze $V_{n,4}$, we shall focus on

$$\begin{aligned}
&\widetilde{V_{n,4}} \\
&= E^*\Big[\frac{\sum_t Y_t^{*2} - Var^*(T_n^*)}{Var(T_n)}\Big]^2 \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4} E^*\Big\{\Big[\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1} w_t Z_t^T Z_s w_s w_{s'} Z_{s'}^T Z_t w_t - \sum_{t=2}^{n}\sum_{s=1}^{t-1} Z_t^T Z_s Z_s^T Z_t\Big]^2\Big\} \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4} E^*\Big[\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{s_1=1}^{t'-1}\sum_{s_1'=1}^{t'-1} Z_t^T Z_s Z_{s'}^T Z_t w_t^2 w_{t'}^2 w_s w_{s'} w_{s_1'} w_{s_1} Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'} \\
&- 2\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{s_1=1}^{t'-1} Z_t^T Z_s Z_{s'}^T Z_t w_t^2 w_s w_{s'} Z_{t'}^T Z_{s_1} Z_{s_1}^T Z_{t'} \\
&+ \sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1} Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}\Big] \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4}\Big[\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{s_1=1}^{t'-1}\sum_{s_1'=1}^{t'-1} Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'} E(w_t^2 w_{t'}^2 w_s w_{s'} w_{s_1'} w_{s_1}) \\
&- \sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1} Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}\Big].
\end{aligned}$$

Then

$$\begin{aligned}
&E(\widetilde{V_{n,4}}) \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4}\Big[\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t-1}\sum_{s_1=1}^{t'-1}\sum_{s_1'=1}^{t'-1} E(Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'}) E(w_t^2 w_{t'}^2 w_s w_{s'} w_{s_1'} w_{s_1}) \\
&- \sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1} E(Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'})\Big] \\
&= I_{n,1} - I_{n,2}.
\end{aligned}$$

For $I_{n,1}$, for different values of $\{t, t', s, s', s_1, s_1'\}$, $E(w_t^2 w_{t'}^2 w_s w_{s'} w_{s_1'} w_{s_1}) \neq 0$ when there are at least two equalities among $\{t, t', s, s', s_1, s_1'\}$. Then we can write

$$\begin{aligned}
I_{n,1} &= \frac{4}{n^2(n-1)^2||\Sigma||_F^4}\sum_{\Omega} E(Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'}) E(w_t^2 w_{t'}^2 w_s w_{s'} w_{s_1'} w_{s_1}) \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4}\Big(\sum_{\Omega_2} + \sum_{\Omega_3} + \sum_{\Omega_4}\Big) E(Z_t^T Z_s Z_{s'}^T Z_t Z_{t'}^T Z_{s_1'} Z_{s_1}^T Z_{t'}) E(w_t^2 w_{t'}^2 w_s w_{s'} w_{s_1'} w_{s_1}),
\end{aligned}$$

where Ω_j is defined in the proof of Theorem 3.3.1.

(1) When $(t, t', s, s', s_1, s_1') \in \Omega_2$. Then $E(w_t^2 w_{t'}^2 w_s w_{s'} w_{s_1'} w_{s_1}) = E(w_1^2)E(w_2^2)E(w_3^2)E(w_4^2)$. There are two kinds of cases, and we use one example for each case.

(1a) When $s = s' \neq s_1 = s_1'$, $t \neq t', t \neq s_1, t' \neq s$, then we have

$$\left| \frac{4}{n^2(n-1)^2||\Sigma||_F^4} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \sum_{t'=2, t'\neq t, t'\neq s}^{n} \sum_{s_1=1, s_1\neq s, s_1 \neq t}^{t'-1} E(Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s_1} Z_{s_1}^T Z_{t'}) - 1 \right| \leq \frac{K}{n}$$

in view of Equation (C.1.5)
(1b) When $s = s_1 \neq s' = s_1'$, $t \neq t'$, then we have

$$\left| \frac{4}{n^2(n-1)^2||\Sigma||_F^4} \sum_{t=2}^{n} \sum_{t'=2, t'\neq t}^{n} \sum_{s=1}^{t\wedge t'-1} \sum_{s'=1, s'\neq s}^{t\wedge t'-1} E(Z_t^T Z_s Z_{s_1'}^T Z_t Z_{t'}^T Z_{s_1'} Z_s^T Z_{t'}) \right| \leq K \frac{tr(\Sigma^4)}{||\Sigma||_F^4},$$

in view of Equation (C.1.6)
(2) When $(t, t', s, s', s_1, s_1') \in \Omega_3$ or Ω_4, then $E(w_t^2 w_{t'}^2 w_s w_{s'} w_{s_1'} w_{s_1}) = E(w_1^4)E(w_2^2)E(w_3^2)$ or $E(w_1^4)E(w_2^4)$. Thus we have three or two summations for the term, which is bounded by $K\frac{D_2(Z)^4}{n}$.

For $I_{n,2}$, we have different cases for different values of $\{t, t', s, s'\}$
(1') When none of them are equal, then it is the same as the case **(1a)** for $I_{n,1}$, we have

$$|\sum_{t=2}^{n} \sum_{t'=2, t'\neq t}^{n} \sum_{s=1, s\neq t'}^{t-1} \sum_{s'=1, s'\neq s, s'\neq t}^{t'-1} E(Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}) - 1| \leq \frac{K}{n}.$$

(2') When there is at least one equality, there are at most three summations for $I_{n,2}$, thus the term is bounded by $K\frac{D_2(Z)^4}{n}$.
We combine all the cases and get

$$E(\widetilde{V_{n,4}}) \leq K\left(\frac{tr(\Sigma^4)}{||\Sigma||_F^4} + \frac{D_2(Z)^4 + 1}{n}\right),$$

which implies

$$V_{n,4} = O_p\left(\frac{tr(\Sigma^4)}{||\Sigma||_F^4} + \frac{D_2(Z)^4 + 1}{n}\right),$$

in view of Proposition 3.2.2. The conclusion follows.

C.1.6 Proof of Theorem 3.4.1

According to Lahiri's [59] Conditional Slutsky's Theorem (see Lemma 4.1 on page 77). The conclusion follows from Theorem 3.3.2 and the following statement:

$$P^*\left(\left|\frac{\widehat{Var^*(T_n^*)}}{Var^*(T_n^*)} - 1\right| > \epsilon\right) \xrightarrow{p} 0. \tag{C.1.7}$$

To show Equation (C.1.7), it suffices to show

$$E^*\left[\frac{\widehat{Var^*(T_n^*)} - Var^*(T_n^*)}{Var^*(T_n^*)}\right]^2 \xrightarrow{p} 0.$$

In light of the fact that $\frac{Var^*(T_n^*)}{Var(T_n)} \xrightarrow{p} 1$, it suffices to show

$$E^*\left[\frac{\widehat{Var^*(T_n^*)} - Var^*(T_n^*)}{Var(T_n)}\right]^2 \xrightarrow{p} 0,$$

which follows from the following statement by Markov's inequality.

$$E\left\{E^*\left[\frac{\widehat{Var^*(T_n^*)} - Var^*(T_n^*)}{Var(T_n)}\right]^2\right\} \to 0. \tag{C.1.8}$$

To show Equation (C.1.8), we note that

$$\begin{aligned}
E^*[\widehat{Var^*(T_n^*)} - Var^*(T_n^*)]^2 &= E^*\left[\frac{1}{\binom{n}{2}^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} Z_t^T Z_s Z_s^T Z_t w_t^2 w_s^2 - \frac{1}{\binom{n}{2}^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} Z_t^T Z_s Z_s^T Z_t\right]^2 \\
&= \frac{1}{\binom{n}{2}^4}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2}^{n}\sum_{s'=1}^{t'-1} Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}[E(w_t^2 w_s^2 w_{t'}^2 w_{s'}^2) - 1].
\end{aligned}$$

Then

$$\begin{aligned}
E\left\{E^*\left[\frac{\widehat{Var^*(T_n^*)} - Var^*(T_n^*)}{Var(T_n)}\right]^2\right\} &= \frac{4}{n^2(n-1)^2||\Sigma||_F^4} E\left\{\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2}^{n}\sum_{s'=1}^{t'-1}[E(w_t^2 w_s^2 w_{t'}^2 w_{s'}^2) - 1]\right. \\
&\quad \left. \times Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'}\right\} \\
&= \frac{4}{n^2(n-1)^2||\Sigma||_F^4} E(V_n).
\end{aligned}$$

For $E(V_n)$, we have the following cases:

(1) When none of $\{t, t', s, s'\}$ are equal to each other, we have

$$\sum_{t=2}^{n}\sum_{s=1}^{t-1}\sum_{t'=2, t'\neq t, t'\neq s}^{n}\sum_{s'=1, s'\neq s, s'\neq t}^{t'-1} E(Z_t^T Z_s Z_s^T Z_t Z_{t'}^T Z_{s'} Z_{s'}^T Z_{t'})[E(w_t^2)E(w_s^2)E(w_{t'}^2)E(w_{s'}^2) - 1] = 0.$$

(2) When there is at least one equality among (t, t', s, s'), then the four summations reduce to three or less. The summand is bounded by $D_2(Z)^4||\Sigma||_F^4$ by Equation (C.1.4). Therefore the term is bounded by $K\frac{D_2(Z)^4}{n}$, which converges to zero under Assumption 3.2.1.

Thus we can combine all the cases and get the result.

C.1.7 Proof of Theorem 3.5.1

$$\begin{aligned}
\hat{T_n} &= \frac{1}{\binom{n}{2}}\sum_{t=2}^{n}\sum_{s=1}^{t-1}[e_t - (\hat{\beta} - \beta)^T x_t] Z_t^T Z_s [e_s - x_s^T(\hat{\beta} - \beta)] \\
&= T_n - \frac{1}{\binom{n}{2}}\sum_{t=2}^{n}\sum_{s=1}^{t-1}(\hat{\beta} - \beta)^T A_t^T Z_s^* \\
&\quad - \frac{1}{\binom{n}{2}}\sum_{t=2}^{n}\sum_{s=1}^{t-1} Z_t^{*T} A_t(\hat{\beta} - \beta) + \frac{1}{\binom{n}{2}}\sum_{t=2}^{n}\sum_{s=1}^{t-1}(\hat{\beta} - \beta)^T A_t^T A_s(\hat{\beta} - \beta)
\end{aligned} \tag{C.1.9}$$

Here we can assume $\sqrt{n}(\hat{\beta}-\beta) = O_p(1)$ according to the Assumption 4(a) of Anatolyev et al. [51]. We need to show the last three terms are negligible with respect to $\sqrt{Var(T_n)}$.

For the last term in Equation (C.1.9), we need to show that

$$\frac{1}{\binom{n}{2}}(\hat{\beta}-\beta)^T \sum_{t=2}^{n}\sum_{s=1}^{t-1} A_t^T A_s(\hat{\beta}-\beta) = o_p(\sqrt{Var(T_n)}).$$

Since $Var(T_n) = \frac{1}{\binom{n}{2}}||\Sigma^*||_F^2$, then it suffices to show that

$$E\left[\frac{1}{\binom{n}{2}^{\frac{1}{2}}}\frac{1}{n||\Sigma^*||_F}||\sum_{t=2}^{n}\sum_{s=1}^{t-1} A_t^T A_s||_F\right]^2 \to 0.$$

Since

$$\begin{aligned}
&E\left[\frac{1}{\binom{n}{2}^{\frac{1}{2}}}\frac{1}{n||\Sigma^*||_F}||\sum_{t=2}^{n}\sum_{s=1}^{t-1} A_t^T A_s||_F\right]^2 \\
&= \frac{1}{\binom{n}{2}}\frac{1}{n^2||\Sigma^*||_F^2}E\left[||\sum_{t=2}^{n}\sum_{s=1}^{t-1} A_t^T A_s||_F^2\right] \\
&= \frac{1}{\binom{n}{2}}\frac{1}{n^2||\Sigma^*||_F^2}\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1} E[tr(A_t^T A_s A_{s'}^T A_{t'})] \\
&= I_1 + I_2 + I_3 + I_4 + I_5,
\end{aligned} \tag{C.1.10}$$

let $\bar{\Omega}$ denote the set of indices with no particular equality constraints among $\{t, t', s, s'\}$, i.e.

$$\bar{\Omega} = \{(t, t', s, s')|t, t' = 2, ..., n; s = 1, ..., t-1; s' = 1, ..., t'-1\}.$$

Then we can write $\bar{\Omega} = \bar{\Omega}_1 \cup \bar{\Omega}_2 \cup \bar{\Omega}_3 \cup \bar{\Omega}_4 \cup \bar{\Omega}_5$, where

$$\bar{\Omega}_1 = \{(t, t', s, s')|\text{there is no equality among } (t, t', s, s') \in \Omega\}$$

$$\bar{\Omega}_2 = \{(t, t', s, s')|t = t', s \neq s'\}$$

$$\bar{\Omega}_3 = \{(t, t', s, s')|t \neq t', s = s'\}$$

$$\bar{\Omega}_4 = \{(t, t', s, s')|t = s' \text{ or t'=s}\}$$

$$\bar{\Omega}_5 = \{(t, t', s, s')|t = t', s = s'\}.$$

Thus $I_j = \frac{1}{\binom{n}{2}}\frac{1}{n^2||\Sigma^*||_F^2}\sum_{\bar{\Omega}_j} E[tr(A_t^T A_s A_{s'}^T A_{t'})]$.

(1) When $\{t, t', s, s'\} \in \bar{\Omega}_1$, $I_1 = O\left(\frac{||A^T A||_F^2}{||\Sigma^*||_F^2}\right) = o(1)$ according to Equation (3.5.1) in Assumption 3.5.3.

(2) When $\{t, t', s, s'\} \in \bar{\Omega}_2$ or $\bar{\Omega}_3$, $I_2 = I_3 = O\left(\frac{tr(AA^T\Sigma_a)}{n||\Sigma^*||_F^2}\right)$ according to Equation (3.5.1) and Equation (3.5.3) in Assumption 3.5.3. Since

$$\frac{tr(AA^T\Sigma_a)}{n||\Sigma^*||_F^2} \leq \frac{||A^T A||_F}{||\Sigma^*||_F}\frac{||\Sigma_a||_F}{n||\Sigma^*||_F} = o(1),$$

then we have $I_2 = I_3 = O\left(\frac{tr(AA^T\Sigma_a)}{n||\Sigma^*||_F^2}\right) = o(1)$.

(3) When $\{t, t', s, s'\} \in \bar{\Omega}_4$, $I_4 = O\left(\frac{tr(E(A_1^T A_2 A_1^T)A)}{n||\Sigma^*||_F^2}\right) = o(1)$ according to Equation (3.5.2) in Assumption 3.5.3.

(4)$\{t, t', s, s'\} \in \bar{\Omega}_5$, $I_5 = O\left(\frac{||\Sigma_a||_F^2}{n^2||\Sigma^*||_F^2}\right) = o(1)$ according to Equation (3.5.3) in Assumption 3.5.3.

Thus the term in Equation (C.1.10) goes to 0 according to Assumption 3.5.3.

For the second term in Equation (C.1.9),

$$\begin{aligned} E\left[\frac{1}{\binom{n}{2}^{\frac{1}{2}}}\frac{1}{\sqrt{n}||\Sigma^*||_F}||\sum_{t=2}^{n}\sum_{s=1}^{t-1} A_t^T Z_s^* ||_F\right]^2 &= \frac{1}{\binom{n}{2}}\frac{1}{n||\Sigma^*||_F^2}\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1} E[tr(A_t^T Z_s^* Z_{s'}^{*T} A_{t'})] \\ &= I_1' + I_2' + I_3' + I_4' + I_5', \end{aligned} \tag{C.1.11}$$

where $I_j' = \frac{1}{\binom{n}{2}}\frac{1}{n||\Sigma^*||_F^2}\sum_{\bar{\Omega}_j} E[tr(A_t^T Z_s^* Z_{s'}^{*T} A_{t'})]$.

(1) When $\{t, t', s, s'\} \in \bar{\Omega}_1$ and $\bar{\Omega}_2$, $I_1' = I_2' = 0$.

(2) When $\{t, t', s, s'\} \in \bar{\Omega}_3$, $I_3' = O\left(\frac{tr(\Sigma^* AA^T)}{||\Sigma^*||_F^2}\right) = o(1)$ according to Equation (3.5.4) in Assumption 3.5.3.

(3) When $\{t, t', s, s'\} \in \bar{\Omega}_4$, $I_4' = 0$.

(4) When $\{t, t', s, s'\} \in \bar{\Omega}_5$, $I_5' = O\left(\frac{tr(\Sigma^*\Sigma_a)}{n||\Sigma^*||_F^2}\right) = o(1)$ according to Equation (3.5.5) in Assumption 3.5.3.

Thus the term in Equation (C.1.11) goes to 0 according to Assumption 3.5.3.

Similarly, the third term over $\sqrt{Var(T_n)}$ in Equation (C.1.9) also converges to 0 in probability.

C.1.8 Proof of Theorem 3.5.2

$$\begin{aligned} \hat{T_n^*} &= \frac{1}{\binom{n}{2}}\sum_{t=2}^{n}\sum_{s=1}^{t-1} w_t[e_t - (\hat{\beta}-\beta)^T x_t] Z_t^T Z_s [e_s - x_s^T(\hat{\beta}-\beta)] w_s \\ &= T_n^* - \frac{1}{\binom{n}{2}}\sum_{t=2}^{n}\sum_{s=1}^{t-1}(\hat{\beta}-\beta)^T w_t x_t Z_t^T Z_s e_s w_s \\ &\quad - \frac{1}{\binom{n}{2}}\sum_{t=2}^{n}\sum_{s=1}^{t-1} w_t e_t Z_t^T Z_s x_s^T w_s(\hat{\beta}-\beta) + \frac{1}{\binom{n}{2}}\sum_{t=2}^{n}\sum_{s=1}^{t-1}(\hat{\beta}-\beta)^T w_t x_t Z_t^T Z_s x_s^T w_s(\hat{\beta}-\beta) \end{aligned} \tag{C.1.12}$$

We assume $\mathcal{F}_n' = \sigma(Z_1e_1, ..., Z_ne_n) = \sigma(Z_1^*, ..., Z_n^*)$ and $\mathcal{F}_n'' = \sigma(Z_1x_1^T, ..., Z_nx_n^T) = \sigma(A_1, ..., A_n)$, and we need to show that

$$\frac{\hat{T_n^*}}{\sqrt{Var(T_n)}} = \frac{T_n^*}{\sqrt{Var(T_n)}} + o_p(1).$$

For the last term in Equation (C.1.12), we need to show

$$E^*\left[\frac{1}{\binom{n}{2}^{\frac{1}{2}}}\frac{1}{n||\Sigma^*||_F}||\sum_{t=2}^{n}\sum_{s=1}^{t-1} w_t A_t^T A_s w_s||_F\right]^2 \xrightarrow{p} 0.$$

Since the above term is non-negative, it suffices to show

$$E\left[E^*\left[\frac{1}{\binom{n}{2}^{\frac{1}{2}}}\frac{1}{n||\Sigma^*||_F}||\sum_{t=2}^{n}\sum_{s=1}^{t-1} w_t A_t^T A_s w_s||_F\right]^2\right] \to 0$$

$$
\begin{aligned}
&E^*\Big[\frac{1}{\binom{n}{2}^{\frac{1}{2}}}\frac{1}{n||\Sigma^*||_F}||\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TA_sw_s||_F\Big]^2\\
&=E\Big[\frac{1}{\binom{n}{2}^{\frac{1}{2}}}\frac{1}{n||\Sigma^*||_F}||\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TA_sw_s||_F|\mathcal{F}_n''\Big]^2\\
&=\frac{1}{\binom{n}{2}}\frac{1}{n^2||\Sigma^*||_F^2}E\Big[||\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TA_sw_s||_F^2|\mathcal{F}_n''\Big]\\
&=\frac{1}{\binom{n}{2}}\frac{1}{n^2||\Sigma^*||_F^2}\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1}tr(E(w_tA_t^TA_sw_sw_{s'}A_{s'}^TA_{t'}w_{t'}|\mathcal{F}_n''))\\
&=\frac{1}{\binom{n}{2}}\frac{1}{n^2||\Sigma^*||_F^2}\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1}tr(A_t^TA_sA_{s'}^TA_{t'}E(w_tw_sw_{s'}w_{t'}))\\
&=\frac{1}{\binom{n}{2}}\frac{1}{n^2||\Sigma^*||_F^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}tr(A_t^TA_sA_s^TA_t).
\end{aligned}
$$

Then by Equation (3.5.3) in Assumption 3.5.3, we have

$$
E\Big[E^*\Big[\frac{1}{\binom{n}{2}^{\frac{1}{2}}}\frac{1}{n||\Sigma^*||_F}\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TA_sw_s\Big]^2\Big]=O\Big(\frac{||\Sigma_a||_F^2}{n^2||\Sigma^*||_F^2}\Big)=o(1).
$$

For the second term in Equation (C.1.12), we need to show

$$
E^*\Big[\frac{1}{\binom{n}{2}}\frac{1}{\sqrt{nVar(T_n)}}||\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TZ_s^*w_s||_F\Big]^2\xrightarrow{p}0.
$$

Since the above term is non-negative, it suffices to show that

$$
E\Big[E^*\Big[\frac{1}{\binom{n}{2}}\frac{1}{\sqrt{nVar(T_n)}}||\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TZ_s^*w_s||_F]^2\Big]\to 0
$$

$$
\begin{aligned}
&E^*\Big[\frac{1}{\binom{n}{2}}\frac{1}{\sqrt{nVar(T_n)}}||\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TZ_s^*w_s||_F\Big]^2\\
&=\frac{1}{\binom{n}{2}}\frac{1}{n||\Sigma^*||_F^2}E\Big[||\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TZ_s^*w_s||_F|\mathcal{F}_n',\mathcal{F}_n''\Big]^2\\
&=\frac{1}{\binom{n}{2}}\frac{1}{n||\Sigma^*||_F^2}\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1}tr(E(w_tA_t^TZ_s^*w_sw_{s'}Z_{s'}^{*T}A_{t'}w_{t'}|\mathcal{F}_n',\mathcal{F}_n''))\\
&=\frac{1}{\binom{n}{2}}\frac{1}{n||\Sigma^*||_F^2}\sum_{t=2}^{n}\sum_{t'=2}^{n}\sum_{s=1}^{t-1}\sum_{s'=1}^{t'-1}tr(A_t^TZ_s^*Z_{s'}^{*T}A_{t'}E(w_tw_sw_{s'}w_{t'}))\\
&=\frac{1}{\binom{n}{2}}\frac{1}{n||\Sigma^*||_F^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}tr(A_t^TZ_s^*Z_s^{*T}A_t).
\end{aligned}
$$

Thus by Equation (3.5.5) in Assumption 3.5.3, we can get

$$E\left[E^*\left[\frac{1}{\binom{n}{2}}\frac{1}{\sqrt{nVar(T_n)}}\sum_{t=2}^{n}\sum_{s=1}^{t-1}w_tA_t^TZ_s^*w_s]^2\right] = O\left(\frac{tr(\Sigma^*\Sigma_a)}{n||\Sigma^*||_F^2}\right) = o(1).$$

C.1.9 Proof of Theorem 3.5.3

Let $D_2(\tilde{Z})^4 := E\frac{|\tilde{Z_1}^T\tilde{Z_1}|^4}{||B||_F^4}$, we just need to show that based on Assumption 3.2.2 and 3.5.5,

$$\frac{D_2(\tilde{Z})^4}{n} = o(1)$$

and

$$\frac{tr(B^4)}{||B||_F^4},$$

then we can use the results in Section 3.3.

For $\frac{D_2(\tilde{Z})^4}{n}$, since

$$\begin{aligned}||B||_F^4 &= [tr(B^2)]^2\\ &= \left[tr\left(E\left(\frac{Z_1Z_1^T}{||Z_1||^2}\right)E\left(\frac{Z_2Z_2^T}{||Z_2||^2}\right)\right)\right]^2\\ &= \left[E\left(tr\left(\frac{Z_1Z_1^TZ_2Z_2^T}{||Z_1||^2||Z_2||^2}\right)\right)\right]^2\\ &= \left[E\left(\frac{(Z_1^TZ_2)^2}{||Z_1||^2||Z_2||^2}\right)\right]^2\\ &= E^2[(\tilde{Z_1}^T\tilde{Z_2})^2],\end{aligned}$$

thus

$$\begin{aligned}\frac{D_2(\tilde{Z})^4}{n} &= \frac{E\left[\left(\frac{(Z_1^TZ_2)^2}{||Z_1||^2||Z_2||^2}\right)^2\right]}{n\left[E\left(\frac{(Z_1^TZ_2)^2}{||Z_1||^2||Z_2||^2}\right)\right]^2}\\ &= \frac{E[(\tilde{Z_1}^T\tilde{Z_2})^4]}{nE^2[(\tilde{Z_1}^T\tilde{Z_2})^2]}.\end{aligned}$$

For $\frac{tr(B^4)}{||B||_F^4}$,

$$\begin{aligned}\frac{tr(B^4)}{||B||_F^4} &= \frac{E(tr(B\tilde{Z_2}\tilde{Z_2}^TB\tilde{Z_1}\tilde{Z_1}^T))}{E(tr(B\tilde{Z_1}\tilde{Z_1}^T))E(tr(B\tilde{Z_1}\tilde{Z_1}^T))}\\ &= \frac{E[(\tilde{Z_1}^TB\tilde{Z_2})^2]}{E^2(\tilde{Z_1}^TB\tilde{Z_1})}.\end{aligned}$$

Thus according to Assumption 3.2.2 and 3.5.5, we can directly apply the results in Wang et al. [54]:

$$E[(\tilde{Z_1}^T\tilde{Z_2})^4] = O(1)E^2[(\tilde{Z_1}^T\tilde{Z_2})^2]$$

$$E[(\tilde{Z_1}^T B\tilde{Z}_2)^2] = o(1)(E^2(\tilde{Z_1}^T B\tilde{Z}_1)).$$

C.1.10 Proof of Proposition 3.5.1

For $\frac{D_2(W)^4}{n}$, since

$$W_t = \begin{bmatrix} Z_{t1}Z_{t2} \\ Z_{t1}Z_{t3} \\ ... \\ Z_{t(d-1)}Z_{td} \end{bmatrix}$$

and

$$W_tW_t^T = \begin{bmatrix} Z_{t1}^2Z_{t2}^2 & Z_{t1}^2Z_{t2}Z_{t3} & ... & Z_{t1}Z_{t2}Z_{t(d-1)}Z_{td} \\ Z_{t1}^2Z_{t2}Z_{t3} & ... & ... & ... \\ ... & & & \\ Z_{t1}Z_{t2}Z_{t(d-1)}Z_{td} & ... & ... & Z_{t(d-1)}^2Z_{td}^2 \end{bmatrix}$$

Thus if we assume that $Z_{ti} \perp Z_{tj}$ for $i \neq j$:

$$|W_t^T W_s|^4 = |\sum_{j=2}^{d}\sum_{i=1}^{j-1} Z_{ti}Z_{tj}Z_{si}Z_{sj}|^4$$

$$E[|W_t^T W_s|^4] = \sum_{j_1=2}^{d}\sum_{i_1=1}^{j_1-1}\sum_{j_2=2}^{d}\sum_{i_2=1}^{j_2-1}\sum_{j_3=2}^{d}\sum_{i_3=1}^{j_3-1}\sum_{j_4=2}^{d}\sum_{i_4=1}^{j_4-1} E^2(Z_{ti_1}Z_{ti_2}Z_{ti_3}Z_{ti_4}Z_{tj_1}Z_{tj_2}Z_{tj_3}Z_{tj_4})$$

$$\Sigma_w = E(W_tW_t^T) = \begin{bmatrix} E(Z_{t1}^2Z_{t2}^2) & 0 & ... & 0 \\ 0 & ... & ... & ... \\ ... & & & \\ 0 & ... & ... & E(Z_{t(d-1)}^2Z_{td}^2) \end{bmatrix}$$

$$||\Sigma_w||_F^4 = [\sum_{i=2}^{d}\sum_{j=1}^{i-1} E^2(Z_{ti}^2Z_{tj}^2)]^2$$

$$\frac{D_2(W)^4}{n} = \frac{E[(\sum_{j=2}^{d}\sum_{i=1}^{j-1} Z_{ti}Z_{tj}Z_{si}Z_{sj})^4]}{n[\sum_{i=2}^{d}\sum_{j=1}^{i-1} E^2(Z_{ti}^2Z_{tj}^2)]^2}.$$

Then for $E[|W_t^T W_s|^4]$, since $Z_{ti} \perp Z_{tj}$ for $i \neq j$, then we can reduce it to at most four summations and the summand is bounded by C_1^4, thus $E[|W_t^T W_s|^4] \leq Cd^4$. Similarly, $||\Sigma_w||_F^4 \asymp d^4$, then clearly $\frac{D_2(W)^4}{n} \to 0$.

For $\frac{tr(\Sigma_w^4)}{||\Sigma_w||_F^4}$, we have

$$\Sigma_w^4 = \begin{bmatrix} E^4(Z_{t1}^2 Z_{t2}^2) & 0 & ... & 0 \\ 0 & ... & ... & ... \\ ... & & & \\ 0 & ... & ... & E^4(Z_{t(d-1)}^2 Z_{td}^2) \end{bmatrix}$$

$$tr(\Sigma_w^4) = \sum_{i=2}^{d} \sum_{j=1}^{i-1} E^4(Z_{ti}^2 Z_{tj}^2)$$

$$\frac{tr(\Sigma_w^4)}{||\Sigma_w||_F^4} = \frac{\sum_{i=2}^{d} \sum_{j=1}^{i-1} E^4(Z_{ti}^2 Z_{tj}^2)}{[\sum_{i=2}^{d} \sum_{j=1}^{i-1} E^2(Z_{ti}^2 Z_{tj}^2)]^2} \to 0.$$

Thus we have $tr(\Sigma_w^4) \asymp d^2$, thus $\frac{tr(\Sigma_w^4)}{||\Sigma_w||_F^4} \asymp \frac{1}{d^2} \to 0$.

C.2 Simulation Results

Table C.1: Comparison of the Approximate Errors (%) under Independent $\chi^2 - 1$ Data

	$n = 50$				
	$d = 5$	$d = 20$	$d = 50$	$d = 100$	$d = 300$
$a = 0.90$					
$\varepsilon_{T_{us,n}}$	1.29	0.17	0.06	0.44	0.22
$\varepsilon_{T_{s,n}}$	0.05	0.89	0.80	0.21	0.45
$\varepsilon_{T_{us,m}}$	0.71	0.11	0.17	0.21	0.07
$\varepsilon_{T_{s,m}}$	0.38	0.52	0.51	0.06	0.16
ε_{V_p}	0.13	1.99	2.51	2.37	3.52
ε_{F_s}	6.35	9.42	9.91	10.00	10.00
ε_{Q_n}	1.34	4.76	7.02	8.44	9.94
ε_{Q_m}	0.29	2.99	5.00	6.29	9.43
$a = 0.95$					
$\varepsilon_{T_{us,n}}$	0.65	0.20	0.37	0.16	0.09
$\varepsilon_{T_{s,n}}$	0.10	0.58	0.19	0.12	0.04
$\varepsilon_{T_{us,m}}$	0.40	0.32	0.30	0.14	0.09
$\varepsilon_{T_{s,m}}$	0.38	0.43	0.29	0.31	0.04
ε_{V_p}	0.52	1.89	1.73	1.82	2.78
ε_{F_s}	3.97	4.92	5.00	5.00	5.00
ε_{Q_n}	1.38	3.40	4.30	4.69	5.00
ε_{Q_m}	0.64	2.35	3.26	3.94	4.93
	$n = 100$				
	$d = 5$	$d = 20$	$d = 50$	$d = 100$	$d = 300$
$a = 0.90$					
$\varepsilon_{T_{us,n}}$	0.58	0.32	0.62	0.61	0.52
$\varepsilon_{T_{s,n}}$	0.25	0.22	0.24	0.31	0.20
$\varepsilon_{T_{us,m}}$	0.36	0.17	0.39	0.46	0.39
$\varepsilon_{T_{s,m}}$	0.20	0.05	0.30	0.49	0.21
ε_{V_p}	0.18	0.86	0.96	0.92	1.23
ε_{F_s}	5.96	8.90	9.77	9.99	10.00
ε_{Q_n}	0.90	2.35	4.19	6.39	9.41
ε_{Q_m}	0.21	1.27	2.56	4.00	7.62
$a = 0.95$					
$\varepsilon_{T_{us,n}}$	0.51	0.02	0.00	0.04	0.10
$\varepsilon_{T_{s,n}}$	0.13	0.24	0.04	0.10	0.11
$\varepsilon_{T_{us,m}}$	0.44	0.14	0.05	0.03	0.07
$\varepsilon_{T_{s,m}}$	0.12	0.04	0.14	0.04	0.00
ε_{V_p}	0.29	1.24	1.32	1.37	1.58
ε_{F_s}	3.64	4.86	4.99	5.00	5.00
ε_{Q_n}	0.81	2.15	3.18	4.22	4.96
ε_{Q_m}	0.14	1.34	2.14	3.10	4.52

Table C.2: Comparison of the Approximate Errors (%) Under Linear Model

	$n = 50$							
	$\beta = 0.6$				$\beta = 2$			
	$d = 5$	$d = 20$	$d = 50$	$d = 100$	$d = 5$	$d = 20$	$d = 50$	$d = 100$
$a = 0.900$								
$\varepsilon_{T_{us,n}}$	0.26	0.44	0.34	0.80	1.58	0.88	0.88	0.66
$\varepsilon_{T_{s,n}}$	0.94	0.56	1.26	0.12	0.88	0.24	0.26	0.14
$\varepsilon_{T_{us,m}}$	0.06	0.16	0.84	0.56	1.24	0.18	0.68	0.56
$\varepsilon_{T_{s,m}}$	1.00	0.46	1.28	0.04	0.82	0.02	0.30	0.32
ε_{V_p}	0.48	0.64	0.46	1.20	1.98	0.74	0.36	0.24
ε_{F_s}	4.20	4.16	5.04	4.66	5.66	9.14	9.92	10.00
ε_{Q_n}	0.40	0.04	0.94	0.02	0.32	2.96	6.16	8.14
ε_{Q_m}	0.22	0.08	1.04	0.02	0.56	1.50	3.74	6.04
$a = 0.950$								
$\varepsilon_{T_{us,n}}$	0.20	0.16	0.72	0.16	0.76	0.34	0.20	0.78
$\varepsilon_{T_{s,n}}$	0.08	0.08	0.58	0.24	0.48	0.24	0.08	0.48
$\varepsilon_{T_{us,m}}$	0.04	0.02	0.66	0.22	0.74	0.40	0.02	0.54
$\varepsilon_{T_{s,m}}$	0.16	0.16	0.60	0.40	0.44	0.20	0.02	0.46
ε_{V_p}	0.34	0.14	0.02	0.84	1.10	0.22	0.60	0.20
ε_{F_s}	3.48	4.78	5.00	5.00	3.28	4.76	4.98	5.00
ε_{Q_n}	0.62	2.26	3.76	4.70	0.10	2.12	4.06	4.78
ε_{Q_m}	0.36	1.46	2.54	3.88	0.16	1.12	3.02	4.02
	$n = 100$							
	$\beta = 0.6$				$\beta = 2$			
	$d = 5$	$d = 20$	$d = 50$	$d = 100$	$d = 5$	$d = 20$	$d = 50$	$d = 100$
$a = 0.900$								
$\varepsilon_{T_{us,n}}$	0.48	0.60	0.58	0.38	0.26	0.70	0.30	1.24
$\varepsilon_{T_{s,n}}$	0.42	0.42	0.34	0.14	0.42	0.32	0.42	1.02
$\varepsilon_{T_{us,m}}$	0.34	0.68	0.50	0.24	0.58	0.62	0.28	0.98
$\varepsilon_{T_{s,m}}$	0.14	0.56	0.34	0.18	0.58	0.46	0.36	0.86
ε_{V_p}	0.66	0.90	0.56	0.04	0.22	0.54	0.42	0.62
ε_{F_s}	5.94	8.44	9.78	9.98	6.00	8.28	9.58	10.00
ε_{Q_n}	0.06	1.30	3.54	6.10	0.70	1.24	3.88	5.92
ε_{Q_m}	0.20	0.32	1.94	3.54	0.72	0.30	2.82	3.74
$a = 0.950$								
$\varepsilon_{T_{us,n}}$	0.86	0.54	0.36	0.34	0.40	0.72	0.12	0.28
$\varepsilon_{T_{s,n}}$	0.70	0.30	0.28	0.46	0.50	0.60	0.24	0.26
$\varepsilon_{T_{us,m}}$	0.84	0.26	0.22	0.58	0.48	0.74	0.12	0.14
$\varepsilon_{T_{s,m}}$	0.76	0.26	0.12	0.78	0.52	0.72	0.18	0.14
ε_{V_p}	0.94	0.22	0.12	0.78	0.52	0.62	0.44	0.28
ε_{F_s}	3.02	4.50	4.96	5.00	3.64	4.54	4.94	5.00
ε_{Q_n}	0.42	1.14	2.96	4.02	0.88	0.80	3.14	3.96
ε_{Q_m}	0.40	0.54	1.82	2.90	0.70	0.18	1.90	2.66

Table C.3: Comparison of the Approximate Errors (%) Under Factor Model

	$n = 50$							
	$b = 0.05$				$b = 0.5$			
	$d = 5$	$d = 20$	$d = 50$	$d = 100$	$d = 5$	$d = 20$	$d = 50$	$d = 100$
$a = 0.900$								
$\varepsilon_{T_{us,n}}$	1.38	0.52	0.66	0.02	1.12	0.44	2.42	1.48
$\varepsilon_{T_{s,n}}$	0.62	0.10	0.10	0.48	0.08	0.60	1.38	0.80
$\varepsilon_{T_{us,m}}$	1.18	0.26	0.66	0.04	0.84	0.04	2.24	1.42
$\varepsilon_{T_{s,m}}$	0.82	0.10	0.32	0.14	0.08	0.92	1.06	0.48
ε_{V_p}	1.50	0.28	0.26	0.74	0.98	0.02	1.90	1.44
ε_{F_s}	6.06	9.26	9.86	9.98	5.66	6.52	5.82	6.00
ε_{Q_n}	0.26	3.12	5.62	8.46	0.24	1.36	0.24	0.56
ε_{Q_m}	0.60	1.70	3.72	6.26	0.24	0.60	0.62	0.02
$a = 0.950$								
$\varepsilon_{T_{us,n}}$	0.56	0.42	0.66	0.16	0.54	1.10	2.08	2.10
$\varepsilon_{T_{s,n}}$	0.20	0.26	0.52	0.00	0.06	0.66	1.70	1.76
$\varepsilon_{T_{us,m}}$	0.70	0.22	0.34	0.00	0.34	0.78	1.76	2.10
$\varepsilon_{T_{s,m}}$	0.38	0.06	0.28	0.02	0.02	0.28	1.18	1.66
ε_{V_p}	0.90	0.30	4.40	0.72	0.34	0.82	2.08	2.48
ε_{F_s}	3.60	4.84	5.00	5.00	2.94	3.58	3.08	3.16
ε_{Q_n}	0.58	2.56	0.94	4.74	0.40	0.62	0.10	0.32
ε_{Q_m}	0.04	1.50	1.04	4.14	0.10	0.14	0.70	1.06
	$n = 100$							
	$b = 0.05$				$b = 0.5$			
	$d = 5$	$d = 20$	$d = 50$	$d = 100$	$d = 5$	$d = 20$	$d = 50$	$d = 100$
$a = 0.900$								
$\varepsilon_{T_{us,n}}$	0.58	0.26	0.02	0.56	0.12	0.74	0.48	0.64
$\varepsilon_{T_{s,n}}$	0.24	0.66	0.14	0.22	0.54	0.06	0.14	0.02
$\varepsilon_{T_{us,m}}$	0.28	0.40	0.10	0.52	0.00	0.40	0.06	0.44
$\varepsilon_{T_{s,m}}$	0.26	0.50	0.20	0.30	0.32	0.20	0.66	0.24
ε_{V_p}	0.54	0.44	0.36	0.06	0.04	0.52	0.22	0.50
ε_{F_s}	5.98	8.80	9.80	9.98	5.04	5.40	5.76	5.60
ε_{Q_n}	0.22	2.50	3.94	6.14	0.42	0.40	0.68	0.58
ε_{Q_m}	0.08	1.40	2.22	3.64	0.36	0.14	0.52	0.08
$a = 0.950$								
$\varepsilon_{T_{us,n}}$	0.22	0.16	0.28	0.00	0.46	0.76	1.42	1.16
$\varepsilon_{T_{s,n}}$	0.28	0.18	0.30	0.02	0.10	0.44	1.12	0.80
$\varepsilon_{T_{us,m}}$	0.20	0.10	0.34	0.08	0.28	0.68	1.36	1.06
$\varepsilon_{T_{s,m}}$	0.30	0.06	0.34	0.00	0.02	0.28	0.92	0.32
ε_{V_p}	0.34	0.06	0.02	0.54	0.38	0.58	1.56	1.24
ε_{F_s}	3.56	4.60	4.98	5.00	2.94	2.68	2.60	2.60
ε_{Q_n}	0.70	1.22	2.20	3.92	0.14	0.08	0.60	0.20
ε_{Q_m}	0.48	0.64	1.34	2.80	0.08	0.26	0.86	0.40

Table C.4: Comparison of the Approximate Errors (%) under DGP1

	$n = 30$				
	$d = 20$	$d = 50$	$d = 100$	$d = 300$	$d = 1000$
$a = 0.90$					
$\varepsilon_{T_{us,n}}$	0.06	0.02	1.48	0.08	0.36
$\varepsilon_{T_{s,n}}$	0.54	0.86	0.80	0.78	0.18
ε_{W_n}	0.26	0.12	1.86	0.28	0.26
$a = 0.95$					
$\varepsilon_{T_{us,n}}$	0.36	0.16	0.24	0.64	0.04
$\varepsilon_{T_{s,n}}$	0.26	0.54	0.40	0.40	0.12
ε_{W_n}	2.52	1.78	1.24	1.56	0.64
$a = 0.99$					
$\varepsilon_{T_{us,n}}$	0.06	0.01	0.06	0.31	0.22
$\varepsilon_{T_{s,n}}$	0.28	0.29	0.36	0.08	0.28
ε_{W_n}	1.46	1.17	0.95	0.47	0.43

Table C.5: Comparison of the Approximate Errors (%) under DGP2

	$n = 30$				
	$d = 20$	$d = 50$	$d = 100$	$d = 300$	$d = 1000$
$a = 0.90$					
$\varepsilon_{T_{us,n}}$	0.48	0.86	0.90	0.49	1.09
$\varepsilon_{T_{s,n}}$	0.39	0.04	0.13	0.56	0.30
ε_{W_n}	0.53	0.54	0.38	0.05	0.45
$a = 0.95$					
$\varepsilon_{T_{us,n}}$	0.02	0.32	0.17	0.62	0.11
$\varepsilon_{T_{s,n}}$	0.24	0.13	0.17	0.21	0.29
ε_{W_n}	1.39	1.65	1.63	2.06	1.78
$a = 0.99$					
$\varepsilon_{T_{us,n}}$	0.02	0.32	0.17	0.62	0.11
$\varepsilon_{T_{s,n}}$	0.24	0.13	0.17	0.21	0.29
ε_{W_n}	1.39	1.65	1.63	2.06	1.78

Table C.6: Comparison of the Approximate Errors (%) under DGP3

	$n = 30$				
	$d = 20$	$d = 50$	$d = 100$	$d = 300$	$d = 1000$
$a = 0.90$					
$\varepsilon_{T_{us,n}}$	1.19	0.69	0.74	0.27	0.89
$\varepsilon_{T_{s,n}}$	0.08	0.29	0.40	0.71	0.42
ε_{W_n}	0.82	0.30	0.13	0.05	0.64
$a = 0.95$					
$\varepsilon_{T_{us,n}}$	0.35	0.73	0.40	0.27	0.32
$\varepsilon_{T_{s,n}}$	0.10	0.22	0.08	0.21	0.11
ε_{W_n}	1.79	2.28	2.16	2.60	2.65
$a = 0.99$					
$\varepsilon_{T_{us,n}}$	0.16	0.09	0.19	0.14	0.27
$\varepsilon_{T_{s,n}}$	0.25	0.20	0.43	0.25	0.38
ε_{W_n}	2.41	2.40	3.15	3.63	3.85

Table C.7: Comparison of the Approximate Errors (%) under DGP4

	$n = 30$				
	$d = 20$	$d = 50$	$d = 100$	$d = 300$	$d = 1000$
$a = 0.90$					
$\varepsilon_{T_{us,n}}$	0.68	0.66	0.08	0.80	0.15
$\varepsilon_{T_{s,n}}$	0.44	0.42	0.97	0.42	0.96
ε_{W_n}	0.64	0.29	0.33	0.46	0.06
$a = 0.95$					
$\varepsilon_{T_{us,n}}$	0.27	0.57	0.13	0.45	0.66
$\varepsilon_{T_{s,n}}$	0.10	0.25	0.23	0.15	0.10
ε_{W_n}	1.92	2.27	2.11	2.70	2.87
$a = 0.99$					
$\varepsilon_{T_{us,n}}$	0.17	0.14	0.06	0.02	0.24
$\varepsilon_{T_{s,n}}$	0.40	0.25	0.15	0.06	0.39
ε_{W_n}	2.25	2.74	3.25	3.29	3.60

Table C.8: Comparison of the Approximate Errors (%) under DGP5

	$n = 30$				
	$d = 20$	$d = 50$	$d = 100$	$d = 300$	$d = 1000$
$a = 0.90$					
$\varepsilon_{T_{us,n}}$	0.47	0.67	0.92	0.81	1.19
$\varepsilon_{T_{s,n}}$	0.24	0.33	0.23	0.29	0.23
ε_{W_n}	0.82	1.04	1.15	0.63	0.99
$a = 0.95$					
$\varepsilon_{T_{us,n}}$	0.33	0.67	0.56	0.30	0.07
$\varepsilon_{T_{s,n}}$	0.02	0.19	0.16	0.06	0.28
ε_{W_n}	2.43	2.62	2.04	1.42	0.49
$a = 0.99$					
$\varepsilon_{T_{us,n}}$	0.17	0.14	0.06	0.02	0.24
$\varepsilon_{T_{s,n}}$	0.40	0.25	0.15	0.06	0.39
ε_{W_n}	2.25	2.74	3.25	3.29	3.60

Table C.9: Comparison of the Approximate Errors (%) for Tests of Covariance under DGP1

	$n = 30$				
	$d = 20$	$d = 50$	$d = 100$	$d = 300$	$d = 1000$
$a = 0.90$					
$\varepsilon_{E_{us,n}}$	0.23	0.03	0.17	0.42	0.80
$\varepsilon_{E_{s,n}}$	1.51	1.47	1.60	1.64	2.23
ε_{U_n}	1.04	1.36	1.31	1.37	1.25
ε_{V_n}	1.72	1.62	1.51	1.46	1.23
$a = 0.95$					
$\varepsilon_{E_{us,n}}$	1.18	0.82	0.88	1.06	1.02
$\varepsilon_{E_{s,n}}$	1.38	1.02	1.09	1.09	1.20
ε_{U_n}	1.37	1.26	1.25	1.54	1.50
ε_{V_n}	1.98	1.61	1.36	1.56	1.57
$a = 0.99$					
$\varepsilon_{E_{us,n}}$	0.77	0.81	0.66	0.72	0.65
$\varepsilon_{E_{s,n}}$	0.37	0.55	0.37	0.30	0.35
ε_{U_n}	1.00	0.72	0.75	0.77	0.77
ε_{V_n}	1.55	0.96	0.73	0.81	0.86

Table C.10: Comparison of the Approximate Errors (%) for Tests of Covariance under DGP2

	$n = 30$				
	$d = 20$	$d = 50$	$d = 100$	$d = 300$	$d = 1000$
$a = 0.90$					
$\varepsilon_{E_{us,n}}$	0.40	0.40	0.12	0.39	0.36
$\varepsilon_{E_{s,n}}$	1.34	1.87	1.25	1.48	1.58
ε_{U_n}	2.11	1.62	1.81	1.45	1.81
ε_{V_n}	3.35	2.10	2.17	1.56	1.83
$a = 0.95$					
$\varepsilon_{E_{us,n}}$	0.70	1.26	0.58	1.06	0.99
$\varepsilon_{E_{s,n}}$	0.98	1.29	0.74	1.13	1.03
ε_{U_n}	2.66	1.82	2.04	1.64	1.48
ε_{V_n}	3.79	2.18	2.24	1.62	1.50
$a = 0.99$					
$\varepsilon_{E_{us,n}}$	0.64	0.70	0.77	0.68	0.67
$\varepsilon_{E_{s,n}}$	0.23	0.32	0.39	0.42	0.30
ε_{U_n}	1.67	1.33	0.84	0.72	0.85
ε_{V_n}	2.19	1.43	0.93	0.82	0.88