Essays in Econometrics

Min Lin

Doctor of Philosophy

University of York

Economics

July 2022

Abstract

This thesis consists of four independent chapters that focus on two increasingly popular models in the field of econometrics, namely Mixed Data Sampling (MIDAS) models and spatial econometric models.

Chapter 1 introduces the MIDAS approach to a panel data model. It presents the asymptotic distributions of the nonlinear least squares (NLS) estimator for the panel-MIDAS model and the least squares (LS) estimator for the standard LS model based on predetermined weights. The utility of the panel-MIDAS model is demonstrated by a Monte Carlo simulation and an empirical application.

Chapter 2 presents an application of the MIDAS model to study the economic policy uncertainty (EPU) effect on mortality in the US. The results indicate a significant negative correlation between EPU and mortality. The accompanying analysis reveals that changes in risky health behaviors are consistent with fluctuations in mortality.

Chapter 3 proposes a spatial quasi-limited information maximum likelihood (SQLIML) estimation for a cross-sectional spatial lag model (SLM) with additional endogenous variables X. It derives the asymptotic properties of the SQLIML estimator and compares them with the spatial two-stage least squares (S2SLS) estimator. It also introduces a test of the null that X is exogenous in the SLM model. The above theoretical aspects are explored via a Monte Carlo study and an empirical application.

Chapter 4 presents an application of the spatial econometric model to study the minimum wage effect on the gender wage gap across income levels in the US. The results reveal that minimum wage increases in a given state narrow the 10th-60th percentile gender wage gap in its own state and its neighboring states. However, there is no evidence that the gender wage gap in the upper tail of the wage distribution is associated with minimum wages.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my sincere gratitude to those who have continuously contributed their support to me during my Ph.D. study. First of all, I am extremely grateful to my supervisor Dr. Michael Thornton for his patience, motivation, and immense knowledge. I have enjoyed and benefited as your student.

Moreover, I am hugely indebted to the members of my advisory panel, Prof. Laura Coroneo and Prof. Nigel Rice, for sharing their knowledge and providing comments on my current thesis. My thanks also go to Prof. Yongcheol Shin and Prof. Takashi Yamagata for their helpful comments on Chapter 3. Their helpful comments and suggestions steered me in the right direction and helped me successfully complete this thesis.

I am also thankful to my friends and fellow Ph.D. students for their friendship, Shuyi Lou, Jiaxin Peng, Yanting Chen, Songshan Li, Rui Lin, Chaowen Zheng, and Hussain Akseer. I have learned from you every day. The funding from the Department through a three-year studentship is also much acknowledged.

Last but not least, my thanks must go to my mum, my dad, and my brother Han for their incredible support. This thesis would not have been possible without your great encouragement. I dedicate this thesis to you. Thank you.

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

An earlier version of Chapter 3 was presented at the Thursday Workshop at the University of York in December 2021 and at the Programme of the York-UM Economics Workshop in September 2022.

An earlier version of Chapter 4 was presented at the 8th SidE Workshop for Ph.D. Students in Econometrics and Empirical Economics in September 2020.

Min Lin

York, July 2022

Introduction

This thesis comprises four chapters within the field of econometrics. It mainly focuses on two econometric models, the Mixed Data Sampling (MIDAS) model and the spatial econometric model. The first two chapters focus on the MIDAS model, and the last two chapters focus on the spatial econometric model.

One dilemma faced by researchers is how to deal with mixed-frequency observations in a regression model. This is often the case for macroeconomic variables. As an example, Gross Domestic Product (GDP), one of the key indicators of a nation's economic status, is measured quarterly in the US and it is released with a significant delay, while many other variables are available more timely, such as monthly unemployment rate, daily stock returns. This frequency imbalance presents a significant challenge when estimating low-frequency variables with high-frequency variables. Several methods have been proposed to tackle this issue. Among those, the temporally aggregating approach is the most commonly used, where high-frequency variables are temporally aggregated to low-frequency variables using a predetermined weighting scheme. However, one can expect that the temporally aggregating approach will lose high-frequency information in the high-frequency data if predetermined weights are not properly chosen. Instead, Ghysels, Santa-Clara, and Valkanov (2004) propose the MIDAS approach, where the aggregating weights are determined by data, allowing to retain most of the information in high-frequency variables.

One attractive feature of the MIDAS model is that it can deal with mixed-frequency data without a substantial loss of degrees of freedom. It can explain the low-frequency variable based on the high-frequency variable and its lags. One can expect that there would be a large number of parameters to be estimated if the number of high-frequency lags is large. To address this parameter proliferation, in the MIDAS model, a weighting function that depends on a few parameters is used to capture the weights of lagged high-frequency observations. Since their introduction by Ghysels, Santa-Clara, and Valkanov (2004, 2005), the MIDAS approach has been developed theoretically and empirically. The development in application and extension of the MIDAS model can be found in Ghysels, Santa-Clara, and Valkanov (2006), Andreou, Ghysels, and Kourtellos (2010, 2011), Guérin and Marcellino (2013), Bai, Ghysels, and Wright (2013), and Miller (2014), among others. The rule of thumb that all data are sampled at the same frequency is even more restrictive when the panel data is used. While the literature on the MIDAS approach is expansive, relatively little work has been done on the extension of the MIDAS approach to panels.

Chapter 1 contributes to the literature by introducing the MIDAS framework into the fixed-effects models and random-effects models and making the first attempt to establish the asymptotic distributions of their nonlinear least squares (NLS) estimators. In particular, it derives the conditions for the consistency and the asymptotic normality of the MIDAS-NLS estimator and shows that the least squares (LS) estimator of the standard LS model based on predetermined weights is inconsistent. In addition, it proposes a Wald test for the null of flat weights in the panel-MIDAS model. The finite sample properties of estimators are studied via a Monte Carlo simulation. In comparing the size of the bias and the efficiency gains, the results show that the MIDAS-NLS estimator outperforms the LS estimator in the presence of mixed-frequency data. An empirical application to a model of Okun's law illustrates the usefulness of the proposed model. The empirical findings indicate that the estimation accuracy is significantly improved by the use of the panel-MIDAS model.

Chapter 2 presents an application of the MIDAS model to the following questions: (i) How does economic policy uncertainty (EPU) affect mortality? (ii) Does this effect differ by gender and age? (iii) Could lifestyle changes be the mechanism through which EPU affects mortality? Using time-series data from 1960-2013 in the US, the analysis suggests a significant negative correlation between EPU and mortality of all sex and age groups, with particularly strong pronounced for the old (65-84-year-olds). A 1% point increase in EPU is significantly associated with a 0.4410% decline in mortality rates of 65-84-year-olds. It is noteworthy that women are more sensitive to changes in EPU than men, with a greater elasticity of -0.2821 and a longer-lasting effect of 57 months. To further explore the mechanisms behind this reduction in mortality when uncertainty increases, microdata from the Behavioral Risk Factor Surveillance System (BRFSS) are used to examine how risky health behaviors respond to changes in EPU. Using BRFSS data, we construct variables: the prevalence rate of current drinkers, the prevalence rate of current smokers, the prevalence rate of insufficient exercise, and the prevalence rate of overweight and obesity as indicators of risky health behaviors. The accompanying analysis reveals that increased EPU is associated with decreased prevalence rates of risky health behaviors, partly explaining the health improvement during high uncertainty periods.

Another dilemma faced by researchers is how to capture the potential relationships and interactions between spatial units. It is not possible for traditional econometric models to deal with these features, as the number of parameters would be greater than the number of observations. This has motivated the introduction of spatial econometric models that accommodate co-dependencies across units. In the spatial econometric model, the information about the dependence between spatial units is incorporated into a spatial weight matrix. It captures the spatial structure of the data and usually needs to be specified in advance. In particular, each element of this matrix represents the strength of the spatial proximity between units. Therefore, due to its ability to model spatial dependencies at a low cost in terms of degrees of freedom, the spatial econometric model is highly attractive to researchers. Early development in the estimation and application of the spatial econometric model and its further research directions can be found in Anselin (1988, 2010), Anselin and Florax (1995), Anselin and Rey (1997), Pinkse and Slade (2010), among others.

A simple version of the spatial econometric model is known as the spatial lag model (SLM) introduced by Cliff and Ord (1973, 1981), which augments the linear model by adding a spatial lag term. Theoretically, the spatial lag term is endogenous, and several estimation methods have been proposed to address its endogeneity. Among these estimation methods, the maximum likelihood (ML) (Ord, 1975) is the most widely used. Notice that all of these estimation methods considered above for the SLM model are only applicable when the spatial lag variable is the only endogenous variable. In other words, they fail to address the potential endogeneity of the explanatory variables X. However, in practice, one may expect that some of the other explanatory variables are endogenous as well. Therefore, it is necessary to consider a SLM model that allows for additional endogenous RHS variables X. The main contribution of Chapter 3 is to propose a new approach to estimating this model.

The model of interest in Chapter 3 is a cross-sectional SLM model with additional endogenous variables. It presents a spatial quasi-limited information maximum likelihood (SQLIML) estimation for this model and shows that the SQLIML estimator is consistent and asymptotically normal under certain conditions. This estimator extends Ord's (1975) ML estimator for SLM models and Wooldridge's (2014) control function estimator for non-spatial models. In particular, as with the ML estimator, it controls for the endogeneity of the spatial lag term by using a determinant of a matrix that depends on the spatial lag parameter and the sample size, and as with the control function estimator, it controls for the endogeneity of X by adding the reduced form error of X to the SLM model. In addition, this chapter introduces a regression-based test of the null that X is exogenous and provides a Monte Carlo study to compare the performance of the spatial two-stage least squares (S2SLS) estimator and the SQLIML estimator. The Monte Carlo simulation results show that the proposed SQLIML estimator outperforms the S2SLS estimator, especially for models with strong endogeneity and weak instruments, and indicate the correct size and good power of the proposed exogeneity test. The usefulness of the SQLIML estimation is demonstrated by an empirical application to revisit the driving under the influence (DUI) arrest rate model in US counties by Drukker, Prucha, and Raciborski (2013). In this literature, a key independent variable is the number of sworn officers, which is argued to be correlated with the alcohol-related arrest rate, thereby leading to an additional endogeneity problem. For purposes of comparison, the DUI arrest rate model is estimated by S2SLS and SQLIML methods. The results suggest that the sign and significance of the SQLIML estimators are the same as those of the S2SLS estimators. However, in comparing the variance of estimators and the mean squared error of models, the SQLIML estimation method provides better estimates than the S2SLS estimation method.

A major focus of the spatial econometric model is spatial spillover. In the spatial context, spatial spillover defines that changes in one unit exert impacts on other units. In contrast to the traditional econometric model that limits spillover to zero, an attractive feature of the spatial econometric model is that spatial spillover can be accounted for. This has led to the widespread application of the spatial econometric model in regional science and urban economics.

Chapter 4 presents an application of the spatial econometric model to examine the effect of the minimum wage on the gender wage gap with two specific objectives in mind. The first is to quantify the extent to which minimum wage hikes contribute to changes in the gender wage gap at different income levels. The second is to assess the spillover effects of the minimum wage on the gender wage gap of neighboring regions. Measuring the minimum wage spillover effect is important because it enables an accurate estimate of the minimum wage effect. If the minimum wage spillover effect does exist, then an important implication for policy-makers is that the collaboration between adjacent regions contributes to developing effective minimum wage policies to address wage inequality. Using US state-level data during 1979-2018, the analysis provides evidence of a spillover effect of the minimum wage on the gender wage gap. Specifically, own-state minimum wage increases are expected to narrow the 10th-60th percentile gender wage gap in its own state and its neighboring states, with particularly strong pronounced for those in the lower tail of the wage distribution. Consistent with previous studies, this analysis indicates that the minimum wage is insignificantly associated with the gender wage gap in the upper tail of the wage distribution.

This thesis is organized as follows. Chapter 1 introduces the panel-MIDAS model and derives the asymptotic properties and the finite sample properties of its NLS estimator. Chapter 2 applies the MIDAS model to explore the EPU effects on health status. Chapter 3 proposes the SQLIML estimation for the SLM model with additional endogenous variables and establishes the consistency and the asymptotic normality of the SQLIML estimator. Chapter 4 applies the spatial econometric model to measure the size of the minimum wage effect on gender wage inequality. The last chapter concludes the thesis. Detailed proofs of Chapters 1 and 3 are presented in Appendices A and B, respectively.

Chapter 1

Asymptotic Distributions of Nonlinear Least Squares Estimators for Panel-MIDAS Models

1.1 Introduction

A time-series regression model generally requires that all data are sampled at the same frequency. When a time series regression involves mixed-frequency data, the common practice is to achieve the same frequency data by temporally aggregating high frequency to low frequency using a predetermined weighting scheme. However, one may expect that the inappropriate user-chosen components may lead to incorrect inferences. Marcellino (1999) points out that empirical properties, such as exogeneity, Granger causality, and cointegration, depend on temporal aggregation. Other important contributions to the temporal aggregation effects on estimating and testing include Engle (1969), Tiao (1972), Wei (1978), Granger (1987), Breitung and Swanson (2002), and Ghysels and Miller (2015). The limitation of the temporal aggregation approach has motivated the introduction of the Mixed Data Sampling (MIDAS) approach (Ghysels, Santa-Clara, and Valkanov, 2005, 2006).

The MIDAS approach proposes a data-driven method to aggregate high-frequency data into low-frequency data. It can explain the low-frequency variable by the high-frequency variable and its lags. The structure of the weighting scheme is data-driven. To reduce the number of parameters in the MIDAS model, a weighting function of a low-dimensional vector θ is used to capture the weights of lagged high-frequency variables. Due to its ability to deal with mixed-frequency data at a relatively low parametric cost, the MIDAS approach has attracted considerable attention recently. It has been used in a wide range of applications, particularly in macroeconomics and financial economics (see, Clements and Galvão, 2008; Ghysels and Wright, 2009; Alessi et al., 2014; Ghysels and Marcellino, 2018; Babii, Ghysels, and Striaukas, 2021). It also has been further extended into various time series models including the smooth transition MIDAS model (Galvão, 2013), the Markov-switching MIDAS model (Guérin and Marcellino, 2013), the GARCH-MIDAS model (Engle, Ghysels, and Sohn, 2013), the ECM-MIDAS model (Götz, Hecq, and Urbain, 2014), the cointegrated MIDAS model (Miller, 2014), the unrestricted MIDAS model (Foroni, Marcellino, and Schumacher, 2015), and the mixed-frequency Markov-switching VAR model (Foroni, Guérin, and Marcellino, 2015).

Nevertheless, studies on the extension of the MIDAS model to panels are rare. Khalaf et al. (2021) make the first thorough attempt to introduce the MIDAS model into the dynamic panel framework. In particular, they extend MIDAS to the Arellano-Bond model and estimate this model using the GMM method. Further, based on the dynamic panel MIDAS model of Khalaf et al. (2021) and the Markov-switching MIDAS model of Guérin and Marcellino (2013), Casarin et al. (2018) develop a panel Markov-Switching unrestricted MIDAS regression that is suitable for analysis with Bayesian methods.

Encouraged by Khalaf et al. (2021), we extend the MIDAS model into the panel data context with fixed effects and random effects. The parameters of the time series MIDAS model can be estimated by the nonlinear least squares (NLS) estimation method, and the asymptotic and finite sample properties of its NLS estimator are established in Andreou, Ghysels, and Kourtellos (2010). Based on the success of Andreou, Ghysels, and Kourtellos (2010), we decided to take a step forward and explore the performance of the NLS estimator in the panel-MIDAS model. To the best of my knowledge, this paper is the first to derive the theoretical properties of the MIDAS-NLS estimator in panel models. If the MIDAS NLS estimator outperforms the least squares (LS) estimator of the linear model with a predetermined weighting scheme, its improved power can be applied to better explain economic activities.

The growing interest in the extension of the MIDAS approach to panels is partly due to the increased availability of datasets containing observations across a collection of individuals. There are several reasons why the extension of the MIDAS approach from time series to panels is interesting. First, the panel data contains more information and more variation among variables. With more degrees of freedom and more sample variability, the panel-MIDAS model is expected to have more accurate inferences of model parameters than the time-series MIDAS model. Second, the asymptotic normality of the MIDAS-NLS estimator when the lag order of the high-frequency variable is larger than its aggregation horizon does not hold in the time-series model, but it holds in the panel model. Therefore, compared to the time series MIDAS model, the panel-MIDAS model allows us to study the dynamic behavior of the economy with consistent parameter estimation.

The aim of this paper is threefold. First, this paper explores the NLS estimator of the panel-MIDAS model. It assumes that the Data Generating Process (DGP) is the mixed data sampling process where the dependent variable is sampled at a low frequency, while the independent variable is sampled at a high frequency. We shall consider both the fixed-effects and the random-effects specification of the individual effects and highlight their differences for estimation and inference. We establish the conditions for the consistency and the asymptotic normality of the MIDAS-NLS estimator of the fixed-effects and the random-effects model, where the time dimension, T, horizon aggregation, m, and high-frequency lag order, K, are fixed, and the cross-sectional dimension, N, grows without bound. Indeed, the MIDAS-NLS estimators are $\sqrt{N}$-consistent.

Second, we compare the performance of the estimators of the panel-MIDAS model and the standard LS model based on predetermined weights. We borrow the decomposition for the time series MIDAS regression from Andreou, Ghysels, and Kourtellos (2010) to decompose the conditional mean into an aggregated term in the LS model and a nonlinear term, which is the difference between the aggregated term in the MIDAS model and the aggregated term in the LS model. Using this decomposition, we evaluate the asymptotic bias of the LS estimator and compare the asymptotic variance of the LS estimator with the MIDAS-NLS estimator. Besides, we provide several examples of models and conduct the Monte Carlo simulation with high-frequency data being a MA(1) or an AR(1) process to assess the asymptotic and finite sample properties of the estimators. Analytical and numerical results for the bias and the relative efficiency of these two estimators are derived. Our results show that in the presence of mixed-frequency data, the standard approach of aggregating equally the high-frequency variable can yield a biased LS estimator if the high-frequency process is serially correlated. This bias depends on the level of aggregation horizon, the pattern of true weighting schemes, and the number of high-frequency lags. In comparing the root mean squared error (RMSE) of MIDAS-NLS and LS estimators, we uncover that the MIDAS-NLS estimator is more efficient relative to the LS estimator, particularly for models with an AR(1) process in combination with a high level of aggregation horizon, a large number of high-frequency lags, and rapid decaying weights. We illustrate the power of the Wald test for the null of the flat weighting scheme via the Monte Carlo simulation. The simulation results indicate the good power of this test.

Third, in an empirical application, we use the panel-MIDAS model to relate annual GDP growth rates to monthly unemployment rate changes in the Metropolitan Statistical Areas (MSA) of the US from 2004 to 2020. In comparing the R^2 values of the MIDAS model and the LS model based on equal weights, we find that the use of the panel-MIDAS model significantly improves estimation accuracy. Our Wald test results reject the null hypothesis of a flat-weighting scheme for unemployment growth in the MIDAS model.

The rest of this paper is organized as follows. In Section 1.2, we introduce the panel-MIDAS model. In Section 1.3, we derive the asymptotic properties of the MIDAS-NLS estimator and propose the Wald test for the flat weighting scheme in both the fixed-effects model and the random-effects model. In Sections 1.4 and 1.5, we compare the performance of the MIDAS-NLS estimator and the LS estimator when the high-frequency regressor is a MA(1) or an AR(1) process. In Section 1.6, we provide an empirical application, and the last section concludes. The proofs are collected in Appendix A.

1.2 Panel-MIDAS Model

Consider two processes $\left\{Y_{it}, \boldsymbol{X}^{(m)}_{i,t/m}\right\}$ for individuals $i = 1, 2, \ldots, N$, where Y_{it} is a low-frequency process that is observed at $t = 1, 2, \ldots, T$, $\boldsymbol{X}^{(m)}_{i,t/m} = (X^{1(m)}_{i,t/m}, X^{2(m)}_{i,t/m}, \ldots, X^{p(m)}_{i,t/m})$ is a $1 \times p$ vector of high-frequency process that is observed m times between $t-1$ and t. $\left\{\boldsymbol{Y}_i, \boldsymbol{X}^{(m)}_i\right\}$ are assumed to be independent and identically distributed across i. Following the idea of the MIDAS approach, we develop the panel-MIDAS model:

$$Y_{it} = \boldsymbol{X}_{it}(\boldsymbol{\theta})\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}, \tag{1.2.1}$$

where α_i is the unobservable individual effects which could be either fixed or random, $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional column vector of parameters, and $\boldsymbol{X}_{it}(\boldsymbol{\theta}) = (X^1_{it}(\boldsymbol{\theta}^1), X^2_{it}(\boldsymbol{\theta}^2), \ldots, X^p_{it}(\boldsymbol{\theta}^p))$ is a $1 \times p$ vector of aggregated high-frequency data with

$$X^j_{it}(\boldsymbol{\theta}^j) = \sum_{k=1}^{K} w_k(\boldsymbol{\theta}^j) X^{j(m)}_{i,t-(k-1)/m}, \; j = 1, \ldots, p,$$

where K is the maximum lag order of high-frequency variables, which can be shorter or greater than m[1]. $X_{i,t-k/m}^{j(m)}$ represents the k-th observation of $\boldsymbol{X}_i^{j(m)}$ that we look into the past from the most recent observation, $w_k(\boldsymbol{\theta}^j)$ is the aggregation weighting function that is parameterized as a function of a low-dimensional vector $\boldsymbol{\theta}^j$. The weighting function can have a number of functional forms. Among those, the Beta function and the Exponential Almon function are the most commonly used. A Beta function with two parameters $\boldsymbol{\theta}^j = (\theta_1^j, \theta_2^j)'$ is defined as

$$w_k(\boldsymbol{\theta}^j) = \frac{f(\frac{k}{K}, \theta_1^j; \theta_2^j)}{\sum\limits_{k=1}^{K} f(\frac{k}{K}, \theta_1^j; \theta_2^j)},$$

where

$$f(a, \theta_1^j; \theta_2^j) = \frac{(a)^{\theta_1^j - 1}(1-a)^{\theta_2^j - 1}\Gamma(\theta_1^j + \theta_2^j)}{\Gamma(\theta_1^j)\Gamma(\theta_2^j)}, \ \Gamma(\theta) = \int_0^\infty e^{-x} x^{\theta-1} dx.$$

An Exponential Almon function with two parameters $\boldsymbol{\theta}^j = (\theta_1^j, \theta_2^j)'$ is defined as

$$w_k(\boldsymbol{\theta}^j) = \frac{e^{\theta_1^j k + \theta_2^j k^2}}{\sum\limits_{k=1}^{K} e^{\theta_1^j k + \theta_2^j k^2}}, \tag{1.2.2}$$

where the weights add up to 1 in the sense that $\sum\limits_{k=1}^{K} w_k(\boldsymbol{\theta}^j) = 1$. In this paper, we focus on a two-parameter Exponential Almon function. Ghysels, Sinko, and Valkanov (2007) point out that even though the Exponential Almon polynomial function has only two parameters, it is still flexible enough to take various weighting shapes. As shown in Figure 1.1, we can obtain the rapid decaying weights and slow decaying weights corresponding to $\theta_1 = 0$, $\theta_2 = -0.2$ and $\theta_1 = 0$, $\theta_2 = -0.02$, respectively. The equal weights can be developed when $\theta_1 = \theta_2 = 0$. The exponential Almon weights can also produce a hump shape, which emerges for $\theta_1 = 0.2$, $\theta_2 = -0.02$. The usage of Exponential Almon polynomials enables us to fit a large number of lags with only two parameters and obtain more degrees of freedom.

[1] For simplicity and without loss of generality, we assume that all high-frequency variables are sampled at the same frequency and have the same lag orders.

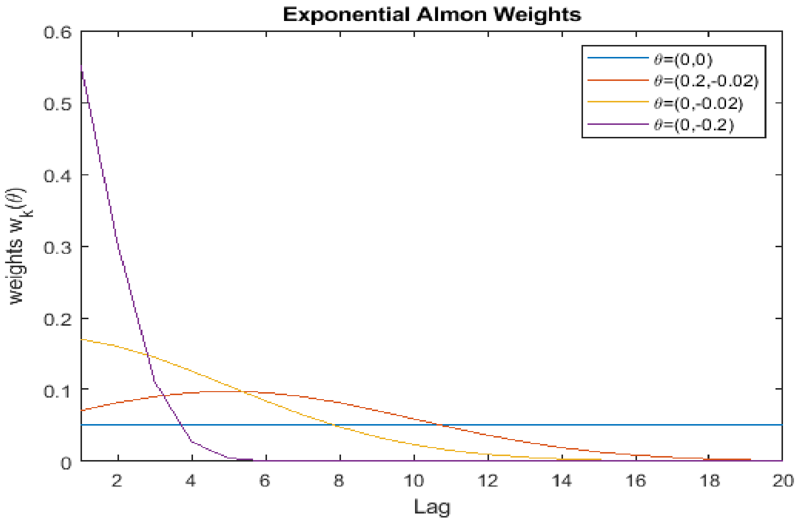

Figure 1.1: Exponential Almon Polynomial Weighting Function

1.3 Asymptotic Properties of the MIDAS-NLS Estimator

The objective of this subsection is to establish the consistency and the asymptotic normality of the NLS estimator of the panel-MIDAS model, denoted as MIDAS-NLS. We consider two cases: (i) α_i are the fixed effects; (ii) α_i are the random effects. Our asymptotics are based on the case where m, K, and T are fixed, and $N \to \infty$.

1.3.1 Fixed-Effects Model

The cross-section equation in the panel-MIDAS model (1.2.1) is

$$\boldsymbol{Y}_i = \boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{\beta} + \alpha_i \boldsymbol{l}_T + \boldsymbol{\varepsilon}_i, \tag{1.3.1}$$

where $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})'$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})'$ are $T \times 1$ column vectors. $\boldsymbol{l}_T$ is a $T \times 1$ vector of ones. $\boldsymbol{X}_i(\boldsymbol{\theta})$ is a $T \times p$ matrix of aggregated variable, defined by

$$\underset{T\times p}{\boldsymbol{X}_i(\boldsymbol{\theta})} = \begin{pmatrix} X_{i1}^1(\boldsymbol{\theta}^1) & \cdots & X_{i1}^p(\boldsymbol{\theta}^p) \\ \vdots & \ddots & \vdots \\ X_{iT}^1(\boldsymbol{\theta}^1) & \cdots & X_{iT}^p(\boldsymbol{\theta}^p) \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{X}_{i1}(\boldsymbol{\theta}) \\ \vdots \\ \boldsymbol{X}_{iT}(\boldsymbol{\theta}) \end{pmatrix} = (\boldsymbol{X}_i^1(\boldsymbol{\theta}^1), \ldots, \boldsymbol{X}_i^p(\boldsymbol{\theta}^p)).$$

First, we consider the panel-MIDAS model with fixed effects (FE). In this case, following the standard practice, we eliminated α_i from the model (1.3.1) by demeaning the variables, namely,

$$\tilde{\boldsymbol{Y}}_i = \tilde{\boldsymbol{X}}_i(\boldsymbol{\theta})\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}_i. \tag{1.3.2}$$

Define $\boldsymbol{H}_T = \boldsymbol{I}_T - \boldsymbol{l}_T(\boldsymbol{l}_T'\boldsymbol{l}_T)^{-1}\boldsymbol{l}_T'$. The demeaned equation (1.3.2) is obtained by premultiplying equation (1.3.1) by $\boldsymbol{H}_T$. Specifically, $\boldsymbol{H}_T\boldsymbol{Y}_i = \tilde{\boldsymbol{Y}}_i$, $\boldsymbol{H}_T\boldsymbol{l}_T = \boldsymbol{0}$, $\boldsymbol{H}_T\boldsymbol{X}_i(\boldsymbol{\theta}) = \tilde{\boldsymbol{X}}_i(\boldsymbol{\theta})$, and $\boldsymbol{H}_T\boldsymbol{\varepsilon}_i = \tilde{\boldsymbol{\varepsilon}}_i$.

Define $f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}) = \boldsymbol{H}_T f(\boldsymbol{X}_i;\boldsymbol{\delta})$ and $f(\boldsymbol{X}_i;\boldsymbol{\delta}) = \boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{\beta}$ with $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\theta}')'$. Let $\boldsymbol{\delta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\theta}_0')'$ be the true parameter vector, where $\boldsymbol{\beta}_0 = (\beta_{10}, \beta_{20}, \ldots, \beta_{p0})'$, $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_0^{1\prime}, \ldots, \boldsymbol{\theta}_0^{p\prime})'$ with $\boldsymbol{\theta}_0^j = (\theta_{1,0}^j, \theta_{2,0}^j)'$. The equilibrium vector $\boldsymbol{Y}_i$ is

$$\tilde{\boldsymbol{Y}}_i = f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) + \tilde{\boldsymbol{\varepsilon}}_i. \tag{1.3.3}$$

To provide a rigorous analysis of the MIDAS-NLS estimator, basic regularity conditions are assumed below.

Assumption 1: $\boldsymbol{\delta}_0$ is the true value of $\boldsymbol{\delta}$ and is an interior point of the parameter space $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a compact set.

Assumption 2: $\mathbb{E}(\varepsilon_{it} \mid \boldsymbol{X}_i^{(m)}, \alpha_i) = 0$ and $\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \mid \boldsymbol{X}_i^{(m)}, \alpha_i) = \sigma_\varepsilon^2\boldsymbol{I}_T$.

Assumption 3: $\mathbb{E}(\tilde{\boldsymbol{Y}}_i{}'\tilde{\boldsymbol{Y}}_i) < \infty$ and $\mathbb{E}\underset{\boldsymbol{\delta}\in\boldsymbol{\Theta}}{sup}[f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})'f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})] < \infty$.

Assumption 4: $\mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'})$ exists and is non-singular.

Assumption 1 provides a restriction on the parameter space. Assumption 2 is a basic assumption of errors. Assumption 3 is a sufficient condition for the consistency of the MIDAS-NLS estimator, ensuring that the uniform law of large numbers (LLN) holds. Assumption 4 ensures that there is no multicollinearity in the regressors.

The nonlinear least-squares estimator $\hat{\boldsymbol{\delta}}$ is the minimizer of the sum of squared residuals:

$$Q(\boldsymbol{\delta}) = \sum_{i=1}^{N} (\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}))'(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})). \tag{1.3.4}$$

Let $Q_0(\boldsymbol{\delta}) = \mathbb{E}(\frac{Q(\boldsymbol{\delta})}{N})$. The consistency of $\hat{\boldsymbol{\delta}}$ follows from the identifiable uniqueness and uniform convergence. We need to show that (i) $Q_0(\boldsymbol{\delta})$ is uniquely minimized at the true value $\boldsymbol{\delta}_0$, and (ii) $\frac{1}{N}Q(\boldsymbol{\delta}) \xrightarrow{p} Q_0(\boldsymbol{\delta})$ uniformly in $\boldsymbol{\delta} \in \boldsymbol{\Theta}$.

Theorem 1.1. *Under Assumptions 1-3, $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$.*

Proof. See Appendix A.1. □

The asymptotic distribution of the MIDAS-NLS estimator is obtained from the Taylor series expansion of $\frac{\partial Q(\hat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}} = \mathbf{0}$ at $\boldsymbol{\delta}_0$. The first-order derivatives of (1.3.4) at $\boldsymbol{\delta}_0$ are

$$\begin{cases} \frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\beta}} &= -\frac{2}{\sqrt{N}}\sum_{i=1}^{N} \tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)'\tilde{\boldsymbol{\varepsilon}}_i, \\ \frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}} &= -\frac{2}{\sqrt{N}}\sum_{i=1}^{N} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}}\tilde{\boldsymbol{\varepsilon}}_i, \end{cases}$$

where $\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \boldsymbol{\theta}^1} \\ \vdots \\ \frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \boldsymbol{\theta}^p} \end{pmatrix} = \begin{pmatrix} \frac{\partial \boldsymbol{X}_i^1(\boldsymbol{\theta}^1)'}{\partial \boldsymbol{\theta}^1}\beta_1 \\ \vdots \\ \frac{\partial \boldsymbol{X}_i^p(\boldsymbol{\theta}^p)'}{\partial \boldsymbol{\theta}^p}\beta_p \end{pmatrix}$, with $\frac{\partial \boldsymbol{X}_i^j(\boldsymbol{\theta}^j)'}{\partial \boldsymbol{\theta}^j} = \begin{pmatrix} \frac{\partial \boldsymbol{X}_i^j(\boldsymbol{\theta}^j)'}{\partial \theta_1^j} \\ \frac{\partial \boldsymbol{X}_i^j(\boldsymbol{\theta}^j)'}{\partial \theta_2^j} \end{pmatrix}$ and

$$\frac{\partial \boldsymbol{X}_i^j(\boldsymbol{\theta}^j)}{\partial \theta_1^j} = \begin{pmatrix} \sum_{k=1}^{K} \frac{\partial w_k(\boldsymbol{\theta}^j)}{\partial \theta_1^j} X_{i,1-(k-1)/m}^{j(m)} \\ \vdots \\ \sum_{k=1}^{K} \frac{\partial w_k(\boldsymbol{\theta}^j)}{\partial \theta_1^j} X_{i,T-(k-1)/m}^{j(m)} \end{pmatrix}, \quad \frac{\partial \boldsymbol{X}_i^j(\boldsymbol{\theta}^j)}{\partial \theta_2^j} = \begin{pmatrix} \sum_{k=1}^{K} \frac{\partial w_k(\boldsymbol{\theta}^j)}{\partial \theta_2^j} X_{i,1-(k-1)/m}^{j(m)} \\ \vdots \\ \sum_{k=1}^{K} \frac{\partial w_k(\boldsymbol{\theta}^j)}{\partial \theta_2^j} X_{i,T-(k-1)/m}^{j(m)} \end{pmatrix},$$

where for a two-parameter Exponential Almon polynomial function given by the equation (1.2.2), we have

$$\begin{cases} \frac{\partial w_k(\boldsymbol{\theta}^j)}{\partial \theta_1^j} = \frac{e^{\theta_1^j k + \theta_2^j k^2}(k \sum_{k=1}^{K} e^{\theta_1^j k + \theta_2^j k^2} - \sum_{k=1}^{K} k e^{\theta_1^j k + \theta_2^j k^2})}{(\sum_{k=1}^{K} e^{\theta_1^j k + \theta_2^j k^2})^2}, \\ \frac{\partial w_k(\boldsymbol{\theta}^j)}{\partial \theta_2^j} = \frac{e^{\theta_1^j k + \theta_2^j k^2}(k^2 \sum_{k=1}^{K} e^{\theta_1^j k + \theta_2^j k^2} - \sum_{k=1}^{K} k^2 e^{\theta_1^j k + \theta_2^j k^2})}{(\sum_{k=1}^{K} e^{\theta_1^j k + \theta_2^j k^2})^2}. \end{cases} \tag{1.3.5}$$

The first-order derivatives of (1.3.4) at $\boldsymbol{\delta}_0$ appear in linear forms of $\boldsymbol{\varepsilon}_i$, then we can apply the central limit theorem to $\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}}$ to derive the asymptotic distribution of the estimator.

Theorem 1.2. *Under Assumptions 1-4, the asymptotic normality of the MIDAS estimator* $\hat{\boldsymbol{\delta}}$ *is given by*

$$\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0}, \sigma_\varepsilon^2(\mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}))^{-1}),$$

and the asymptotic variance of $\hat{\boldsymbol{\beta}}$ *is given by*

$$AVar(\hat{\boldsymbol{\beta}}) = \frac{\sigma_\varepsilon^2}{N}(G_{11}(\boldsymbol{\theta}_0) - G_{12}(\boldsymbol{\theta}_0)G_{22}(\boldsymbol{\theta}_0)^{-1}G_{12}(\boldsymbol{\theta}_0)')^{-1}, \tag{1.3.6}$$

where $G_{11}(\boldsymbol{\theta}_0) = \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)'\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0))$, $G_{12}(\boldsymbol{\theta}_0) = \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)'\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'})$, *and* $G_{22}(\boldsymbol{\theta}_0) =$

$\mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'})$.

Proof. See Appendix A.2. □

1.3.2 Random-Effects Model

If a_i is treated as the random effects (RE), then the panel-MIDAS model (1.3.1) can be written as

$$\boldsymbol{Y}_i = \boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{\beta} + \boldsymbol{u}_i, \tag{1.3.7}$$

where $\boldsymbol{u}_i = \alpha_i \boldsymbol{l}_T + \boldsymbol{\varepsilon}_i$. Define the variance-covariance matrix of $\boldsymbol{u}_i$ as $\boldsymbol{\Omega} = \mathbb{E}(\boldsymbol{u}_i\boldsymbol{u}_i')$. In this case, premultiplying (1.3.7) by $\boldsymbol{\Omega}^{-1/2}$, the transformed model is

$$\tilde{\boldsymbol{Y}}_i = \tilde{\boldsymbol{X}}_i(\boldsymbol{\theta})\boldsymbol{\beta} + \tilde{\boldsymbol{u}}_i, \tag{1.3.8}$$

where $\tilde{\boldsymbol{Y}}_i = \boldsymbol{\Omega}^{-1/2}\boldsymbol{Y}_i$, $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}) = \boldsymbol{\Omega}^{-1/2}\boldsymbol{X}_i(\boldsymbol{\theta})$, and $\tilde{\boldsymbol{u}}_i = \boldsymbol{\Omega}^{-1/2}\boldsymbol{u}_i$. Similarly, define $f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}) = \boldsymbol{\Omega}^{-1/2}f(\boldsymbol{X}_i;\boldsymbol{\delta})$ and $f(\boldsymbol{X}_i;\boldsymbol{\delta}) = \boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{\beta}$ with $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\theta}')'$. Let $\boldsymbol{\delta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\theta}_0')'$ be the true parameter vector, the equilibrium vector $\boldsymbol{Y}_i$ is

$$\tilde{\boldsymbol{Y}}_i = f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) + \tilde{\boldsymbol{u}}_i. \tag{1.3.9}$$

The main results of this subsection are presented in the following theorem. Before we proceed, we list some additional assumptions necessary for deriving the asymptotic distribution of the MIDAS-NLS estimator of the RE model.

Assumption 2': (a) $\mathbb{E}(\varepsilon_{it} \mid \boldsymbol{X}_i^{(m)}, \alpha_i) = 0$ and $\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \mid \boldsymbol{X}_i^{(m)}, \alpha_i) = \sigma_\varepsilon^2 \boldsymbol{I}_N$; (b) $\mathbb{E}(\alpha_i \mid \boldsymbol{X}_i^{(m)}) = 0$ and $\mathbb{E}(\alpha_i^2 \mid \boldsymbol{X}_i^{(m)}) = \sigma_\alpha^2$.

Assumption 3': $\mathbb{E}(\boldsymbol{Y}_i\boldsymbol{\Omega}^{-1}\boldsymbol{Y}_i) < \infty$ and $\mathbb{E}\underset{\boldsymbol{\delta}\in\boldsymbol{\Theta}}{sup}[f(\boldsymbol{X}_i;\boldsymbol{\delta})'\boldsymbol{\Omega}^{-1}f(\boldsymbol{X}_i;\boldsymbol{\delta})] < \infty$.

Assumption 4': $\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\Omega}^{-1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'})$ exists and is non-singular.

Under Assumption 2', the variance of u_{it} is $\sigma_u^2 = \sigma_\alpha^2 + \sigma_\varepsilon^2$ and its error covariance matrix is $\boldsymbol{\Omega} = \sigma_\alpha^2 \boldsymbol{l}_T \boldsymbol{l}_T' + \sigma_\varepsilon^2 \boldsymbol{I}_T$. Obviously, when $\sigma_\alpha^2 = 0$, there are no individual effects in the model, the RE model becomes the pooled data model. Given that $\boldsymbol{H}_T = \boldsymbol{I}_T - \boldsymbol{l}_T(\boldsymbol{l}_T'\boldsymbol{l}_T)^{-1}\boldsymbol{l}_T'$ in the FE model, $\boldsymbol{\Omega}^{-1} = \frac{1}{\sigma_\varepsilon^2}(\boldsymbol{I}_T - \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2 + T\sigma_\alpha^2}\boldsymbol{l}_T\boldsymbol{l}_T')$ in the RE model, the RE model becomes the FE model if T is large or σ_α^2 is large.

Theorem 1.3. *Under Assumptions 1,2', 3', and 4', the MIDAS estimator $\hat{\boldsymbol{\delta}}$ is consistent, and the asymptotic normality of $\hat{\boldsymbol{\delta}}$ is given by*

$$\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0}, (\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\Omega}^{-1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}))^{-1}),$$

and the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is given by

$$AVar(\hat{\boldsymbol{\beta}}) = \frac{(G_{11}(\boldsymbol{\theta}_0) - G_{12}(\boldsymbol{\theta}_0)G_{22}(\boldsymbol{\theta}_0)^{-1}G_{12}(\boldsymbol{\theta}_0)')^{-1}}{N}, \tag{1.3.10}$$

where $G_{11}(\boldsymbol{\theta}_0) = \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}_0)'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i(\boldsymbol{\theta}_0))$, $G_{12}(\boldsymbol{\theta}_0) = \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}_0)'\boldsymbol{\Omega}^{-1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'})$, and $G_{22}(\boldsymbol{\theta}_0) = \mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}}\boldsymbol{\Omega}^{-1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'})$.

Proof. See Appendix A.3. □

Remark 1.1. *The above theorem is derived based on the assumption that the error covariance matrix $\boldsymbol{\Omega}$ is known. However, in practice, $\boldsymbol{\Omega}$ depends on two unknown parameters, namely σ_α^2 and σ_ε^2. To implement the RE procedure, we need to estimate σ_α^2 and σ_ε^2. Here, we first estimate σ_u^2. A consistent estimator of σ_u^2 is*

$$\hat{\sigma}_u^2 = \frac{1}{NT - 3p}\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{u}_{it}^2,$$

where $\hat{u}_{it} = Y_{it} - \boldsymbol{X}_{it}(\hat{\boldsymbol{\theta}}_{PONLS})\hat{\boldsymbol{\beta}}_{PONLS}$ is the MIDAS-NLS residuals from the pooled data *model, and a consistent estimator of* σ_ε^2 *is*

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)-3p}\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{\varepsilon}_{it}^2,$$

where $\hat{\varepsilon}_{it}$ *is the MIDAS-NLS residuals from the FE model, and consequently, we can form* $\hat{\sigma}_\alpha^2 = \hat{\sigma}_u^2 - \hat{\sigma}_\varepsilon^2$*. Finally,* $\boldsymbol{\Omega}$ *can be estimated by*

$$\hat{\boldsymbol{\Omega}} = \hat{\sigma}_\alpha^2 \boldsymbol{l}_T \boldsymbol{l}_T' + \hat{\sigma}_\varepsilon^2 \boldsymbol{I}_T. \tag{1.3.11}$$

Once we obtain $\hat{\boldsymbol{\Omega}}$*, we insert* $\hat{\boldsymbol{\Omega}}$ *into Theorem 1.3 in place of* $\boldsymbol{\Omega}$*.*

1.3.3 Test for the Flat Weighting Scheme

The flat temporal aggregation is the most commonly used in the literature. To justify the validity of the flat weighting scheme, the objective of this subsection is to discuss how to test the null hypothesis: $H_0 : \boldsymbol{\theta}_0 = \mathbf{0}$. This test can take the form of the Wald test. The Wald test statistic can be written as

$$W = \hat{\boldsymbol{\theta}}'[\boldsymbol{R}(\hat{\boldsymbol{V}}/N)\boldsymbol{R}']^{-1}\hat{\boldsymbol{\theta}} \sim \chi_{2p}^2, \tag{1.3.12}$$

where $\boldsymbol{R}$ "picks up" the parameters $\boldsymbol{\theta}$ and is a $2p \times 3p$ matrix generated from a $3p \times 3p$ identity matrix by removing the first p rows. In the FE model,

$$\hat{\boldsymbol{V}} = \hat{\sigma}_\varepsilon^2 (\mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\hat{\boldsymbol{\delta}})'}{\partial \boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\hat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}'}))^{-1},$$

and in the RE model,

$$\hat{\boldsymbol{V}} = \mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\hat{\boldsymbol{\delta}})'}{\partial \boldsymbol{\delta}}\hat{\boldsymbol{\Omega}}^{-1}\frac{\partial f(\boldsymbol{X}_i;\hat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}'}))^{-1},$$

where $\hat{\sigma}_{\varepsilon}^2$ and $\hat{\boldsymbol{\Omega}}$ are given in Remark 1.1. We reject H_0 if $W > \chi^2_{2p}$, and failure to reject the null hypothesis suggests the validity of the flat weighting scheme.

1.4 Comparison with the LS Model

1.4.1 Asymptotic Properties of the LS Estimator

One traditional method to deal with the mixed-frequency data is to aggregate high-frequency observations to low frequency using a predetermined weighting scheme $\boldsymbol{\pi}$. Then the regression model (1.2.1) becomes

$$Y_{it} = \boldsymbol{X}_{it}(\boldsymbol{\pi})\boldsymbol{\beta}^* + \alpha_i^* + \varepsilon_{it}^*. \tag{1.4.1}$$

where $\boldsymbol{\pi} = (\boldsymbol{\pi}^{1\prime}, \boldsymbol{\pi}^{2\prime}, \dots, \boldsymbol{\pi}^{p\prime})'$ with $\boldsymbol{\pi}^j = (\pi_1^j, \pi_2^j, \dots, \pi_m^j)'$ such that $\pi_k^j \geq 0$ and $\sum_{k=1}^{K} \pi_k^j = 1$. $\boldsymbol{X}_{it}(\boldsymbol{\pi}) = (X_{it}^1(\boldsymbol{\pi}^1), X_{it}^2(\boldsymbol{\pi}^2), \dots, X_{it}^p(\boldsymbol{\pi}^p))$ is a $1 \times p$ vector of aggregated high-frequency data with

$$X_{it}^j(\boldsymbol{\pi}^j) = \sum_{k=1}^{K} \pi_k^j X_{i,t-(k-1)/m}^{j(m)},\ j = 1, \dots, p.$$

The model (1.4.1) is a linear LS model where the weighting schemes do not depend on any unknown parameters.

Suppose that the correct specification is in the MIDAS form (1.2.1), but we mistakenly estimate the linear LS model (1.4.1). To assess the consequences of this misspecification, we decompose $\boldsymbol{X}_{it}(\boldsymbol{\theta})$ in (1.2.1) into $\boldsymbol{X}_{it}(\boldsymbol{\pi})$ and one nonlinear term, $\boldsymbol{X}_{it}(\boldsymbol{\theta}^*)$, then the model (1.2.1) becomes

$$Y_{it} = \boldsymbol{X}_{it}(\boldsymbol{\pi})\boldsymbol{\beta} + \boldsymbol{X}_{it}(\boldsymbol{\theta}^*)\boldsymbol{\beta} + \alpha_i + \boldsymbol{\varepsilon}_{it}, \tag{1.4.2}$$

where $\boldsymbol{X}_{it}(\boldsymbol{\theta}^*) = (X_{it}^1(\boldsymbol{\theta}^{1*}), X_{it}^2(\boldsymbol{\theta}^{2*}), \dots, X_{it}^p(\boldsymbol{\theta}^{p*}))$ is a $1 \times p$ vector with

$$X_{it}^j(\boldsymbol{\theta}^{j*}) = \sum_{k=1}^{K} (w_k(\boldsymbol{\theta}^j) - \pi_k^j) X_{i,t-(k-1)/m}^{j(m)}, \; j = 1, \dots, p.$$

Eq.(1.4.2) implies that the LS model yields a nonlinear omitted variable term, $\boldsymbol{X}_{it}(\boldsymbol{\theta}^*)$, and therefore may lead to a biased and inefficient estimator. The objective of this subsection is to study the asymptotic bias of the LS estimator of $\boldsymbol{\beta}^*$ and derive its asymptotic variance.

Fixed-Effects Model

If a_i is treated as the fixed effects, the demeaned equation of (1.4.2) at $\boldsymbol{\delta}_0$ is

$$\tilde{\boldsymbol{Y}}_i = \tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})\boldsymbol{\beta}_0 + \tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)\boldsymbol{\beta}_0 + \tilde{\boldsymbol{\varepsilon}}_i. \tag{1.4.3}$$

where $\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi}) = \boldsymbol{H}_T \boldsymbol{X}_i(\boldsymbol{\pi})$ and $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*) = \boldsymbol{H}_T \boldsymbol{X}_i(\boldsymbol{\theta}_0^*)$. Since the transformed LS model (1.4.1) can be estimated by OLS: $\hat{\boldsymbol{\beta}}^* = (\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}(\boldsymbol{\pi}))^{-1}\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{Y}}$, plug (1.4.3) into $\hat{\boldsymbol{\beta}}^*$ yields the estimates

$$\hat{\boldsymbol{\beta}}^* = \boldsymbol{\beta}_0 + (\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}(\boldsymbol{\pi}))^{-1}(\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}(\boldsymbol{\theta}_0^*))\boldsymbol{\beta}_0 + (\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}(\boldsymbol{\pi}))^{-1}\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{\varepsilon}}.$$

One additional assumption necessary for the asymptotic properties of the LS estimator is summarized below.

Assumption 5: $\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi}))$ exists and is non-singular.

Theorem 1.4. *Under Assumptions 2 and 5, the asymptotic bias of $\hat{\boldsymbol{\beta}}^*$ is given by*

$$ABias(\hat{\boldsymbol{\beta}}^*; \boldsymbol{\beta}_0) = (\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi}))^{-1}\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*))\boldsymbol{\beta}_0, \tag{1.4.4}$$

and its asymptotic variance is given by

$$AVar(\hat{\boldsymbol{\beta}}^*) = \frac{\sigma_\varepsilon^2}{N}(\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})))^{-1}. \tag{1.4.5}$$

Proof. See Appendix A.4. □

Eq.(1.4.4) implies that the asymptotic bias of the LS estimator is directly related to the correlation between $\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})$ and $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)$. It gives us a simple way to determine the sign of bias. If $\boldsymbol{\beta}_0 > \mathbf{0}$, $\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})$ and $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)$ are positively correlated, the asymptotic bias is positive. Conversely, if $\boldsymbol{\beta}_0 > \mathbf{0}$, but $\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})$ and $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)$ are negatively correlated, the asymptotic bias is negative. Besides, the bias formula (1.4.4) suggests that the greater difference between $\boldsymbol{\pi}$ and $w(\boldsymbol{\theta}_0)$ leads to a greater bias. It is easy to show that omitting $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)$ does not lead to biased estimate of $\boldsymbol{\beta}^*$ in two cases: (i) when $\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})$ and $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)$ are orthogonal in the sense that $\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)) = \mathbf{0}$, regardless of the value of the true weighting scheme $\boldsymbol{\theta}_0$; (ii) The true weighting scheme is $\boldsymbol{\pi}$ in the sense that $\tilde{\boldsymbol{X}}(\boldsymbol{\theta}_0^*) = \mathbf{0}$. Eq.(1.4.5) shows that the asymptotic variance of the LS estimator does not depend on the omitted variable, implying that its variance stays the same regardless of the values of $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$.

By (1.3.6) and (1.4.5), it is difficult to determine whether the MIDAS-NLS estimator or the LS estimator is more efficient since their asymptotic variance ratio would depend on the nonlinear function $f(\boldsymbol{X}_i; \boldsymbol{\delta}_0)$ and its derivatives.

Random-Effects Model

Similarly, if a_i is treated as the random effects, the demeaned equation of (1.4.2) at $\boldsymbol{\delta}_0$ is

$$\tilde{\boldsymbol{Y}}_i = \tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})\boldsymbol{\beta}_0 + \tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)\boldsymbol{\beta}_0 + \tilde{\boldsymbol{u}}_i, \tag{1.4.6}$$

where $\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi}) = \boldsymbol{\Omega}^{-1/2}\boldsymbol{X}_i(\boldsymbol{\pi})$ and $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*) = \boldsymbol{\Omega}^{-1/2}\boldsymbol{X}_i(\boldsymbol{\theta}_0^*)$.

For the RE model, we impose the following additional regularity condition and the asymptotic properties of the LS estimator.

Assumption 5': $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i(\boldsymbol{\pi}))$ exists and is non-singular.

Theorem 1.5. *Under Assumptions 2' and 5', the asymptotic bias of* $\hat{\boldsymbol{\beta}}^*$ *is given by*

$$ABias(\hat{\boldsymbol{\beta}}^*;\boldsymbol{\beta}_0) = \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i(\boldsymbol{\pi})))^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i(\boldsymbol{\theta}_0^*))\boldsymbol{\beta}_0, \tag{1.4.7}$$

and its asymptotic variance is given by

$$AVar(\hat{\boldsymbol{\beta}}^*) = \frac{1}{N}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i(\boldsymbol{\pi})))^{-1}. \tag{1.4.8}$$

Remark 1.2. *The asymptotic bias of the LS estimator in the RE model is consistent with that in the FE model, suggesting that the LS estimator is biased unless the true weighting scheme is* $\boldsymbol{\pi}$ *or* $\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})$ *and* $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)$ *are orthogonal. (1.4.7) and (1.4.8) show that the asymptotic properties of* $\hat{\boldsymbol{\beta}}^*$ *are directly related to the variance-covariance matrix of* $\boldsymbol{u}_i$*. The consistent estimator of* $\boldsymbol{\Omega}$ *is given in (1.3.11).*

1.4.2 Example

In this subsection, we derive the analytical results for the asymptotic bias of the LS estimator and the asymptotic variance ratio of the LS and MIDAS-NLS estimators where the high-frequency regressor $\{X_{i,t}^{(m)}\}$ follows a MA(1) process or an AR(1) process. For simplicity, we consider the case of a single regressor with $K = \{m, 2m\}$, the panel-MIDAS model is given by

$$Y_{it} = X_{it}(\boldsymbol{\pi})\beta + X_{it}(\boldsymbol{\theta}^*)\beta + \alpha_i + \varepsilon_{it},$$

where $\alpha_i \mid \boldsymbol{X}_i^{(m)} \sim IID(0,\sigma_\alpha^2)$, $\varepsilon_{it} \mid (\boldsymbol{X}_i^{(m)}, \alpha_i) \sim IID(0,\sigma_\varepsilon^2)$, $X_{it}(\boldsymbol{\pi}) = \sum_{j=1}^{K} \pi_j x_{i,t-(j-1)/m}^{(m)}$, and $X_{it}(\boldsymbol{\theta}^*) = \sum_{j=1}^{K} w_j(\boldsymbol{\theta}^*) x_{i,t-(j-1)/m}^{(m)}$ with $w_j(\boldsymbol{\theta}^*) = w_j(\boldsymbol{\theta}) - \pi_j$.

MA(1) process

Let high-frequency regressor $\{X_{i,t}^{(m)}\}$ follows a stationary MA(1) given by $X_{i,t}^{(m)} = e_{i,t}^{(m)} + \rho e_{i,t-1/m}^{(m)}$, where $e_{i,t}^{(m)} \sim IID(0, \sigma_e^2)$, then $X_{it}(\boldsymbol{\pi})$ and $X_{it}(\boldsymbol{\theta}^*)$ are given by

$$\begin{cases} X_{it}(\boldsymbol{\pi}) = \pi_1 e_{i,t}^{(m)} + \sum\limits_{j=1}^{K-1} (\rho\pi_j + \pi_{j+1}) e_{i,t-j/m}^{(m)} + \rho\pi_K e_{i,t-K/m}^{(m)}, \\ X_{it}(\boldsymbol{\theta}^*) = w_1(\boldsymbol{\theta}^*) e_{i,t}^{(m)} + \sum\limits_{j=1}^{K-1} (\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*)) e_{i,t-j/m}^{(m)} + \rho w_K(\boldsymbol{\theta}^*) e_{i,t-K/m}^{(m)}. \end{cases}$$

Proposition 1.1. *In the FE model, let high-frequency regressor $\{X_{i,t}^{(m)}\}$ follows a MA(1) process, the asymptotic bias of the LS estimator is given by*

$$ABias(\hat{\beta}^*; \beta) = \gamma\beta, \tag{1.4.9}$$

where $\gamma = \frac{\gamma_1}{\gamma_2}$ with $\gamma_1 = tr(\boldsymbol{H}_T \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})'))$ and $\gamma_2 = tr(\boldsymbol{H}_T \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})'))$. When $K = m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} = \begin{cases} \sigma_e^2[\pi_1 w_1(\boldsymbol{\theta}^*) + \sum\limits_{j=1}^{m-1} (\rho\pi_j + \pi_{j+1})(\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*)) + \rho^2 \pi_m w_m(\boldsymbol{\theta}^*)] & \text{if } r = k \\ \sigma_e^2 \rho \pi_m w_1(\boldsymbol{\theta}^*) & \text{if } r = k+1 \\ \sigma_e^2 \rho \pi_1 w_m(\boldsymbol{\theta}^*) & \text{if } r = k-1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} = \begin{cases} \sigma_e^2[\pi_1^2 + \sum\limits_{j=1}^{m-1} (\rho\pi_j + \pi_{j+1})^2 + \rho^2 \pi_m^2] & \text{if } r = k \\ \sigma_e^2 \rho \pi_1 \pi_m & \text{if } r = k+1 \text{ or } r = k-1 \\ 0 & \text{otherwise} \end{cases}$$

When $K = 2m$,

$$
\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =
\begin{cases}
\sigma_e^2[\pi_1 w_1(\boldsymbol{\theta}^*) + \sum_{j=1}^{2m-1} (\rho\pi_j + \pi_{j+1})(\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*)) + \rho^2 \pi_{2m} w_{2m}(\boldsymbol{\theta}^*)] & \textit{if } r = k \\
\sigma_e^2[w_1(\boldsymbol{\theta}^*)(\rho\pi_m + \pi_{m+1}) + \sum_{j=1}^{m-1} (\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*)) & \\
\times(\rho\pi_{m+j} + \pi_{m+j+1}) + \rho\pi_{2m}(\rho w_m(\boldsymbol{\theta}^*) + w_{m+1}(\boldsymbol{\theta}^*))] & \textit{if } r = k+1 \\
\sigma_e^2[\pi_1(\rho w_m(\boldsymbol{\theta}^*) + w_{m+1}(\boldsymbol{\theta}^*)) + \sum_{j=1}^{m-1} (\rho w_{m+j}(\boldsymbol{\theta}^*) + w_{m+j+1}(\boldsymbol{\theta}^*)) & \\
\times(\rho\pi_j + \pi_{j+1}) + \rho w_{2m}(\boldsymbol{\theta}^*)(\rho\pi_m + \pi_{m+1})] & \textit{if } r = k-1 \\
\sigma_e^2 \rho \pi_{2m} w_1(\boldsymbol{\theta}^*) & \textit{if } r = k+2 \\
\sigma_e^2 \rho \pi_1 w_{2m}(\boldsymbol{\theta}^*) & \textit{if } r = k-2 \\
0 & \textit{otherwise}
\end{cases}
$$

and

$$
\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =
\begin{cases}
\sigma_e^2[\pi_1^2 + \sum_{j=1}^{2m-1} (\rho\pi_j + \pi_{j+1})^2 + \rho^2\pi_{2m}^2] & \textit{if } r = k \\
\sigma_e^2[\pi_1(\rho\pi_m + \pi_{m+1}) + \sum_{j=1}^{m-1} (\rho\pi_j + \pi_{j+1})(\rho\pi_{m+j} + \pi_{m+j+1}) & \\
+\rho\pi_{2m}(\rho\pi_m + \pi_{m+1})] & \textit{if } r = k+1 \textit{ or } r = k-1 \\
\sigma_e^2 \rho \pi_1 \pi_{2m} & \textit{if } r = k+2 \textit{ or } r = k-2 \\
0 & \textit{otherwise}
\end{cases}
$$

By (1.3.6) and (1.4.5), the asymptotic variance ratio of the MIDAS-NLS estimator and the LS estimator is given by

$$
\frac{AVar(\hat{\beta})}{AVar(\hat{\beta}^*)} = V_\beta / V_{\beta^*}, \tag{1.4.10}
$$

where $V_{\beta^*} = \frac{\sigma_\varepsilon^2}{N}(tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')))^{-1}$ *and* $V_\beta = \frac{\sigma_\varepsilon^2}{N}(V_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V'}_{12})^{-1}$ *with*

$$\begin{aligned} V_{11} = & \ tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')), \\ \boldsymbol{V}_{12} = & \begin{pmatrix} tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\boldsymbol{X}_i(\boldsymbol{\theta})')) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\boldsymbol{X}_i(\boldsymbol{\theta})')) \end{pmatrix}, \\ \boldsymbol{V}_{22} = & \begin{pmatrix} tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_1})) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_1})) \\ tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_2})) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_2})) \end{pmatrix}. \end{aligned}$$

When $K = m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} = \begin{cases} \sigma_e^2[w_1(\boldsymbol{\theta})^2 + \sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta}) + w_{j+1}(\boldsymbol{\theta}))^2 + \rho^2 w_m(\boldsymbol{\theta})^2] & if\ r = k \\ \rho w_m(\boldsymbol{\theta})w_1(\boldsymbol{\theta})\sigma_e^2 & if\ r = k+1\ or\ r = k-1 \\ 0 & otherwise \end{cases}$$

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} = \begin{cases} \sigma_e^2[w_1(\boldsymbol{\theta})\frac{\partial w_1(\boldsymbol{\theta})}{\partial\theta_s} + \sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta}) + w_{j+1}(\boldsymbol{\theta}))(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial\theta_s} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial\theta_s}) & \\ +\rho^2 w_m(\boldsymbol{\theta})\frac{\partial w_m(\boldsymbol{\theta})}{\partial\theta_s}]\beta & if\ r = k \\ \sigma_e^2\rho w_m(\boldsymbol{\theta})\frac{\partial w_1(\boldsymbol{\theta})}{\partial\theta_s}\beta & if\ r = k+1 \\ \sigma_e^2\rho w_1(\boldsymbol{\theta})\frac{\partial w_m(\boldsymbol{\theta})}{\partial\theta_s}\beta & if\ r = k-1 \\ 0 & otherwise \end{cases}$$

and

$$\mathbb{E}(\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_l}\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_s})_{kr} =$$

$$\begin{cases} \sigma_e^2[\frac{\partial w_1(\boldsymbol{\theta})}{\partial\theta_l}\frac{\partial w_1(\boldsymbol{\theta})}{\partial\theta_s}+\sum_{j=1}^{m-1}(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial\theta_l}+\frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial\theta_l})(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial\theta_s}+\frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial\theta_s}) & \\ +\rho^2\frac{\partial w_m(\boldsymbol{\theta})}{\partial\theta_l}\frac{\partial w_m(\boldsymbol{\theta})}{\partial\theta_s}]\beta^2 & \text{if } r=k \\ \sigma_e^2\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial\theta_s}\frac{\partial w_1(\boldsymbol{\theta})}{\partial\theta_l}\beta^2 & \text{if } r=k+1 \\ \sigma_e^2\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial\theta_l}\frac{\partial w_1(\boldsymbol{\theta})}{\partial\theta_s}\beta^2 & \text{if } r=k-1 \\ 0 & \text{otherwise} \end{cases}$$

When $K = 2m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$

$$\begin{cases} \sigma_e^2[w_1(\boldsymbol{\theta})^2+\sum_{j=1}^{2m-1}(\rho w_j(\boldsymbol{\theta})+w_{j+1}(\boldsymbol{\theta}))^2+\rho^2 w_{2m}(\boldsymbol{\theta})^2] & \text{if } r=k \\ \sigma_e^2[w_1(\boldsymbol{\theta})(\rho w_m(\boldsymbol{\theta})+w_{m+1}(\boldsymbol{\theta}))+\sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta})+w_{j+1}(\boldsymbol{\theta})) & \\ \times(\rho w_{m+j}(\boldsymbol{\theta})+w_{m+j+1}(\boldsymbol{\theta}))+\rho w_{2m}(\boldsymbol{\theta})(\rho w_m(\boldsymbol{\theta})+w_{m+1}(\boldsymbol{\theta}))] & \text{if } r=k+1 \text{ or } r=k-1 \\ \sigma_e^2\rho w_1(\boldsymbol{\theta})w_{2m}(\boldsymbol{\theta}) & \text{if } r=k+2 \text{ or } r=k-2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$
$$\begin{cases} \sigma_e^2[w_1(\boldsymbol{\theta})\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}+\sum_{j=1}^{2m-1}(\rho w_j(\boldsymbol{\theta})+w_{j+1}(\boldsymbol{\theta}))(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ +\rho^2 w_{2m}(\boldsymbol{\theta})\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}]\beta & \textit{if } r=k \\ \sigma_e^2[(\rho w_m(\boldsymbol{\theta})+w_{m+1}(\boldsymbol{\theta}))\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}+\sum_{j=1}^{m-1}(\rho w_{m+j}(\boldsymbol{\theta})+w_{m+j+1}(\boldsymbol{\theta})) & \\ \times(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s})+\rho w_{2m}(\boldsymbol{\theta})(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_s})]\beta & \textit{if } r=k+1 \\ \sigma_e^2[w_1(\boldsymbol{\theta})(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_s})+\sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta})+w_{j+1}(\boldsymbol{\theta})) & \\ \times(\rho\frac{\partial w_{m+j}(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{m+j+1}(\boldsymbol{\theta})}{\partial \theta_s})+\rho(\rho w_m(\boldsymbol{\theta})+w_{m+1}(\boldsymbol{\theta}))\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}]\beta & \textit{if } r=k-1 \\ \sigma_e^2\rho w_{2m}(\boldsymbol{\theta})\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}\beta & \textit{if } r=k+2 \\ \sigma_e^2\rho w_1(\boldsymbol{\theta})\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}\beta & \textit{if } r=k-2 \\ 0 & \textit{otherwise} \end{cases}$$

and

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \theta_s})_{kr} =$$
$$\begin{cases} \sigma_e^2[\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_l}\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}+\sum_{j=1}^{2m-1}(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_l}+\frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_l})(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ +\rho^2\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_l}\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}]\beta^2 & \textit{if } r=k \\ \sigma_e^2[(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_s})\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_l}+\sum_{j=1}^{m-1}(\rho\frac{\partial w_{m+j}(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{m+j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ \times(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_l}+\frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_l})+\rho\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_l}+\frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_l})]\beta^2 & \textit{if } r=k+1 \\ \sigma_e^2[\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_l}+\frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_l})+\sum_{j=1}^{m-1}(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ \times(\rho\frac{\partial w_{m+j}(\boldsymbol{\theta})}{\partial \theta_l}+\frac{\partial w_{m+j+1}(\boldsymbol{\theta})}{\partial \theta_l})+\rho(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s}+\frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_s})\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_l}]\beta^2 & \textit{if } r=k-1 \\ \sigma_e^2\rho\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_l}\beta^2 & \textit{if } r=k+2 \\ \sigma_e^2\rho\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_l}\beta^2 & \textit{if } r=k-2 \\ 0 & \textit{otherwise} \end{cases}$$

for all $s, l = 1, 2$.

Proof. See Appendix A.5. □

Proposition 1.2. *In the RE model, let high-frequency regressor* $\{X_{i,t}^{(m)}\}$ *follows a MA(1) process, the asymptotic bias of the LS estimator is given by*

$$ABias(\hat{\beta}^*; \beta) = \gamma\beta,$$

where $\gamma = \frac{\gamma_1}{\gamma_2}$ *with* $\gamma_1 = tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})'))$ *and* $\gamma_2 = tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})'))$. *By (1.3.10) and (1.4.8), the asymptotic variance ratio of the MIDAS-NLS estimator and the LS estimator is given by*

$$\frac{AVar(\hat{\beta})}{AVar(\hat{\beta}^*)} = V_\beta / V_{\beta^*},$$

where $V_{\beta^*} = \frac{1}{N}(tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')))^{-1}$ *and* $V_\beta = \frac{1}{N}(V_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{12}')^{-1}$ *with*

$$\begin{aligned}
V_{11} =& \; tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')),\\
\boldsymbol{V}_{12} =& \begin{pmatrix} tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\boldsymbol{X}_i(\boldsymbol{\theta})')) & tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\boldsymbol{X}_i(\boldsymbol{\theta})')) \end{pmatrix},\\
\boldsymbol{V}_{22} =& \begin{pmatrix} tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_1})) & tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_1})) \\ tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_2})) & tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_2})) \end{pmatrix},
\end{aligned}$$

where $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')$, $\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')$, *and* $\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_s})$ *are given in Proposition 1.1.*

Propositions 1.1 and 1.2 show that it is difficult to evaluate the asymptotic bias of the LS estimator and to determine under what conditions the MIDAS-NLS estimator is asymptotically more efficient than the LS estimator. To explore how $\boldsymbol{\theta}$, ρ, and K affect the asymptotic properties of estimators, we plot the figures for (1.4.9) and (1.4.10) as a function of m. We consider three cases. In the first case, we let $\rho = 0.8$, $K = m$, and consider four values of $\boldsymbol{\theta} = \{(0, -0.2), (0, -0.02), (0, -0.0002), (0, 0)\}$, which correspond to the rapid decaying weights, the slow decaying weights, near-flat weights, and the flat weights. In the second case, we let $\boldsymbol{\theta} = (0, -0.2)$, $K = m$, and consider four values of $\rho = \{0, 0.2, 0.5, 0.8\}$. In the third case, we let $\rho = 0.8$, $\boldsymbol{\theta} = (0, -0.2)$, and consider two values of $K = \{m, 2m\}$. In these cases, we let $\beta = 2$, $T = 10$, m ranges from 3 to 200, and we let $\boldsymbol{\pi}$ be the flat weights in the sense that $\boldsymbol{\pi} = (1/m, 1/m, \ldots, 1/m)'$.

Figure 1.2 plots the asymptotic bias of the LS estimator when the high-frequency regressor is a MA(1) process. We first focus on panel (a). As expected, the bias with flat weights is 0, while the bias with rapid decaying weights is the greatest. The bias with rapid and slow decaying weights is negative and increases in absolute value as m increases, and when m becomes large, it approaches a certain negative value. Panel (b) shows that the LS estimator is unbiased when $\{X_{i,t}^{(m)}\}$ follows an i.i.d process ($\rho = 0$), but it is biased when $\{X_{i,t}^{(m)}\}$ are serially correlated. This bias increases as ρ and m increase. Panel (c) suggests that the bias when $K = m$ and $K = 2m$ have the same trend, but the bias when $K = 2m$ is consistently larger than the bias when $K = m$.

Figure 1.3 explores the role of $\boldsymbol{\theta}$, ρ, and K to the asymptotic variance ratio of MIDAS-NLS and LS estimators. Panel (a) suggests that in the special case of $\boldsymbol{\theta} = \mathbf{0}$, the variance ratio of these two estimators stabilizes at 1. For all non-flat weights, the MIDAS-NLS estimator is asymptotically more efficient than the LS estimator. Their relative efficiency ratio is smaller for the faster-decaying weights and tends to be smaller as m increases. It is interesting to note that for the near-flat weighting scheme, the variance ratio is around 1 for $m < 100$ and then decreases dramatically for $m \geq 100$. Panels (b) and (c) show that the asymptotic variance ratio of estimators is smaller for the case of $K = 2m$, but it is not greatly affected by ρ.

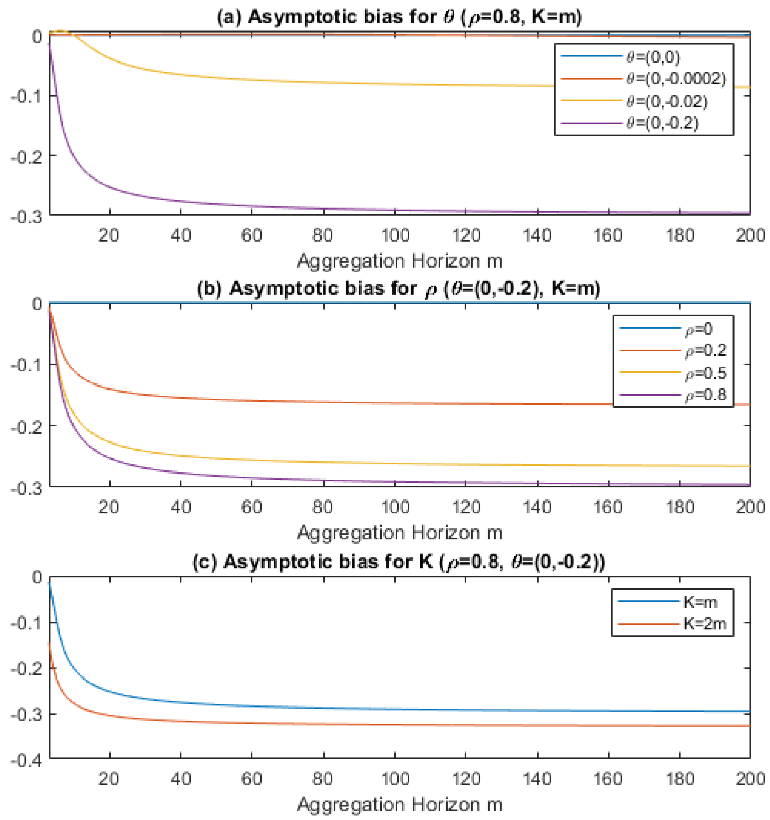

Figure 1.2: The Asymptotic Bias of LS Estimators for Models with a MA(1) Regressor

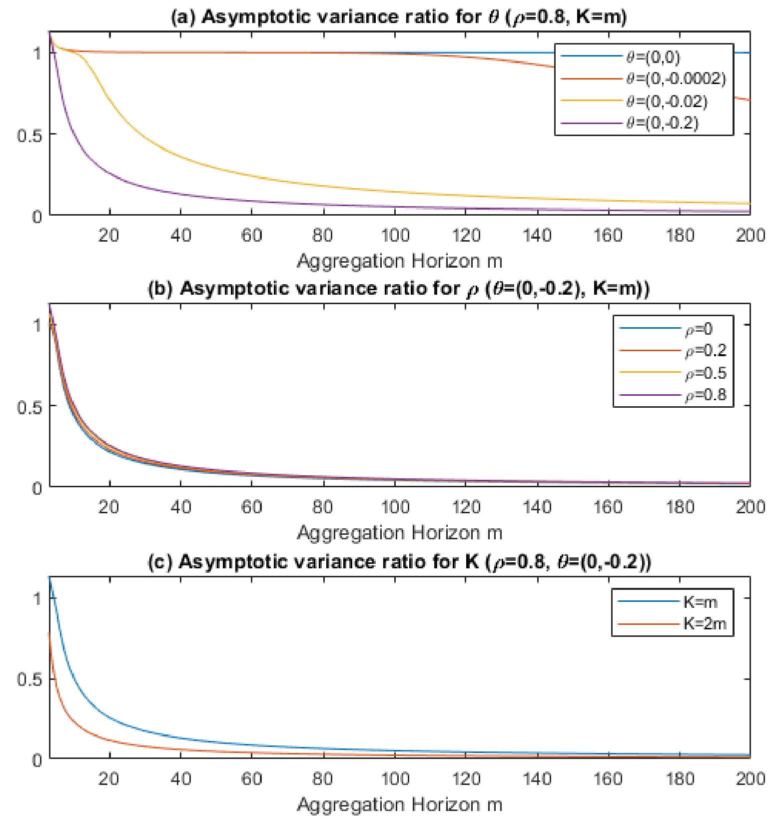

Figure 1.3: The Asymptotic Variance Ratio of MIDAS-NLS and LS Estimators for Models with a MA(1) Regressor

AR(1) process

Let high-frequency regressor $\{X_{i,t}^{(m)}\}$ follows a stationary AR(1) given by $X_{i,t}^{(m)} = \phi X_{i,t-1/m}^{(m)} + e_{i,t}^{(m)}$, where $e_{i,t}^{(m)} \sim IID(0, \sigma_e^2)$, then $X_{it}(\boldsymbol{\pi})$ and $X_{it}(\boldsymbol{\theta}^*)$ are given by

$$\begin{cases} X_{it}(\boldsymbol{\pi}) = \sum_{j=1}^{K} \pi_j \phi^{K-j} x_{i,t-(K-1)/m}^{(m)} + \sum_{j=1}^{K-1} \sum_{q=1}^{j} \pi_q \phi^{j-q} e_{i,t-(j-1)/m}^{(m)}, \\ X_{it}(\boldsymbol{\theta}^*) = \sum_{j=1}^{K-1} w_j(\boldsymbol{\theta}^*)(\phi^{K-j} - 1) x_{i,t-(K-1)/m}^{(m)} + \sum_{j=1}^{K-1} \sum_{q=1}^{j} w_q(\boldsymbol{\theta}^*) \phi^{j-q} e_{i,t-(j-1)/m}^{(m)}. \end{cases}$$

Proposition 1.3. *In the FE model, let high-frequency regressor $\{X_{i,t}^{(m)}\}$ follows an AR(1)* *process, the asymptotic bias of the LS estimator is given by*

$$ABias(\hat{\beta^*};\beta) = \gamma\beta, \tag{1.4.11}$$

where $\gamma = \frac{\gamma_1}{\gamma_2}$ *with* $\gamma_1 = tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})'))$ *and* $\gamma_2 = tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})'))$. *When* $K = m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$
$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m}\pi_j\phi^{m-j})[\sum\limits_{j=1}^{m-1}w_j(\boldsymbol{\theta}^*)(\phi^{m-j}-1)] \\ +\sigma_e^2\sum\limits_{j=1}^{m-1}(\sum\limits_{q=1}^{j}\pi_q\phi^{j-q})(\sum\limits_{q=1}^{j}w_q(\boldsymbol{\theta}^*)\phi^{j-q}) & \textit{if } r = k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m}\pi_j\phi^{m-j})[\sum\limits_{j=1}^{m-1}w_j(\boldsymbol{\theta}^*)(\phi^{m-j}-1)]\phi^{m(k-r)} \\ +\sigma_e^2[\sum\limits_{j=1}^{m-1}w_j(\boldsymbol{\theta}^*)(\phi^{m-j}-1)](\sum\limits_{j=1}^{m-1}\sum\limits_{q=1}^{j}\pi_q\phi^{m(k-r)-m+2j-q}) & \textit{if } r < k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m}\pi_j\phi^{m-j})[\sum\limits_{j=1}^{m-1}w_j(\boldsymbol{\theta}^*)(\phi^{m-j}-1)]\phi^{m(r-k)} \\ +\sigma_e^2(\sum\limits_{j=1}^{m}\pi_j\phi^{m-j})(\sum\limits_{j=1}^{m-1}\sum\limits_{q=1}^{j}w_q(\boldsymbol{\theta}^*)\phi^{m(r-k)-m+2j-q}) & \textit{if } r > k \end{cases}$$

and

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$
$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m}\pi_j\phi^{m-j})^2 + \sigma_e^2\sum\limits_{j=1}^{m-1}(\sum\limits_{q=1}^{j}\pi_q\phi^{j-q})^2 & \textit{if } r = k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m}\pi_j\phi^{m-j})^2\phi^{m|r-k|} + \sigma_e^2(\sum\limits_{j=1}^{m}\pi_j\phi^{m-j})(\sum\limits_{j=1}^{m-1}\sum\limits_{q=1}^{j}\pi_q\phi^{m|r-k|-m+2j-q}) & \textit{otherwise} \end{cases}$$

When $K = 2m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$

$$\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)] \\
+\sigma_e^2\sum_{j=1}^{2m-1}(\sum_{q=1}^{j}\pi_q\phi^{j-q})(\sum_{q=1}^{j}w_q(\boldsymbol{\theta}^*)\phi^{j-q}) & \text{if } r=k \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)]\phi+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{m+j}\pi_q\phi^{m+j-q}) \\
\times(\sum_{q=1}^{j}w_q(\boldsymbol{\theta}^*)\phi^{j-q})+\sigma_e^2(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j}w_q(\boldsymbol{\theta}^*)\phi^{m+2j-q}) & \text{if } r=k+1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)]\phi+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{j}\pi_q\phi^{j-q}) \\
\times(\sum_{q=1}^{m+j}w_q(\boldsymbol{\theta}^*)\phi^{m+j-q})+\sigma_e^2[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)](\sum_{j=0}^{m-1}\sum_{q=1}^{m+j}\pi_q\phi^{m+2j-q}) & \text{if } r=k-1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)]\phi^{m(r-k)} \\
+\sigma_e^2(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j}w_q(\boldsymbol{\theta}^*)\phi^{m(r-k)-2m+2j-q}) & \text{if } r>k+1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)]\phi^{m(k-r)} \\
+\sigma_e^2[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)](\sum_{j=1}^{2m-1}\sum_{q=1}^{j}\pi_q\phi^{m(k-r)-2m+2j-q}) & \text{if } r<k-1
\end{cases}$$

and

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$

$$\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})^2+\sigma_e^2\sum_{j=1}^{2m-1}(\sum_{q=1}^{j}\pi_q\phi^{j-q})^2 & \text{if } r=k \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})^2\phi+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{m+j}\pi_q\phi^{m+j-q}) \\
\times(\sum_{q=1}^{j}\pi_q\phi^{j-q})+\sigma_e^2(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j}\pi_q\phi^{m+2j-q}) & \text{if } r=k+1 \text{ or } r=k-1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})^2\phi^{m|r-k|} \\
+\sigma_e^2(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j}\pi_q\phi^{m|r-k|-2m+2j-q}) & \text{otherwise}
\end{cases}$$

The asymptotic variance ratio of the MIDAS-NLS estimator and the LS estimator is given by

$$\frac{AVar(\hat{\beta})}{AVar(\hat{\beta}^*)} = V_\beta / V_{\beta^*}, \tag{1.4.12}$$

where $V_{\beta^*} = \frac{\sigma_\varepsilon^2}{N}(tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')))^{-1}$ *and* $V_\beta = \frac{\sigma_\varepsilon^2}{N}(V_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V'}_{12})^{-1}$ *with*

$$\begin{aligned}
V_{11} =&\ tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')),\\
\boldsymbol{V}_{12} =&\ \begin{pmatrix} tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\boldsymbol{X}_i(\boldsymbol{\theta})')) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\boldsymbol{X}_i(\boldsymbol{\theta})')) \end{pmatrix},\\
\boldsymbol{V}_{22} =&\ \begin{pmatrix} tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_1})) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_1}))\\ tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_2})) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_2})) \end{pmatrix},
\end{aligned}$$

When $K = m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$
$$\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m} w_j(\boldsymbol{\theta})\phi^{m-j})^2 + \sigma_e^2\sum\limits_{j=1}^{m-1}(\sum\limits_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{j-q})^2 & \textit{if } r = k\\
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m} w_j(\boldsymbol{\theta})\phi^{m-j})^2\phi^{m|r-k|} & \\
+\sigma_e^2(\sum\limits_{j=1}^{m} w_j(\boldsymbol{\theta})\phi^{m-j})(\sum\limits_{j=1}^{m-1}\sum\limits_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{m|r-k|-m+2j-q}) & \textit{otherwise}
\end{cases}$$

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$
$$\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum\limits_{j=1}^{m} w_j(\boldsymbol{\theta})\phi^{m-j})\beta & \\
+\sigma_e^2\sum\limits_{j=1}^{m-1}(\sum\limits_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{j-q})(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta & \textit{if } r = k\\
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum\limits_{j=1}^{m} w_j(\boldsymbol{\theta})\phi^{m-j})\phi^{m(k-r)}\beta & \\
+\sigma_e^2(\sum\limits_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum\limits_{j=1}^{m-1}\sum\limits_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{m(k-r)-m+2j-q})\beta & \textit{if } r < k\\
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum\limits_{j=1}^{m} w_j(\boldsymbol{\theta})\phi^{m-j})\phi^{m(r-k)}\beta & \\
+\sigma_e^2(\sum\limits_{j=1}^{m} w_j(\boldsymbol{\theta})\phi^{m-j})(\sum\limits_{j=1}^{m-1}\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m(r-k)-m+2j-q})\beta & \textit{if } r > k
\end{cases}$$

and

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_s})_{kr} =$$

$$\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m-j})(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})\beta^2 & \\
+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{j-q})(\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta^2 & if\ r=k\\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m-j})(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})\phi^{m(k-r)}\beta^2 & \\
+\sigma_e^2(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m-j})(\sum_{j=1}^{m-1}\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m(k-r)-m+2j-q})\beta^2 & if\ r<k\\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m-j})(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})\phi^{m(r-k)}\beta^2 & \\
+\sigma_e^2(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum_{j=1}^{m-1}\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m(k-r)-m+2j-q})\beta^2 & if\ r>k
\end{cases}$$

When $K = 2m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$

$$\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})^2+\sigma_e^2\sum_{j=1}^{2m-1}(\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{j-q})^2 & if\ r=k\\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})^2\phi+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{m+j}w_q(\boldsymbol{\theta})\phi^{m+j-q}) & \\
\times(\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{j-q})+\sigma_e^2(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j}w_q(\boldsymbol{\theta})\phi^{m+2j-q}) & if\ r=k+1\ or\ r=k-1\\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})^2\phi^{m|r-k|} & \\
+\sigma_e^2(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{m|r-k|-2m+2j-q}) & otherwise
\end{cases}$$

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$
$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})\beta & \\ +\sigma_e^2\sum_{j=1}^{2m-1}(\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{j-q})(\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta & \mathit{if}\ r=k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})\phi\beta+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{m+j}w_q(\boldsymbol{\theta})\phi^{m+j-q}) & \\ \times(\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta+\sigma_e^2(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m+2j-q})\beta & \mathit{if}\ r=k+1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})\phi\beta+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m+j-q}) & \\ \times(\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{j-q})\beta+\sigma_e^2(\sum_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j}w_q(\boldsymbol{\theta})\phi^{m+2j-q})\beta & \mathit{if}\ r=k-1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})\phi^{m(r-k)}\beta & \\ +\sigma_e^2(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m(r-k)-2m+2j-q})\beta & \mathit{if}\ r>k+1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m}w_j(\boldsymbol{\theta})\phi^{2m-j})\phi^{m(k-r)}\beta & \\ +\sigma_e^2(\sum_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{m(k-r)-2m+2j-q})\beta & \mathit{if}\ r<k-1 \end{cases}$$

and

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_s})_{kr} =$$
$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\beta^2 & \\ +\sigma_e^2\sum\limits_{j=1}^{2m-1}(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{j-q})(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta^2 & \text{if } r=k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\phi\beta^2+\sigma_e^2\sum\limits_{j=1}^{m-1}(\sum\limits_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m+j-q}) & \\ \times(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{j-q})\beta^2+\sigma_e^2(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum\limits_{j=0}^{m-1}\sum\limits_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m+2j-q})\beta^2 & \text{if } r=k+1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\phi\beta^2+\sigma_e^2\sum\limits_{j=1}^{m-1}(\sum\limits_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m+j-q}) & \\ \times(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta^2+\sigma_e^2(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=0}^{m-1}\sum\limits_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m+2j-q})\beta^2 & \text{if } r=k-1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\phi^{m(r-k)}\beta^2 & \\ +\sigma_e^2(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum\limits_{j=1}^{2m-1}\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m(r-k)-2m+2j-q})\beta^2 & \text{if } r>k+1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\phi^{m(k-r)}\beta^2 & \\ +\sigma_e^2(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m-1}\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m(k-r)-2m+2j-q})\beta^2 & \text{if } r<k-1 \end{cases}$$

for all $s, l = 1, 2$.

Proof. See Appendix A.6. □

Proposition 1.4. *In the RE model, let high-frequency regressor* $\{X_{i,t}^{(m)}\}$ *follows an AR(1) process, the asymptotic bias of the LS estimator is given by*

$$ABias(\hat{\beta}^*;\beta) = \gamma\beta,$$

where $\gamma = \frac{\gamma_1}{\gamma_2}$ *with* $\gamma_1 = tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})'))$ *and* $\gamma_2 = tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})'))$. *The asymptotic variance ratio of the MIDAS-NLS estimator and the LS estimator is given* *by*

$$\frac{AVar(\hat{\beta})}{AVar(\hat{\beta^*})} = V_\beta / V_{\beta^*},$$

where $V_{\beta^*} = \frac{1}{N}(tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')))^{-1}$ *and* $V_\beta = \frac{1}{N}(V_{11} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{12}')^{-1}$, *with*

$$\begin{aligned}
V_{11} = &\ tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')), \\
\boldsymbol{V}_{12} = &\ \begin{pmatrix} tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_1}\boldsymbol{X}_i(\boldsymbol{\theta})')) & tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_2}\boldsymbol{X}_i(\boldsymbol{\theta})')) \end{pmatrix}, \\
\boldsymbol{V}_{22} = &\ \begin{pmatrix} tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \theta_1})) & tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \theta_1})) \\ tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \theta_2})) & tr(\boldsymbol{\Omega}^{-1}\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \theta_2})) \end{pmatrix},
\end{aligned}$$

where $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')$, $\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')$, *and* $\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \theta_s})$ *are given in Proposition 1.3.*

Again, we plot the figures for (1.4.11) and (1.4.12) to explore the role of $\boldsymbol{\theta}$, ρ, and K in the asymptotic properties of the estimators. Three cases are considered. In the first case, we let $\phi = 0.8$, $K = m$, and consider four values of $\boldsymbol{\theta} = \{(0,-0.2),(0,-0.02),(0,-0.0002),(0,0)\}$. In the second case, we let $\boldsymbol{\theta} = (0,-0.2)$, $K = m$, and consider three values of $\phi = \{0.2, 0.5, 0.8\}$. In the third case, we let $\phi = 0.8$, $\boldsymbol{\theta} = (0,-0.2)$, and consider two values of $K = \{m, 2m\}$. We let $\beta = 2$, $T = 10$, m ranges from 3 to 200, and $\boldsymbol{\pi}$ be the flat weights.

Figures 1.4-1.5 plot the asymptotic bias of the LS estimator and the asymptotic variance ratio of the MIDAS-NLS and LS estimators when the high-frequency regressor is an AR(1) process. Figure 1.4 shows that the bias of the LS estimator increases as m and ϕ increase, and it is the largest for the case of fast decaying weights with $K = 2m$. Consistent with Figure 1.2, Figure 1.4 again shows that the LS estimator is unbiased when the high-frequency regressor is i.i.d. (ϕ=0). Figure 1.5 suggests that the MIDAS-NLS estimator gains the largest efficiency gain for models with fast decaying weights in combination with $\phi = 0.2$ and $K = 2m$.

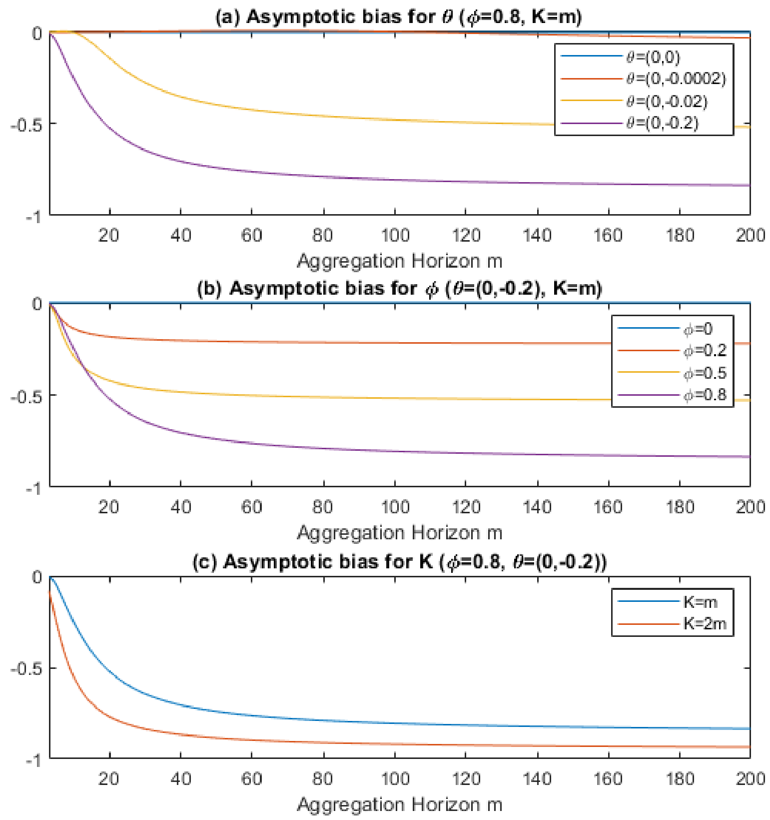

Figure 1.4: The Asymptotic Bias of LS Estimators for Models with an AR(1) Regressor

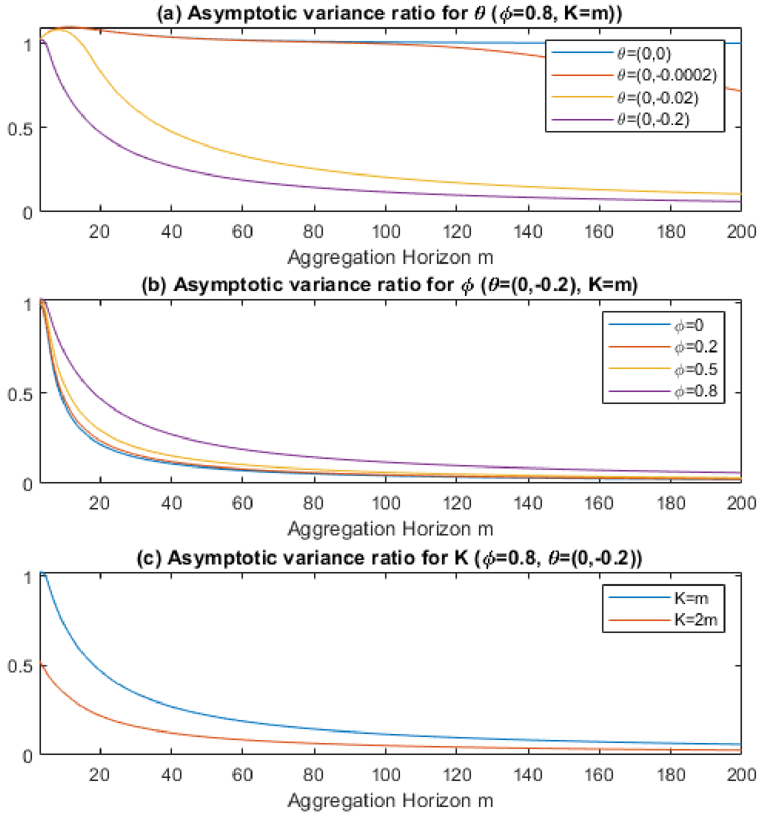

Figure 1.5: The Asymptotic Variance Ratio of MIDAS-NLS and LS Estimators for Models with an AR(1) Regressor

1.5 Finite Sample Properties of MIDAS-NLS and LS Estimators

The objective of the Monte Carlo simulation is to compare the finite sample properties of the LS estimator and the MIDAS-NLS estimator, as well as to assess the power of the hypothesis test of a flat weighting scheme. For simplicity, we consider the case of one high-frequency explanatory variable. The Monte Carlo simulation is based on the following DGP:

$$\begin{aligned} Y_{it} &= X_{it}(\boldsymbol{\theta})\beta + u_{it}, \\ u_{it} &= \alpha_i + \varepsilon_{it}, \end{aligned}$$

where $X_{it}(\boldsymbol{\theta}) = \sum_{k=1}^{K} w_k(\boldsymbol{\theta}) X_{i,t-(k-1)/m}^{(m)}$, $\varepsilon_{it} \mid (\boldsymbol{X}_i^{(m)}, \alpha_i) \sim NID(0, 0.25)$. We set $\beta = 2$ and investigate three aggregation horizons $m = \{12, 60, 120\}$, which correspond to the three aggregation schemes: the monthly-to-annual, the weekdaily-to-quarterly, the weekdaily-to-semiannual, respectively. We consider three values of $\boldsymbol{\theta} = \{(0, -0.2), (0, -0.02), (0, -0.0002)\}$ and two values of $K = \{m, 2m\}$. The various values of m, $\boldsymbol{\theta}$, and K, respectively, enable us to inspect how the LS estimator and the MIDAS-NLS estimator behave when the aggregation horizon, the aggregation weighting scheme, and lag order vary. Two experiments are conducted. In the FE model, the individual effects are assumed to be correlated with $X_{i,t}^{(m)}$ and are given by $\alpha_i = \sqrt{T} \bar{X}_i^{(m)}$, where $\bar{X}_i^{(m)}$ are the within-group means of $X_{i,t}^{(m)}$. In the RE model, α_i are uncorrelated with $X_{i,t}^{(m)}$ and are drawn from i.i.d. $N(0, 0.25)$. We consider three different high-frequency regressors:

- $X_{i,t}^{(m)} \sim NID(0, 1)$.
- $X_{i,t}^{(m)} = e_{i,t}^{(m)} + 0.8 e_{i,t-1/m}^{(m)}$, $e_{i,t}^{(m)} \sim NID(0, 1)$.
- $X_{i,t}^{(m)} = 0.8 X_{i,t-1/m}^{(m)} + e_{i,t}^{(m)}$, $e_{i,t}^{(m)} \sim NID(0, 1)$.

We consider sample sizes $N = \{500, 1000\}$ and $T = 10$. The Monte Carlo experiment is performed with 1,000 simulations. The estimators considered are (i) LS estimators with equal weights; (ii) MIDAS-NLS estimators. To assess the performance of these estimators, we report the following statistics:

- Average bias of MIDAS-NLS and LS estimators: $bias(\hat{\beta}) = \frac{\sum_{m=1}^{M} (\hat{\beta}_m - \beta)}{M}$ and $bias(\hat{\beta}^*) = \frac{\sum_{m=1}^{M} (\hat{\beta}_m^* - \beta)}{M}$, where M is the number of replications. $\hat{\beta}_m$ and $\hat{\beta}_m^*$, respectively, are the MIDAS-NLS estimator and the LS estimator of β in the mth repetition;
- Relative efficiency between MIDAS-NLS and LS estimators: $RMSE = MSE(\hat{\beta})/MSE(\hat{\beta}^*)$, where $MSE(\hat{\beta})$ and $MSE(\hat{\beta}^*)$ are the average MSE of

the MIDAS-NLS and LS given by $MSE(\hat{\beta}) = \sqrt{\frac{\sum_{m=1}^{M}(\hat{\beta}_m-\beta)^2}{M}}$ and $MSE(\hat{\beta}^*) = \sqrt{\frac{\sum_{m=1}^{M}(\hat{\beta}^*_m-\beta)^2}{M}}$, respectively;

- Power of non-flat weighs test: $M^{-1}\sum_{m=1}^{M}\mathbb{I}(W > \chi^2_{0.95,2})$, where $\mathbb{I}$ is the indicator function.

Tables 1.1-1.2 report the average bias of MIDAS-NLS and LS estimators when the high-frequency regressor is an i.i.d, a MA(1), and an AR(1) process for the FE model and the RE model. These results are consistent with the above analytical and numerical results for the asymptotic bias of the estimators. Specifically, the average bias of the MIDAS-NLS estimator lies in the vicinity of zero for all cases. Relative to the MIDAS-NLS estimator, the LS estimator is biased when $\{X^{(m)}_{i,t}\}$ follows a MA(1) or an AR(1) process with rapid and slow decaying weights. This bias tends to be larger with faster-decaying weights, a higher level of aggregation horizon, and a larger number of K, but it is not greatly affected by N. In particular, overall, the bias in the MIDAS model with an AR(1) regressor is more sensitive to changes in K and is greater than that in the MIDAS model with a MA(1) regressor. When $K = m$ and $K = 2m$, the largest bias for the case of an AR(1) regressor is about -0.8 and -0.9, respectively, while the largest bias for the case of a MA(1) regressor is -0.3 and -0.33, respectively.

Table 1.3 displays the efficiency of the MIDAS-NLS estimator in terms of the LS estimator, which we calculate as the MSE ratio. The results for the FE model and for the RE model are quite similar. Two points are noteworthy. First, in line with Figures 1.3-1.5, our results show that for MIDAS models with rapid and slow decaying weights, the MIDAS-NLS estimator is relatively more efficient than the LS estimator, irrespective of the high-frequency regressor. While for MIDAS models with near-flat weights, the MSE ratio between the estimators appears to stabilize at 1 for all $m = 12$, but it is smaller than 1 when m is relatively large (e.g. $m = 120$), which depends on the high-frequency regressor. Second, in all cases, the relative efficiency of the MIDAS-NLS estimator increases with faster decaying weights and higher values of aggregation horizons. This efficiency gain is larger in the case of an AR(1) regressor and in the case of $K = 2m$. In particular, the MIDAS-NLS estimator enjoys the largest efficiency gain with the corresponding MSE ratio of about 0.0040 in MIDAS models with an AR(1) regressor when $\boldsymbol{\theta} = (0, -0.2)$, $m = 120$, $K = 2m$, and $N = 1000$.

Table 1.4 reports the power of the hypothesis test for $H_0 : \boldsymbol{\theta} = \mathbf{0}$. As can be seen, the power of the test is greater than 5% for all cases, and it depends on the value of $\boldsymbol{\theta}$. For near-flat weights, the power increases with m, K, and N. As m becomes large (e.g. $m = 60$), the power tends to 1. While for rapid and slow decaying weights, the power is always 1. Overall, the results indicate the good power of our test for the flat weighting scheme.

1.6 Application: Okun's Law

In this section, we demonstrate the use of the proposed panel-MIDAS model with an Okun's law (1963) application. In particular, we examine the relationship between GDP growth rates and unemployment rates in the US metropolitan statistical areas (MSAs). The panel-MIDAS model with individual effects of Okun's law is as follows

$$\triangle GDP_{i,t} = c + \alpha_i + \beta \sum_{k=1}^{K} w_k(\boldsymbol{\theta}) \triangle U^{(12)}_{i,t-(k-1)/12} + \varepsilon_{i,t}, \tag{1.6.1}$$

where $\triangle GDP_{i,t}$ is the log difference of GDP in MSA i at time t. α_i are the MSA-specific effects. $\triangle U^{(12)}$ is the change in the log unemployment rate from the preceding period. This model is estimated using annual GDP from the Bureau of Economic Analysis and the monthly unemployment rate from the Federal Reserve Bank of St Louis website ($m = 12$). The analysis runs from 2004 to 2020 and covers 367 MSAs. We set $K = \{12, 24\}$, corresponding to 1-year and 2-year monthly lags. To identify the best model between the

Table 1.1: Average Bias of the MIDAS-NLS Estimator on Monte Carlo Simulation

Panel A: Fixed Effects Model

Weighting Scheme	$N = 500$						$N = 1000$					
	$K = m$			$K = 2m$			$K = m$			$K = 2m$		
	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$
IID process												
Rapid Decaying	0.0004	-0.0001	0.0005	-0.0004	0.0003	0.0006	-0.0001	-0.0005	0.0003	-0.0004	0.0006	-0.0001
Slow Decaying	-0.0003	0.0008	0.0007	0.0001	1.22E-05	0.0001	0.0004	0.0001	0.0015	-0.0004	-0.0002	0.0008
Near-Flat	0.0010	-0.0013	-0.0016	0.0018	0.0019	0.0079	-0.0007	0.0002	0.0003	-0.0005	-0.0008	0.0007
MA(1) process												
Rapid Decaying	0.0001	-0.0001	0.0001	0.0004	-0.0002	-1.17E-05	0.0004	0.0002	-0.0001	4.01E-05	-0.0001	0.0001
Slow Decaying	-0.0004	0.0005	-0.0002	0.0002	0.0004	4.45E-05	0.0001	-0.0001	0.0002	0.0002	0.0005	0.0001
Near-Flat	0.0001	0.0007	-0.0005	-0.0014	-0.0002	-0.0008	-0.0003	-0.0002	-0.0001	-0.0004	-0.0004	0.0001
AR(1) process												
Rapid Decaying	-0.0002	0.0003	-0.0005	0.0003	0.0001	-0.0001	-0.0001	4.60e-05	-0.0003	0.0002	0.0001	2.97E-05
Slow Decaying	-0.0001	-0.0001	-0.0001	0.0002	-0.0001	0.0001	1.78e-06	0.0002	1.48e-05	0.0001	-3.05E-05	-0.0001
Near-Flat	-0.0003	-0.0004	0.0010	0.0007	0.0006	0.0005	0.0003	-0.0002	-0.0002	-4.98E-05	0.0003	0.0004

Panel B: Random Effects Model

Weighting Scheme	$N = 500$						$N = 1000$					
	$K = m$			$K = 2m$			$K = m$			$K = 2m$		
	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$
IID process												
Rapid Decaying	0.0004	-0.0003	0.0002	0.0005	0.0002	0.0006	0.0004	0.0001	0.0005	0.0001	-3.20E-05	-0.0002
Slow Decaying	0.0002	0.0002	0.0013	-0.0002	0.0005	0.0006	-0.0006	0.0009	0.0010	0.0004	-0.0013	0.0001
Near-Flat	0.0001	-0.0007	0.0042	0.0025	-0.0008	0.0042	0.0001	-0.0016	-0.0022	-0.0014	-0.0018	0.0002
MA(1) process												
Rapid Decaying	0.0004	-0.0003	-0.0001	-0.0002	0.0003	-0.0001	-0.0001	0.0003	0.0001	0.0002	-0.0001	-0.0003
Slow Decaying	0.0008	0.0001	4.38E-05	-0.0009	0.0007	0.0010	-0.0001	-0.0001	0.0004	0.0003	0.0003	0.0003
Near-Flat	0.0006	-0.0008	-0.0017	-0.0001	0.0012	0.0027	0.0003	-0.0003	-0.0003	-0.0001	-0.0004	0.0004
AR(1) process												
Rapid Decaying	-0.0001	0.0002	-0.0001	-0.0003	-0.0001	0.0001	0.0001	4.47e-05	0.0003	-0.0001	-0.0001	0.0001
Slow Decaying	0.0003	0.0001	0.0003	-0.0001	2.29E-05	0.0001	-0.0001	-9.78e-07	-0.0001	0.0001	3.52E-06	-0.0003
Near-Flat	0.0001	-0.0009	0.0001	-0.0001	0.0001	0.0001	0.0002	-1.93e-05	0.0002	-1.91E-06	-0.0002	-0.0001

Table 1.2: Average Bias of the LS Estimator on Monte Carlo Simulation

Panel A: Fixed Effects Model												
Weighting Scheme	$N = 500$						$N = 1000$					
	$K = m$			$K = 2m$			$K = m$			$K = 2m$		
	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$
IID process												
Rapid Decaying	-0.0006	-0.0089	-0.0114	0.0005	0.0123	0.0123	0.0006	-0.0014	-0.0016	-0.0014	0.0083	0.0002
Slow Decaying	-0.0007	-0.0008	-0.0023	0.0003	0.0008	0.0072	0.0005	0.0011	0.0034	-0.0016	-0.0027	0.0117
Near-Flat	0.0010	-0.0015	-0.0012	0.0019	0.0017	0.0037	-0.0007	0.0001	0.0009	-0.0005	-0.0010	0.0005
MA(1) process												
Rapid Decaying	-0.2217	-0.2879	-0.3027	-0.2833	-0.3232	-0.3305	-0.2183	-0.2837	-0.3031	-0.2868	-0.3145	-0.3285
Slow Decaying	-0.0096	-0.0755	-0.0857	-0.0541	-0.0956	-0.1025	-0.0090	-0.0729	-0.0834	-0.0538	-0.0937	-0.0995
Near-Flat	0.0004	0.0022	-0.0019	-0.0006	-0.0023	-0.0069	-0.0001	0.0011	-0.0007	0.0004	-0.0007	-0.0053
AR(1) process												
Rapid Decaying	-0.3113	-0.7604	-0.8126	-0.6474	-0.8994	-0.9219	-0.3108	-0.7610	-0.8092	-0.6484	-0.8972	-0.9266
Slow Decaying	-0.0065	-0.4178	-0.4883	-0.2193	-0.5400	-0.5706	-0.0062	-0.4195	-0.4866	-0.2195	-0.5394	-0.5751
Near-Flat	0.0004	0.0092	-0.0010	0.0043	-0.0015	-0.0404	0.0010	0.0093	-0.0018	0.0034	-0.0017	-0.0405
Panel B: Random Effects Model												
Weighting Scheme	$N = 500$						$N = 1000$					
	$K = m$			$K = 2m$			$K = m$			$K = 2m$		
	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$
IID process												
Rapid Decaying	0.0010	-0.0067	0.0073	-0.0020	0.0067	0.0062	0.0016	-0.0033	0.0022	0.0007	-0.0046	-0.0007
Slow Decaying	0.0003	-0.0041	-0.0007	0.0008	0.0013	0.0077	-0.0005	0.0011	0.0007	0.0006	-0.0014	-0.0041
Near-Flat	4.48E-05	-0.0008	0.0039	0.0025	-0.0002	0.0069	-0.0001	-0.0017	-0.0028	-0.0014	-0.0021	-0.0015
MA(1) process												
Rapid Decaying	-0.2166	-0.2864	-0.2856	-0.2808	-0.3180	-0.3201	-0.2178	-0.2891	-0.2915	-0.2781	-0.3140	-0.3273
Slow Decaying	-0.0084	-0.0713	-0.0814	-0.0526	-0.0928	-0.0954	-0.0087	-0.0764	-0.0808	-0.0517	-0.0892	-0.0879
Near-Flat	0.0010	0.0001	-0.0007	0.0006	-0.0001	-0.0038	0.0006	0.0010	-0.0011	0.0007	-0.0012	-0.0041
AR(1) process												
Rapid Decaying	-0.3043	-0.7558	-0.8043	-0.6321	-0.8785	-0.9060	-0.3057	-0.7529	-0.8009	-0.6317	-0.8780	-0.9090
Slow Decaying	-0.0057	-0.4192	-0.4834	-0.2139	-0.5257	-0.5659	-0.0062	-0.4162	-0.4850	-0.2140	-0.5282	-0.5599
Near-Flat	0.0008	0.0084	-0.0019	0.0033	-0.0017	-0.0395	0.0008	0.0095	-0.0015	0.0034	-0.0021	-0.0410

Table 1.3: Relative Efficiency of Estimators (MIDAS-NLS/LS) on Monte Carlo Simulation

Panel A: Fixed Effects Model												
Weighting Scheme	$N = 500$						$N = 1000$					
	$K = m$			$K = 2m$			$K = m$			$K = 2m$		
	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$
IID process												
Rapid Decaying	0.2370	0.1019	0.0691	0.1713	0.0670	0.0489	0.2449	0.1061	0.0682	0.1597	0.0679	0.0452
Slow Decaying	0.7721	0.3080	0.2049	0.5850	0.2013	0.1368	0.7732	0.2959	0.2112	0.5647	0.2023	0.1327
Near-Flat	0.9998	0.9959	0.9526	0.9986	0.9645	0.7102	1.0013	0.9955	0.9697	0.9996	0.9781	0.7226
MA(1) process												
Rapid Decaying	0.0403	0.0296	0.0263	0.0309	0.0255	0.0217	0.0302	0.0226	0.0202	0.0230	0.0191	0.0171
Slow Decaying	0.5781	0.1486	0.1071	0.2444	0.1082	0.0781	0.5242	0.1185	0.1006	0.1881	0.0853	0.0703
Near-Flat	1.0019	0.9810	0.9086	1.0053	0.9479	0.6392	1.0078	0.9795	0.9115	0.9994	0.9621	0.6703
AR(1) process												
Rapid Decaying	0.0168	0.0067	0.0065	0.0082	0.0058	0.0054	0.0127	0.0051	0.0045	0.0059	0.0042	0.0040
Slow Decaying	0.4352	0.0168	0.0141	0.0346	0.0129	0.0121	0.4186	0.0115	0.0101	0.0238	0.0091	0.0083
Near-Flat	1.0372	0.7590	0.6878	0.9351	0.8609	0.3249	1.0203	0.6432	0.6452	0.8928	0.8314	0.2633
Panel B: Random Effects Model												
Weighting Scheme	$N = 500$						$N = 1000$					
	$K = m$			$K = 2m$			$K = m$			$K = 2m$		
	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$
IID process												
Rapid Decaying	0.2594	0.1010	0.0700	0.1674	0.0663	0.0450	0.2442	0.1005	0.0687	0.1658	0.0681	0.0479
Slow Decaying	0.7559	0.3175	0.2215	0.5500	0.2056	0.1333	0.7821	0.3026	0.2052	0.5668	0.2155	0.1397
Near-Flat	1.0006	0.9898	0.9398	0.9996	0.9781	0.7431	0.9998	0.9898	0.9565	1.0011	0.9874	0.7456
MA(1) process												
Rapid Decaying	0.0432	0.0300	0.0270	0.0331	0.0254	0.0225	0.0305	0.0217	0.0203	0.0239	0.0200	0.0181
Slow Decaying	0.5719	0.1517	0.1124	0.2479	0.1053	0.0837	0.5355	0.1147	0.1015	0.1917	0.0923	0.0749
Near-Flat	1.0045	0.9851	0.9181	0.9992	0.9603	0.6911	1.0087	0.9835	0.9155	0.9985	0.9519	0.6825
AR(1) process												
Rapid Decaying	0.0172	0.0067	0.0062	0.0086	0.0061	0.0056	0.0125	0.0048	0.0046	0.0060	0.0042	0.0041
Slow Decaying	0.4368	0.0163	0.0145	0.0338	0.0132	0.0120	0.4228	0.0118	0.0101	0.0234	0.0091	0.0087
Near-Flat	1.0522	0.7940	0.6546	0.9775	0.8431	0.3370	1.0458	0.6525	0.6849	0.9076	0.8194	0.2628

Table 1.4: Power of the Test $H_0 : \boldsymbol{\theta} = \mathbf{0}$

Panel A: Fixed Effects Model

Weighting Scheme	$N = 500$						$N = 1000$					
	$K = m$			$K = 2m$			$K = m$			$K = 2m$		
	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$
IID process												
Rapid Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Slow Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Near-Flat	0.112	1	1	0.451	1	1	0.176	1	1	0.746	1	1
MA(1) process												
Rapid Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Slow Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Near-Flat	0.232	1	1	0.894	1	1	0.358	1	1	0.998	1	1
AR(1) process												
Rapid Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Slow Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Near-Flat	0.463	1	1	1	1	1	0.682	1	1	1	1	1

Panel B: Random Effects Model

Weighting Scheme	$N = 500$						$N = 1000$					
	$K = m$			$K = 2m$			$K = m$			$K = 2m$		
	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$	$m = 12$	$m = 60$	$m = 120$
IID process												
Rapid Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Slow Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Near-Flat	0.094	1	1	0.418	1	1	0.159	1	1	0.682	1	1
MA(1) process												
Rapid Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Slow Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Near-Flat	0.177	1	1	0.865	1	1	0.322	1	1	0.994	1	1
AR(1) process												
Rapid Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Slow Decaying	1	1	1	1	1	1	1	1	1	1	1	1
Near-Flat	0.368	1	1	1	1	1	0.675	1	1	1	1	1

Table 1.5: Parameter Estimates of the Determinants of GDP Growth

Regressor	MIDAS Model	LS Model	Wald test	Relative R^2	Hausman test
$K = 12$	-0.08585***	-0.08837***	6.6552**	1.0065	1.2157
	(0.00319)	(0.00222)			
$K = 24$	-0.08278***	-0.10448***	28.5030***	1.4252	0.2523
	(0.00308)	(0.00324)			

Note: *** p<0.01, ** p<0.05, * p<0.1. The sample runs from 2004 to 2020 and includes 367 metropolitan statistical areas. Standard errors are reported in parentheses. The Wald test tests the null of the flat weighting scheme in the MIDAS model. The relative R^2 value is calculated as the R^2 ratio between the MIDAS model and the LS model. The Hausman tests test the null of the random-effects model.

fixed-effects MIDAS model and the random-effects MIDAS model, we perform a Hausman test. Besides, we also compare the empirical performance of the panel-MIDAS model with the simple LS model based on the flat weighting scheme.

The estimation results of the model (1.6.1) are presented in Table 1.5. The insignificance of the Hausman test points to a random-effects MIDAS model for $K = 12, 24$. In column 1, we present the estimated unemployment coefficients for the random-effects MIDAS model, and in column 2, we present the estimated unemployment coefficients for the standard random-effects model. We also calculate the Wald test statistics of the flat weighting scheme and the R-squared ratio between the MIDAS model and the LS model in the next two columns. Our results suggest that unemployment growth is always significantly and inversely associated with GDP growth, in line with Okun's law. However, when using the flat weighting scheme, the unemployment effect on GDP growth is overestimated by 2.94% and 26.21% for $K = 12$ and $K = 24$, respectively. The Wald test results reject the null hypothesis of the flat weighting scheme in the MIDAS model at the 1% significance level. Figure 1.6 shows that the polynomial weights of unemployment growth in the MIDAS model are in fact U-shaped. Besides, the relative R-squared values indicate that the MIDAS model has a better fit than the LS model, particularly when $K = 24$.

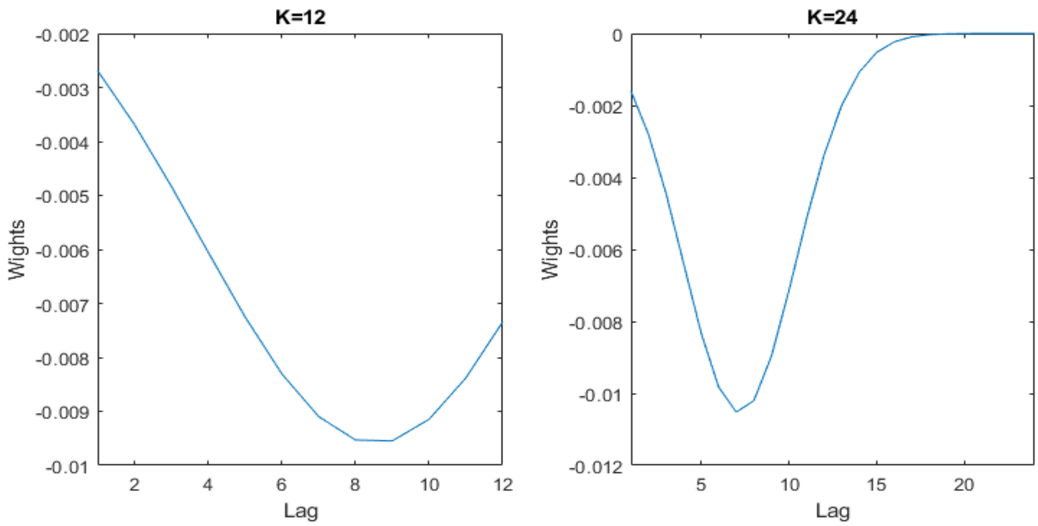

Figure 1.6: Polynomial Weights of Unemployment Growth on GDP Growth

1.7 Discussion

The study of Andreou, Ghysels, and Kourtellos (2010) considered the asymptotic distributions of the NLS estimator of a time-series MIDAS model. Their work largely motivated our work in a panel setup. This paper aims to derive the asymptotic properties of the MIDAS-NLS estimator and the LS estimator in the fixed-effects model and the random-effects models. We compare the finite sample properties of estimators via the Monte Carlo simulation. Our results reveal that the LS estimator is inefficient in the MIDAS model with an i.i.d regressor, and it is biased and inefficient in the MIDAS model with a MA(1) or an AR(1) regressor. Our empirical application revisits Okun's law and confirms the good performance of the panel-MIDAS model.

There remains an interesting challenge to be addressed. One main limitation of the proposed panel-MIDAS model is that it relies on the assumption that all individuals in a variable share the same weighting scheme. Our proposed MIDAS-NLS estimation is no longer applicable if the weighting scheme differs across individuals. Further work is required to develop a more general model by including an individual-specific weighting scheme. This remains a topic of ongoing research.

Chapter 2

The Effect of Economic Policy Uncertainty on Health Status

2.1 Introduction

Can economic policy uncertainty affect mortality? How long does this effect last? In this paper, we approach these main questions by studying the role of economic policy uncertainty in the variations in mortality rates. It is well recognized that one aim of implementing economic policies is to maintain and improve citizens' health status. However, the implementation of policies is often fraught with uncertainty about economic consequences. Evidence has shown that such policy-related uncertainty could in turn lead to increased suicide rates (Antonakakis and Gupta, 2017; Vandoros, Avendano, and Kawachi, 2019; Vandoros and Kawachi, 2021). Although economic policy uncertainty has received attention in the literature on suicide, no study, to my knowledge, has explored the impact of economic policy uncertainty on mortality, let alone gender and age differences in this effect.

In this paper, we fill this gap in the literature by examining the impact of economic policy uncertainty on sex- and age-specific mortality in the US. This analysis uses the historical index of economic policy uncertainty (EPU) to proxy the uncertainty over the scope and direction of future economic policy. Using time-series data from 1960 to 2013 in the US, we uncover a negative relationship between EPU and mortality rates of all sex and age groups. In particular, women are more sensitive to changes in EPU than men, with a greater elasticity of -0.2821 and a longer-lasting effect of 57 months. In comparing the magnitude of the EPU coefficient on mortality, we find that individuals aged 65-84 are the most sensitive to changes in EPU, while those aged 55-64 are the least sensitive to changes in EPU. A 1% point increase in EPU is predicted to decrease the mortality of 65-84-year-olds by 0.4410%, but the same growth in EPU is predicted to decrease the mortality of 55-64-year-olds by 0.1558%.

Previous studies of the relationship between economic conditions and health have shown that changes in health behaviors could be a driver of procyclical fluctuations in mortality (Ruhm, 1995, 2000). Motivated by this finding, we explore whether changes in health behaviors could also act as channels through which EPU affects mortality. Using the Behavioral Risk Factor Surveillance System (BRFSS) data for 1987-2013, our results reveal that individuals are more likely to follow a healthy lifestyle during high uncertainty periods. Specifically, as uncertainty mounts, individuals are more likely to reduce tobacco consumption, tend to exercise, and be in a healthy weight range.

This paper introduces the MIDAS model to identify the EPU effect on health outcomes and health behaviors. The model used in previous studies on macroeconomics and health is mostly limited to the case where variables are sampled at the same frequency, such as the fixed-effect model (Ruhm, 2000, 2015, 2016), the error correlation model (Laporte, 2004), and the VAR model (Nicolini, 2007). However, our data set comprises data at three different frequencies (monthly, quarterly, and annual). To make full use of high-frequency data, we use the MIDAS model (Ghysels, Santa-Clara, and Valkanov, 2004), which estimates a regression combining data with different frequencies, to conduct our analysis. Given that EPU may take a long time to affect mortality and health behaviors, another attractive feature of the MIDAS model is its ability to allow us to consider the long-term effects of EPU on health outcomes by controlling for several years' lagged terms of monthly EPU.

The rest of this paper is organized as follows. Section 2.2 summarizes the previous literature, and Section 2.3 introduces the data employed in this paper. Section 2.4 provides a conceptual model of the potential association between EPU and health. We then introduce the empirical methodology and present the estimation results in Sections 2.5 and 2.6. In Section 2.7, we run the robustness analysis, and the last section concludes.

2.2 Literature Review

Studies addressing the issue of the impact of macroeconomic conditions on health have received significant attention since Brenner's (1973, 1975, 1979, 1987, 2005) path-breaking work. Using aggregate time-series data, Brenner uncovers the counter-cyclical pattern of mortality. However, Brenner's studies have been widely criticized by other scholars who fail to replicate his results in other countries and periods (Marshall and Funch, 1979; Forbes and McGregor, 1984; Wagstaff, 1985; Laporte, 2004). Contrary to Brenner's results, Ruhm (2000) finds the opposite. Using a panel data model, he uncovers that an increase in the unemployment rate is associated with a decline in the total mortality rate of approximately 0.5% over the period 1972–1991 in the US. His following research also confirms the procyclicality of the total mortality rate (Ruhm, 2015, 2016; Heutel and Ruhm, 2016). Much subsequent research also confirms Ruhm's findings for other countries (Neumayer, 2004; Granados, 2005; Gerdtham and Ruhm, 2006; Tapia Granados and Ionides, 2017).

Apart from the association between economic conditions and health status, a growing number of studies have been conducted focusing on how health status responds to economic uncertainty. For example, Ferrie et al. (2005) investigate the role of uncertainty about job security on health among British white-collar civil servants. They suggest that self-reported job insecurity is a predictor of poor self-rated health and minor psychiatric morbidity. Similarly, using longitudinal data, Burgard, Brand, and House (2009) find that persistent perceived job insecurity is significantly and substantively associated with poorer self-rated health in the American's Changing Lives (ACL) and Midlife in the United States (MIDUS) samples, and it is a predictor of depressive symptoms among ACL samples. Antonakakis and Gupta (2017), to my knowledge, is the first study to explore the relationship between policy-related uncertainty and suicide mortality. Using data from 1950 to 2013 in the US, they find that economic policy uncertainty increases male suicide mortality in the youngest and oldest segments. Studies using a short-run analysis come to a similar conclusion suggesting a positive association between economic policy uncertainty and suicide in the US (Vandoros and Kawachi, 2021) and the UK (Vandoros, Avendano, and Kawachi, 2019).

From an economic perspective, uncertainty is closely related to the inability to assess the economy's current state and unpredictable future economic outlook. Higher uncertainty matters because it has a long-range impact on consumers and companies in many aspects. For example, in times of high uncertainty, consumers may reduce spending until uncertainty has declined and increase precautionary savings (Carroll, 1997; Bloom, 2014). Similarly, high uncertainty gives companies an incentive to postpone or cancel their investment and hiring decisions (Dixit, 1994; Leahy and Whited, 1996) since reversing investment and decisions is costly. Therefore, they prefer to wait until uncertainty falls and new information becomes available.

Recent studies suggest that economic policy uncertainty is counter-cyclical - rises in recessions and declines in booms. At the onset of an economic recession, policy-makers try to respond. They may implement some new policies that have not been tested, which may cause higher uncertainty since economic agents might be uncertain about whether these policies could work. For example, Baker, Bloom, and Davis (2012) find that the US EPU index rises with the implementation of the Troubled Asset Relief Program (TARP), which was a new program created in 2008 to address the subprime mortgage crisis. Using data from January 1985 to October 2013, Bloom (2014) also confirms the countercyclicality of the US EPU index - rising by 51% in recessions.

2.3 Data

The nation's health status is difficult to measure directly. Following previous studies (Behrman and Deolalikar, 1988; Ruhm, 2000; Brainerd, 2001; Stuckler et al., 2009, 2011; Nordentoft et al., 2013; Breuer, 2015; Case and Deaton, 2015), we use mortality rates as an indicator of health status.

Annual mortality data are collected from the Centers for Disease Control and Prevention (CDC), which contains age-adjusted total mortality rate, age-adjusted sex-specific mortality rate, and crude death rate counts in each specific stratum of age and sex per 100,000 inhabitants. Using the CDC mortality data, we construct dependent variables: age-adjusted total mortality rate, age-adjusted sex-specific mortality rates, and crude death rates of four age groups: 15-24, 25-54, 55-64, and 65-84-year-olds for 1960-2013.

The EPU index has been examined to be a good indicator to capture the perceived uncertainty about economic policy (Pástor and Veronesi, 2013). Recent studies for the US market also demonstrate that EPU (i) predicts future recessions (Karnizova and Li, 2014); (ii) affects asset prices (Brogaard and Detzel, 2015; Da, Engelberg, and Gao, 2015); and (iii) affects corporate investment (Kang, Lee, and Ratti, 2014; Wang, Chen, and Huang, 2014; Gulen and Ion, 2016). This paper uses the historical EPU index to proxy the economic policy uncertainty.

The US historical EPU index is constructed by Baker, Bloom, and Davis (2016) based on newspaper coverage frequency and is sampled at a monthly frequency. The measurement of US historical EPU is based on the information contained in leading US newspapers. There are two overlapping sets of newspapers employed to construct this index. From 1900 to 1984, the set is comprised of six newspapers: the Wall Street Journal, the New York Times, the Washington Post, the Chicago Tribune, the LA Times, and the Boston Globe. From 1985 onwards, the set is expanded to cover four additional newspapers: USA Today, the Miami Herald, the Dallas Morning Tribune, and the San Francisco Chronicle. To build the index, Baker, Bloom, and Davis rely on archives for these newspapers to search for the terms related to economic and policy uncertainty. These data can be collected from *www.policyuncertainty.com*.

Our analysis also experimented with including the quarterly personal consumption expenditures for services in health care (PCEHC), quarterly GDP growth rate, annual divorce rate, and annual fertility rate as control variables. Data on PCEHC and GDP growth rate are obtained from the Federal Reserve Economic Data (FRED), data on the divorce rate are obtained from the CDC, and data on the fertility rate are obtained from the World Bank. The sample period is from 1960 to 2013. Summary statistics for these data are presented in Table 2.1.

2.4 Conceptual Model

Figure 2.1 gives a first indication of the association between EPU and total mortality rates in the US from 1960 to 2013 using national annual data. Given that EPU is likely to take many years to have an effect on health outcomes, we examine both the contemporaneous and long-term effects of EPU on mortality by controlling for 1-year lagged, 4-year lagged, and 8-year lagged terms of monthly EPU.

In the top of Figure 2.1, the annual EPU is created by averaging the 1-year lagged monthly EPU; in the middle of Figure 2.1, the annual EPU is created by averaging the 4-year lagged monthly EPU; and in the bottom of Figure 2.1, the annual EPU is created

Table 2.1: Summary Statistics for Selected Variables

Variables	Mean	Std. Dev.
EPU	122.989	48.334
PCEHC	43.773	32.013
GDP growth rate	6.728	4.117
Divorce rate	4.002	0.915
Fertility rate	2.158	0.482
Type of Mortality Rate (per 100,000)		
Total	1006.576	189.905
Sex-specific		
Male	1258.076	249.001
Female	816.391	143.268
Age-specific		
15-24-year-olds	98.041	17.984
25-54-year-olds	289.268	63.322
55-64-year-olds	1267.764	300.211
65-84-year-olds	4099.930	722.634

by averaging the 8-year lagged monthly EPU. All variables have been detrended with a linear trend and standardized to have a mean of zero and a standard deviation of one. As can be seen, the figure suggests an inverse relationship between EPU and total mortality rates. A regression of standardized detrended mortality on 1 year's, 4 years', and 8 years' worth of lagged monthly EPU yields the EPU coefficient of -0.5098 with a robust standard error of 0.1191 (p=0.000), -0.6982 with a robust standard error of 0.1135 (p=0.000), and -0.8349 with a robust standard error of 0.0713 (p=0.000), respectively. As can be seen, the relationship between EPU and mortality becomes stronger as the number of lagged EPU increases.

The aim of this section is to explore how risky health behaviors respond to changes in EPU. As shown in Figure 2.2, there are four channels through which EPU affects health behaviors and health outcomes, which are detailed below.

Decreased Household Spending The first mechanism is that increased uncertainty induces households to increase precautionary savings, which itself may affect

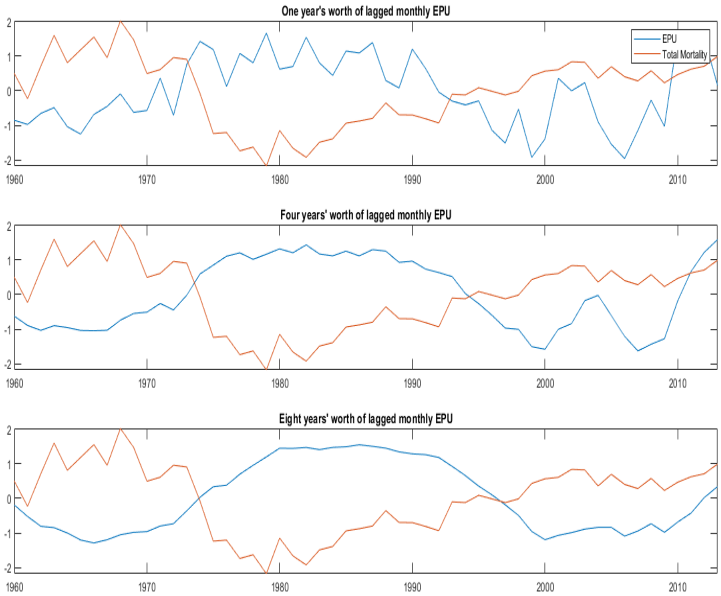

Figure 2.1: Detrended and Standardized EPU and Total Mortality Rates, 1960-2013

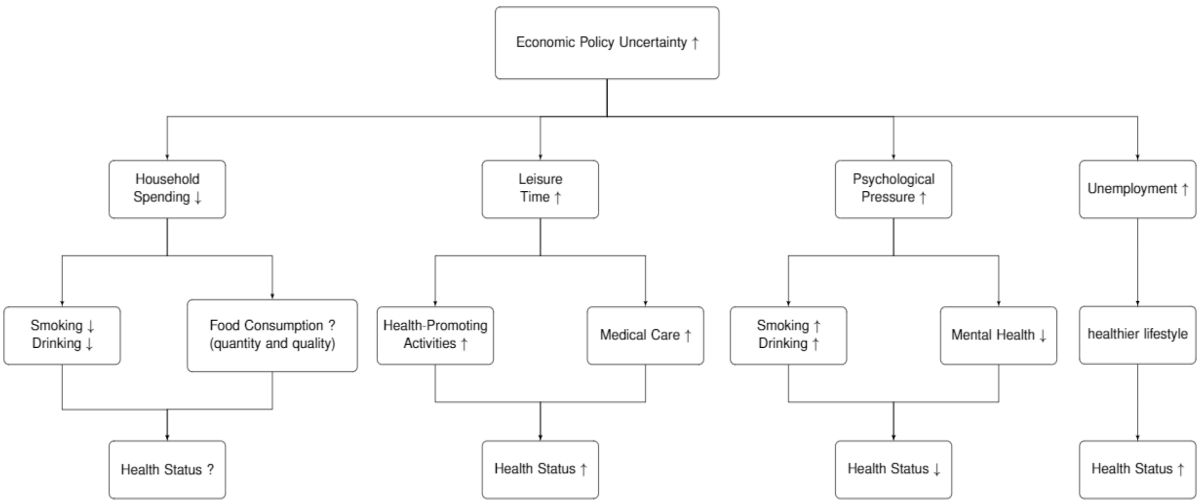

Figure 2.2: Flow Chart Mechanisms

householder's consumption decisions related to drinking, smoking, and food expenditure. For example, as budgets tighten, increased uncertainty would be associated with reduced alcohol and tobacco use. To slash food spending, individuals would also be more likely to cook at home instead of eating out. The British Nutrition Foundation suggests that homemade meals tend to have lower levels of fat, saturated fat, added sugars, and higher levels of fruit and vegetables compared with meals made outside of the home, indicating that increased uncertainty might reduce the prevalence of overweight and obesity due to lower intake of dietary fat and sugar.

However, on the other hand, decreased household spending might be associated with a decline in both the quantity and quality of food consumption. Assuming health-related foods are normal goods, consumers might reduce their spending on such goods on a tighter budget. Hence, in this mechanism, the effect of economic policy uncertainty on health is not clear.

Increased Leisure Time Economic policy uncertainty is associated with lower levels of activeness of economic activities (Bloom, 2014; Handley and Limao, 2015), leading to reduced business time and increased leisure time for workers. When leisure time increases, individuals would have more time for health-promoting activities (i.e., exercise) and more time to invest in medical care. As a result, the lifestyle would be more healthy during periods of heightened uncertainty, thereby enhancing health.

Increased Psychological Pressure Since increased economic policy uncertainty induces firms to reduce hiring decisions, higher uncertainty may lead to greater psychological pressure on workers who are directly involved in the labor market due to worsening employment positions, which may increase the use of alcohol and tobacco. As suggested by Cooper et al. (1992), stress-producing life situations exacerbate the consumption of alcohol. Hence, from this point, higher uncertainty might damage health.

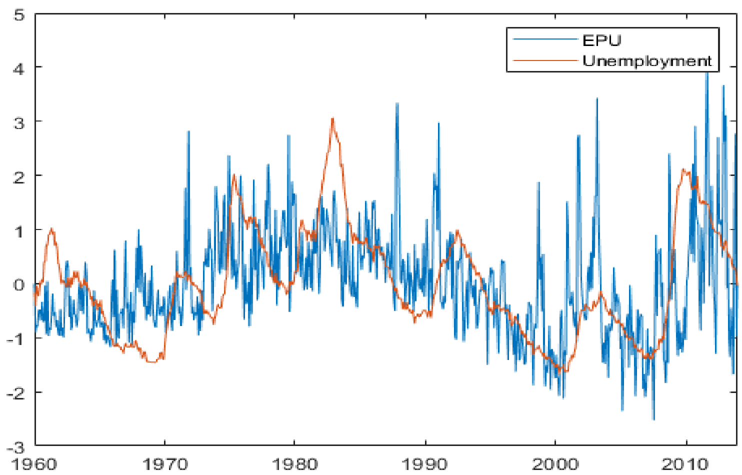

Figure 2.3: Detrended and Standardized EPU and Unemployment, 1960-2013

Increased Unemployment As mentioned, high uncertainty may boost unemployment since companies would postpone or cancel their hiring decisions when uncertainty increases. Figure 2.3 suggesting the positive association between EPU and unemployment[1] provides additional evidence for this finding. Plus, previous studies have demonstrated that individuals tend to have a healthier lifestyle during economic bad times (Ruhm, 1995, 2000, 2005; Dehejia and Lleras-Muney, 2004). Therefore, we may conclude that the increase in EPU can lead to a reduction in mortality through increased unemployment and a healthier lifestyle.

2.5 Empirical Framework

2.5.1 MIDAS Model

It is a rule of thumb that, in traditional regression models, the combination of dependent and independent variables with different frequencies is not allowed. When faced with mixed-frequency data, one typical strategy is to achieve the same frequency series by temporally aggregating high frequency to low frequency. For instance, to investigate the

[1]A regression of standardized detrended unemployment on EPU yields the EPU coefficient of 0.4023 with a robust standard error of 0.0377 (p=0.000).

relationship between annual suicide rates and monthly EPU, Antonakakis and Gupta (2017) create annual EPU by averaging monthly EPU. However, in this process, it will lose high-frequency information due to the loss of high-frequency data. A second strategy is to temporally interpolate low frequency to high frequency. However, the selection of appropriate interpolation methods is difficult. In practice, different interpolation methods can produce different values of high-frequency points. The use of an inappropriate method can result in distorted resulting high-frequency information.

To make the best use of high-frequency information, Ghysels, Santa-Clara, and Valkanov (2004) propose the MIDAS approach, which allows running a regression combining time series data sampled at different frequencies. Consider two time series $\{Y_t, t = 0, 1, 2, \ldots, T\}$ and $\{X_t^{(m)}, t = 0, 1/m, 2/m, \ldots, T-1/m, T\}$, which are available at low and high frequency, respectively. Suppose that Y_t is available at a low frequency (say, annual), called the reference interval, then the number m in superscript on $X_t^{(m)}$ denotes the times X_t (say, monthly or $m = 12$) is sampled in this reference interval. In other words, the left-hand side observation Y_t can be projected onto a history-lagged observation $X_{t-k/m}^{(m)}$, where $X_{t-k/m}^{(m)}$ represents the k-th high-frequency observation X we look into the past from the end-of-period observation. For example, in an annual/monthly example, if $k = 6$, $X_{t-6/12}^{(12)}$ denotes June's X-value at time t. Notation by the help of this example is provided below the table.

Notation		$t = 2013$, $m = 12$	
Y_t	Y_{2013}	X_t	$X_{2013,Dec}$
/	/	$X_{t-1/12}$	$X_{2013,Nov}$
/	/	$\vdots$	$\vdots$
/	/	$X_{t-11/12}$	$X_{2013,Jan}$
Y_{t-1}	Y_{2012}	X_{t-1}	$X_{2012,Dec}$

Theoretically, the basic idea behind the MIDAS model is to temporally aggregate high-frequency variables into low-frequency variables. The simple MIDAS model is presented as follows:

$$Y_t = \alpha + \beta W(L^{1/m}, \theta) X_t^{(m)} + \varepsilon_t, \tag{2.5.1}$$

where

$$W(L^{1/m}, \theta) = \sum_{k=1}^{K} w(k;\theta) L^{(k-1)/m},$$

where $w(k;\theta)$ is the lagged coefficient parameterized as a function of low-dimensional vector θ. $L^{1/m}$ is a lag operator such that $L^{k/m} X_t^{(m)} = X_{t-k/m}^{(m)}$. K denotes the maximum lag order of high-frequency observation and can be shorter or greater than m. ε_t is the error term.

In the specification (2.5.1), the number of the lags of $X_t^{(m)}$ is possibly large. For example, if the annual sampling frequency data Y_t is affected by 3 years' worth of lagged monthly $X_t^{(m)}$'s, 3×12 ($K = 36$) lagged coefficients need to be estimated. A solution to address this parameter proliferation is to use a known function $w(k;\theta)$ to capture the weights of lagged variables so that the slope coefficient β_i would capture all effects of $X_t^{(m)}$'s on Y_t.

Next, we discuss two parameterizations of $w(k;\theta)$ with only two parameters $\theta = [\theta_1; \theta_2]$. The first one is known as the "Exponential Almon Lag" defined as

$$w(k;\theta) = \frac{e^{\theta_1 k + \theta_2 k^2}}{\sum_{k=1}^{K} e^{\theta_1 k + \theta_2 k^2}}. \tag{2.5.2}$$

The second one is known as the "Beta Lag" defined as

$$w(k;\theta) = \frac{f(\frac{k}{K}, \theta_1; \theta_2)}{\sum_{k=1}^{K} f(\frac{k}{K}, \theta_1; \theta_2)}, \tag{2.5.3}$$

where

$$f(a, \theta_1; \theta_2) = \frac{(a)^{\theta_1 - 1}(1-a)^{\theta_2 - 1}\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)}, \ \Gamma(\theta) = \int_0^\infty e^{-x} x^{\theta - 1} dx.$$

In specifications (2.5.2) and (2.5.3), weights add up to 1, i.e. $\sum_{k=1}^{K} w(k;\theta) = 1$. The usage of the polynomial functions enables us to fit a large number of lags with only two parameters and obtain more degrees of freedom. Ghysels, Santa-Clara, and Valkanov (2004) point out that even though the Exponential Almon polynomial function and the Beta polynomial function depend on only two parameters, they are still flexible enough to take various weighting shapes (Figure 2.4). In the remainder of this paper, we use the MIDAS model with Beta lag polynomials to conduct our analysis.

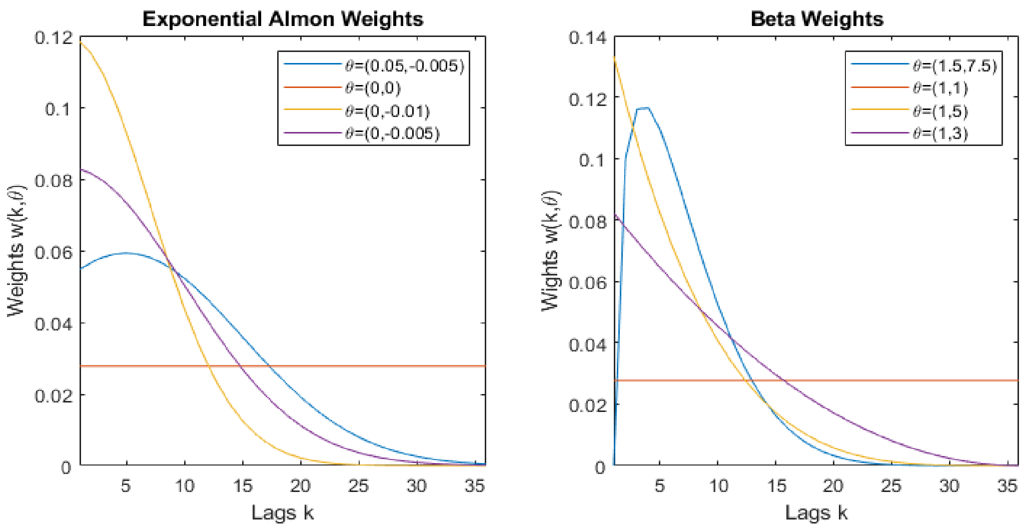

Figure 2.4: Weighting Functions for Exponential Almon Polynomials and Beta Polynomials

2.5.2 Estimation of EPU Effects on Health Status

Using the MIDAS model, our baseline equation to capture the effect of EPU on mortality rates is organized as follows:

$$\begin{aligned} M_t = \;& \alpha_1 + \alpha_2 \sum_{k=1}^{K_m} w(k;\theta_m) E^{(12)}_{t-(k-1)/12} + \alpha_3 \sum_{k=1}^{K_q} w(k;\theta_{q_1}) P^{(4)}_{t-(k-1)/4} \\ & + \alpha_4 \sum_{k=1}^{K_q} w(k;\theta_{q_2}) Growth^{(4)}_{t-(k-1)/4} + \alpha_5 Div_t + \alpha_6 Fer_t + \varepsilon_t, \end{aligned} \tag{2.5.4}$$

where M is the log mortality rate, $E^{(12)}$ is the log EPU, $P^{(4)}$ is the log PCEHC, $Growth^{(4)}$ is the GDP growth rate, Div is the divorce rate, and Fer is the fertility rate. The data set is comprised of three different frequency data shown below the table.

Frequency	Variable
Annual	Mortality rates, Divorce rate, Fertility rate
Quarterly	PCEHC ($m = 4$), GDP growth rate ($m = 4$)
Monthly	EPU ($m = 12$)

In the model (2.5.4), K_m refers to the lag order of EPU, and K_q refers to the lag order of PCEHC and GDP growth rates. To make our work easier to undertake, we let PCEHC and GDP growth rates have the same lag orders, and we restrict our analysis to the number of K_m and K_q ranging from 12-96 and 4-32, respectively. The optimal lag determination is based on the Akaike Information Criterion (AIC). Besides, the computation is conducted by Matlab[2]. To compare the estimation results of the MIDAS model and the simple OLS model, we convert monthly and quarterly data to their annual averages with the same lag horizon as the MIDAS.

2.5.3 Estimation of EPU Effects on Risky Health Behaviors

We next use US microdata to explore the EPU effects on risky health behaviors. This accompanying analysis aims to show that changes in individual lifestyles are consistent with fluctuations in mortality and covers the period 1987-2013. Microdata are obtained from the Behavioral Risk Factor Surveillance System (BRFSS), which is conducted by the Centers for Disease Control and Prevention (CDC). The BRFSS collects data annually about US adults' information on health-related risk behaviors, preventive health practices, and chronic health conditions.

The outcomes studied include the following: alcohol and cigarette use, physical activity, and Body Mass Index (BMI). Alcohol consumption is analyzed using a dummy variable indicating whether or not respondents are current drinkers. In our analysis, respondents who reported having consumed at least one alcoholic beverage within the past 30 days are

[2] We gratefully acknowledge the help of Eric Ghysels who provided his MATLAB code for MIDAS models. See Eric Ghysels' homepage at http://eghysels.web.unc.edu/ for links.

Table 2.2: Summary Statistics of Risky Heath Behaviors

Variables	Mean	Std. Dev.	Variables	Mean	Std. Dev.	Variables	Mean	Std. Dev.
Current Drinker			Physical Inactivity			Overweight or Obese		
Both sexes	0.5285	0.0169	Both sexes	0.2715	0.0307	Both sexes	0.5503	0.0718
Male	0.6094	0.0190	Male	0.2488	0.0321	Male	0.6334	0.0643
Female	0.4532	0.0183	Female	0.2927	0.0294	Female	0.4694	0.0778
Current Smoker			Overweight			Obese		
Both sexes	0.2124	0.0261	Both sexes	0.3549	0.0158	Both sexes	0.1953	0.0587
Male	0.2348	0.0241	Male	0.4356	0.0127	Male	0.1976	0.0617
Female	0.1915	0.0289	Female	0.2764	0.0229	Female	0.1930	0.0558

defined as current drinkers. Following the BRFSS measure of smoking status, we create a dummy variable indicating whether or not respondents are current smokers based on their responses to the questions: "Have you smoked at least 100 cigarettes in your entire life?" and "Do you now smoke cigarettes every day, some days, or not at all?". Respondents who answered "yes" to the first question and "every day or some days" to the second question are defined as current smokers. The BRFSS survey includes one core question related to physical activity: "During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?". We create a dummy variable to estimate self-reported physical inactivity that is defined as responding "no" to this question. Based on the BMI data, dummy variables indicating whether respondents are overweight or obese are created. Overweight is defined as a BMI of 25.00 to 29.9, and obesity is defined as a BMI of 30.0 to 99.8. Respondents with missing, don't know and refused answers are excluded from the analysis.

Using these dummy variables, we construct dependent variables in each sex grouping (male, female, or pooled) for 1987-2013, weighting individual observations by their BRFSS design weight: the prevalence of current drinkers, the prevalence of current smokers, the prevalence of physical inactivity, the prevalence of overweight, the prevalence of overweight or obesity, and the prevalence of obesity. Summary statistics of risky health behaviors are reported in Table 2.2.

The MIDAS model to study the EPU effects on risky health behaviors is organized as follows:

$$\begin{aligned} HB_t = \; & \alpha_1 + \alpha_2 \sum_{k=1}^{K_m} w(k;\theta_m) E^{(12)}_{t-(k-1)/12} + \alpha_3 \sum_{k=1}^{K_q} w(k;\theta_{q_1}) P^{(4)}_{t-(k-1)/4} \\ & + \alpha_4 \sum_{k=1}^{K_q} w(k;\theta_{q_2}) Growth^{(4)}_{t-(k-1)/4} + \alpha_5 Div_t + \alpha_6 Fer_t + \varepsilon_t, \end{aligned} \tag{2.5.5}$$

where HB refers to the log risky health behaviors, other notations are the same as those described in the model (2.5.4). Again, we let PCEHC and GDP growth rates have the same lag orders, and we restrict the number of K_m and K_q changing from 12-96 and 4-32, respectively. The optimal lag determination is based on the AIC.

2.6 Empirical Results

2.6.1 Unit Root Test

Theoretically, one of the assumptions in a time series regression is that variables are stationary. Otherwise, it may lead to a so-called spurious regression. Newbold and Granger (1974) point out that two non-stationary variables are spuriously related owing to the fact that they are both trended. In this case, estimators and test statistics may be misleading. Therefore, it is necessary to test whether a series is stationary or not before estimation. This objective can be attained by the augmented Dickey-Fuller (ADF) unit root test - if a time series possesses a unit root, the series is not stationary.

The common solution to deal with the non-stationary series is to difference the series until stationarity is achieved. However, in the process, long-run information might be discarded, and only the short-run model is estimated. A real breakthrough came with the concept of cointegration, a term introduced by Engle and Granger (1987). Consider two independent I(1) variables, Y_t and X_t. If there exists a certain value β such that $Y_t - \beta X_t$ is stationary. Then we say Y_t and X_t are cointegrated with cointegrating vector

Table 2.3: ADF Unit Root Test Statistic on Variables in the Model (2.5.4)

Variables	At Level	First Difference	Orders of Integration
EPU	-1.7063	-10.6600***	I(1)
PCEHC	-1.2713	-2.8164*	I(1)
GDP growth rate	-1.6902	-6.3731***	I(1)
Divorce rate	-3.0662**		I(0)
Fertility Rate	-4.5432***		I(0)
Type of Mortality			
Total	-0.3538	-8.6387***	I(1)
Sex-specific			
Male	0.8759	-8.2760***	I(1)
Female	-1.2121	-8.4210***	I(1)
Age-specific			
15-24-year-olds	1.0961	-5.1715***	I(1)
25-54-year-olds	-1.0821	-3.5189**	I(1)
55-64-year-olds	-0.5944	-3.4410**	I(1)
65-84-year-olds	0.0865	-2.6049*	I(1)

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The ADF test tests the null of a unit root. All test regressions contain a constant term. A time series is said to be integrated of order d, denoted I(d), if its d-th difference is stationary.

$(1, -\beta)'$. If the data are cointegrated, then we can estimate a model using the level of the data. The cointegrating regression enables us to capture the long-run equilibrating relationship between variables. One of the most popular tests for cointegration based on a single equation is the EG residual-based ADF test (Engle and Granger, 1987). This test first obtains the residuals from the regression and then uses the ADF test to examine whether the residuals are stationary. If the residuals are stationary, then the variables are cointegrated.

Since mortality rates, EPU, PCEHC, and GDP growth rate in the model (2.5.4) are non-stationary in levels (Table 2.3), it is necessary to test the presence of unit root on their residuals to ensure that there is no spurious regression.

2.6.2 Total Mortality Rates

Column 1 of Table 2.4 summarizes estimation results of models where total mortality rates are dependent variables. The simple OLS model, in column 1a, suggests the inverse association between total mortality rates and EPU. Specifically, a 1% point increase in EPU is significantly associated with a 0.2666% decrease in total mortality rates. The Engle-Granger residual-based ADF test also finds strong evidence of linear cointegration, implying that there is a long-run relationship between EPU and total mortality. As expected, the increase in PCEHC reduces total mortality rates, while the increase in divorce rates increases total mortality rates. A 1% point increase in PCEHC is predicted to reduce the total mortality rate by 0.1095%, while the same growth in divorce rates is predicted to increase the total mortality rate by 0.2327%. However, there is no evidence that the GDP growth rate and the fertility rate are significantly associated with total mortality rates.

In column 1b, we present estimation results using the MIDAS model where a sufficient number of lagged values of EPU is included in the model. Consistent with the results of the simple OLS model, EPU is negatively correlated with total mortality rates. The use of the MIDAS model witnesses the decreased EPU effect on total mortality rates. In the OLS model, a 1% point increase in EPU reduces predicted mortality by 0.2666%; this decreases to 0.2568% when using the MIDAS model. Among other control variables, when using the MIDAS model, we see an increase in the estimated PCEHC effect, the attenuation of the estimated divorce effect, and a statistical significance of the GDP growth rate coefficient. The optimal lag order of EPU is 57, suggesting that the effect of EPU on mortality can last for over 4 years. The R^2 values suggest that the MIDAS model has a better fit than the simple OLS model.

Table 2.4: Parameter Estimates of the Determinants of Total and Sex-specific Mortality

Variables	(1) Total		(2) Male		(3) Female	
	(a)	(b)	(a)	(b)	(a)	(b)
EPU	-0.2666***	-0.2568***	-0.2264***	-0.1860***	-0.2819***	-0.2821***
PCEHC	-0.1095***	-0.1138***	-0.1713***	-0.1726***	-0.0722***	-0.0717***
GDP growth	-0.0077	-0.0102**	-0.0158**	-0.0139**	-0.0055	-0.0034
Divorce	0.2327***	0.2235***	0.3595***	0.3151***	0.1381***	0.1628***
Fertility	0.0178	0.0259	-0.0040	0.0176	0.0162	0.0288
K_m		57		33		57
K_q		32		25		31
R^2	0.9824	0.9864	0.9760	0.9809	0.9828	0.9858
EG test	0.0018***	0.0001***	0.0019***	0.0008***	0.0005***	0.0001***

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variables are log age-adjusted total mortality rates and sex-specific mortality rates. Log EPU, log PCEHC, GDP growth rate, log divorce rate, and fertility rate are independent variables. Column (a) reports the estimation results of simple OLS models based on the flat weighting scheme but with the same lag horizon as the MIDAS, and column (b) reports the estimation results of MIDAS models (2.5.4). K_m is the lag order of EPU ranging from 12-96, and K_q is the lag order of PCEHC and GDP growth rate ranging from 4-32. The p-value of the Engle-Granger (EG) residual-based ADF test indicates that estimation results are not spurious.

2.6.3 Sex-specific Mortality Rates

Next, we examine whether men and women respond differently to changes in economic policy uncertainty. Columns 2-3 of Table 2.4 reports the results of this analysis using two different models. The first (column a) is the simple OLS model, and the second (column b) is the MIDAS model. The EG test results indicate a long-run relationship between EPU and sex-specific mortality rates in these two models.

In the simple OLS model (column a), we observe a greater EPU effect for women than men. A 1% point increase in EPU predicted a 0.2264% reduction in male mortality rates and a 0.2819% reduction in female mortality rates. An increase in PCEHC is significantly associated with a decline in mortality, with the greatest impact on men. A 1% point increase in PCEHC is predicted to reduce the mortality rate of males by 0.1713%, which compares to a decrease of 0.0722% for females. A one-point increase in GDP growth rate is estimated to reduce male mortality by 1.58%, but it is insignificantly associated with female mortality. The result related to divorce rates is in line with previous studies indicating that men are more vulnerable to the adverse effects of divorce (Shor et al., 2012; Antonakakis and Gupta, 2017). Consistent with the results for total mortality above, fertility rates do not have a significant effect on mortality.

In the MIDAS model (column b), based on the magnitude of the EPU coefficients and the persistence of the EPU impact, our results suggest that women are more sensitive to changes in EPU than men. Specifically, the elasticity for female mortality is -0.2821, which is 1.5 times as large for male mortality (-0.1860), and the optimal lag order of EPU for female mortality is 57, which is 2 years more than that for male mortality. Among other control variables, when using the MIDAS model, we see a slight change in the estimated PCEHC effect on mortality, the attenuation of the estimated GDP growth effect and divorce effect on male mortality, and the increase in the estimated divorce effect on female mortality.

2.6.4 Age-specific Morality Rates

Table 2.5 reports information on age-specific mortality of four age groups: 15-24, 25-54, 55-64, and 65-84-year-olds. In the simple OLS model (column a), EPU is negatively correlated with mortality rates of all age groups, with the greatest impact for individuals aged 65-84-year-olds. A 1% point increase in EPU is expected to decrease the mortality rate of 65-84-year-olds by 0.4589%, which compares to decreases of 0.4156%, 0.3628%, and 0.1227% for 15-24, 25-54, and 55-64-years-olds. PCEHC has the expected effect on mortality, with higher PCEHC being associated with lower mortality of all age groups, although the association is not significant for the mortality of 15-24-year-olds. A one-point increase in GDP growth is predicted to reduce the mortality of 25-64-year-olds by 2.0%-3.0%. The divorce rate is positively correlated with mortality rates of 15-24 and 55-84-year-olds. This conclusion is in line with previous literature indicating that divorce is accompanied by a wide range of poor health outcomes (Burgoa et al., 1998;

Sbarra, Law, and Portley, 2011; Antonakakis and Gupta, 2017). Interestingly, divorce is found to have no significant effect on mortality rates of prime-working-age individuals (25-54-year-olds). The possible explanation can be summarized as follows. People of those ages have the highest participation rate in the labor force, whereby they are most likely to end up not getting enough time for themselves and suffer much more work-related stress, being divorced helps them escape from responsibilities that come with marriage and give them more time for themselves in socializing and relaxing, thereby contributing to health improvement. When the health improvements resulting from divorce are offset by the health deterioration caused by divorce, one may expect an insignificant effect of divorce on mortality. The results related to the fertility rate also suggest that an increase in the fertility rate is significantly associated with a decline in mortality rates of 25-54-year-olds. This finding supports the existing literature suggesting that parity is the protection against several diseases (see, Vecchia and Franceschi, 1991; Erlandsson et al., 2002).

Our results in the MIDAS model (column b) suggest that the EPU effect on mortality lasts the longest for 15-54-year-olds, with a lag order of 93, while the EPU effect on mortality is greatest for 65-84-year-olds, with an elasticity of -0.4410. Besides, using the MIDAS model increases the significance of the (negative) PCEHC coefficient on mortality rates of 15-24-year-olds. Again, the EG residual-based ADF test results uncover the long-run relationship between mortality and EPU, and the higher value of R^2 suggests that the usage of the MIDAS model significantly improves estimation accuracy.

2.6.5 Risky Health Behaviors

Table 2.6 reports the marginal effects of EPU on risky health behaviors. These findings are consistent with the conceptual model in Section 2.4 and possibly explain the health improvements that accompany economic uncertainty.

There is no statistically significant relationship between EPU and the prevalence of

Table 2.5: Parameter Estimates of the Determinants of Age-specific Mortality

Variables	15-24 Year Olds		25-54 Year Olds	
	(a)	(b)	(a)	(b)
EPU	-0.4156***	-0.3812***	-0.3628***	-0.2300***
PCEHC	-0.0599	-0.0771**	-0.1279***	-0.1436***
GDP growth	0.0003	-0.0059*	-0.0281***	-0.0238***
Divorce	0.6014***	0.5939***	0.0007	-0.0317
Fertility	0.0176	0.0320	-0.1544***	-0.0711**
K_m		93		93
K_q		17		29
R^2	0.9283	0.9421	0.9722	0.9756
EG test	0.0000***	0.0000***	0.0001***	0.0008***
Variables	55-64 Year Olds		65-84-Year-Olds	
	(a)	(b)	(a)	(b)
EPU	-0.1227**	-0.1558***	-0.4589***	-0.4410***
PCEHC	-0.2485***	-0.2416***	-0.0346*	-0.0379**
GDP growth	-0.0230***	-0.0282***	-0.0020	-0.0021
Divorce	0.3754***	0.3508***	0.4007***	0.4302***
Fertility	0.0247	0.0156	0.0228	0.0453
K_m		32		57
K_q		31		15
R^2	0.9807	0.9870	0.9717	0.9767
EG test	0.0230**	0.0019***	0.0000***	0.0002***

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Dependent variables are log age-specific mortality rates. Log EPU, log PCEHC, GDP growth rate, log divorce rate and fertility rate are independent variables. Column (a) reports the estimation results of simple OLS models based on the flat weighting scheme but with the same lag horizon as the MIDAS, and column (b) reports the estimation results of MIDAS models (2.5.4). K_m is the lag order of EPU ranging from 12-96, and K_q is the lag order of PCEHC and GDP growth rate ranging from 4-32. The p-value of the Engle-Granger residual-based ADF test indicates that estimation results are not spurious.

current drinkers at the 10% significance level. According to Figure 2.2, this may imply that the decrease in alcohol consumption due to reduced household expenditures may be offset by the increase in alcohol consumption caused by increased work-related stress, leading to no significant effect of EPU on the prevalence of current drinkers.

The EPU coefficients are negative for the prevalence of current smokers. A 1% point increase in EPU is significantly associated with a 0.7727% and 0.7886% decrease in the prevalence of current smokers for males and females. A possible explanation is that as uncertainty mounts, individuals may curtail their spending on tobacco consumption to increase precautionary savings. Since tobacco use is the leading risk for preventable death in the US and causes about one-fifth of deaths each year, decreased current smoker prevalence is likely to indicate better health.

Overweight and obesity together are the second leading cause of preventable death in the US. Thus, death rates might decrease in times of high uncertainty periods because individuals are more likely to be in the healthy weight range. Specifically, a 1% point increase in EPU is predicted to reduce the prevalence of overweight by 0.2213%, the prevalence of overweight or obesity by 0.3560%, and the prevalence of obesity by 0.4288%.

The prevalence rate of overweight and obesity may decrease since people are likely to engage in physical activities. A 1% point increase in EPU lowers the expected prevalence of physical inactivity by a statistically significant 0.4290%, 0.4375%, and 0.4527% for both sexes, males, and females, respectively. According to the conceptual model, this may be due to increased leisure time, making it possible to undertake more health-producing activities.

2.7 Robustness Check

In this section, we conduct additional robustness checks on our results in Table 2.4. Since the weighting function is a key in the MIDAS model, the choice of it may affect

Table 2.6: EPU Coefficients on Risky Health Behaviors

Gender	Current Drinker	Current Smoker	Physical Inactivity
Both sexes	0.0634	-0.7625***	-0.4290***
Male	0.0369	-0.7727***	-0.4375***
Female	0.1349*	-0.7886***	-0.4527***
Gender	**Overweight**	**Overweight or Obesity**	**Obesity**
Both sexes	-0.2213***	-0.3560***	-0.4288***
Male	-0.1700***	-0.3364***	-0.5993***
Female	-0.3766***	-0.2989***	-0.4756**

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variables are the log prevalence of risky health behaviors. Log EPU, log PCEHC, GDP growth rate, log divorce rate, and fertility rate are independent variables. The lag order of EPU ranges from 12-96, and the lag order of PCEHC and GDP growth rate ranges from 4-32.

our estimation results. We, therefore, explore whether our estimation results are robust to the choice of the weighting function. To this end, we re-estimate the model (2.5.4) using the MIDAS approach with Exponential Almon lag polynomials. The results of this analysis, presented in Table 2.7A, again confirm that EPU is inversely associated with mortality rates. However, when using the Exponential Almon lag polynomials, the EPU coefficient on male mortality rates increases, and men and women respond similarly to changes in EPU.

Second, there is a possibility that the relationship between mortality and economic uncertainty is driven by reverse causation. Given that the decline in mortality promotes economic growth (Kalemli-Ozcan, 2002), one may expect a decrease in economic policy uncertainty that is associated with economic upturns when mortality declines. To deal with this potential endogeneity problem, we re-estimate the model (2.5.4) by using lagged explanatory variables:

$$\begin{aligned} M_t = \; & \alpha_1 + \alpha_2 \sum_{k=1}^{K_m} w(k;\theta_m) E^{(12)}_{t-1-(k-1)/12} + \alpha_3 \sum_{k=1}^{K_q} w(k;\theta_{q_1}) P^{(4)}_{t-1-(k-1)/4} \\ & + \alpha_4 \sum_{k=1}^{K_q} w(k;\theta_{q_2}) Growth^{(4)}_{t-1-(k-1)/4} + \alpha_5 Div_{t-1} + \alpha_6 Fer_{t-1} + \varepsilon_t, \end{aligned} \tag{2.7.1}$$

where M_t is the log sex-specific mortality rate, $E^{(12)}_{t-1}$, $P^{(4)}_{t-1}$, $Growth^{(4)}_{t-1}$, Div_{t-1} and Fer_{t-1}

Table 2.7: Robustness Check to Specification Changes

Panel A: Exponential Almon Polynomials			
Determinants	Total	Male	Female
EPU	-0.2235***	-0.2982***	-0.3050***
PCEHC	-0.1437***	-0.1232***	-0.0674***
GDP growth	-0.0195***	-0.0072***	-0.0064**
Divorce	0.2610***	0.3212***	0.1860***
Fertility	0.0106	0.0061	0.0270
Panel B: Lagged Explanatory Variables			
Determinants	Total	Male	Female
EPU_1	-0.2817***	-0.1940***	-0.3107***
PCEHC_1	-0.1139***	-0.1589***	-0.0696***
GDP growth_1	-0.0120***	-0.0110*	-0.0075**
Divorce_1	0.1989***	0.2998***	0.1851***
Fertility_1	-0.0045	0.0400	0.0213

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Dependent variables are log total and sex-specific mortality rates. The lag order of EPU ranges from 12-96, and the lag order of PCEHC and GDP growth rate ranges from 4-32.

are the one-year lag of log EPU, one-year lag of log PCEHC, one-year lag of the GDP growth rate, one-year lag of log divorce rate, and one-year lag of fertility rate. The estimation results of the model (2.7.1), presented in Table 2.7B, provide qualitatively similar effects of EPU on mortality rates, with coefficients of -0.2817, -0.1940, and -0.3107 for both sexes, males, and females, respectively. The results related to other controlled factors of mortality are broadly in line with those in Table 2.4. This analysis suggests that our baseline results remain largely robust to this specification change.

2.8 Conclusion

This paper uses time-series data for 1960-2013 to study the EPU effects on sex-specific and age-specific mortality rates using the MIDAS model. Our analysis uncovers a significant negative correlation between EPU and mortality rates of all sex and age groups. This relationship is strongest for the old (65-84-year-olds), with an EPU elasticity of -

0.4410. Our accompanying analysis suggests that the reason health improves during high uncertainty periods is that individuals are more likely to engage in a healthy lifestyle. The R^2 values indicate that using the MIDAS model significantly improves the estimation accuracy.

This paper makes contributions to three strands of the literature. First, this study is expected to have an influential effect on the methodology investigating the relationship between health outcomes and macroeconomic conditions. More accurate results can be obtained via the MIDAS model since variables with several different frequencies are allowed to be taken, making the best use of high-frequency information. To my knowledge, this study is the first to use the MIDAS model to analyze the effect of EPU on health status and individual health behaviors. Second, although the literature on the relationship between business cycles and health outcomes is expansive, literature on the impact of economic policy uncertainty on health status is rare. This paper contributes to the study of sex- and age-specificities in the impact of EPU on mortality. Third, this paper contributes to the literature by exploring health behaviors as the channels through which EPU affects mortality.

Several policy implications can be generated from our results. First, since we expect that the decline in EPU leads to a greater increase in the mortality rate of 65-84-year-olds, interventions to reduce mortality should be directed more towards this age group when economic policy uncertainty declines. Second, since increased economic policy uncertainty is associated with decreased tobacco consumption, decreased prevalence of overweight and obesity, and increased suicide mortality, then an important implication for policymakers and hospitals is that when uncertainty increases, they should pay more attention to how to improve citizens' mental health and how to decrease mortality from suicide rather than from tobacco, overweight, and obesity. Last, our results uncover that women are more sensitive to changes in EPU than men and that individuals aged 65-84 respond more heavily to changes in EPU than those of working age, suggesting that the direct effect of labor market participation can not fully explain this pattern.

Chapter 3

The Spatial Quasi-Limited Information Maximum Likelihood Estimation for Spatial Lag Models with Additional Endogenous Variables

3.1 Introduction

Spatial econometrics is a subset of econometrics that is concerned with spatial effects (Anselin, 1988; Anselin, Le Gallo, and Jayet, 2006; LeSage and Pace, 2009). Spatial effects occur when the geographical closeness of observations affects the correlation between observations. Originally, most of the work in spatial econometrics was inspired by empirical economic problems caused by spatial dependence in cross-sectional and panel data (Cliff and Ord, 1973; Paelinck and Klaassen, 1979; Upton and Fingleton, 1985; Anselin and Florax, 1995; Anselin and Bera, 1998). In spatial econometrics, the information about the dependence between spatial units is incorporated into the spatial weight matrix. It captures the spatial structure of the data and usually needs to be specified in advance.

Spatial dependence can be introduced in the dependent variable, leading to a so-called spatial lag model (SLM). Over the past few decades, the SLM model has flourished, and the estimation of the SLM model has received a great deal of attention. Since the spatial lag variable, Wy, is endogenous to the model, the method of ordinary least-squares (OLS) is no longer applicable. The main consequence of applying OLS is that the estimators of the parameters will be biased and inconsistent. Several methods to estimate the SLM model are available in the literature. One method that has been widely used is the maximum likelihood (ML), which was first derived by Ord (1975). The conditions for the consistency and the asymptotic normality of the ML estimator are established by Lee (2002, 2004). Other alternative methods have been considered including the instrumental variables (IV) (Anselin, 1988), the generalized method of moments (GMM) (Kelejian and Prucha, 1998, 1999), the quasi-maximum likelihood (Lee, 2004), and the Bayesian Markov Chain Monte Carlo method (Bayesian MCMC) (LeSage, 1997; LeSage and Pace, 2009).

All of the above estimators for the SLM model are derived and their asymptotic properties are established under the assumption of exogeneity of the regressors X. However, this assumption may not always be fulfilled. Conversely, in the context of many empirical applications, the simultaneous presence of spatial lag and additional endogenous variables is a common occurrence. The concern is that if the endogeneity of the regressors is ignored, one may draw false conclusions about the spatial effects. This additional endogeneity problem may derive from the measurement errors of the variables, omitted variables, or simultaneity between the dependent and explanatory variables. For example, Anselin and Lozano-Gracia (2008) study the effect of improved air quality on house prices, where air quality variables are obtained using the interpolated air pollution measures. Anselin and Lozano-Gracia argue that these measures may lead to the "error in variable" problem, thereby leading to an additional endogeneity problem in the SLM model. Similarly, Dall'Erba and Le Gallo (2008) use spatial econometric methods to assess the impact of structural funds on the convergence process of European regions. In their study, the Hausman test results demonstrate the presence of the endogeneity of structural funds in the SLM model.

The estimation of the SLM model that allows for endogenous predictors has been theoretically motivated. Several estimation methods are considered. Among those, the spatial two-stage least squares (S2SLS) estimation is the most commonly used. The asymptotic distribution of this estimator when the number of instruments grows with the sample size is derived by Liu and Lee (2013). Based on instrumental variables (IV) and GMM, Fingleton and Le Gallo (2008) and Drukker, Egger, and Prucha (2013) consider, respectively, the feasible generalized spatial two-stage least squares (FGS2SLS) estimator and the two-step GMM/IV estimator to account for additional endogenous variables in the spatial dependence model. In a recent paper, Liu and Saraiva (2015) propose a new set of quadratic moment conditions for the GMM estimator and suggest that their proposed estimator is more efficient than other IV-based estimators in the literature. Besides, Kelejian and Prucha (2004) point out that the SLM model with additional endogenous variables can also be viewed as an equation in a system of simultaneous equations. They introduce the generalized spatial two-stage least squares (GS2SLS) estimation and the generalized spatial three-stage least squares (GS3SLS) estimation for the limited and full information estimation of the system, respectively. Moreover, instead of the IV-based estimation, Liu (2012) proposes the limited information maximum likelihood (LIML) estimation in the presence of many IVs, where the estimator is estimated based on the joint normality of the errors in the SLM model and the reduced form equations. He shows that the LIML estimator is only consistent when the number of IVs increases at a slower rate than the sample size.

In this paper, we propose a new spatial quasi-limited information maximum likelihood (SQLIML) estimation for a SLM with additional endogenous variables X. This method extends Ord's (1975) ML estimator and Wooldridge's (2014) control function estimator to account for the spatial lag and additional endogenous variables in a model. In particular, it goes beyond Ord (1975) by allowing additional endogenous variables in the SLM and goes beyond Wooldridge (2014) by allowing the presence of spatial lag in a model. One key assumption of our proposed SQLIML estimator is that the reduced form of additional endogenous variables has additive errors that depend on the errors in the SLM model. As with the control function approach of Wooldridge (2014), these additive errors are added to the SLM model to control for the endogeneity of X.

One main difference between Liu's (2012) LIML estimation and the proposed SQLIML estimation is that Liu's LIML estimation is based on the original SLM model, while the proposed SQLIML estimation is based on the modified SLM model produced by adding reduced form errors under the joint normality of errors in this modified SLM model and the reduced form equations. Our proposed SQLIML estimation has some decisive advantages over Liu's (2012) LIML estimation. First, the proposed SQLIML estimation addresses the endogeneity problem by adding reduced form errors as regressors. Theoretically, the significance of coefficients of these reduced form errors indicates the endogeneity of X. Our proposed SQLIML estimation therefore can give a regression-based exogeneity test of X. Second, the proposed SQLIML estimator is computationally simpler than Liu's (2012) LIML estimator. In Liu (2012), the concentrated log-likelihood function contains the log difference of the terms that depend on the errors and the hat matrix for exogenous variables, as well as the log-determinant of the product of matrices that depend on the spatial dependence parameter and the number of endogenous variables. This would complicate its computation considerably.

Moreover, we establish the consistency and asymptotic normality of the SQLIML estimator. We compare the finite sample properties of the S2SLS estimator and the SQLIML estimator via the Monte Carlo simulation for models with different values of instrument strength, degrees of endogeneity, and different numbers of instruments. In comparing the bias and root mean squared error (RMSE) of the SQLIML and S2SLS estimators, we find that the proposed SQLIML estimator has a better performance than the S2SLS estimator in models with strong endogeneity and weak instruments. Besides, the simulation results indicate the correct size and good power of the proposed exogeneity tests.

Lastly, in an empirical application, we revisit the driving under the influence (DUI) arrest rate model of Drukker, Prucha, and Raciborski (2013), where the alcohol-related arrest rate is the dependent variable and the number of sworn officers is treated as an endogenous explanatory variable. Following Drukker et al. (2013), we use a dummy variable that takes on 1 if a county government faces an election, as an instrument. We then compare the mean squared error (MSE) of models estimated via the S2SLS and SQLIML approaches, as well as the standard errors of their corresponding estimators of the endogenous variable and the spatial lag term. Consistent with the Monte Carlo simulation, our results show that the SQLIML estimator is relatively more efficient than the S2SLS estimator. The MSE values indicate that the use of SQLIML estimation significantly improves estimation accuracy.

The rest of this paper is organized as follows. In Section 3.2, we introduce the SLM with additional endogenous variables, together with the SQLIML estimation method that is used. In Section 3.3, we derive the conditions for the consistency and the asymptotic normality of the SQLIML estimator. In Section 3.4, we conduct the Monte Carlo simulations to assess the finite sample properties of the S2SLS estimator and the SQLIML estimator, and we summarize the Monte Carlo simulation results. In Section 3.5, we provide an empirical application, and the last section concludes. The proofs are collected in Appendix B.

3.2 The Model

3.2.1 The Model and Endogenous Regressors

Consider a cross-sectional spatial lag model (SLM) with additional endogenous regressors, $\boldsymbol{X}_1$:

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}, \tag{3.2.1}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)'$ and $\boldsymbol{u} = (u_1, u_2, \ldots, u_N)'$ are $N \times 1$ column vectors, and $\{u_i\}$ are independent and identically distributed with mean 0 and variance σ_u^2 for all i. $\boldsymbol{W}$ is a $N \times N$ spatial weights matrix with zero diagonal elements. Let $\boldsymbol{X} \equiv (\boldsymbol{X}_1, \boldsymbol{X}_2)$ be a $N \times K$ matrix of variables, where $\boldsymbol{X}_1$ is a $N \times K_1$ matrix of stochastic endogenous regressors, $\boldsymbol{X}_2$ is a $N \times K_2$ matrix of exogenous regressors, defined by

$$\underset{N\times K_1}{\boldsymbol{X}_1} = \begin{pmatrix} X_{11}^1 & \cdots & X_{11}^{K_1} \\ \vdots & \ddots & \vdots \\ X_{1N}^1 & \cdots & X_{1N}^{K_1} \end{pmatrix} \equiv (\boldsymbol{X}_1^1, \ldots, \boldsymbol{X}_1^{K_1})$$

and

$$\underset{N\times K_2}{\boldsymbol{X}_2} = \begin{pmatrix} X_{21}^1 & \cdots & X_{21}^{K_1} \\ \vdots & \ddots & \vdots \\ X_{2N}^1 & \cdots & X_{2N}^{K_1} \end{pmatrix} \equiv (\boldsymbol{X}_2^1, \ldots, \boldsymbol{X}_2^{K_2}).$$

Define $\boldsymbol{A}(\rho) = \boldsymbol{I}_N - \rho\boldsymbol{W}$, where $\boldsymbol{I}_N$ is an identity matrix of size N. The regression model (3.2.1) becomes

$$\boldsymbol{y} = \boldsymbol{A}(\rho)^{-1}(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u}). \tag{3.2.2}$$

To deal with endogeneity of $\boldsymbol{X}_1$, we need to find the $N \times L_1$ matrix of instrumental variables, denoted $\boldsymbol{Z}_1$, where $L_1 \geq K_1$. Let $\boldsymbol{Z} \equiv (\boldsymbol{Z}_1, \boldsymbol{X}_2)$ be a $N \times (L_1 + K_2)$ matrix of constant variables,

$$\underset{N\times L}{\boldsymbol{Z}} = \begin{pmatrix} Z_1^1 & \cdots & Z_1^L \\ \vdots & \ddots & \vdots \\ Z_N^1 & \cdots & Z_N^L \end{pmatrix} \equiv (\underbrace{\boldsymbol{Z}^1, \ldots, \boldsymbol{Z}^{L_1}}_{\boldsymbol{Z}_1}, \underbrace{\boldsymbol{Z}^{L_1+1}, \ldots, \boldsymbol{Z}^L}_{\boldsymbol{X}_2}),$$

where $L = L_1 + K_2$. $\boldsymbol{Z}$ satisfies the following condition.

Assumption 1: The linear reduced form equation of $\boldsymbol{X}_1$ with $\boldsymbol{Z}$ is:

$$\begin{aligned} \boldsymbol{X}_1 &= \boldsymbol{Z\delta} + \boldsymbol{v}, \\ \mathbb{E}(\boldsymbol{v}) &= \boldsymbol{0}, \end{aligned} \tag{3.2.3}$$

where $\boldsymbol{\delta}$ is a $L \times K_1$ matrix of parameters, $\boldsymbol{v}$ is a $N \times K_1$ matrix of errors, defined by

$$\underset{L\times K_1}{\boldsymbol{\delta}} = \begin{pmatrix} \delta_1^1 & \cdots & \delta_1^{K_1} \\ \vdots & \ddots & \vdots \\ \delta_L^1 & \cdots & \delta_L^{K_1} \end{pmatrix} \equiv (\boldsymbol{\delta}^1, \ldots, \boldsymbol{\delta}^{K_1})$$

and

$$\underset{N\times K_1}{\boldsymbol{v}} = \begin{pmatrix} v_1^1 & \cdots & v_1^{K_1} \\ \vdots & \ddots & \vdots \\ v_N^1 & \cdots & v_N^{K_1} \end{pmatrix} \equiv \begin{pmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_N \end{pmatrix} \equiv (\boldsymbol{v}^1, \ldots, \boldsymbol{v}^{K_1}).$$

Under Assumption 1, the endogeneity of $\boldsymbol{X}_1$ is fully reflected in $\mathbb{E}(\boldsymbol{v}'\boldsymbol{u})$ since by (3.2.3) we have

$$\mathbb{E}(\boldsymbol{X}_1'\boldsymbol{u}) = \boldsymbol{\delta}'\boldsymbol{Z}'\mathbb{E}(\boldsymbol{u}) + \mathbb{E}(\boldsymbol{v}'\boldsymbol{u}) = \mathbb{E}(\boldsymbol{v}'\boldsymbol{u}), \tag{3.2.4}$$

which implies that the endogeneity of $\boldsymbol{X}_1$ occurs if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are correlated. The following assumption states the association between $\boldsymbol{u}$ and $\boldsymbol{v}$.

Assumption 2: Assume the following:

(a) A linear projection of $\boldsymbol{u}$ on $\boldsymbol{v}$:

$$\begin{aligned} \boldsymbol{u} &= \boldsymbol{v}\boldsymbol{\lambda} + \boldsymbol{e}, \\ \mathbb{E}(\boldsymbol{e} \mid \boldsymbol{v}) &= \mathbf{0}, \end{aligned} \tag{3.2.5}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{K_1})'$ is a $K_1 \times 1$ vector of parameters with

$$\boldsymbol{\lambda} = [\mathbb{E}(\boldsymbol{v}'\boldsymbol{v})]^{-1}\mathbb{E}(\boldsymbol{v}'\boldsymbol{u}). \tag{3.2.6}$$

(b) $(e_i, \boldsymbol{v}_i)'$ are i.i.d. cross i with zero mean and finite variance:

$$(e_i, \boldsymbol{v}_i)' \sim IID\left(\begin{pmatrix} 0 \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_e^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} \end{pmatrix} \right),$$

where $\boldsymbol{\Sigma} = \mathbb{E}(\boldsymbol{v}_i'\boldsymbol{v}_i)$ is a $K_1 \times K_1$ variance-covariance matrix of $\boldsymbol{v}_i$ and is symmetric, defined by

$$\underset{K_1 \times K_1}{\boldsymbol{\Sigma}} = \begin{pmatrix} \sigma_{v1}^2 & * & \cdots & * \\ \sigma_{v2v1} & \sigma_{v2}^2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{vK_1v1} & \sigma_{vK_1v2} & \cdots & \sigma_{vK_1}^2 \end{pmatrix}.$$

Under Assumptions 1-2, $\boldsymbol{e}$ is uncorrelated with $\boldsymbol{X}_1$ and $\boldsymbol{v}$. By equations (3.2.4) and (3.2.6), it is straightforward to show that $\boldsymbol{\lambda}$ captures the severity of endogeneity of $\boldsymbol{X}_1$. In particular, when $\boldsymbol{\lambda} = \mathbf{0}$, $\boldsymbol{X}_1$ is exogenous because $\mathbb{E}(\boldsymbol{v}'\boldsymbol{u}) = \mathbf{0}$. Hence, a t-test of $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ is a test of the null that $\boldsymbol{X}_1$ is exogenous.

Plugging (3.2.5) into equation (3.2.2), we have

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{A}(\rho)^{-1}(\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{v}\boldsymbol{\lambda} + \boldsymbol{e}), \\ \boldsymbol{X}_1 &= \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{v}, \end{aligned} \tag{3.2.7}$$

where the reduced form error, $\boldsymbol{v}$, is now the regressor in the model of $\boldsymbol{y}$. As just noted, the new error term, $\boldsymbol{e}$, is uncorrelated with $\boldsymbol{X}_1$ and $\boldsymbol{v}$, implying that that adding $\boldsymbol{v}$ as a regressor addresses the endogeneity of $\boldsymbol{X}_1$.

3.2.2 Spatial Quasi-Limited Information Maximum Likelihood Estimation

Denote $\boldsymbol{\theta} = (\sigma_e^2, \text{vech}(\boldsymbol{\Sigma})', \boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\lambda}', \text{vec}(\boldsymbol{\delta})', \rho)'$, where $\text{vec}(\boldsymbol{\delta}) = (\delta_1^1, \ldots, \delta_L^1, \ldots, \delta_1^{K_1}, \ldots, \delta_L^{K_1})'$ and $\text{vech}(\boldsymbol{\Sigma}) = (\sigma_{v1}^2, \ldots, \sigma_{K_1 v1}, \sigma_{v2}^2 \ldots, \sigma_{K_1 v2}, \ldots, \sigma_{vK_1}^2)'$. Let $\boldsymbol{e}(\boldsymbol{\phi}) = \boldsymbol{A}(\rho)\boldsymbol{y} - \boldsymbol{M}(\boldsymbol{\delta})\boldsymbol{\alpha}$ and $\boldsymbol{v}(\boldsymbol{\delta}) = \boldsymbol{X}_1 - \boldsymbol{Z}\boldsymbol{\delta}$ with $\boldsymbol{\phi} = (\boldsymbol{\alpha}', \text{vec}(\boldsymbol{\delta})', \rho)'$, $\boldsymbol{\alpha} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\lambda}')'$, $\boldsymbol{M}(\boldsymbol{\delta}) = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{v}(\boldsymbol{\delta}))$, and $\boldsymbol{v}(\boldsymbol{\delta}) = (\boldsymbol{v}_1(\boldsymbol{\delta})', \ldots, \boldsymbol{v}_N(\boldsymbol{\delta})')'$. The log-likelihood function of (3.2.7), as if the errors are normally distributed, is

$$lnL(\boldsymbol{\theta}) = -\frac{N(K_1+1)}{2} ln(2\pi) - \frac{N}{2} ln\sigma_e^2 - \frac{N}{2} ln\left|\boldsymbol{\Sigma}\right| - \frac{\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{e}(\boldsymbol{\phi})}{2\sigma_e^2} - \sum_{i=1}^{N} \frac{\boldsymbol{v}_i(\boldsymbol{\delta})\boldsymbol{\Sigma}^{-1}\boldsymbol{v}_i(\boldsymbol{\delta})'}{2} + ln\left|\boldsymbol{A}(\rho)\right|, \tag{3.2.8}$$

where $ln\left|\boldsymbol{A}(\rho)\right|$ stands for the natural logs of the absolute value of the determinant of $\boldsymbol{A}(\rho)$. If the errors are truly normal, the log-likelihood function (3.2.8) is the exact one and the estimators from it are the spatial limited information maximum likelihood (SLIML) estimators. If the errors are not really normal, but their elements are i.i.d., (3.2.8) is a quasi-likelihood function and the estimators from it are the spatial quasi-limited information maximum likelihood (SQLIML) estimators.

The first order conditions of (3.2.8) are

$$\begin{cases} \frac{\partial lnL(\boldsymbol{\theta})}{\partial \sigma_e^2} &= -\frac{N}{2\sigma_e^2} + \frac{\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{e}(\boldsymbol{\phi})}{2\sigma_e^4}, \\ \frac{\partial lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma}^{-1}} &= -\frac{N}{2}\boldsymbol{\Sigma} + \frac{1}{2}\boldsymbol{v}(\boldsymbol{\delta})'\boldsymbol{v}(\boldsymbol{\delta}), \\ \frac{\partial lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\alpha}} &= \frac{\boldsymbol{M}(\boldsymbol{\delta})'\boldsymbol{e}(\boldsymbol{\phi})}{\sigma_e^2}, \\ \frac{\partial lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\delta}} &= -\frac{\boldsymbol{Z}'\boldsymbol{e}(\boldsymbol{\phi})}{\sigma_e^2}\boldsymbol{\lambda}' + \boldsymbol{Z}'\boldsymbol{v}\boldsymbol{\Sigma}^{-1}, \\ \frac{\partial lnL(\boldsymbol{\theta})}{\partial \rho} &= \frac{1}{\sigma_e^2}(\boldsymbol{W}\boldsymbol{y})'\boldsymbol{e}(\boldsymbol{\phi}) - tr(\boldsymbol{W}\boldsymbol{A}(\rho)^{-1}). \end{cases} \tag{3.2.9}$$

As the equation $\frac{\partial lnL(\boldsymbol{\theta})}{\partial \rho} = 0$ is highly nonlinear, solving for ρ is difficult. Computationally and analytically, we work with the concentrated log-likelihood function of ρ. Given ρ, see Appendix B.1, the estimators of σ_e^2, $\boldsymbol{\Sigma}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\delta}$ are given by

$$\begin{cases} \hat{\sigma}_e^2(\rho) &= \frac{1}{N}\boldsymbol{y}'\boldsymbol{A}(\rho)'\hat{\boldsymbol{H}}\boldsymbol{A}(\rho)\boldsymbol{y} \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{N}\boldsymbol{X}_1'\boldsymbol{P}\boldsymbol{X}_1 \\ \hat{\boldsymbol{\alpha}}(\rho) &= (\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{M}(\hat{\boldsymbol{\delta}}))^{-1}\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{A}(\rho)\boldsymbol{y} \\ \hat{\boldsymbol{\delta}} &= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}_1, \end{cases} \tag{3.2.10}$$

where $\hat{\boldsymbol{H}} = \boldsymbol{I}_N - \boldsymbol{M}(\hat{\boldsymbol{\delta}})(\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{M}(\hat{\boldsymbol{\delta}}))^{-1}\boldsymbol{M}(\hat{\boldsymbol{\delta}})'$, $\boldsymbol{M}(\hat{\boldsymbol{\delta}}) = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{P}\boldsymbol{X}_1)$, and $\boldsymbol{P} = \boldsymbol{I}_N - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$. Plugging $\hat{\sigma}_e^2(\rho)$, $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\alpha}}(\rho)$, and $\hat{\boldsymbol{\delta}}$ into (3.2.8), we obtain the concentrated log-likelihood function of ρ:

$$lnL(\rho) = -\frac{N(K_1+1)}{2}(ln(2\pi)+1) - \frac{N}{2}ln\hat{\sigma}_e^2(\rho) - \frac{N}{2}ln\left|\hat{\boldsymbol{\Sigma}}\right| + ln\left|\boldsymbol{A}(\rho)\right|.$$

The estimator $\hat{\rho}$ is obtained from a numerical optimization of the concentrated log-likelihood function. Following Elhorst (2014), ρ lies between $1/w_{max}$ and $1/w_{min}$, where w_{min} and w_{max} are the minimum and maximum eigenvalues of $\boldsymbol{W}$. Once we obtain $\hat{\rho}$, we can obtain $\hat{\sigma}_e^2$ and $\hat{\boldsymbol{\alpha}}$.

It is noteworthy that $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\Sigma}}$ are the ML estimators for the reduced form equation of $\boldsymbol{X}_1$ given by (3.2.3) and do not depend on the value of ρ. Hence, estimating $\boldsymbol{\theta}$ can be considered as a two-step procedure: (1) Estimate $\boldsymbol{\delta}$ and $\boldsymbol{\Sigma}$ by ML estimation of the reduced form equation of $\boldsymbol{X}_1$. Obtain the residuals $\hat{\boldsymbol{v}}$ given by $\hat{\boldsymbol{v}} = \boldsymbol{X}_1 - \boldsymbol{Z}\hat{\boldsymbol{\delta}}$. (2) Using $\hat{\boldsymbol{v}}$ in place of $\boldsymbol{v}$, estimate σ_e^2, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, and ρ by ML estimation of the model (3.2.7).

3.3 Asymptotic Properties of SQLIML Estimators

In this section, we focus on the asymptotic properties of the SQLIML estimator. In particular, we establish that the SQLIML estimator is consistent and asymptotically normal.

3.3.1 Assumptions

Let $\boldsymbol{\theta}_0 = (\sigma_{e0}^2, \text{vech}(\boldsymbol{\Sigma}_0)', \boldsymbol{\beta}_{10}', \boldsymbol{\beta}_{20}', \boldsymbol{\lambda}_0', \text{vec}(\boldsymbol{\delta}_0)', \rho_0)'$ be the true parameter vector, where $\boldsymbol{\lambda}_0 = (\lambda_{10}, \lambda_{20}, \ldots, \lambda_{K_1 0})'$, $\boldsymbol{\delta}_0 = (\boldsymbol{\delta}_0^1, \ldots, \boldsymbol{\delta}_0^{K_1})$. Let $\boldsymbol{\phi}_0 = (\boldsymbol{\alpha}_0', \text{vec}(\boldsymbol{\delta}_0)', \rho_0)'$ with $\boldsymbol{\alpha}_0 = (\boldsymbol{\beta}_{10}', \boldsymbol{\beta}_{20}', \boldsymbol{\lambda}_0')'$, the true models for $\boldsymbol{y}$ and $\boldsymbol{X}_1$ are

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{A}(\rho_0)^{-1}(\boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0 + \boldsymbol{e}_0), \\ \boldsymbol{X}_1 &= \boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0, \end{aligned} \tag{3.3.1}$$

where $\boldsymbol{e}_0 \equiv \boldsymbol{e}(\boldsymbol{\phi}_0)$ and $\boldsymbol{v}_0 \equiv \boldsymbol{v}(\boldsymbol{\delta}_0) = (\boldsymbol{v}_0^1, \ldots, \boldsymbol{v}_0^{K_1}) = \begin{pmatrix} \boldsymbol{v}_{10} \\ \vdots \\ \boldsymbol{v}_{N0} \end{pmatrix}$. Define $\boldsymbol{G}(\rho) = \boldsymbol{W}\boldsymbol{A}(\rho)^{-1}$ and let $\boldsymbol{G} \equiv \boldsymbol{G}(\rho_0)$ be its true value. Note that $\boldsymbol{A}(\rho_0)^{-1} = \boldsymbol{I}_N + \rho_0 \boldsymbol{G}$. It immediately follows from (3.3.1) that

$$\boldsymbol{y} = \boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0 + \rho_0 \boldsymbol{G}\boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0 + \boldsymbol{A}(\rho_0)^{-1}\boldsymbol{e}_0. \tag{3.3.2}$$

We extend the conditions given by Lee (2004), who derives the asymptotic properties of the QML estimator for the SLM model, to study the consistency and the asymptotic normality of the SQLIML estimator given by (3.2.10).

Assumption 3: The moment of the errors $\mathbb{E}(|e_{i0}|^{4+\varepsilon}) < \infty$ for some $\varepsilon > 0$, and $\mathbb{E}(|v_{i0}^j|^{4+\gamma_j}) < \infty$ for some $\gamma_j > 0$ for all $i = 1, \ldots, N$ and $j = 1, \ldots, K_1$.

Assumption 4: The true value ρ_0 is an interior point of the parameter space $\boldsymbol{\Lambda}$, where

$\boldsymbol{\Lambda}$ is a compact set.

Assumption 5: The elements of $\boldsymbol{Z}$ are uniformly bounded for all N. The $\lim_{N\to\infty} \frac{1}{N}(\boldsymbol{Z}'\boldsymbol{Z})$ exists and is non-singular.

Assumption 6: $\operatorname{plim}_{N\to\infty} \frac{\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0)}{N}$ exists and is non-singular.

Assumption 7: $\lim_{N\to\infty} \frac{1}{N}[\mathbb{E}(\boldsymbol{J}_1'\boldsymbol{J}_1) - \mathbb{E}(\boldsymbol{J}_1'\boldsymbol{M}(\boldsymbol{\delta}_0))(\mathbb{E}[\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0)])^{-1}\mathbb{E}(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{J}_1)] > 0$ for all $\rho \neq \rho_0$, where $\boldsymbol{J}_1 = \boldsymbol{A}(\rho)\boldsymbol{A}(\rho_0)^{-1}\boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0$.

Assumption 8: $\boldsymbol{A}(\rho_0)$ is non-singular.

Assumption 9: $\{\boldsymbol{W}\}$ and $\{\boldsymbol{A}(\rho_0)^{-1}\}$ are uniformly bounded in both row and column sums for all N.

Assumption 10: $\{\boldsymbol{A}(\rho)^{-1}\}$ is uniformly bounded in either row or column sums, uniformly in $\rho \in \boldsymbol{\Lambda}$.

Assumption 11: Assume that the spatial weight matrix satisfies:

(a) The elements w_{ij} of $\boldsymbol{W}$ are at most of order $(1/h_N)$ uniformly in all i, j, where the rate sequence h_N can be bounded or divergent.

(b) The rate $h_N/N \to 0$ as $N \to \infty$.

Assumption 3 is a basic assumption of errors in the spatial econometrics literature. Since the computation of the asymptotic distribution of the estimators involves quadratic forms of errors, the existence of the fourth-order moment of errors is required to ensure the finite variance of its quadratic forms. Assumption 4 provides a restriction on the parameter space. The compactness of the parameter space is vital when we work with the concentrated log-likelihood function ρ.

Assumptions 5-6 ensure that there is no multicollinearity in the regressors and are sufficient conditions to assure that $\boldsymbol{\theta}_0$ is identified. Assumption 7 is an identification condition for ρ and σ_e^2. A similar assumption appears in Li (2017). Under Assumption

8, the SLM model has the reduced form given by (3.3.1), and so the conditional variance of the true value $\boldsymbol{y}$ exists:

$$Var(\boldsymbol{y} \mid \boldsymbol{X}_1, \boldsymbol{Z}) = \sigma_{e0}^2 \boldsymbol{A}(\rho_0)^{-1} \boldsymbol{A}(\rho_0)^{-1\prime}.$$

Note that for a spatial weight matrix $\boldsymbol{W}$ that is not row-standardized, if $w_{min} < 0$ and $w_{max} > 0$, then a sufficient condition for the non-singularity of $\boldsymbol{A}(\rho)$ is that $1/w_{min} < \rho < 1/w_{max}$, where w_{min} and w_{max} are the minimum and maximum eigenvalues of $\boldsymbol{W}$. For a spatial weight matrix $\boldsymbol{W}$ that is row-standardized, a sufficient condition for the non-singularity of $\boldsymbol{A}(\rho)$ is that $-1 < \rho < 1$.

With Assumption 9, the conditional variance of $\boldsymbol{y}$ is bounded as $N \to \infty$. The uniform boundedness of $\boldsymbol{W}$ is a sufficient condition to produce asymptotic results of consistent estimators (Kelejian and Prucha, 1998; Kapoor, Kelejian, and Prucha, 2007). The uniform boundedness of $\boldsymbol{A}(\rho)^{-1}$ at ρ_0 implies that $\boldsymbol{A}(\rho)^{-1}$ is uniformly bounded in both row and column sums in the neighborhood of ρ_0. Recall that $\boldsymbol{G}(\rho) = \boldsymbol{W}\boldsymbol{A}(\rho)^{-1}$, Assumption 9 implies that $\boldsymbol{G}$ is uniformly bounded in both row and column sums in the neighborhood of ρ_0.

Assumption 10 addresses the nonlinearity of $ln\,|\boldsymbol{A}(\rho)|$. Lee (2002, 2004) shows that if $\boldsymbol{W}$ is row-standardized, $\boldsymbol{\Lambda}$ can be considered as a closed subset of $(-1, 1)$; if $\boldsymbol{W}$ is not row-standardized, $\boldsymbol{\Lambda}$ can be a closed subset of $(-1/ \mid w_{min} \mid, 1/w_{max})$.

Assumption 11 states that the rate at which w_{ij} increases as N increases can be bounded or divergent, but excludes cases where w_{ij} diverges to infinity at a rate equal to or faster than the rate of the sample size N because the SQLIML (SLIML) estimators are inconsistent for those cases. This assumption covers the spatial weight matrices whose elements can be negative and those that might not be row-standardized, and it ensures that the spatial correlation can be limited to a manageable degree. The most commonly used spatial weight matrix in the literature is the contiguity-based spatial weight matrix, where $w_{ij} = 1$ if units i and j have a contiguous border, and that is row-normalized subsequently. Assumption 11 is satisfied when each unit has a limited number of neighbors even when the total number of units increases to infinity because h_N is bounded. This particular case rules out cases where each unit has infinitely many neighbors. For example, if all units are assumed to be neighbors of each other. In this case, all off-diagonal elements of $\boldsymbol{W}$ is $w_{ij} = 1/(N-1)$, and therefore $h_N = (N-1)$ and $h_N/N \to 1$ as $N \to \infty$. With Assumption 9, Assumption 11 implies that the elements $\boldsymbol{G}_{ij}$ of $\boldsymbol{G}$ have $O(1/h_N)$ uniformly in all i, j.

3.3.2 Consistency and Asymptotic Normality

Let $Q(\rho) = max_{\sigma_e^2,\boldsymbol{\Sigma},\boldsymbol{\alpha},\boldsymbol{\delta}}\mathbb{E}(lnL(\boldsymbol{\theta}))$. As in Lee (2004), maximize $\mathbb{E}(lnL(\boldsymbol{\theta}))$ with respect to σ_e^2, $\boldsymbol{\Sigma}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\delta}$, we get the following solutions

$$\boldsymbol{\delta}^* = \mathbb{E}(\hat{\boldsymbol{\delta}}) = \mathbb{E}(\boldsymbol{\delta}_0 + (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{v}_0) = \boldsymbol{\delta}_0,$$

$$\boldsymbol{\Sigma}^* = \frac{1}{N}\mathbb{E}((\boldsymbol{X}_1 - \boldsymbol{Z}\boldsymbol{\delta}^*)'(\boldsymbol{X}_1 - \boldsymbol{Z}\boldsymbol{\delta}^*)) = \boldsymbol{\Sigma}_0,$$

$$\begin{aligned} \boldsymbol{\alpha}^*(\rho) &= (\mathbb{E}[\boldsymbol{M}(\boldsymbol{\delta}^*)'\boldsymbol{M}(\boldsymbol{\delta}^*)])^{-1}\mathbb{E}(\boldsymbol{M}(\boldsymbol{\delta}^*)'\boldsymbol{A}(\rho)\boldsymbol{y}) \\ &= (\mathbb{E}[\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0)])^{-1}\mathbb{E}(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{A}(\rho)\boldsymbol{A}(\rho_0)^{-1}\boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0), \end{aligned}$$

and

$$\begin{aligned} \sigma_e^{*2}(\rho) &= \tfrac{1}{N}\mathbb{E}([\boldsymbol{A}(\rho)\boldsymbol{y} - \boldsymbol{M}(\boldsymbol{\delta}^*)\boldsymbol{\alpha}^*(\rho)]'[\boldsymbol{A}(\rho)\boldsymbol{y} - \boldsymbol{M}(\boldsymbol{\delta}^*)\boldsymbol{\alpha}^*(\rho)]) \\ &= \tfrac{1}{N}\mathbb{E}[(\boldsymbol{J}_1 + \boldsymbol{J}_2 - \boldsymbol{M}(\boldsymbol{\delta}_0)(\mathbb{E}[\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0)])^{-1}\mathbb{E}(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{J}_1))' \\ &\quad \times(\boldsymbol{J}_1 + \boldsymbol{J}_2 - \boldsymbol{M}(\boldsymbol{\delta}_0)(\mathbb{E}[\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0)])^{-1}\mathbb{E}(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{J}_1))] \\ &= \tfrac{1}{N}[\mathbb{E}(\boldsymbol{J}_1'\boldsymbol{J}_1) - \mathbb{E}(\boldsymbol{J}_1'\boldsymbol{M}(\boldsymbol{\delta}_0))(\mathbb{E}[\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0)])^{-1}\mathbb{E}(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{J}_1) \\ &\quad +\sigma_{e0}^2 tr(\boldsymbol{A}(\rho_0)^{-1\prime}\boldsymbol{A}(\rho)'\boldsymbol{A}(\rho)\boldsymbol{A}(\rho_0)^{-1})], \end{aligned}$$

where $\boldsymbol{J}_2 = \boldsymbol{A}(\rho)\boldsymbol{A}(\rho_0)^{-1}\boldsymbol{e}_0$. Substitute $\boldsymbol{\Sigma}^*$ and $\sigma_e^{*2}(\rho)$ into concentrated log-likelihood function, we have

$$Q(\rho) = -\frac{N(K_1+1)}{2}(ln(2\pi)+1) - \frac{N}{2}ln\sigma_e^{*2}(\rho) - \frac{N}{2}ln\left|\boldsymbol{\Sigma}_0\right| + ln\left|\boldsymbol{A}(\rho)\right|.$$

To demonstrate that the SQLIML estimator $\hat{\boldsymbol{\theta}}$ is consistent, we need to show that (i)$\frac{1}{N}lnL(\rho) \xrightarrow{p} \frac{1}{N}Q(\rho)$ uniformly in $\rho \in \boldsymbol{\Lambda}$, and (ii) $\frac{1}{N}Q(\rho)$ is uniquely maximized at the true value ρ_0 (White, 1996). The next theorem states the consistency of the SQLIML estimator.

Theorem 3.1. *Under Assumptions 1-11, $\hat{\boldsymbol{\theta}}$ given in (3.2.10) is a consistent estimator of $\boldsymbol{\theta}_0$.*

Proof. See Appendix B.2. □

The asymptotic distribution of the SQLIML estimator is obtained from the Taylor series expansion of $\frac{\partial lnL(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} = \boldsymbol{0}$ at $\boldsymbol{\theta}_0$. The first-order derivatives of the log-likelihood function at $\boldsymbol{\theta}_0$ are

$$\begin{cases} \frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2} &= \frac{1}{2\sigma_{e0}^4\sqrt{N}}(\boldsymbol{e}_0'\boldsymbol{e}_0 - N\sigma_{e0}^2) \\ \frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\Sigma}^{-1}} &= \frac{1}{2\sqrt{N}}(\boldsymbol{v}_0'\boldsymbol{v}_0 - N\boldsymbol{\Sigma}_0) \\ \frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1} &= \frac{1}{\sigma_{e0}^2\sqrt{N}}\boldsymbol{X}_1'\boldsymbol{e}_0 \\ \frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2} &= \frac{1}{\sigma_{e0}^2\sqrt{N}}\boldsymbol{X}_2'\boldsymbol{e}_0 \\ \frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}} &= \frac{1}{\sigma_{e0}^2\sqrt{N}}\boldsymbol{v}_0'\boldsymbol{e}_0 \\ \frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta})}{\partial\boldsymbol{\delta}} &= -\frac{\boldsymbol{Z}'\boldsymbol{e}_0}{\sigma_{e0}^2\sqrt{N}}\boldsymbol{\lambda}_0' + \frac{\boldsymbol{Z}'\boldsymbol{v}_0\boldsymbol{\Sigma}_0^{-1}}{\sqrt{N}} \\ \frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\rho} &= \frac{1}{\sigma_{e0}^2\sqrt{N}}(\boldsymbol{G}\boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0)'\boldsymbol{e}_0 + \frac{1}{\sigma_{e0}^2\sqrt{N}}(\boldsymbol{e}_0'\boldsymbol{G}\boldsymbol{e}_0 - \sigma_{e0}^2 tr(\boldsymbol{G})) \end{cases}$$

Following Assumptions 3 and 7, $\boldsymbol{v}_0$ and $\boldsymbol{e}_0$ have finite fourth moments, and $\boldsymbol{Z}$ is uniformly bounded. Hence, a central limit theorem can be applied to $\frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$. For the case $\{h_N\}$

being a bounded process, we can use the central limit theorem for linear-quadratic forms introduced in Kelejian and Prucha (2001). For the case $\{h_N\}$ being a divergent process, we can use the Kolmogorov central limit theorem to $\frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$. The variance matrix of $\frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$ is

$$Var(\frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}) = \begin{cases} -\mathbb{E}(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}), & \text{if } (e_i,\boldsymbol{v}_i) \text{ are normally distributed} \\ -\mathbb{E}(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}) + \boldsymbol{\Omega}_{\boldsymbol{\theta},N}, & \text{if } (e_i,\boldsymbol{v}_i) \text{ are i.i.d} \end{cases}$$

where

$$\boldsymbol{\Omega}_{\boldsymbol{\theta},N} = \mathbb{E}(\frac{1}{N}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}) + \mathbb{E}(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}).$$

Given the above results and assumptions, the next theorem states that the SQLIML estimator (3.2.10) is $\sqrt{N}$-consistent and follows the asymptotic normal distribution.

Theorem 3.2. *Under Assumptions 1-11, if* $(e_i,\boldsymbol{v}_i)$ *are normal distributed, the asymptotically normality of* $\hat{\boldsymbol{\theta}} = (\hat{\sigma}_e^2, vech(\hat{\boldsymbol{\Sigma}})', \hat{\boldsymbol{\beta}}_1{}', \hat{\boldsymbol{\beta}}_2{}', \hat{\boldsymbol{\lambda}}', vec(\hat{\boldsymbol{\delta}})', \hat{\rho})'$ *is*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = -\lim_{N\to\infty} \mathbb{E}(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})$ *and is given by*

$$
-\mathbb{E}\left(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) =
\begin{bmatrix}
\frac{1}{2\sigma_{e0}^4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{N\sigma_{e0}^2}tr(\boldsymbol{G}) \\
* & \frac{\boldsymbol{D}'(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\Sigma}_0^{-1})\boldsymbol{D}}{2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
* & * & \frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta}_0+N\boldsymbol{\Sigma}_0}{N\sigma_{e0}^2} & \frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & -\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2} & \frac{(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})+tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} \\
* & * & * & \frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \mathbf{0} & -\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_2'\boldsymbol{Z})}{N\sigma_{e0}^2} & \frac{\boldsymbol{X}_2'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \\
* & * & * & * & \frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & \mathbf{0} & \frac{tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} \\
* & * & * & * & * & \frac{(\boldsymbol{\lambda}_0\boldsymbol{\lambda}_0')\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2} + \frac{\boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N} & -\frac{\boldsymbol{\lambda}_0\otimes(\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}))}{N\sigma_{e0}^2} \\
* & * & * & * & * & * & \frac{(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})+(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)'tr(\boldsymbol{G}'\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} + \frac{1}{N}tr(\boldsymbol{G}\boldsymbol{G}+\boldsymbol{G}'\boldsymbol{G})
\end{bmatrix},
$$

where $\otimes$ *is the Kronecker product.* $\boldsymbol{D}$ *is the duplication matrix of* $\boldsymbol{\Sigma}$*, defined in Magnus and Neudecker (2019).*

Proof. See Appendix B.3. □

Theorems 3.1-3.2 are valid for both bounded and divergent $\{h_N\}$. For the case $\{h_N\}$

being a divergent process, the matrix $\boldsymbol{\Sigma_\theta}$ becomes

$$\boldsymbol{\Sigma_\theta} = \begin{bmatrix} \frac{1}{2\sigma_{e0}^4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \\ * & \frac{\boldsymbol{D}'(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\Sigma}_0^{-1})\boldsymbol{D}}{2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ * & * & \frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta}_0+N\boldsymbol{\Sigma}_0}{N\sigma_{e0}^2} & \frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & -\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2} & \frac{(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \\ * & * & * & \frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \mathbf{0} & -\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_2'\boldsymbol{Z})}{N\sigma_{e0}^2} & \frac{\boldsymbol{X}_2'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \\ * & * & * & * & \frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & \mathbf{0} & \mathbf{0} \\ * & * & * & * & * & \frac{(\boldsymbol{\lambda}_0\boldsymbol{\lambda}_0')\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2}+\frac{\boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N} & -\frac{\boldsymbol{\lambda}_0\otimes(\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}))}{N\sigma_{e0}^2} \\ * & * & * & * & * & * & \frac{(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \end{bmatrix}.$$

When $\{h_N\}$ is divergent, the SLIML's $\hat{\sigma}_e^2$ and $\hat{\boldsymbol{\lambda}}$ are asymptotically independent of $\hat{\rho}$ because $\underset{N\to\infty}{lim}\, h_N = \infty$ and $\boldsymbol{G}_{ij} = O(1/h_N)$, and consequently, $\underset{N\to\infty}{lim}\, \frac{1}{N}tr(\boldsymbol{G}) = 0$. Whereas $\hat{\sigma}_e^2$ and $\hat{\boldsymbol{\lambda}}$ are asymptotically dependent on $\hat{\rho}$ when $\{h_N\}$ is bounded because $\underset{N\to\infty}{lim}\, \frac{1}{N}tr(\boldsymbol{G})$ may not be zero.

Remark 3.1. *Theorem 3.2 only provides the asymptotic distribution of the SLIML estimator. When the errors $\boldsymbol{v}_0^1, \boldsymbol{v}_0^2, \dots \boldsymbol{v}_0^{K_1}$ are pairwise correlated, the asymptotic variance of the SQLIML is difficult to obtain because terms $\mathbb{E}(\frac{1}{N}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial vech(\boldsymbol{\Sigma})}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial vech(\boldsymbol{\Sigma})'})$ and $\mathbb{E}(\frac{1}{N}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial vech(\boldsymbol{\Sigma})}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial vec(\boldsymbol{\delta})'})$, respectively, contain the expectation of $vec(\boldsymbol{v}_0'\boldsymbol{v}_0)vec(\boldsymbol{v}_0'\boldsymbol{v}_0)'$ and $vec(\boldsymbol{v}_0'\boldsymbol{v}_0)vec(\boldsymbol{v}_0)'$.*

Next, we consider the special case that the errors $\boldsymbol{v}^1, \boldsymbol{v}^2, \dots \boldsymbol{v}^{K_1}$ are pairwise uncorrelated, then the variance-covariance matrix of $\boldsymbol{v}_i$ becomes

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{v1}^2 & & & \\ & \sigma_{v2}^2 & & \\ & & \ddots & \\ & & & \sigma_{vK_1}^2 \end{pmatrix},$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix with $\sigma_{vij} = 0$. In this case, $\lambda_j = \mathbb{E}(\boldsymbol{v}^{j\prime}\boldsymbol{u})/\mathbb{E}(\boldsymbol{v}^{j\prime}\boldsymbol{v}^j)$ for all j in (3.2.5). Plus, because $\mathbb{E}(\boldsymbol{X}_1^{j\prime}\boldsymbol{u}) = \mathbb{E}(\boldsymbol{v}^j\boldsymbol{u})$, the significance test on $\hat{\lambda}_j$ can be used to test for endogeneity of $\boldsymbol{X}_1^j$. Denote $\boldsymbol{\theta} = (\sigma_e^2, \boldsymbol{\sigma}_v^{2\prime}, \boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\lambda}', vec(\boldsymbol{\delta})', \rho)'$, where

$\boldsymbol{\sigma}_v^2 = (\sigma_{v1}^2, \sigma_{v2}^2, \ldots, \sigma_{vK_1}^2)'$. *From (3.2.10), it is straightforward to get that*

$$\begin{cases} \hat{\sigma}_e^2(\rho) &= \frac{1}{N}\boldsymbol{y}'\boldsymbol{A}(\rho)'\hat{\boldsymbol{H}}\boldsymbol{A}(\rho)\boldsymbol{y} \\ \hat{\sigma}_{vj}^2 &= \frac{1}{N}\boldsymbol{X}_1^{j}{}'\boldsymbol{P}\boldsymbol{X}_1^{j},\ j = 1, \ldots, K_1 \\ \hat{\boldsymbol{\alpha}}(\rho) &= (\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{M}(\hat{\boldsymbol{\delta}}))^{-1}\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{A}(\rho)\boldsymbol{y} \\ \hat{\boldsymbol{\delta}} &= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}_1 \end{cases} \tag{3.3.3}$$

and

$$\hat{\rho} = \underset{\rho}{argmax}(-\frac{N(K_1+1)}{2}(ln(2\pi)+1) - \frac{N}{2}ln\hat{\sigma}_e^2(\rho) - \sum_{j=1}^{K_1}\frac{N}{2}ln\hat{\sigma}_{vj}^2 + ln\,|\boldsymbol{I}_N - \rho\boldsymbol{W}|).$$

Theorem 3.3. *Under Assumptions 1-11, in the event that the errors $\boldsymbol{v}_0^1, \boldsymbol{v}_0^2, \ldots \boldsymbol{v}_0^{K_1}$ are pairwise uncorrelated, asymptotically normality of $\hat{\boldsymbol{\theta}}$ is*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}).$$

If $(e_i, \boldsymbol{v}_i)$ are normal distributed, then

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}),$$

where $\boldsymbol{\Sigma_\theta} = -\lim\limits_{N\to\infty} \mathbb{E}(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})$, $\boldsymbol{\Omega_\theta} = \lim\limits_{N\to\infty} \boldsymbol{\Omega}_{\boldsymbol{\theta},N}$ *given by*

$$-\mathbb{E}(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}) =$$

$$\begin{bmatrix}
\frac{1}{2\sigma_{e0}^4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{1}{N\sigma_{e0}^2}tr(\boldsymbol{G}) \\
* & \frac{\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_0^{-1}}{2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
* & * & \frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta}_0+N\boldsymbol{\Sigma}_0}{N\sigma_{e0}^2} & \frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & -\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2} & \frac{(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})+tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} \\
* & * & * & \frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \mathbf{0} & -\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_2'\boldsymbol{Z})}{N\sigma_{e0}^2} & \frac{\boldsymbol{X}_2'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \\
* & * & * & * & \frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & \mathbf{0} & \frac{tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} \\
* & * & * & * & * & \frac{(\boldsymbol{\lambda}_0\boldsymbol{\lambda}_0')\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2} + \frac{\boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N} & -\frac{\boldsymbol{\lambda}_0\otimes(\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}))}{N\sigma_{e0}^2} \\
* & * & * & * & * & * & \frac{(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})+(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)'tr(\boldsymbol{G}'\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} + \frac{1}{N}tr(\boldsymbol{G}\boldsymbol{G}+\boldsymbol{G}'\boldsymbol{G})
\end{bmatrix}$$

and

$$
\boldsymbol{\Omega}_{\boldsymbol{\theta},N} =
\left[\begin{array}{ccccccc}
\frac{\mu_{e0}^4-3\sigma_{e0}^4}{4\sigma_{e0}^8} & \mathbf{0} & \frac{\mu_{e0}^3}{2N\sigma_{e0}^6}\boldsymbol{l}_N'\boldsymbol{Z}\boldsymbol{\delta}_0 & \frac{\mu_{e0}^3}{2N\sigma_{e0}^6}\boldsymbol{l}_N'\boldsymbol{X}_2 & \mathbf{0} & -\frac{\mu_{e0}^3\boldsymbol{\lambda}_0'\otimes(\boldsymbol{l}_N'\boldsymbol{Z})}{2N\sigma_{e0}^6} & \frac{1}{2N\sigma_{e0}^6}[\mu_{e0}^3\boldsymbol{l}_N'(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{G}\boldsymbol{X}_2\boldsymbol{\beta}_{20})+(\mu_{e0}^4-3\sigma_{e0}^4)tr(\boldsymbol{G})] \\
* & diag(\frac{\mu_{v10}^4-3\sigma_{v10}^4}{4\sigma_{v10}^8},\ldots,\frac{\mu_{vK_10}^4-3\sigma_{vK_10}^4}{4\sigma_{vK_10}^8}) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{diag(\boldsymbol{\zeta}_{v0})\otimes\boldsymbol{l}_N'\boldsymbol{Z}}{2N} & \mathbf{0} \\
* & * & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{\mu_{e0}^3}{N\sigma_{e0}^4}\sum_{i=1}^N \boldsymbol{G}_{ii}(\boldsymbol{\delta}_0'\boldsymbol{Z}')_i \\
* & * & * & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{\mu_{e0}^3}{N\sigma_{e0}^4}\sum_{i=1}^N \boldsymbol{G}_{ii}(\boldsymbol{X}_2')_i \\
* & * & * & * & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
* & * & * & * & * & \mathbf{0} & -\frac{\mu_{e0}^3}{N\sigma_{e0}^4}\boldsymbol{\lambda}_0\otimes\sum_{i=1}^N \boldsymbol{G}_{ii}(\boldsymbol{Z}')_i \\
* & * & * & * & * & * & \frac{2\mu_{e0}^3}{N\sigma_{e0}^4}\sum_{i=1}^N \boldsymbol{G}_{ii}((\boldsymbol{G}\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{G}\boldsymbol{X}_2\boldsymbol{\beta}_{20})')_i+\frac{(\mu_{e0}^4-3\sigma_{e0}^4)}{N\sigma_{e0}^4}\sum_{i=1}^N \boldsymbol{G}_{ii}^2
\end{array}\right],
$$

where $\mu_{vj0}^m = \mathbb{E}[(v_{i0}^j)^m]$, $\mu_{e0}^m = \mathbb{E}[(e_{i0})^m]$, $m = 2,3,4$ *represent the second, third, fourth moments of* $\{v_{i0}^j\}$ *and* $\{e_{i0}\}$. $\boldsymbol{l}_N$ *is an N-vector of ones,* $diag(\boldsymbol{a}) \equiv \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_N \end{bmatrix}$ *is a diagonal matrix with the elements of vector* $\boldsymbol{a}$. $(\boldsymbol{A})_i$ *is the ith column of* $\boldsymbol{A}$, $\boldsymbol{G}_{ij}$ *is the (i,j)th element of* $\boldsymbol{G}$, $\boldsymbol{\zeta}_{v0} = (\zeta_{v10},\ldots,\zeta_{vK_10})$ *with* $\zeta_{vj0} = \mu_{vj0}^3/\sigma_{vj0}^6$.

Proof. See Appendix B.4. □

For the case $\{h_N\}$ being a divergent process, the matrices $\boldsymbol{\Sigma_\theta}$ and $\boldsymbol{\Omega_\theta}$ become

$$\boldsymbol{\Sigma_\theta} =$$
$$\begin{bmatrix} \frac{1}{2\sigma_{e0}^4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 0 \\ * & \frac{\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\Sigma}_0^{-1}}{2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ * & * & \frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta}_0+N\boldsymbol{\Sigma}_0}{N\sigma_{e0}^2} & \frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & -\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2} & \frac{(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \\ * & * & * & \frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \mathbf{0} & -\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_2'\boldsymbol{Z})}{N\sigma_{e0}^2} & \frac{\boldsymbol{X}_2'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \\ * & * & * & * & \frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & \mathbf{0} & \mathbf{0} \\ * & * & * & * & * & \frac{(\boldsymbol{\lambda}_0\boldsymbol{\lambda}_0')\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2}+\frac{\boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N} & -\frac{\boldsymbol{\lambda}_0\otimes(\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}))}{N\sigma_{e0}^2} \\ * & * & * & * & * & * & \frac{(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \end{bmatrix}$$

and

$$\boldsymbol{\Omega_\theta} =$$
$$\begin{bmatrix} \frac{\mu_{e0}^4-3\sigma_{e0}^4}{4\sigma_{e0}^8} & \mathbf{0} & \frac{\mu_{e0}^3\boldsymbol{l}_N'\boldsymbol{Z}\boldsymbol{\delta}_0}{2N\sigma_{e0}^6} & \frac{\mu_{e0}^3\boldsymbol{l}_N'\boldsymbol{X}_2}{2N\sigma_{e0}^6} & \mathbf{0} & -\frac{\mu_{e0}^3\boldsymbol{\lambda}_0'\otimes\boldsymbol{l}_N'\boldsymbol{Z}}{2N\sigma_{e0}^6} & \frac{\mu_{e0}^3\boldsymbol{l}_N'(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{G}\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{2N\sigma_{e0}^6} \\ * & \mathrm{diag}(\frac{\mu_{v1}^4-3\sigma_{v10}^4}{4\sigma_{v10}^8},\dots,\frac{\mu_{vK_1}^4-3\sigma_{vK_10}^4}{4\sigma_{vK_10}^8}) & \mathbf{0} & \mathbf{0} & \mathbf{0} & \frac{\mathrm{diag}(\boldsymbol{\zeta}_{v0})\otimes\boldsymbol{l}_N'\boldsymbol{Z}}{2N} & \mathbf{0} \\ * & * & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ * & * & * & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ * & * & * & * & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ * & * & * & * & * & \mathbf{0} & \mathbf{0} \\ * & * & * & * & * & * & 0 \end{bmatrix}.$$

3.4 Finite Sample Properties of Estimators

The objective of the Monte Carlo simulation is twofold. First, we compare the finite sample properties of the spatial 2SLS (S2SLS) estimators and the proposed SQLIML estimators. Second, we assess the size and power of the test for the exogeneity of $\boldsymbol{X}$, that is $H_0 : \boldsymbol{\lambda} = \mathbf{0}$.

3.4.1 S2SLS Estimator

We next review the S2SLS estimation approach. The S2SLS estimation is a straightforward method for estimating the SLM with additional endogenous variables. Instruments are needed for the spatial lag and additional endogenous variables. For the spatial lag, several studies suggest the selection of optimal instruments based on the reduced form of the model specification (Kelejian and Prucha, 1998, 1999; Lee, 2003; Anselin, Le Gallo, and Jayet, 2006). However, when additional endogenous variables are included in the model, the system determining $\boldsymbol{y}$ is not completely specified, and one can not obtain optimal instruments for the spatial lag.

To obtain a full systems specification, the instruments for additional endogenous variables are needed. Using traditional notations, we rewrite (3.2.1) as

$$\begin{aligned}\boldsymbol{y} &= \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u} \\ &= \boldsymbol{B}\boldsymbol{\gamma} + \boldsymbol{u},\end{aligned}$$

with $\boldsymbol{B} = (\boldsymbol{W}\boldsymbol{y}, \boldsymbol{X}_1, \boldsymbol{X}_2)$ and $\boldsymbol{\gamma} = (\rho, \boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$. In the S2SLS estimates, suppose that the endogenous variables $\boldsymbol{X}_1$ are determined by a set of exogenous variables $\boldsymbol{Z}$ given in (3.2.3), the instrument matrix $\boldsymbol{Q}$ for $\boldsymbol{B}$ is

$$\boldsymbol{Q} = (\boldsymbol{Z}, \boldsymbol{W}\boldsymbol{Z}, \boldsymbol{W}^2\boldsymbol{Z}, \ldots),$$

where $\boldsymbol{Q}$ is a $N \times q$ matrix, with $q \geq 1 + K$, and $\boldsymbol{Q}$ consists of the exogenous variables and multiple orders of their spatial lags. In general, the spatial lag order is set to 2. Therefore, the instrument matrix $\boldsymbol{Q}$ is

$$\boldsymbol{Q} = (\boldsymbol{Z}, \boldsymbol{W}\boldsymbol{Z}, \boldsymbol{W}^2\boldsymbol{Z}).$$

With the instrument matrix, the S2SLS estimator is

$$\hat{\gamma}_{S2SLS} = [\boldsymbol{B}'\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{B}]^{-1}\boldsymbol{B}'\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{y},$$

and its asymptotic variance matrix is

$$AsyVar(\hat{\gamma}_{S2SLS}) = \hat{\sigma}_u^2[\boldsymbol{B}'\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{Q})^{-1}\boldsymbol{Q}'\boldsymbol{B}]^{-1},$$

where $\hat{\sigma}_u^2 = (\boldsymbol{y} - \boldsymbol{B}\hat{\boldsymbol{\gamma}}_{S2SLS})'(\boldsymbol{y} - \boldsymbol{B}\hat{\boldsymbol{\gamma}}_{S2SLS})/(N-1-K)$.

3.4.2 Monte Carlo Simulation

For simplicity, we consider the case of a single endogenous regressor $\boldsymbol{X}$, the Monte Carlo simulation is based on the following DGP

$$\boldsymbol{y} = (\boldsymbol{I}_N - \rho \boldsymbol{W})^{-1}(\boldsymbol{X}\beta + \boldsymbol{u}), \tag{3.4.1}$$

where $\rho = 0.5$ and $\beta = 1$. $\boldsymbol{W}$ is a row-standardized spatial weight matrix based on 10-nearest neighbors, where $w_{i,i-5} = \ldots = w_{i,i-1} = w_{i,i+1} = \ldots w_{i,i+5} = 1/10$ for $i = 1, \ldots, N$ (wrap around to the end or the start of the row if the index lies outside the matrix boundaries), and 0 otherwise. We generate the reduced form equation of $\boldsymbol{X}$ and a linear projection of $\boldsymbol{u}$ as follows:

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{v}, \\ \boldsymbol{u} &= \lambda \boldsymbol{v} + \boldsymbol{e}, \end{aligned} \tag{3.4.2}$$

where $\boldsymbol{e}$ and $\boldsymbol{v}$ are generated from independent t-distributions with υ degrees of freedom, $\boldsymbol{Z}$ is a $N \times L$ matrix of instruments for $\boldsymbol{X}$, $\boldsymbol{\delta} = c\boldsymbol{l}_L$. We use a normal distribution for the instruments, and we set $\upsilon = 5$, $c = \{0.3, 0.5, 0.8\}$, $L = \{1, 4, 8, 16\}$, and $\lambda = \{0.1, 0.5, 1, 1.5\}$. The various values of c and λ enable us to inspect how the SQLIML estimator behaves when the strength of instruments varies and when the degree of endogeneity of $\boldsymbol{X}$ varies, respectively. Besides, the various value of L allows us to consider the case of just-identified models and over-identified models.

We consider sample sizes $N = \{500, 1000\}$. The Monte Carlo experiment was performed with 1,000 simulations. To assess the performance of SQLIML and S2SLS estimators, we report the following statistics in Tables 3.1-3.5:

- Average bias of SQLIML and S2SLS estimators of β: $bias(\hat{\beta}) = \frac{\sum_{m=1}^{M}(\hat{\beta}_m - \beta)}{M}$, where M is the number of replications and $\hat{\beta}_m$ is the estimated value of β in the mth repetition;

- Average bias of SQLIML and S2SLS estimators of ρ: $bias(\hat{\rho}) = \frac{\sum_{m=1}^{M}(\hat{\rho}_m - \rho)}{M}$, where M is the number of replications and $\hat{\rho}_m$ is the estimated value of ρ in the mth repetition;

- Relative efficiency between SQLIML and S2SLS estimators of β: $RE(\hat{\beta}) = RMSE(\hat{\beta}_{SQLIML})/RMSE(\hat{\beta}_{S2SLS})$, where $RMSE(\hat{\beta})$ is the RMSE given by $RMSE(\hat{\beta}) = \sqrt{\frac{\sum_{m=1}^{M}(\hat{\beta}_m - \beta)^2}{M}}$;

- Relative efficiency between SQLIML and S2SLS estimators of ρ: $RE(\hat{\rho}) = RMSE(\hat{\rho}_{SQLIML})/RMSE(\hat{\rho}_{S2SLS})$, where $RMSE(\hat{\rho})$ is the RMSE given by $RMSE(\hat{\rho}) = \sqrt{\frac{\sum_{m=1}^{M}(\hat{\rho}_m - \rho)^2}{M}}$;

- Power of exogeneity test: $M^{-1}\sum_{m=1}^{M}\mathbb{I}(\left|\frac{\hat{\lambda}_m}{std(\hat{\lambda}_m)}\right| > t_{0.975})$, where $\mathbb{I}$ is the indicator function.

Panels A-B of Tables 3.1-3.4 display the average bias of SQLIML and S2SLS estimators, and panels C display the corresponding efficiency of SQLIML in terms of S2SLS, which we calculate as the RMSE ratio. We first focus on the estimator of β. Overall, it is evident the SQLIML estimator has a noticeably lower bias and enjoys a larger efficiency gain than the S2SLS estimator, especially for models with weak instruments and strong endogeneity. Turning to the estimator of ρ. The results for the S2SLS estimator are encouraging. Two points are noteworthy. First, compared to the SQLIML estimator, the S2SLS estimator is biased, implying that the S2SLS estimation may fail to correct for the bias of ρ. This finding supports Lee (2004), suggesting that the IV-based approach to estimating ρ is not applicable if all the spatial regressors are irrelevant. Besides, for all cases, the RMSE ratio values indicate that the SQLIML estimator is much more efficient than the S2SLS estimator. Second, in general, we find that compared to the S2SLS estimator, the performance of the SQLIML estimator improves in terms of bias

Table 3.1: Results on Monte Carlo Simulation with 1 Instrument

Panel A: Average Bias of $\hat{\beta}_{SQLIML}$ and $\hat{\rho}_{SQLIML}$												
	$\hat{\beta}_{SQLIML}$						$\hat{\rho}_{SQLIML}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	-0.0025	-0.0040	-0.0026	-0.0018	0.0009	0.0004	-0.0066	-0.0070	-0.0076	0.0001	-0.0017	-0.0015
$\lambda = 0.5$	-0.0220	-0.0088	-0.0033	-0.0095	0.0055	-0.0070	-0.0058	-0.0053	-0.0021	-0.0020	-0.0016	-0.0021
$\lambda = 1$	-0.0335	-0.0165	-0.0045	-0.0163	-0.0050	-0.0016	-0.0026	-0.0041	-0.0043	-0.0023	-0.0017	-0.0011
$\lambda = 1.5$	-0.0565	-0.0160	-0.0055	-0.0344	-0.0130	0.0005	-0.0008	-0.0035	-0.0032	-0.0012	-0.0011	-0.0019
Panel B: Average Bias of $\hat{\beta}_{S2SLS}$ and $\hat{\rho}_{S2SLS}$												
	$\hat{\beta}_{S2SLS}$						$\hat{\rho}_{S2SLS}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	-0.0157	-0.0097	-0.0052	-0.0080	-0.0033	-0.0014	0.0449	0.0120	-0.0404	0.0374	-0.0237	-0.0122
$\lambda = 0.5$	-0.0192	-0.0103	-0.0066	-0.0109	0.0006	-0.0085	0.1318	0.0447	-0.0141	0.0519	-0.0211	-0.0227
$\lambda = 1$	-0.0149	-0.0190	-0.0084	-0.0028	-0.0076	-0.0040	0.2014	0.0459	-0.0121	0.0888	0.0242	-0.0220
$\lambda = 1.5$	0.0148	-0.0087	-0.0087	-0.0297	-0.0161	-0.0024	0.3448	0.1265	0.0135	0.0785	0.0380	-0.0175
Panel C: Relative Efficiency of $\hat{\beta}_{SQLIML}/\hat{\beta}_{S2SLS}$ and $\hat{\rho}_{SQLIML}/\hat{\rho}_{S2SLS}$												
	$\hat{\beta}_{SQLIML}/\hat{\beta}_{S2SLS}$						$\hat{\rho}_{SQLIML}/\hat{\rho}_{S2SLS}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	0.6780	0.9713	0.9748	0.9531	0.9770	0.9907	0.0321	0.1402	0.2491	0.0744	0.1653	0.2822
$\lambda = 0.5$	0.9127	0.8276	0.9700	0.9367	0.9606	0.9761	0.0554	0.0542	0.1917	0.0637	0.1257	0.2044
$\lambda = 1$	0.9642	0.9009	0.9479	0.9185	0.8492	0.9825	0.0405	0.0480	0.1106	0.0117	0.0349	0.1349
$\lambda = 1.5$	0.6487	0.8910	0.9422	0.6981	0.9321	0.9796	0.0140	0.0363	0.0807	0.0114	0.0438	0.0847

and RMSE values as λ increases and as c decreases. In summary, our results show that the proposed SQLIML estimator outperforms the S2SLS estimator, especially for models with strong endogeneity and weak instruments.

Table 3.5 displays the size and power of the exogeneity test with a significance level of 5%. As can be seen, for all cases, the empirical sizes are very close to the nominal value of 5%. Turning now to the power results, we find that $\boldsymbol{X}$ is detected as endogenous even when λ is small. In particular, the power of this test increases as c, λ, N, and L increase. Overall, our results indicate the correct size and good power of our exogeneity test.

Table 3.2: Results on Monte Carlo Simulation with 4 Instrument

	Panel A: Average Bias of $\hat{\beta}_{SQLIML}$ and $\hat{\rho}_{SQLIML}$											
	$\hat{\beta}_{SQLIML}$						$\hat{\rho}_{SQLIML}$					
	$N=500$			$N=1000$			$N=500$			$N=1000$		
	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$
$\lambda=0.1$	-0.0067	-0.0004	-0.0003	0.0019	-0.0017	0.0016	-0.0021	-0.0039	-0.0041	-0.0019	-0.0014	-0.0025
$\lambda=0.5$	0.0089	0.0052	0.0019	0.0048	0.0040	0.0015	-0.0040	-0.0043	-0.0049	-0.0028	-0.0016	-0.0008
$\lambda=1$	0.0276	0.0046	0.0019	0.0107	0.0039	0.0027	-0.0053	-0.0041	-0.0022	-0.0005	-0.0014	-0.0016
$\lambda=1.5$	0.0312	0.0096	0.0036	0.0143	0.0053	0.0029	-0.0049	-0.0011	-0.0011	0.0002	-0.0001	-0.0009
	Panel B: Average Bias of $\hat{\beta}_{S2SLS}$ and $\hat{\rho}_{S2SLS}$											
	$\hat{\beta}_{S2SLS}$						$\hat{\rho}_{S2SLS}$					
	$N=500$			$N=1000$			$N=500$			$N=1000$		
	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$
$\lambda=0.1$	-0.0162	-0.0043	-0.0022	-0.0030	-0.0042	0.0007	0.1186	0.0515	0.0199	0.0679	0.0302	0.0096
$\lambda=0.5$	0.0165	0.0082	0.0021	0.0074	0.0041	0.0016	0.1552	0.0661	0.0320	0.0959	0.0419	0.0158
$\lambda=1$	0.0594	0.0131	0.0049	0.0244	0.0081	0.0033	0.1880	0.1115	0.0520	0.1355	0.0664	0.0283
$\lambda=1.5$	0.0787	0.0242	0.0075	0.0374	0.0111	0.0045	0.2381	0.1562	0.0810	0.1862	0.0975	0.0421
	Panel C: Relative Efficiency of $\hat{\beta}_{SQLIML}/\hat{\beta}_{S2SLS}$ and $\hat{\rho}_{SQLIML}/\hat{\rho}_{S2SLS}$											
	$\hat{\beta}_{SQLIML}/\hat{\beta}_{S2SLS}$						$\hat{\rho}_{SQLIML}/\hat{\rho}_{S2SLS}$					
	$N=500$			$N=1000$			$N=500$			$N=1000$		
	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$	$c=0.3$	$c=0.5$	$c=0.8$
$\lambda=0.1$	0.9922	1.0023	0.9872	1.0009	0.9925	0.9872	0.2651	0.3906	0.5447	0.2408	0.4098	0.5688
$\lambda=0.5$	0.9859	0.9949	0.9843	0.9604	0.9880	0.9839	0.2007	0.3272	0.4788	0.2019	0.3213	0.4648
$\lambda=1$	0.9286	0.9844	0.9796	0.9782	0.9706	0.9877	0.1532	0.2227	0.3273	0.1411	0.2232	0.3337
$\lambda=1.5$	0.9457	0.9864	0.9947	0.9634	0.9797	0.9835	0.1168	0.1481	0.2197	0.0898	0.1391	0.2320

Table 3.3: Results on Monte Carlo Simulation with 8 Instruments

Panel A: Average Bias of $\hat{\beta}_{SQLIML}$ and $\hat{\rho}_{SQLIML}$												
	$\hat{\beta}_{SQLIML}$						$\hat{\rho}_{SQLIML}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	0.0025	-0.0023	0.0003	0.0010	0.0003	0.0002	-0.0068	-0.0052	-0.0035	-0.0026	-0.0026	-0.0013
$\lambda = 0.5$	0.0154	0.0059	0.0010	0.0099	0.0025	0.0008	-0.0067	-0.0035	-0.0026	-0.0020	-0.0014	-0.0002
$\lambda = 1$	0.0264	0.0111	0.0024	0.0159	0.0065	0.0019	-0.0046	-0.0023	-0.0021	-0.0008	-0.0014	-0.0012
$\lambda = 1.5$	0.0401	0.0141	0.0055	0.0225	0.0086	0.0054	-0.0030	-0.0018	-0.0014	-0.0027	-0.0024	-0.0005

Panel B: Average Bias of $\hat{\beta}_{S2SLS}$ and $\hat{\rho}_{S2SLS}$												
	$\hat{\beta}_{S2SLS}$						$\hat{\rho}_{S2SLS}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	-0.0060	-0.0059	-0.0016	-0.0044	-0.0020	-0.0006	0.1249	0.0612	0.0281	0.0774	0.0331	0.0138
$\lambda = 0.5$	0.0248	0.0089	0.0018	0.0130	0.0037	0.0012	0.1465	0.0817	0.0395	0.1033	0.0475	0.0216
$\lambda = 1$	0.0571	0.0212	0.0053	0.0289	0.0106	0.0033	0.1940	0.1212	0.0641	0.1456	0.0749	0.0314
$\lambda = 1.5$	0.0913	0.0301	0.0106	0.0443	0.0145	0.0073	0.2369	0.1641	0.0904	0.1937	0.1132	0.0508

Panel C: Relative Efficiency of $\hat{\beta}_{SQLIML}/\hat{\beta}_{S2SLS}$ and $\hat{\rho}_{SQLIML}/\hat{\rho}_{S2SLS}$												
	$\hat{\beta}_{SQLIML}/\hat{\beta}_{S2SLS}$						$\hat{\rho}_{SQLIML}/\hat{\rho}_{S2SLS}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	0.9899	0.9918	0.9914	0.9814	0.9945	0.9992	0.2974	0.4850	0.6771	0.3181	0.5270	0.6831
$\lambda = 0.5$	0.9539	0.9703	0.9916	0.9598	0.9874	0.9896	0.2509	0.3601	0.5258	0.2327	0.3862	0.5596
$\lambda = 1$	0.8900	0.9456	0.9802	0.9248	0.9681	0.9876	0.1775	0.2241	0.3596	0.1437	0.2423	0.3995
$\lambda = 1.5$	0.8520	0.9421	0.9723	0.9103	0.9563	0.9769	0.1279	0.1630	0.2461	0.1056	0.1616	0.2572

Table 3.4: Results on Monte Carlo Simulation with 16 Instruments

	Panel A: Average Bias of $\hat{\beta}_{SQLIML}$ and $\hat{\rho}_{SQLIML}$											
	$\hat{\beta}_{SQLIML}$						$\hat{\rho}_{SQLIML}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	0.0014	2.14E-05	0.0003	0.0020	0.0007	0.0012	-0.0044	-0.0042	-0.0034	-0.0015	-0.0013	-0.0020
$\lambda = 0.5$	0.0149	0.0069	0.0030	0.0089	0.0023	0.0003	-0.0045	-0.0022	-0.0021	-0.0020	-0.0016	-0.0008
$\lambda = 1$	0.0308	0.0123	0.0050	0.0158	0.0068	0.0019	-0.0023	-0.0020	-0.0008	-0.0012	-0.0018	-0.0011
$\lambda = 1.5$	0.0479	0.0175	0.0085	0.0226	0.0077	0.0031	-0.0012	-0.0010	-0.0014	-0.0011	-0.0009	-0.0014
	Panel B: Average Bias of $\hat{\beta}_{S2SLS}$ and $\hat{\rho}_{S2SLS}$											
	$\hat{\beta}_{S2SLS}$						$\hat{\rho}_{S2SLS}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	-0.0081	-0.0036	-0.0012	-0.0036	-0.0015	0.0004	0.1249	0.0629	0.0267	0.0823	0.0346	0.0135
$\lambda = 0.5$	0.0246	0.0101	0.0043	0.0133	0.0037	0.0008	0.1528	0.0874	0.0416	0.1077	0.0519	0.0231
$\lambda = 1$	0.0616	0.0225	0.0085	0.0312	0.0114	0.0036	0.1970	0.1241	0.0673	0.1506	0.0767	0.0368
$\lambda = 1.5$	0.0965	0.0338	0.0146	0.0459	0.0149	0.0054	0.2358	0.1681	0.0928	0.1956	0.1137	0.0556
	Panel C: Relative Efficiency of $\hat{\beta}_{SQLIML}/\hat{\beta}_{S2SLS}$ and $\hat{\rho}_{SQLIML}/\hat{\rho}_{S2SLS}$											
	$\hat{\beta}_{SQLIML}/\hat{\beta}_{S2SLS}$						$\hat{\rho}_{SQLIML}/\hat{\rho}_{S2SLS}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$\lambda = 0.1$	0.9953	0.9906	0.9973	0.9880	0.9901	1.0017	0.3123	0.4901	0.6905	0.3096	0.5270	0.7547
$\lambda = 0.5$	0.9294	0.9598	0.9740	0.9746	0.9901	0.9850	0.2423	0.3478	0.5088	0.2451	0.4016	0.5649
$\lambda = 1$	0.7940	0.9009	0.9544	0.8725	0.9398	0.9813	0.1753	0.2393	0.3465	0.1558	0.2471	0.3844
$\lambda = 1.5$	0.7498	0.8675	0.9340	0.8390	0.9285	0.9624	0.1306	0.1664	0.2389	0.1041	0.1588	0.2596

Table 3.5: Size and Power of Exogeneity Test

	$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$L = 1$						
$\lambda = 0$	0.041	0.044	0.042	0.031	0.043	0.046
$\lambda = 0.1$	0.073	0.148	0.251	0.103	0.228	0.374
$\lambda = 0.5$	0.645	0.960	0.999	0.929	0.998	1
$\lambda = 1$	0.994	1	1	1	1	1
$\lambda = 1.5$	1	1	1	1	1	1
$L = 4$						
$\lambda = 0$	0.049	0.053	0.036	0.046	0.046	0.052
$\lambda = 0.1$	0.188	0.338	0.439	0.278	0.492	0.629
$\lambda = 0.5$	0.989	0.999	1	1	1	1
$\lambda = 1$	1	1	1	1	1	1
$\lambda = 1.5$	1	1	1	1	1	1
$L = 8$						
$\lambda = 0$	0.046	0.053	0.052	0.042	0.048	0.047
$\lambda = 0.1$	0.276	0.414	0.491	0.412	0.581	0.695
$\lambda = 0.5$	1	1	1	1	1	1
$\lambda = 1$	1	1	1	1	1	1
$\lambda = 1.5$	1	1	1	1	1	1
$L = 16$						
$\lambda = 0$	0.045	0.050	0.049	0.054	0.044	0.056
$\lambda = 0.1$	0.364	0.475	0.530	0.548	0.669	0.741
$\lambda = 0.5$	0.998	1	1	1	1	1
$\lambda = 1$	1	1	1	1	1	1
$\lambda = 1.5$	1	1	1	1	1	1

3.5 Application: A Spatial Lag Model of DUI Arrest Rates

In this section, we revisit the SLM model of DUI arrest rates by Drukker, Prucha, and Raciborski (2013), which studies the impact of alcohol prohibition in a county on the DUI arrest rates, to demonstrate the use of the proposed SQLIML estimation. In particular, we apply the SQLIML procedure to re-estimate the DUI arrest model. The model is organized as follows

$$dui = \rho \boldsymbol{W} \times dui + \beta_0 + \beta_1 \times nondui + \beta_2 \times dry + \beta_3 \times vehicles + \beta_4 \times police + \boldsymbol{u}, \quad (3.5.1)$$

where *dui* is the alcohol-related arrest rate per 100,000 daily vehicle miles traveled (DVMT), *nondui* is the nonalcohol-related arrest rate per 100,000 DVMT, *dry* is the dummy variable for a county that prohibits alcohol sales within its borders, *vehicles* is the number of registered vehicles per 1,000 residents, and *police* is the number of sworn officers per 100,000 DVMT. $\boldsymbol{W}$ is the contiguity-based spatial weighting matrix for the U.S. counties. The data used covers 1,422 US counties in 2008 and can be found at *http://www.stata-press.com/data/r15/dui_southern.dta*.

Drukker, Prucha, and Raciborski (2013) treat *police* as endogenous since the size of the police force may be a function of the DUI arrest rate. To deal with the endogeneity of *police*, Drukker, Prucha, and Raciborski (2013) use the variable *election*, which is a dummy variable equal to 1 if a county government faces an election and 0 otherwise, as the instrument, and apply the S2SLS procedure for estimating the DUI arrest model.

In Table 3.6, we provide S2SLS and SQLIML estimates of the DUI arrest model, as well as their corresponding MSE values. Our interests are the coefficients of the spatial lag term and *police*. In both S2SLS and SQLIML results, the estimate of ρ is significantly positive, suggesting that the alcohol-related arrest rate in a county directly affects the alcohol-related arrest rate in its neighboring counties. As expected, the *police* coefficient indicates that the increase in the proportionate number of sworn officers leads to a reduction in the arrest rate. Consistent with Drukker, Prucha, and Raciborski (2013), all independent variables are significant with the exception of *nondui*.

Importantly, the standard errors of the coefficients on the spatial lag term and the *police* variable estimated via the SQLIML procedure are smaller than those estimated via the S2SLS procedure. These results are consistent with the Monte Carlo simulation results, suggesting that the SQLIML coefficients on the spatial lag term and the endogenous regressors enjoy larger efficiency gains. Additionally, the significance of λ validates the strong endogeneity of the *police* variable. The MSE values indicate that the use of SQLIML estimation improves the estimation accuracy.

Table 3.6: Estimation Results on the DUI Arrest Model

	S2SLS Estimates		SQLIML Estimates	
	coef	std err	coef	std err
police	-1.3233***	0.1211	-1.5138***	0.0198
nondui	-0.0019	0.0027	-0.0024	0.0035
vehicles	0.0919***	0.0048	0.1007***	0.0011
dry	0.4832***	0.0800	0.4477***	0.1206
constant	9.2390***	1.1334	14.1083***	0.2404
$\boldsymbol{W} \times dui$	0.3909***	0.0205	0.2150***	0.0035
λ			1.2581***	0.0097
MSE	0.7981		0.2630	

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variables are alcohol-related arrest rates per 100,000 DVMT. *police* is the number of sworn officers per 100,000 DVMT, *nondui* is the nonalcohol-related arrest rate per 100,000 DVMT, *vehicles* is the number of registered vehicles per 1,000 residents, *dry* is the dummy variable for a county that prohibits alcohol sales within its border. $\boldsymbol{W}$ is the contiguity-based spatial weighting matrix for the U.S. counties ($N = 1,422$). The variable *election* is an instrument for the endogenous variable *police*, where *election* is a dummy variable equal to 1 if a county government faces an election and 0 otherwise. λ captures the severity of endogeneity of *police*.

3.6 Discussion

The aim of this paper is to introduce the SQLIML estimation of a model including the spatial lag and additional endogenous variables. We derive the asymptotic properties of the SQLIML estimator and discuss its finite sample properties via a Monte Carlo simulation. The Monte Carlo simulation results indicate that the SQLIML estimator performs better than the S2SLS estimator, especially for models with strong endogeneity and weak instruments. We demonstrate the usefulness of our proposed estimation in an application to revisit the DUI arrest rate model of Drukker, Prucha, and Raciborski (2013).

There are several attractive features of the SQLIML estimation with respect to the S2SLS estimation. First, one main drawback of the S2SLS estimation method is that the instruments for the spatial lag variable are required. In the S2SLS procedure, the optimal instrument matrix for the spatial lag variable consists of the spatial lags of all exogenous variables in the model, but the S2SLS estimation may break down when all the spatial regressors are really irrelevant. Instead of applying the IV approach, we rely on the ML estimation procedure to deal with the endogeneity caused by the spatial lag variable by computing the log-determinant of the Jacobian matrix in the log-likelihood function. Second, our proposed SQLIML estimation provides a test of exogeneity of $\boldsymbol{X}_1$.

However, there is a limitation of the SQLIML estimation. The SQLIML estimation relies on the assumption that there must be a linear relationship between the reduced form errors of $\boldsymbol{X}_1$ and the errors in the SLM model. If this assumption is not satisfied, see Tables B.1-B.3, our proposed SQLIML estimator may be inefficient and the proposed exogeneity test may suffer from size distortion.

Chapter 4

The Spatial Effect of the Minimum Wage on the Gender Wage Gap

4.1 Introduction

Alleviation of the gender wage gap holds significance for government authorities and policymakers. Past decades have witnessed great efforts to close the gender wage gap in the U.S., yet it still prevails. Recent U.S. Census Bureau data shows that the 2018 female-to-male earnings ratio among full-time and year-round workers was 0.816 in the U.S., which increased 34.4% from 0.607 in 1960 (Figure 4.1).

Evidence has shown that minimum wage policy can be an instrument to narrow the gender wage gap, as it can compress the lower tail of the wage distribution (Dinardo, Fortin, and Lemieux, 1996; Blau and Kahn, 2003). Given the over-representation of women in low-wage occupations and sectors, minimum wages are thus expected to have the largest impact on the gender wage gap at bottom of the wage distribution. However, the size of the minimum wage effect on the gender wage gap at different income levels is poorly understood. In this paper, we contribute to the literature by quantifying the extent to which minimum wage hikes affect the gender wage gap across income levels.

Another contribution of this study concerns the methodology. One limitation of models applied in previous studies on minimum wages and gender wage gaps is that they fail to consider the potential spatial dependence in data, such as the fixed-effect model (Blau and Kahn, 1997, 2003) and the difference-in-difference model (Robinson, 2005). However, it is commonly observed that geographical units are not independent but spatially dependent. As stated by Tobler (1970), "everything is related to everything else, but near things are more related than distant things" (Tobler's First Law of Geography). It is important to consider spatial correlation when analyzing observations relating to regions. Failure to capture the potential spatial correlation can lead to biased and inconsistent estimation results. To address this issue, we use the spatial econometric model that accommodates spatial dependence to conduct our analysis.

The second research question of this study concerns the spillover effect of the minimum wage on the gender wage gap in the U.S. states, which can be answered with the spatial econometric model. Investigating spillover effects is crucial in assessing the impact of the minimum wage policy and is fundamental to understanding the jurisdictional interactions on minimum wages. However, surprisingly, so far, no study has tackled this issue. By answering this question, our study contributes to the literature by filling this gap. If minimum wage spillover effects do exist, then the implication is that local governments should facilitate ties in implementing minimum wage policy to create better ways to reduce inequality.

There are several reasons that one state's minimum wage might affect the gender wage gap of its neighboring states. First, low-wage workers might be attracted to commute across state borders to neighboring states that have higher minimum wages (Johnson, 2014). This implies that low-wage workers in one state can benefit from a higher minimum wage in its neighbors, which in turn contributes to the narrowing of the gender wage gap in that state. The second mechanism is through inter-jurisdictional competition on minimum wages. States are clearly concerned about how their minimum wage rates compare with those of their neighbors. It is likely due to the fear of driving away workers to neighboring states that have higher minimum wages. The competition for workers may force states to increase their minimum wages in response to higher minimum wages in neighboring states. Another reason state minimum wages move in reaction to changes in other states is that voters may judge incumbent behavior relative to the behavior of neighboring politicians. It is known as the "yardstick competition" (Besley and Case, 1995). Currently, every state in the US directly elects local legislators and governors, and voters evaluate their incumbents based on performance comparison with neighboring states. If labor standards are an important part of performance evaluation, incumbents would be motivated to compete on minimum wages.

The main objective of this paper is to quantify the extent to which the gender wage gap at different income levels responds to in-state and out-of-state minimum wage changes. Using state-level data from 1979 to 2018, our analysis uncovers the presence of the spillover effect of the minimum wage on neighboring states. In particular, an increase in the minimum wage in a given state contributes to closing the 10th-60th income percentile gender wage gap in its own state and its neighboring states, with the greatest impact on low-income workers. Specifically, a 1% point increase in the minimum wage is expected to reduce the 10th percentile gender wage gap by 0.00246 units, with a decrease of 0.00151 units in its own state and a decrease of 0.00095 units in its neighboring states. This finding supports the existing literature suggesting that low-income women can benefit the most from minimum wage increases. However, our results reveal that the minimum wage is insignificantly related to the gender wage gap in the upper tail of the wage distribution.

The rest of this paper is organized as follows. In Sections 4.2 and 4.3, we summarize the previous literature and introduce the minimum wage policy in the US. In Section 4.4, the data on gender wage gaps and minimum wages are introduced. In Section 4.5, we discuss the spatial model selection approach, present empirical results, present empirical results,

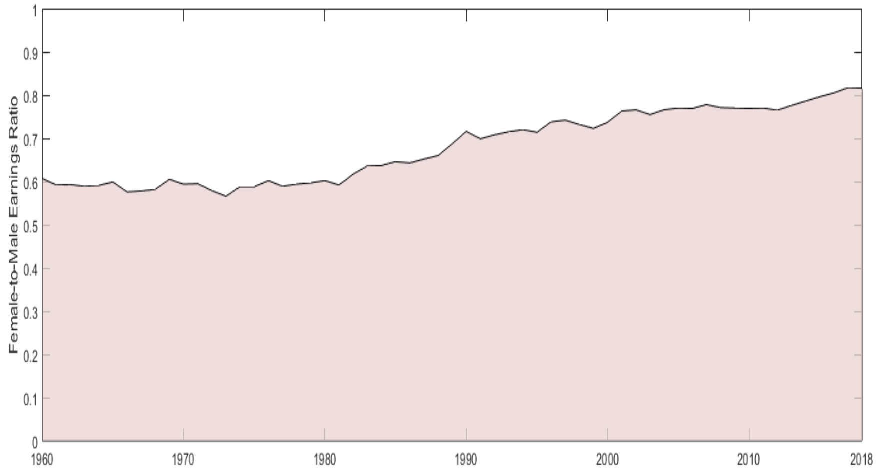

Figure 4.1: Female-to-Male Earnings Ratio of Full-Time and Year-Round Workers in the U.S., 1960-2018
Source: U.S. Census Bureau, Current Population Survey.

and perform the dynamic analysis. In Section 4.6, we conduct the robustness check, and the last section concludes.

4.2 Literature Review

Recent studies have demonstrated that closing gender wage gaps can bring economic benefits (Schober and Winter-Ebmer, 2011; Cavalcanti and Tavares, 2016). The fact that gender wage gap suppresses women's lifetime wage, leading to higher poverty rates among women. Hartmann, Hayes, and Clark (2014) summarize the analysis of the 2010-2012 Current Population Survey Annual Social and Economic Supplement and briefly explain how gender pay equality can reduce poverty and enhance the U.S. economy. They find that increasing women's wages to match men's wages can drastically cut the poverty rate for all working women, working single mothers, and single women from 8.1% to 3.9%, 28.7% to 15%, and 11% to 4.6%, respectively. Besides, the U.S. economy would produce an additional income of $447.6 billion if women receive equal pay; this represents 2.9% of the 2012 U.S. GDP. It can be explained by the fact that earnings inequality for women leads to the misallocation of human capital and causes women to work at less productive firms than they otherwise would, thereby hindering economic growth.

Much research has attempted to explore the factors that contribute to gender wage inequality. First, discrimination may explain why women are paid less than men (Cain, 1986; Altonji and Blank, 1999; Hallward-Driemeier, Rijkers, and Waxman, 2017). A 2017 Pew Research Center survey finds that women (42%) in the U.S. are roughly twice as likely as men (22%) to say they have suffered gender discrimination at work and about 25% of working women said they had been paid less than a man who was doing the same job, but only 5% of working men reported having a similar experience (Parker and Funk, 2017). Second, women's concentration in low-paid occupations may play a role in shaping the gender wage gap (Gunderson, 1975; England, 1992; Korkeamäki and Kyyrä, 2006). The U.S. Bureau of Labor Statistics reports that among workers paid at hourly rates in 2019, about 3% of women with a wage at or below the federal minimum wage compared with 1% of men (BLS, 2020). Third, labor-force attachment and human capital (schooling and work experience) may be important factors contributing to the gender wage gap (Mincer and Polachek, 1974; Goldin, 2014; Blau and Kahn, 2017). Blau and Kahn (2017) point out that due to family responsibilities, women are expected to experience shorter and discontinuous working lives, and thus have a lower motivation to invest in on-the-job training than men. As a result, the smaller human-capital investments and reduced labor-force experience will lower women's wages.

There are a number of studies suggesting that the minimum wage policy is an effective tool to reduce the wage differential between men and women. For example, Blau and Kahn (1997) indicate that the sharp decline in the real value of the federal minimum wage is an institutional factor explaining the widening gender wage gap among low-skill workers in the U.S. from 1979-1988. A follow-up analysis (Blau and Kahn, 2003) for 22 countries from 1985-1994 uncovers an inverse association between the gender wage gap and the "bite" of the minimum wage - the minimum wage level relative to the average wage. Similarly, Angel-Urdinola (2008) runs the set of simulations and finds that the introduction of a minimum wage in Macedonia decreases the gender wage gap from 15% to 23%.

Evidence also shows that minimum wage effects on the gender wage gap vary across the wage distribution. For instance, Robinson (2005), using the difference-in-difference estimation, finds that the National Minimum Wage (NMW) impact on the gender wage gap in British is greatest for workers paid below the NMW. Hallward-Driemeier, Rijkers, and Waxman (2017) use manufacturing firm-level census data from 1995 to 2006 to examine the minimum wage role in gender wage gaps among production workers in Indonesia. They uncover that the minimum wage effects on the gender wage gap are highly associated with educational attainment and wage distribution. In particular, minimum wage increases are correlated to exacerbated gender wage gaps among the least-educated workers.

4.3 Minimum Wage Policy in the U.S.

The U.S. federal minimum wage was first introduced in 1938, and since then, it was increased several times by Congress. Figure 4.2 shows the nominal and real value of the federal minimum wage in the U.S. from 1979 to 2018. As can be seen, the U.S. has witnessed a continuous increase in the nominal value of the federal minimum wage. In 1979, the U.S. federal minimum wage was set at $2.90 per hour, and in 2018, it was raised to $7.25 per hour. The real minimum wage increased in jumps when the nominal minimum wage was increased; however, it decreased over time as it was adjusted for inflation. Although the nominal minimum wage was at a historical high in 2018, the 2018 real minimum wage ($7.25) is about 38.3% lower than the 1979 real minimum wage ($10.03) when adjusted for 2018 inflation. The real value of the federal minimum wage got its highest point in 1979 and fell to its lowest point in 2007 during the whole sample years.

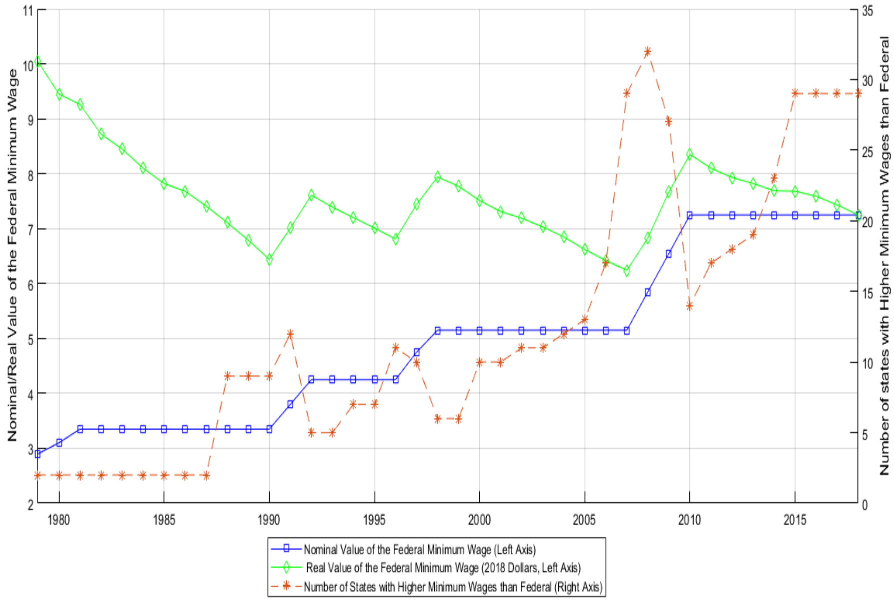

Figure 4.2: Real and Nominal Value of the Federal Minimum Wage in the U.S., 1979-2018

There is a federal minimum wage for all states in the U.S., but each state is free to set its own minimum wage. One aim of the minimum wage is to improve the living standards of workers at the lower tail of the wage distribution. Since the U.S. living standards vary across states, the U.S. does not set one minimum wage applicable to all states. Instead, many states set their own minimum wages. By 1979, 41 out of 50 states had set their own minimum wage, and by 2018, that number had risen to 45.

State minimum wages can be higher than, equal to, or lower than the federal minimum wage. When states have no minimum wage law or a minimum wage lower than the federal minimum wage, the employee is entitled to the federal minimum wage. When states have a minimum wage higher than the federal one, the employee is entitled to the state minimum wage. States increase their minimum wages automatically based on the local cost of living or due to approved state legislation or citizen ballot initiatives. There is an apparent increase in the number of states with minimum wages higher than the federal minimum wage between 1979 and 2018 (Figure 4.2). By 1979, only 2 out of 50 states with minimum wages higher than the federal minimum wage, and by 2018, that

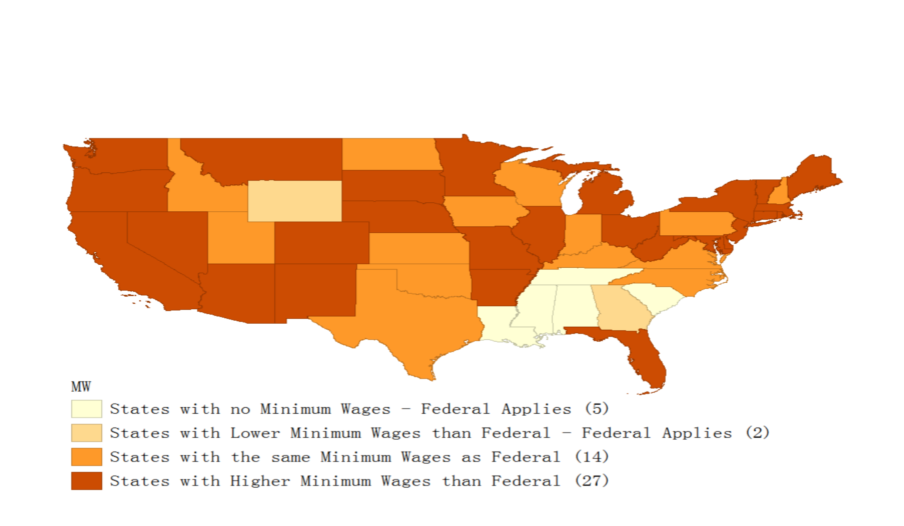

Figure 4.3: Spatial Distribution of the State Minimum Wage in 2018

number had risen to 29.

Figure 4.3 illustrates the spatial distribution of the state minimum wage in 2018 (excluding Alaska and Hawaii). The figure shows that the level of the minimum wage varies considerably between states. In 2018, the state minimum wage was highest in Washington ($11.5 per hour), Massachusetts ($11 per hour), and California ($11 per hour), and it was lowest in Georgia ($5.15 per hour) and Wyoming ($5.15 per hour).

4.4 Data and Descriptive Statistics

4.4.1 Gender wage Gap

Data on U.S. household earnings from 1979 to 2018 is collected from the Current Population Survey Merged Outgoing Rotation Group (CPS MORG), which contains individual data on hours worked, earnings, and some demographic information on sex, age, schooling, marital status, etc. We define hourly wages as reported hourly wages for those who are paid by the hour and weekly earnings divided by hours worked in the prior week for those who are not. Following the study of Autor, Manning, and Smith (2016), we exclude self-employed individuals, regardless of whether their businesses are

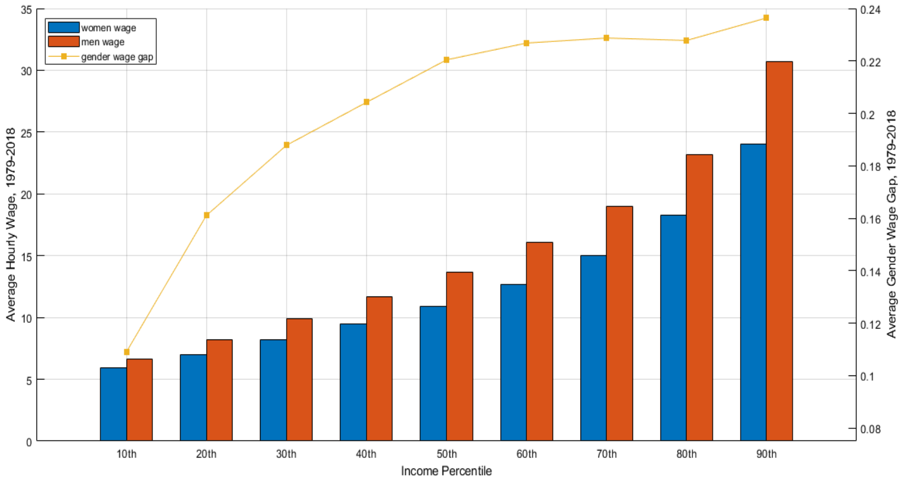

Figure 4.4: Average Hourly Wages for Men and Women and the Average Gender Wage Gap for Selected Income Percentiles, U.S., 1979-2018

incorporated, and exclude those with earnings imputed by the BLS, and we multiply top-coded weekly earnings by 1.5. Using these microdata, we calculate the 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, and 90th percentiles of state wage distributions by sex using the CPS sampling weight multiplied by weekly hours worked, and then calculate their corresponding gender wage gaps in the 48 contiguous states from 1979-2018. Table 4.1 reports summary statistics for this data. We categorize the 10th, 20th, and 30th percentiles as low-income percentiles, the 40th, 50th, and 60th percentiles as middle-income percentiles, and the 70th, 80th, and 90th percentiles as high-income percentiles.

Figure 4.4 plots U.S. average hourly wages for men and women by income percentile, as well as the average gender wage gap from 1979 to 2018. As expected, women are paid less than peer men at every income level, and women at different income levels experience differing gaps in pay. On average, the gender wage gap is largest among high-paid workers and lowest among low-paid workers. Specifically, over the period 1979 to 2018, the gender wage gap is largest at the 90th income percentile, where women make only around \$0.7636 for every dollar a peer man makes, while it is smallest at the 10th percentile, where women make only around \$0.8910 for every dollar a peer man makes.

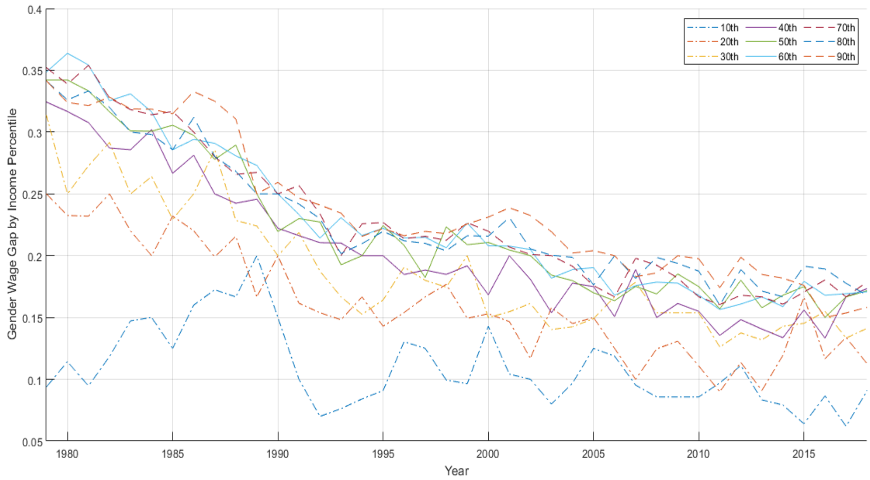

Figure 4.5: Gender Wage Gap for Selected Income Percentiles, U.S., 1979-2018

But how does the gender wage gap vary over time? To answer this question, we plot the gender wage gap by income percentile in the U.S. from 1979 to 2018 (Figure 4.5). Here we see that the gender wage gap decreased gradually at all selected percentiles from 1979 to 2018. The largest reduction occurred at the 90th percentile, where the gender wage gap fell 53.7% to 0.158. The 60th percentile is next, with a 51.0% decrease to 0.171. The smallest reduction occurred at the 10th percentile, with a 3.2% decrease to 0.091.

4.4.2 Effective Minimum Wage

The U.S. federal minimum wage was flat at $7.25 per hour from 2010-2018 (Figure 4.2). However, since the wage levels and consumer price index varies in the 2010s, this minimum wage rate of $7.25 should have heterogeneous effects during this period. Similarly, since the cost of living varies across states and the same rate of the minimum wage may also have different effects on states that differ in terms of labor productivity, prices, or wage levels, a direct comparison of absolute levels of the minimum wage is not meaningful.

In this paper, following Lee's (1999) study, we use the effective minimum wage as a proxy for the bindingness of the minimum wage. The effective minimum wage is defined to be the log difference between the median hourly wage and the applicable federal or

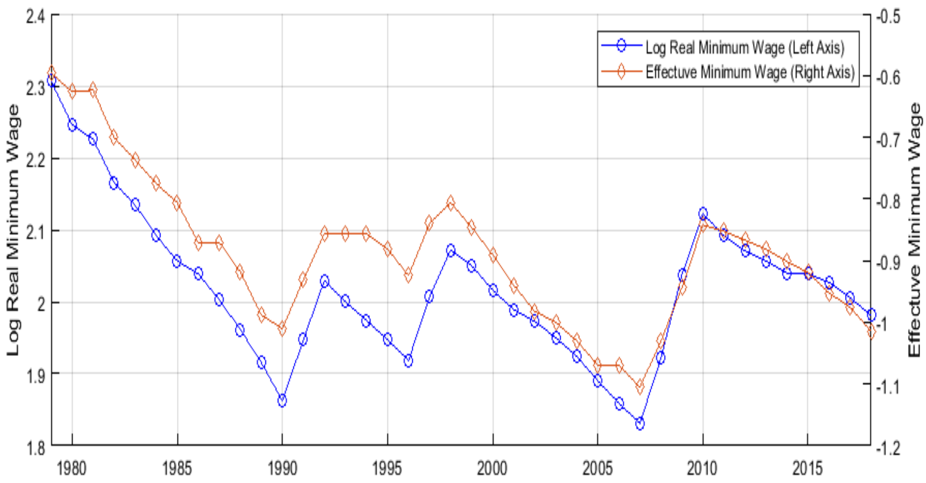

Figure 4.6: Real Value of the Federal Minimum Wage and the Effective Minimum Wage, U.S., 1979-2018

state minimum wage. Autor, Manning, and Smith (2016) suggest that, in this setup, a more binding minimum wage is a minimum wage that is closer to the median wage, and it potentially affects a larger number of employees, while a less binding minimum wage is a minimum wage that is a long way from the median wage and its impact therefore potentially low. Figure 4.6 shows the real value of the federal minimum wage (2018 dollars) and the effective minimum wage in the U.S. from 1979 to 2018. As can be seen, the real minimum wage shares quite a similar trend with the effective minimum wage.

Data on minimum wages are collected from the U.S. Department of Labor. Demographic data are collected from the Centers for Disease Control and Prevention (CDC). The sample includes the 48 contiguous states from 1979 to 2018. Summary statistics for this data are presented in Table 4.1.

Table 4.1: Summary Statistics

	Mean	Std. Dev.
Gender Wage Gap		
p(10)	0.113	0.050
p(20)	0.166	0.060
p(30)	0.195	0.065
p(40)	0.214	0.069
p(50)	0.226	0.073
p(60)	0.233	0.074
p(70)	0.235	0.073
p(80)	0.232	0.071
p(90)	0.235	0.074
State Effective Minimum Wage	-0.821	0.157
State Unemployment Rate (%)	5.873	2.085
State Population Shares		
15-44 years old	0.436	0.034
45-64 years old	0.223	0.037
65-84 years old	0.115	0.017
Female	0.510	0.007
White	0.854	0.094

4.5 Empirical Framework

4.5.1 Testing for Spatial Dependence in the Gender Wage Gap

Recent studies focusing on the spatial distribution of the gender wage gap have demonstrated that the gender wage gap may be spatially autocorrelated. For example, using the Local Indication of Spatial Association (LISA) analysis, Manesh et al. (2020) find that the gender wage gap in construction-related occupations in the U.S. is spatially autocorrelated at local levels.

The gender wage gap may be spatially autocorrelated due to geographic reasons. There are several ways in which geography affects the gender wage gap. For example, Wiseman and Dutta (2016) uncover that the median gender wage gap in the U.S. is significantly higher in states with greater belief and belonging, as religiosity is often associated with more traditional views about gender roles. McCall (1988, 2001) finds that regions in the

U.S. specializing in high-tech services and high-tech manufacturing experienced a widen average gender wage gap than low-tech ones.

In this subsection, we perform Moran's I test to verify the global spatial autocorrelation in the gender wage gap each year. This test was first introduced by Moran (1950). Moran's I statistic measures the correlation coefficient for the relationship between the gender wage gap in a state and the gender wage gap in its surrounding states. To conduct this test, we first need to construct a spatial weight matrix W to measure state neighborliness. The Moran's I statistic is defined as

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(G_i - \bar{G})(G_j - \bar{G})}{\sum_i (G_i - \bar{G})^2},$$

where n is the number of states (i.e., $n = 48$), w_{ij} is the (i,j)th element of W, G_i is the gender wage gap in state i, and $\bar{G}$ is the average gender wage gap for the 48 states. Moran's I has an expected value and standard deviation under the normality assumption. Based on this assumption, Moran's I values can be transformed to z-scores and the corresponding p-values can be calculated. The significance of Moran's I values indicates that the data is spatially autocorrelated.

In this paper, we use the queen contiguity-based spatial weight matrix, which is defined by $w_{ij} = 1$ if states i and j have a common vertex and $w_{ij} = 0$ otherwise, for our analysis. The Moran's I statistic results on the gender wage gap are presented in Table 4.2. Our results indicate a positive spatial autocorrelation in the 10th-70th percentile gender wage gap in most sample years and indicate a non-negligible significant spatial autocorrelation in the 80th and 90th percentile gender wage gap in a few sample years among U.S. states. Because so many periods have significant autocorrelation, it is necessary to use a spatial econometric model in the following analysis.

Table 4.2: Global Moran's I Statistic of the Gender Wage Gap for Selected Income Percentiles

Year	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
1979	0.310***	0.434***	0.444***	0.380***	0.388***	0.305***	0.233***	0.135*	-0.042
1980	0.228***	0.280***	0.299***	0.374***	0.313***	0.277***	0.149**	0.025	-0.017
1981	0.203**	0.210***	0.245***	0.248***	0.268***	0.219***	0.175**	0.035	-0.168*
1982	0.277***	0.226***	0.260***	0.248***	0.343***	0.274***	0.250***	0.054	-0.079
1983	0.380***	0.217***	0.431***	0.364***	0.245***	0.098	-0.030	0.017	-0.006
1984	0.349***	0.226***	0.247***	0.184**	0.258***	0.127*	-0.012	-0.058	-0.099
1985	0.059	0.082	0.152**	0.281***	0.293***	0.259***	0.155**	-0.046	-0.122
1986	0.161**	0.093	0.108*	0.174**	0.178**	0.169**	0.241***	0.044	-0.031
1987	0.335***	0.103*	0.110*	0.077	-0.023	0.120*	0.013	-0.093	-0.006
1988	-0.022	-0.045	-0.003	-0.066	-0.038	-0.021	-0.141	-0.179*	0.025
1989	0.054	0.145**	0.216***	0.138*	-0.047	-0.026	-0.018	0.022	0.109*
1990	0.130*	0.227***	0.247***	0.274***	0.244***	0.063	0.019	-0.021	-0.221**
1991	0.291***	0.314***	0.185**	0.214***	0.157**	0.102	0.044	-0.082	-0.076
1992	0.210***	0.114*	0.091	0.171**	0.097	0.061	0.118*	0.044	-0.016
1993	0.079	0.070	0.116*	0.144**	0.051	0.030	0.139*	0.118*	0.044
1994	0.095	0.086	0.119*	0.202**	0.186**	0.099	0.097	0.192**	0.122*
1995	0.270***	0.190**	0.281***	0.186**	0.196**	0.157**	0.202**	0.138**	0.117*
1996	0.196**	0.123*	0.232***	0.208***	0.224***	0.130*	0.103	0.085	-0.021
1997	0.252***	0.310***	0.267***	0.204***	0.266***	0.312***	0.186**	0.107*	-0.063
1998	0.096	0.065	0.030	0.075	0.146**	0.089	0.115*	-0.071	-0.206**
1999	0.127*	0.061	0.162**	0.086	0.058	0.190**	0.238***	0.147**	-0.020
2000	0.166**	0.028	0.019	-0.014	0.068	0.066	-0.013	-0.085	-0.232**
2001	0.023	0.124*	-0.031	0.090	0.109*	0.173**	0.022	-0.112	-0.219**
2002	0.113*	0.030	0.076	-0.062	-0.058	0.055	-0.080	-0.072	-0.144
2003	0.098	0.039	-0.028	-0.130	-0.033	-0.041	-0.142	-0.215**	-0.113
2004	0.146**	-0.079	0.087	-0.080	-0.075	-0.041	-0.098	-0.198**	-0.013
2005	0.145**	-0.002	0.011	0.047	0.094	-0.012	0.015	-0.052	0.128*
2006	0.209***	0.074	-0.031	-0.040	-0.086	-0.096	-0.122	-0.123	-0.036
2007	0.040	0.034	-0.078	-0.136	-0.108	-0.149*	-0.088	-0.108	-0.007
2008	0.032	-0.032	-0.020	-0.014	-0.055	-0.104	-0.003	-0.025	0.090
2009	-0.208**	0.010	-0.116	-0.089	-0.067	-0.032	-0.038	0.024	0.010
2010	0.119*	0.140**	-0.076	-0.024	-0.003	0.012	0.038	-0.073	0.072
2011	-0.022	0.047	-0.102	0.031	-0.054	-0.123	-0.088	-0.052	0.123*
2012	-0.080	-0.019	0.043	-0.042	-0.085	-0.025	-0.030	0.069	0.128*
2013	0.100	0.050	0.077	0.002	-0.146*	-0.034	-0.111	-0.007	0.177**
2014	0.073	0.238***	-0.047	0.044	0.018	0.038	0.025	0.127*	0.170**
2015	0.170**	-0.052	0.054	0.088	0.071	-0.005	0.067	0.038	0.083
2016	0.131*	0.024	0.041	-0.026	-0.090	-0.107	-0.166*	-0.097	0.111*
2017	0.136*	0.014	0.083	-0.001	-0.026	0.014	0.104	0.003	0.035
2018	0.020	0.024	0.085	0.006	-0.082	0.075	0.168**	0.221***	0.379***

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Moran's I measures overall spatial autocorrelation in the gender wage gap. We use the queen contiguity-based spatial weight matrix to conduct this analysis.

4.5.2 Empirical Model

Spatial Model Selection

Spatial econometrics can be used to cope with spatial effects. One can distinguish between two spatial effects, i.e. spatial dependence and spatial heterogeneity (Anselin, 1988). As suggested by Elhorst (2014), three different types of spatial models can be used to estimate the dependence between the observations, namely, the spatial lag model (SLM), the spatial error model (SEM), and the spatial Durbin model (SDM). Variables related to location, distance, and arrangement are treated in the spatial econometric model.

This paper starts by considering a SDM model to conduct our analysis. The SDM model is characterized by the hypothesis that the variation of the dependent variable for one agent can be explained by its neighboring dependent variables and its neighboring explanatory variables:

$$\begin{aligned} G_{st}(p) = {} & \alpha_s(p) + \gamma_t(p) + \delta(p) \sum_{j=1}^{48} w_{sj} \times G_{jt}(p) + \beta_1(p)[MW_{st} - w_{st}^{50}] + \beta_2(p)U_{st} + \beta_3(p)X_{st} \\ & + \theta_1(p) \sum_{j=1}^{48} w_{sj} \times [MW_{jt} - w_{jt}^{50}] + \theta_2(p) \sum_{j=1}^{48} w_{sj} \times U_{jt} + \theta_3(p) \sum_{j=1}^{48} w_{sj} \times X_{jt} + u_{st}(p), \end{aligned} \tag{4.5.1}$$

where $G_{st}(p)$ denotes the gender wage gap at percentile p in state s at time t, $MW_{st} - w_{st}^{50}$ is the effective minimum wage, where MW_{st} is the log minimum wage, and w_{st}^{50} is the log median hourly wage for that state-year. State-specific fixed effects are represented by α_s; time-period fixed effects are represented by γ_t. X is a vector of demographic characteristics controlling for the share of the state population who were female, white, and aged 15-44, 45-64, and 65-84 years old. To control for much of the economic fluctuations, we also add the state-level log unemployment rate, U_{st}, as a control. The model (4.5.1) is estimated by first transforming the model variables to remove two-way fixed effects, and then using the maximum likelihood method on these transformed variables.

The model (4.5.1) also controls for a spatially lagged dependent variable, $W \times G$, and spatially lagged explanatory variables, $W \times [MW - w^{50}]$, $W \times X$, and $W \times U$. Through spatially lagged dependent variables, the level of the gender wage gap of a particular state is jointly determined with that of neighboring states (Anselin, Le Gallo, and Jayet, 2006). Through spatially lagged explanatory variables, the level of the gender wage gap of a particular state depends on the explanatory variables of its neighboring states.

One advantage of the SDM is that it can obtain consistent estimators, even if the true data-generating process is an SLM or an SEM (Elhorst, 2014). The SLM model involves dependence in the dependent variable ($W \times G$), and the SEM model involves dependence in the disturbances ($W \times \varepsilon$). Define $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$. First, when $\boldsymbol{\theta} = \mathbf{0}$, it leads to the SLM model:

$$G_{st} = \alpha_s + \gamma_t + \delta \sum_{j=1}^{48} w_{sj} \times G_{jt} + \beta_1[MW_{st} - w_{st}^{50}] + \beta_2 U_{st} + \beta_3 X_{st} + u_{st}. \quad (4.5.2)$$

Second, when $\boldsymbol{\theta} = -\delta\boldsymbol{\beta}$, the model (4.5.1) can be written as

$$\begin{aligned} G_{st} = \;& \alpha_s + \gamma_t + \delta \sum_{j=1}^{48} w_{sj}\{G_{jt} - \beta_1[MW_{jt} - w_{jt}^{50}] - \beta_2 U_{jt} - \beta_3 X_{jt}\} \\ & + \beta_1[MW_{st} - w_{st}^{50}] + \beta_2 U_{st} + \beta_3 X_{st} + u_{st}, \end{aligned} \quad (4.5.3)$$

by noting $\varepsilon_{jt} = G_{jt} - \beta_1[MW_{jt} - w_{jt}^{50}] - \beta_2 U_{jt} - \beta_3 X_{jt}$, it leads to the SEM model:

$$\begin{aligned} G_{st} &= \alpha_s + \gamma_t + \beta_1[MW_{st} - w_{st}^{50}] + \beta_2 U_{st} + \beta_3 X_{st} + \varepsilon_{st} \\ \varepsilon_{st} &= \delta \sum_{j=1}^{48} w_{sj} \times \varepsilon_{jt} + u_{st}. \end{aligned} \quad (4.5.4)$$

Next, we use the spatial model selection approach suggested by Elhorst (2014) to select a spatial econometric model. Our strategy is first to run the non-spatial model, then apply the classical and robust Lagrange Multiplier (LM) tests to choose an appropriate model among the SLM, the SEM, and the non-spatial model. Another important issue in determining an appropriate model is to explore whether the model needs to control for the state or time-period fixed effects. To achieve that, we perform the likelihood ratio (LR) tests to test the null of jointly insignificant state fixed effects or jointly insignificant time-period fixed effects[1].

To test for spatial interaction effects in a panel data model, Anselin, Le Gallo, and Jayet (2006) proposed classical LM tests to test the null of no spatially lagged dependent variables (LM-lag) and the null of no spatially autocorrelated error terms (LM-error). Elhorst (2010) also developed the robust counterparts of these LM tests, i.e. robust LM-lag and robust LM-error. In principle, if LM-lag rejects the null, but LM-error can not reject the null, then the SLM model is adopted. Similarly, if LM-error rejects the null, but LM-lag can not reject the null, then the SEM model is adopted.

The diagnostic test results are reported in Table 4.3. According to these results, the LR test results point to models with state and time-period fixed effects for all income percentiles. Given that state and time-period fixed effects are included, robust LM tests reject the hypothesis of no spatially lagged dependent variables and the hypothesis of no spatially autocorrelated error terms for all percentiles. Up to this point, the test results indicate that the spatial model is more appropriate to describe our data than the non-spatial model.

However, Elhorst (2014) suggests that if a non-spatial model is rejected in favor of the SLM model or the SEM model based on the (robust) LM tests, one needs to consider the SDM model. Using the SDM estimation results, one can test the null of $H_0 : \boldsymbol{\theta} = \mathbf{0}$ and $H_0 : \boldsymbol{\theta} + \delta\boldsymbol{\beta} = \mathbf{0}$. As discussed above, the first one tests whether the SDM can be simplified to the SLM, and the second one tests whether the SDM can be simplified to the SEM. If both hypotheses are rejected, then the SDM can better describe the data. If the first hypothesis is accepted, then the SLM can better describe the data, given that the (robust) LM tests also pointed to the SLM. If the second hypothesis is accepted, then the

[1] I want to thank Paul Elhorst, who shared his MATLAB code for the classical LM tests, the robust LM tests, and the LR tests. The code can be obtained from https://spatial-panels.com/software/.

Table 4.3: Diagnostic Tests of Models with State and Time-period Fixed Effects

Determinants	p(10)	p(20)	p(30)	p(40)	p(50)
LM-Lag	13.6159***	17.0100***	14.1597***	18.8755***	15.4439***
Robust LM-Lag	46.2718***	43.0232***	32.6234***	13.7812***	18.6308***
LM-Error	2.2572	6.0011**	6.0574**	12.8533***	9.8368***
Robust LM-Error	34.9131***	32.0143***	24.5211***	7.7590***	13.0237***
LR-test state FE	487.0060***	925.2422***	1259.1807***	1537.8268***	1565.2948***
LR-test time FE	116.7499***	466.1513***	685.0177***	856.0976***	892.3441***
Determinants	p(60)	p(70)	p(80)	p(90)	
LM-Lag	3.9117**	1.0768	2.3277	0.0012	
Robust LM-Lag	14.4347***	5.2463**	24.7227***	16.0267***	
LM-Error	2.2223	1.7445	5.1412**	0.7088	
Robust LM-Error	12.7452***	5.9141**	27.5363***	16.7342***	
LR-test state FE	1494.7087***	1086.1469***	804.1226***	528.9726***	
LR-test time FE	881.3777***	683.0363***	552.0589***	484.5458***	

Note: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The (robust) LM-lag tests test the null of no spatial lag dependence in the model. The (robust) LM-error tests test the null of no spatial error dependence in the model. The LR tests test the null that state-specific fixed effects or time-period fixed effects are jointly insignificant.

SEM can better describe the data, given that the (robust) LM tests also pointed to the SEM. These tests can take the form of the Wald test, and the test results are presented in Table 4.4. As can be observed, the null hypothesis that the SDM can be simplified to the SLM and the null hypothesis that the SDM can be simplified to the SEM are both rejected for all percentiles. Therefore, we can conclude that the SDM model (4.5.1) can best describe our data.

Direct and Spillover Effects

One logical consequence of the SDM is that a change in an explanatory variable in a given region might affect the dependent variable in all other regions since the model includes dependent and explanatory variables of other regions (LeSage and Pace, 2009). The simple SDM model is organized as follows:

$$Y_t = \alpha + \gamma_t l_N + \delta W Y_t + X_t\beta + W X_t\theta + u_t,$$

where $\alpha = (\alpha_1, \ldots, \alpha_N)'$ is the vector of spatial specific effects, γ_t is the time-period specific effects, and l_N is a $N \times 1$ vector of ones. The reduced form of this SDM model can be defined by:

$$Y_t = (I_N - \delta W)^{-1}(\alpha + \gamma_t l_N + X_t \beta + W X_t \theta + u_t),$$

where I_N is a $N \times N$ identity matrix. Then, the derivatives of Y with respect to X at time t are:

$$\left[\frac{\partial \mathbb{E}(Y)}{\partial X_1} \cdots \frac{\partial \mathbb{E}(Y)}{\partial X_N}\right]_t = \begin{bmatrix} \frac{\partial \mathbb{E}(Y_1)}{\partial X_1} & \cdots & \frac{\partial \mathbb{E}(Y_1)}{\partial X_N} \\ \vdots & \vdots & \vdots \\ \frac{\partial \mathbb{E}(Y_N)}{\partial X_1} & \cdots & \frac{\partial \mathbb{E}(Y_N)}{\partial X_N} \end{bmatrix} = (I_N - \delta W)^{-1}(\beta I_N + \theta W) = S(W).$$

It follows that the impact of a change in the explanatory variable of a given region j on the expected value of region i is given by $S(W)_{ij}$. This result implies that changes in a characteristic of one region will potentially affect all other regions' outcomes. This type of effect is known as the indirect effect. Similarly, the impact of a change in an explanatory variable of a given region i on the dependent variable of the same region is given by $S(W)_{ii}$. This type of effect is known as the direct effect.

Since the marginal effect of a change in the explanatory variable appears different for all regions, LeSage and Pace (2009) introduce the following scalar summary measures of impacts. The average direct effect, obtained as the mean of diagonal elements of $S(W)$, summarizes the impact of a change in the explanatory variable of a given region on the dependent variable of the same region. The average indirect effect, obtained as the mean row sum of off-diagonal elements $S(W)$, summarizes the impact of a change in an explanatory variable of a given region on the dependent variable of all other regions. The average total effect, defined as the sum of the average direct effect and the average indirect effect, summarizes the impact of a change in an explanatory variable of a given region on the dependent variable of all regions.

4.5.3 Estimation Results

Table 4.4 summarizes the results estimated from the model (4.5.1) for all selected income percentiles. As can be seen, the coefficients on spatially lagged dependent, $W \times G$, are significantly positive for the 20th-50th percentiles, suggesting that the gender wage gap of one state is strongly affected by the gender wage gap of its neighboring states. Paradoxically, we find that the 80th percentile gender wage gap is negatively spatially autocorrelated.

In Table 4.5, we summarize the marginal effect of the effective minimum wage on the gender wage gap for all selected percentiles. We start by considering the total effects. As expected, we find that the minimum wage increase narrows the gender wage gap in the lower tail and middle of the wage distribution. In particular, the minimum wage effect is greater for low-income workers than for middle-income workers. A 1% point increase in the minimum wage will result in 0.00246, 0.00224, and 0.00187 units decline in the gender wage gap at the 10th, 20th, and 30th percentiles, which compares to decreases of 0.00149, 0.00146, and 0.00090 units for the gender wage gap at the 40th, 50th, and 60th percentiles. Since increasing the minimum wage allows for the possibility of increasing wages of workers at the lower tail of the wage distribution, where women are disproportionately represented, low-paid women can benefit the most from the increase in the minimum wage. As a result, a higher minimum wage lowers the gender wage gap. However, there is no evidence showing that the minimum wage affects the gender wage gap in the upper tail of the wage distribution.

The direct and indirect effects suggest that a higher own-state minimum wage narrows the gender wage gap at the 10th-60th percentiles in its own state and its neighboring states. A 1% point increase in the own-state minimum wage is predicted to reduce the 10th-60th percentile gender wage gaps in its own state and its neighboring states by 0.00151-0.00033

Table 4.4: Estimation Results of the Spatial Durbin Model

Determinants	p(10)	p(20)	p(30)	p(40)	p(50)
EMW	-0.151***	-0.117***	-0.100***	-0.100***	-0.076***
unemployment rate	-0.031***	-0.025***	-0.020***	-0.016***	-0.014**
15-44 years old	0.803***	0.364*	0.350**	0.641***	0.590***
45-64 years old	0.786***	0.749***	0.577***	0.726***	0.797***
65-84 years old	-0.169	-0.925***	-0.815***	-0.262	-0.232
female	2.635***	4.087***	4.079***	2.735***	1.974***
white	0.460***	0.172*	0.037	-0.028	-0.135*
W					
EMW	-0.089***	-0.093***	-0.074***	-0.033*	-0.055***
unemployment rate	0.006	0.016	0.015	0.012	0.025**
15-44 years old	0.391	0.818***	0.201	-0.191	-0.022
45-64 years old	-0.538*	0.208	-0.184	-0.457*	-0.488**
65-84 years old	-1.382***	-0.939**	-1.157***	-1.340***	-0.858**
female	5.250***	3.183***	3.077***	3.260***	2.984***
white	0.722***	0.350**	0.451***	0.362***	0.469***
$W \times G$	0.028	0.065**	0.071**	0.104***	0.100***
Wald test spatial lag	64.09***	55.74***	45.54***	32.29***	47.93***
Wald test spatial error	73.99***	65.50***	52.74***	37.51***	52.79***

Determinants	p(60)	p(70)	p(80)	p(90)
EMW	-0.032***	-0.012	0.009	0.009
unemployment rate	-0.004	0.004	0.006	0.009
15-44 years old	0.465***	0.335**	0.261	0.704***
45-64 years old	0.686***	0.629***	0.114	-0.403*
65-84 years old	-0.118	-0.284	-0.221	0.242
female	1.547**	1.431**	1.600**	0.627
white	-0.175**	-0.273***	-0.096	0.506***
W				
EMW	-0.054***	-0.009	0.019	-0.033
unemployment rate	0.005	-0.005	-0.023**	-0.030**
15-44 years old	-0.040	0.280	0.134	-0.705**
45-64 years old	-0.629***	-0.527**	-0.540*	-0.567*
65-84 years old	-0.736**	-0.294	-0.378	-1.719***
female	3.384***	3.964***	4.702***	5.619***
white	0.592***	0.590***	0.567***	1.012***
$W \times G$	0.052	-0.039	-0.071**	-0.037
Wald test spatial lag	47.90***	41.90***	46.74***	47.76***
Wald test spatial error	49.30***	41.05***	43.48***	46.63***

Note: *** p<0.01, ** p<0.05, * p<0.1. Dependent variables are the gender wage gap at all selected income percentiles. EMW is the effective minimum wage. W is the queen contiguity-based spatial weight matrix. The Wald tests test whether the SDM can be simplified to the SLM or the SEM. All models control for state-specific fixed effects and time-period fixed effects.

Table 4.5: Marginal Effects of the Effective Minimum Wage on the Gender Wage Gap

Determinants	p(10)	p(20)	p(30)	p(40)	p(50)
Direct effect of EMW	-0.151***	-0.119***	-0.102***	-0.101***	-0.078***
Indirect effect of EMW	-0.095***	-0.106***	-0.086***	-0.048**	-0.068***
Total effect of EMW	-0.246***	-0.224***	-0.187***	-0.149***	-0.146***
Determinants	p(60)	p(70)	p(80)	p(90)	
Direct effect of EMW	-0.033***	-0.012	0.009	0.009	
Indirect effect of EMW	-0.058***	-0.008	0.017	-0.033	
Total effect of EMW	-0.090***	-0.020	0.026	-0.024	

Note: *** p<0.01, ** p<0.05, * p<0.1. The table presents the marginal effects of the effective minimum wage on the gender wage gap for all selected income percentiles.

and 0.00106-0.00048 units, respectively[2]. Our indirect results reveal that the effect of a state minimum wage can extend beyond its borders. The reasons why a state minimum wage might affect the gender wage gap of its neighbors are discussed in the Introduction section.

4.5.4 Dynamic Effects

The minimum wage has been assumed to have only a contemporaneous effect on the gender wage gap up to this point. However, we may expect that the minimum wage hikes in one state may take some time to have a noticeable effect on the gender wage gap in neighboring states. In other words, the minimum wage spillover effect may not occur instantaneously but is spread over future time periods. To provide information on the dynamic effects of the minimum wage, the model (4.5.1) is estimated with the inclusion of two-year lags of the effective minimum wage[3].

Table 4.6 summarizes the cumulative effect of the effective minimum wage on the gender

[2] Notice that the negative spillover effect is larger than the negative direct effect at the 60th percentile. Since the indirect effect measures the impact of a change in an independent variable of a given region on the dependent variable of all neighboring regions, it is plausible that the average indirect effect is higher than the average direct effect when the parameter of spatial interaction is high (LeSage and Fischer, 2008).

[3] A two-year lag is chosen because the cumulative spillover effects of the minimum wage on the gender wage gap are insignificant for all income percentiles when three years of the lagged effective minimum wage are included.

Table 4.6: Cumulative Effects of the Effective Minimum Wage on the Gender Wage Gap

Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Time t									
Direct effect of EMW	-0.134***	-0.119***	-0.123***	-0.109***	-0.085***	-0.016	-0.002	0.002	0.017
Indirect effect of EMW	-0.086**	-0.034	0.003	0.053	-0.007	-0.015	0.019	0.048	0.040
Total effect of EMW	-0.220***	-0.152***	-0.120***	-0.056	-0.092***	-0.031	0.017	0.050	0.057
Time $t+1$									
Direct effect of EMW	-0.134***	-0.100***	-0.059***	-0.061***	-0.062***	-0.018	-0.003	-0.004	-0.013
Indirect effect of EMW	-0.069*	-0.059	-0.045	0.023	0.012	-0.011	0.029	0.052	0.040
Total effect of EMW	-0.203***	-0.159***	-0.104***	-0.038	-0.050	-0.029	0.026	0.048	0.027
Time $t+2$									
Direct effect of EMW	-0.145***	-0.104***	-0.083***	-0.084***	-0.061***	-0.033**	-0.013	0.020	0.025
Indirect effect of EMW	-0.089***	-0.120***	-0.110***	-0.082***	-0.102***	-0.067***	-0.018	-0.005	-0.081***
Total effect of EMW	-0.234***	-0.224***	-0.193***	-0.166***	-0.163***	-0.100***	-0.031	0.015	-0.056**

Note: *** p<0.01, ** p<0.05, * p<0.1. The table presents the cumulative effects of the effective minimum wage on the gender wage gap for all selected income percentiles.

wage gap for all income percentiles. As expected, the cumulative total effect of the minimum wage over 2 years suggests that the minimum wage reduces the gender wage gap in the lower tail and middle of the wage distribution, with the greatest impact on the lower tail inequality. Three main findings deserve mention. First, compared to the results in the static model (Table 4.5), the indirect effect of the minimum wage only occurs simultaneously for the 10th percentile, while the minimum wage has no immediate effect for the 60th percentile. The contemporaneous direct effect of the minimum wage is a reduction of 0.00134-0.00085 units in the gender wage gap at the 10th-50th percentile. Second, overall, the estimated cumulative direct effect for the 10th-50th percentile decreases over the next year and then increases to a reduction in the gender wage gap of 0.00145-0.00061 units. Lastly, focusing on the cumulative effect of the minimum wage over 2 years, we find that the estimated indirect effect for the 20th-50th percentile and the estimated effect for the 60th percentile become significant, suggesting that it may take 2 years for the minimum wage to affect the 20th-60th percentile gender wage gap of neighboring states. Perhaps more importantly, there is a significant spillover effect of the minimum wage on the gender wage gap at the 90th percentile.

4.6 Robustness Check

4.6.1 Robustness to Model Specification

In this section, we conduct additional robustness checks on our results obtained from the model (4.5.1). One potential concern with the data is that measurement error in the median wage may bias our estimates. Given that the median wage is used to construct the effective minimum wage and is collected from the CPS MORG. If the reported wage involves some measurement error, the effective minimum wage may be endogenous. To deal with this potential endogeneity problem, we re-estimate the model (4.5.1) using the one-year lag of independent variables:

$$\begin{aligned} G_{st} = \;& \alpha_s + \gamma_t + \delta \sum_{j=1}^{48} w_{sj} \times G_{jt} + \beta_1 [MW_{st-1} - w_{st-1}^{50}] + \beta_2 U_{st-1} + \beta_3 X_{st-1} \\ & + \theta_1 \sum_{j=1}^{48} w_{sj} \times [MW_{jt-1} - w_{jt-1}^{50}] + \theta_2 \sum_{j=1}^{48} w_{sj} \times U_{jt-1} + \theta_3 \sum_{j=1}^{48} w_{sj} \times X_{jt-1} + u_{st}, \end{aligned} \tag{4.6.1}$$

where $MW_{st-1} - w_{st-1}^{50}$ is the one-year lag of the effective minimum wage, U_{st-1} is the one-year lag of the log unemployment rate, and X_{st-1} is the one-year lag of other control variables.

Previous studies also suggest that the minimum wage has a significant effect on unemployment (Adie, 1973; Mincer, 1976; Brown, Gilroy, and Kohen, 1982). Since a higher minimum wage increases the labor costs of employers, one might expect that the increase in the minimum wage leads to increased unemployment. Using data from January 1954 to December 1965, Adie (1973) suggests that a 10 percent increase in the federal real minimum wage is predicted to increase the unemployment rate for all teens by 3.62 percent of the prevailing rate. Given that the unemployment rate might be correlated with the effective minimum wage, resulting in the problem of multicollinearity,

Table 4.7: Robustness to Specification Changes

Panel A: Lagged Independent Variables									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.117***	-0.081***	-0.053***	-0.057***	-0.048***	-0.029***	-0.011	0.013	0.010
Indirect effect of *EMW*	-0.084***	-0.109***	-0.114***	-0.086***	-0.085***	-0.060***	-0.010	0.006	-0.052**
Total effect of *EMW*	-0.201***	-0.189***	-0.167***	-0.143***	-0.133***	-0.089***	-0.021	0.018	-0.043*
Panel B: Excluding Unemployment Rate									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.150***	-0.118***	-0.102***	-0.101***	-0.078***	-0.033***	-0.012	0.010	0.010
Indirect effect of *EMW*	-0.094***	-0.105***	-0.086***	-0.048**	-0.069***	-0.058***	-0.008	0.018	-0.031
Total effect of *EMW*	-0.244***	-0.224***	-0.187***	-0.149***	-0.147***	-0.091***	-0.020	0.028	-0.021

Note: *** p<0.01, ** p<0.05, * p<0.1. This table reports the estimation results resulting from models (4.6.1) and (4.6.2), respectively. W is the queen contiguity-based spatial weight matrix. All models control for state-specific fixed effects and time-period fixed effects.

we re-estimate the model (4.5.1) excluding the unemployment rate:

$$\begin{aligned} G_{st} = \quad & \alpha_s + \gamma_t + \delta \sum_{j=1}^{48} w_{sj} \times G_{jt} + \beta_1 [MW_{st} - w_{st}^{50}] + \beta_2 X_{st} \\ & + \theta_1 \sum_{j=1}^{48} w_{sj} \times [MW_{jt} - w_{jt}^{50}] + \theta_2 \sum_{j=1}^{48} w_{sj} \times X_{jt} + u_{st}, \end{aligned} \tag{4.6.2}$$

where all notations are the same as described in the model (4.5.1).

The results based on models (4.6.1) and (4.6.2), presented in Table 4.7, are broadly in line with those obtained from the model (4.5.1). Specifically, when using the lagged independent variables, the total effects of the minimum wage on the 10th-60th percentile gender wage gap are slightly smaller than those in our baseline model, and the spillover effect on the 90th percentile gender wage gap becomes significant. When excluding the unemployment rate, the magnitude, significance, and sign of estimated coefficients are quite similar to that of baseline results. This analysis shows that our baseline empirical results are robust to these specification changes.

4.6.2 Robustness to Sample Changes

Given that minimum wage spillover effects on neighboring states might be driven by a subset of states, we re-estimate the model (4.5.1) by excluding states that have no minimum wages from 1979-2018 (Alabama, Louisiana, Mississippi, South Carolina,

Table 4.8: Robustness to Sample Changes

Panel A: Excluding States with No Minimum Wages									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.147***	-0.109***	-0.093***	-0.087***	-0.069***	-0.023**	-0.005	0.010	0.012
Indirect effect of *EMW*	-0.085***	-0.098***	-0.078***	-0.047**	-0.062***	-0.062***	-0.021	0.002	-0.042*
Total effect of *EMW*	-0.231***	-0.207***	-0.171***	-0.134***	-0.130***	-0.084***	-0.026	0.012	-0.030
Panel B: Excluding 3 States with the Highest Minimum Wages									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.153***	-0.127***	-0.101***	-0.100***	-0.077***	-0.037***	-0.013	0.012	0.010
Indirect effect of *EMW*	-0.111***	-0.129***	-0.114***	-0.069***	-0.079***	-0.067***	-0.024	0.003	-0.036
Total effect of *EMW*	-0.264***	-0.255***	-0.216***	-0.168***	-0.156***	-0.104***	-0.037**	0.016	-0.026
Panel C: Excluding 3 States with the Lowest Minimum Wages									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.146***	-0.116***	-0.101***	-0.100***	-0.076***	-0.029***	-0.011	0.011	0.016
Indirect effect of *EMW*	-0.071***	-0.099***	-0.078***	-0.048**	-0.072***	-0.065***	-0.014	0.014	-0.023
Total effect of *EMW*	-0.217***	-0.215***	-0.179***	-0.148***	-0.148***	-0.094***	-0.026	0.025	-0.007

Note: *** p<0.01, ** p<0.05, * p<0.1. Panels A-C report the estimation results from the model (4.5.1) after excluding some states. All models control for state-specific fixed effects and time-period fixed effects.

Tennessee), or 3 states that have the lowest average minimum wages from 1979-2018 (Georgia, Kansas, Wyoming), or 3 states that have the highest average minimum wages from 1979-2018 (Connecticut, Massachusetts, Oregon). The magnitude of the results of this analysis is slightly different from the magnitude of our baseline results, but these results again indicate that the minimum wage has a significant direct and spillover effect on the 10th-60th percentile gender wage gap (Table 4.8). This analysis suggests that our results are robust to excluding some particular states.

4.6.3 Robustness to Alternative Spatial Weight Matrices

One weakness of the spatial econometric model is that the spatial weight matrix W needs to be specified in advance. However, the economic theory based on the application of spatial econometrics usually does not say much about the specification of W. Therefore, as another robustness check, we explore whether the estimation results are robust to the choice of W. To this end, following the mechanisms for the minimum wage spillovers, we re-estimate the model (4.5.1) with various spatial weight matrices.

Considering that the state gender wage gap might be influenced by the minimum wage of neighboring states, we use two alternative matrices based on the geographic distance to validate our baseline results. The first one is the K-nearest neighbor spatial weight matrix and the second one is the threshold distance spatial weight matrix. The former matrix is defined by $w_{ij} = 1$ if the centroid of state j is one of the K-nearest centroids to state i, and $w_{ij} = 0$ otherwise. We set the value of K as 6 and 7. The latter one is defined by $w_{ij} = 1$ if $0 < d_{ij} \leq 8.5$, where d_{ij} is the Euclidean distance between states i and j, and $w_{ij} = 0$ otherwise.

However, the weight matrix based on geographic distance is rather rough, since it does not consider physical geographic barriers (rivers, mountains, and forests) that may influence the interaction between states. To address this concern, we construct a spatial weight matrix based on the commuting rate, where the weight is calculated by $w_{ij} = c_{ij}$, c_{ij} is the share of workers in state i commuting to state j in 2018[4].

The next weight matrix we propose focuses on the economic distance between states. Economic distance might also be of great importance in shaping the spatial interdependence of the minimum wage, as voters may judge the performance of incumbent politicians by comparing it with other politicians in those states that have similar economic development levels. Here, we use the real GDP as an indicator of economic performance. The spatial weight based on GDP is calculated by $w_{ij} = \frac{1}{|RGDP_i - RGDP_j|}$, where $RGDP_i$ is the log real GDP in state i in 2018.

Lastly, we construct a new matrix to capture both geographic and economic structures by taking the Hadamard product of the queen contiguity-based spatial weight matrix and the GDP-based spatial weight matrix. Note that all spatial weight matrices used in this section are row-normalized.

The results of this analysis, presented in Table 4.9, suggest similar effects of the minimum wage on the gender wage gap compared to those using the contiguity-based spatial weight

[4] The commuting rate data are collected from 2018 1-Year American Community Survey.

Table 4.9: Robustness to Alternative Weight Matrices

Panel A: 6 Nearest Neighbors									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.161***	-0.130***	-0.111***	-0.103***	-0.086***	-0.037***	-0.017	0.006	0.002
Indirect effect of *EMW*	-0.096***	-0.116***	-0.071***	-0.045*	-0.053**	-0.062***	-0.009	0.016	-0.030
Total effect of *EMW*	-0.258***	-0.246***	-0.182***	-0.148***	-0.139***	-0.100***	-0.025	0.022	-0.028
Panel B: 7 Nearest Neighbors									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.154***	-0.123***	-0.105***	-0.100***	-0.082***	-0.036***	-0.017	0.005	0.006
Indirect effect of *EMW*	-0.110***	-0.130***	-0.083***	-0.049**	-0.056**	-0.057***	-0.003	0.020	-0.032
Total effect of *EMW*	-0.264***	-0.253***	-0.188***	-0.149***	-0.137***	-0.092***	-0.020	0.025	-0.026
Panel C: Distance-Band Weights									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.158***	-0.130***	-0.107***	-0.100***	-0.083***	-0.038***	-0.019	0.002	0.001
Indirect effect of *EMW*	-0.122***	-0.118***	-0.083***	-0.047***	-0.055**	-0.049**	0.004	0.040	-0.008
Total effect of *EMW*	-0.281***	-0.248***	-0.191***	-0.147***	-0.137***	-0.087***	-0.015	0.043*	-0.007
Panel D: Commuting-Based Weights									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.149***	-0.124***	-0.105***	-0.098***	-0.078***	-0.034***	-0.011	0.013	0.001
Indirect effect of *EMW*	-0.098***	-0.095***	-0.075***	-0.050***	-0.061***	-0.048***	-0.002	0.004	-0.018
Total effect of *EMW*	-0.247***	-0.219***	-0.180***	-0.148***	-0.139***	-0.081***	-0.014	0.017	-0.017
Panel E: Economic Distance Weights									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.168***	-0.142***	-0.120***	-0.108***	-0.094***	-0.046***	-0.016	0.012	0.006
Indirect effect of *EMW*	-0.074***	-0.085***	-0.088***	-0.076***	-0.078***	-0.067***	-0.050**	-0.021	-0.002
Total effect of *EMW*	-0.242***	-0.228***	-0.208***	-0.184***	-0.171***	-0.113***	-0.066**	-0.009	0.004
Panel F: Geographic and Economic Distance Weights									
Determinants	p(10)	p(20)	p(30)	p(40)	p(50)	p(60)	p(70)	p(80)	p(90)
Direct effect of *EMW*	-0.158***	-0.121***	-0.101***	-0.094***	-0.076***	-0.031***	-0.006	0.012	0.001
Indirect effect of *EMW*	-0.045**	-0.076***	-0.076***	-0.062***	-0.074***	-0.065***	-0.038**	-0.007	-0.007
Total effect of *EMW*	-0.203***	-0.198***	-0.177***	-0.156***	-0.150***	-0.096***	-0.044***	0.005	-0.006

Note: *** p<0.01, ** p<0.05, * p<0.1. This table reports the estimation results of the model (4.5.1) using different types of weight matrices. Spatial weights are all row standardized. All models control for state-specific fixed effects and time-period fixed effects.

matrix reported in Table 4.5, implying that our results are robust to the choice of spatial weight matrices.

4.7 Conclusion

Although the research on the minimum wage and the gender wage gap is expansive, there has been little research on the minimum wage spillover effects and the extent to which the minimum wage affects workers of different income levels. This study extends the literature by using the spatial econometric model to account for the minimum wage effect on the gender wage gap at different income percentiles.

We adopted the diagnostic test based on the approach proposed by Elhorst (2014) to choose the appropriate model. We find that the SDM model with two-way fixed effects best describes data for all selected income percentiles. Our results show that the minimum wage hikes not only contribute to the narrowing of the gender pay gap in the state itself but also in its neighboring states, with a significant effect among the 10th-60th income percentiles. However, we find no evidence that the gender wage gap in the upper tail of the wage distribution is associated with the minimum wage.

Several implications can be generated from this study. First, the minimum wage policy is shown to be an effective tool to reduce the gender wage gap in the lower tail and middle of the wage distribution. Second, the presence of the minimum wage spillover effect suggests that the implementation of the minimum wage policy should take a systematic perspective by taking into account both the direct effect and the indirect effect generated by neighboring regions.

Moreover, many aspects are affecting the gender wage gap that leaves for future study. Future studies can also focus on the spillover effect of the minimum wage on employment and unemployment and the asymmetric effect of the minimum wage on the gender wage gap.

Conclusion and Future Research

This thesis consists of four independent chapters within the field of econometrics. The significance of this thesis is mainly related to two models, the MIDAS model and the spatial econometric model. Chapters 1 and 3 mainly contribute to the field of estimation theory, as they both extend the existing models and establish the asymptotic distributions of their corresponding estimators. To the best of my knowledge, Chapter 2 is the first to capture the sex- and age-specificities in the EPU effect on health status using the MIDAS model, and Chapter 4 is the first to capture the minimum wage spillover effects on the gender wage gap using the spatial econometric model.

Chapters 1 and 2 focus on the MIDAS model that involves data sampled at different frequencies. Chapter 1 extends the traditional MIDAS model to cover the case of panel data and compares the proposed panel-MIDAS model with the traditional LS model based on predetermined weights. It also derives the asymptotic properties and finite sample properties of the NLS estimator for the panel-MIDAS model and the LS estimator for the LS model. The results reveal that in terms of the bias and variance of estimators, the MIDAS-NLS estimator outperforms the LS estimator in presence of the mixed-frequency data. Chapter 2 presents an application of the MIDAS model to study the EPU effect on mortality and risky health behaviors. The results reveal that EPU is inversely associated with mortality rates and suggest that the decline in risky health behaviors during high uncertainty periods can possibly explain this relationship.

Chapters 3 and 4 focus on the spatial econometric model. Chapter 3 extends the traditional SLM model by allowing for the additional endogenous RHS variables. It proposes a new SQLIML estimation approach for this model and establishes the asymptotic distributions of its corresponding estimator. The finite sample properties of the SQLIML estimator and the S2SLS estimator are studied via a Monte Carlo simulation. The results show that the SQLIML estimator outperforms the S2SLS estimator, especially for models with strong endogeneity and weak instruments. Chapter 4 presents an application of the spatial econometric model to capture the spillover effects of the minimum wage on gender wage inequality. The empirical results show that the own-state minimum wage increases will narrow the 10th-60th percentile gender wage gap not only in their own state but to a limited extent also in their neighboring states.

There are several directions for future research to extend the current work. The first research direction is to introduce the MIDAS approach into a spatial panel model. As mentioned earlier, an attractive feature of the MIDAS model is that it can make the best use of the high-frequency information in high-frequency variables to explain low-frequency variables without a large loss of degrees of freedom. Therefore, the combination of the spatial panel model and the MIDAS model is expected to deal with the mixed-frequency data in the spatial panel model at a relatively low parametric cost and is expected to yield more accurate inferences. However, such a topic is still an unexplored area.

The second compelling direction is to extend the proposed SQLIML estimation method in Chapter 3 to more general forms of the SLM model. The current approach is built upon the cross-sectional SLM model with additional endogenous variables. Future research can focus on the extensions of the SQLIML estimation method to control for additional endogeneity in the panel SLM model, the SLM model with spatially autocorrelated errors, or the SDM model. These remain topics of ongoing research.

The third future direction is to apply the proposed SQLIML estimation method to reassess the minimum wage effect on gender wage inequality in Chapter 4. In Chapter 4, the state effective minimum wage is constructed using the state median wage, which is collected from the CPS MORG. However, the CPS MORG data may be subject to measurement errors. If the measurement error in the median wage is correlated with the fluctuation in the gender wage gap, one can expect the existence of the endogeneity problem in the baseline model. Future work can focus on addressing this potential endogeneity problem.

Appendix A

Appendix to Chapter 1

A.1 Proof of Theorem 1.1

Recall that

$$Q(\boldsymbol{\delta}) = \sum_{i=1}^{N} (\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}))'(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})).$$

Define $Q_0(\boldsymbol{\delta}) = \mathbb{E}(\frac{Q(\boldsymbol{\delta})}{N})$, we have

$$\begin{aligned} Q_0(\boldsymbol{\delta}) &= \mathbb{E}(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}))'(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})) \\ &= \mathbb{E}(\tilde{\boldsymbol{\varepsilon}}_i + f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}))'(\tilde{\boldsymbol{\varepsilon}}_i + f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})) \\ &= \mathbb{E}(\tilde{\boldsymbol{\varepsilon}}_i{}'\tilde{\boldsymbol{\varepsilon}}_i) + \sum_{t=1}^{T} \mathbb{E}[(f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}))^2]. \end{aligned}$$

Uniquely Identifiable

To demonstrate the consistency of $\hat{\boldsymbol{\delta}}$, we need to show that $Q_0(\boldsymbol{\delta})$ is uniquely minimized at the true value $\boldsymbol{\delta}_0$. Consider

$$f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}) = \tilde{\boldsymbol{X}}_{it}(\boldsymbol{\theta}_0)\boldsymbol{\beta}_0 - \tilde{\boldsymbol{X}}_{it}(\boldsymbol{\theta})\boldsymbol{\beta}.$$

clearly, for all $\boldsymbol{\delta} \neq \boldsymbol{\delta}_0$ in $\boldsymbol{\Theta}$, $f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}_0) \neq f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta})$. Since $f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})$ is the function of $\boldsymbol{X}_i^{(m)}$,

$f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})$ is the random variable. $f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}_0)\neq f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta})$ means $(f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}_0)-f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}))^2 > 0$ with positive probability, implying that $\mathbb{E}[(f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}_0)-f(\tilde{\boldsymbol{X}}_{it};\boldsymbol{\delta}))^2] > 0$. Therefore, we have

$$Q_0(\boldsymbol{\delta}) > \mathbb{E}(\tilde{\varepsilon}_i{}'\tilde{\varepsilon}_i) \text{ for all } \boldsymbol{\delta}\neq\boldsymbol{\delta}_0 \text{ in } \boldsymbol{\Theta}$$

and

$$Q_0(\boldsymbol{\delta}) = \mathbb{E}(\tilde{\varepsilon}_i{}'\tilde{\varepsilon}_i) \ \text{ if } \boldsymbol{\delta}{=}\boldsymbol{\delta}_0.$$

This implies that $Q_0(\boldsymbol{\delta})$ is uniquely minimized at the true value $\boldsymbol{\delta}_0$. Hence, the identifiable uniqueness must hold.

Uniform Convergence

By Assumption 3,

$$\mathbb{E}\underset{\boldsymbol{\delta}\in\boldsymbol{\Theta}}{sup}[(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}))'(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}))] \leq 2\mathbb{E}(\tilde{\boldsymbol{Y}}_i{}'\tilde{\boldsymbol{Y}}_i) + 2\mathbb{E}\underset{\boldsymbol{\delta}\in\boldsymbol{\Theta}}{sup}[f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})'f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})] < \infty,$$

then the uniform law of large numbers (LLN) gives $\frac{1}{N}Q(\boldsymbol{\delta}) \xrightarrow{p} Q_0(\boldsymbol{\delta})$ uniformly in $\boldsymbol{\delta}\in\boldsymbol{\Theta}$. It follows

$$\begin{aligned}
0 &\leq Q_0(\hat{\boldsymbol{\delta}}) - Q_0(\boldsymbol{\delta}_0) \\
&= Q_0(\hat{\boldsymbol{\delta}}) - Q_0(\boldsymbol{\delta}_0) + \tfrac{Q(\hat{\boldsymbol{\delta}})}{N} - \tfrac{Q(\hat{\boldsymbol{\delta}})}{N} \\
&= Q_0(\hat{\boldsymbol{\delta}}) - \tfrac{Q(\hat{\boldsymbol{\delta}})}{N} + \tfrac{Q(\hat{\boldsymbol{\delta}})}{N} - Q_0(\boldsymbol{\delta}_0) \\
&\leq \mid Q_0(\hat{\boldsymbol{\delta}}) - \tfrac{Q(\hat{\boldsymbol{\delta}})}{N} \mid + \mid \tfrac{Q(\boldsymbol{\delta}_0)}{N} - Q_0(\boldsymbol{\delta}_0) \mid \\
&\leq 2\underset{\boldsymbol{\delta}\in\boldsymbol{\Theta}}{sup} \mid Q_0(\boldsymbol{\delta}) - \tfrac{Q(\boldsymbol{\delta})}{N} \mid \xrightarrow{p} 0,
\end{aligned}$$

where the inequality follows by triangle inequality and by noting $\frac{Q(\hat{\boldsymbol{\delta}})}{N} \leq \frac{Q(\boldsymbol{\delta}_0)}{N}$. It implies that $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$.

A.2 Proof of Theorem 1.2

Given that the first order condition of $Q(\boldsymbol{\delta})$ satisfies $\frac{\partial Q(\hat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}} = \mathbf{0}$. By the mean value theorem, we have that

$$\mathbf{0} = \frac{\partial Q(\hat{\boldsymbol{\delta}})}{\partial \boldsymbol{\delta}} = \frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} + \frac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0),$$

where $\boldsymbol{\delta}_1$ is a point between $\hat{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}_0$. Then inverting yields

$$\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) = -(\frac{1}{N}\frac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'})^{-1}\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}}.$$

Next, we need to prove that (i) $\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}}$ is asymptotically normal, and (ii) $\frac{1}{N}\frac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'}$ converges in probability to a non-singular matrix. We start by focusing on (i):

$$\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} = -\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial f(\tilde{\boldsymbol{X}}_i; \boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\tilde{\boldsymbol{\varepsilon}}_i.$$

By the law of iterated expectations, we have

$$\begin{aligned}
\mathbb{E}(\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\tilde{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\varepsilon}}_i{}'\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}) &= \mathbb{E}(\mathbb{E}(\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'} \mid \tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0))) \\
&= \mathbb{E}(\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\mathbb{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i' \mid \tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0))\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}) \\
&= \sigma_\varepsilon^2\mathbb{E}(\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}),
\end{aligned}$$

here we use the fact that $\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})'}{\partial \boldsymbol{\delta}}\tilde{\boldsymbol{\varepsilon}}_i = \frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \boldsymbol{\delta}}\boldsymbol{H}_T'\boldsymbol{H}_T\boldsymbol{\varepsilon}_i = \frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \boldsymbol{\delta}}\boldsymbol{H}_T'\boldsymbol{\varepsilon}_i = \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta})'}{\partial \boldsymbol{\delta}}\boldsymbol{\varepsilon}_i$. Then, by the central limit theorem (CLT), we have

$$\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} \xrightarrow{d} N(\mathbf{0}, 4\sigma_\varepsilon^2\mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'})).$$

Consider (ii):

$$\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} = -\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial f(\tilde{\boldsymbol{X}}_i; \boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\tilde{\boldsymbol{\varepsilon}}_i.$$

$$\begin{aligned}
\frac{1}{N}\frac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'} &= -\frac{2}{N}\sum_{i=1}^{N}\left(\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)) - \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)}{\partial\boldsymbol{\delta}'}\right)\\
&= \boldsymbol{G}_1 - \frac{2}{N}\sum_{i=1}^{N}\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) + f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1))\\
&= \boldsymbol{G}_1 - \frac{2}{N}\sum_{i=1}^{N}\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\tilde{\boldsymbol{\varepsilon}}_i - \frac{2}{N}\sum_{i=1}^{N}\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}(f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1))\\
&= \boldsymbol{G}_1 - \boldsymbol{G}_2 - \boldsymbol{G}_3,
\end{aligned}$$

where $\boldsymbol{G}_1 = \frac{2}{N}\sum_{i=1}^{N}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)}{\partial\boldsymbol{\delta}'}$, $\boldsymbol{G}_2 = \frac{2}{N}\sum_{i=1}^{N}\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\tilde{\boldsymbol{\varepsilon}}_i$, and $\boldsymbol{G}_3 = \frac{2}{N}\sum_{i=1}^{N}\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}(f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1))$. It is straightforward to show that $\boldsymbol{\delta}_1 \xrightarrow{p} \boldsymbol{\delta}_0$. Since $\boldsymbol{\delta}_1$ lies between $\hat{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}_0$, then $\boldsymbol{\delta}_1$ can be represented as $\boldsymbol{\delta}_1 = \lambda\hat{\boldsymbol{\delta}} + (1-\lambda)\boldsymbol{\delta}_0$, $\exists\lambda \in (0,1)$. Since $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$, we have for any $\delta > 0$

$$\begin{aligned}
\mathbb{P}(\mid \boldsymbol{\delta}_1 - \boldsymbol{\delta}_0 \mid > \delta) &= \mathbb{P}(\mid \lambda\hat{\boldsymbol{\delta}} + (1-\lambda)\boldsymbol{\delta}_0 - \boldsymbol{\delta}_0 \mid > \delta)\\
&= \lambda\mathbb{P}(\mid \hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 \mid > \delta) \to 0,
\end{aligned}$$

which implies $\boldsymbol{\delta}_1 \xrightarrow{p} \boldsymbol{\delta}_0$. Then by the LLN and the continuous mapping theorem (CMT), we have

$$\boldsymbol{G}_1 \xrightarrow{p} 2\mathbb{E}\left(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)}{\partial\boldsymbol{\delta}'}\right) \xrightarrow{p} 2\mathbb{E}\left(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial\boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial\boldsymbol{\delta}'}\right),$$

$$\boldsymbol{G}_2 \xrightarrow{p} \mathbb{E}\left(\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\tilde{\boldsymbol{\varepsilon}}_i\right) = \mathbf{0},$$

and

$$\boldsymbol{G}_3 \xrightarrow{p} \frac{2}{N}\sum_{i=1}^{N}\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}(f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)) = \mathbf{0}.$$

It follows that

$$-\left(\frac{1}{N}\frac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right)^{-1} \xrightarrow{p} -\left(2\mathbb{E}\left(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial\boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial\boldsymbol{\delta}'}\right)\right)^{-1}.$$

Finally, by Slutsky's theorem, we have

$$\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} N(\mathbf{0}, \sigma_\varepsilon^2(\mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}))^{-1}).$$

It is straightforward to show that

$$\begin{aligned}
(\mathbb{E}(\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}} \tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}))^{-1} &= (\mathbb{E}(\begin{pmatrix} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\beta}} \\ \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}} \end{pmatrix} \begin{pmatrix} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\beta}'} & \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'} \end{pmatrix}))^{-1} \\
&= \begin{pmatrix} \mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\beta}} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\beta}'}) & \mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\beta}} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'}) \\ \mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\beta}'}) & \mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'}) \end{pmatrix}^{-1} \\
&= \begin{pmatrix} \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)'\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)) & \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)'\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'}) \\ \mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}}\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)) & \mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'}) \end{pmatrix}^{-1} \\
&= \begin{pmatrix} G_{11}(\boldsymbol{\theta}_0) & G_{12}(\boldsymbol{\theta}_0) \\ G_{12}(\boldsymbol{\theta}_0)' & G_{22}(\boldsymbol{\theta}_0) \end{pmatrix}^{-1} \\
&= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},
\end{aligned}$$

where $G_{11}(\boldsymbol{\theta}_0) = \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)'\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0))$, $G_{12}(\boldsymbol{\theta}_0) = \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0)'\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'})$, and $G_{22}(\boldsymbol{\theta}_0) = \mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}} \frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'})$. Given that $A_{11} = (G_{11}(\boldsymbol{\theta}_0) - G_{12}(\boldsymbol{\theta}_0)G_{22}(\boldsymbol{\theta}_0)^{-1}G_{12}(\boldsymbol{\theta}_0)')^{-1}$, the asymptotic variance of $\hat{\boldsymbol{\beta}}$ is given by $AVar(\hat{\boldsymbol{\beta}}) = \frac{\sigma_\varepsilon^2}{N}A_{11}$.

A.3 Proof of Theorem 1.3

Given that

$$\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) = -(\frac{1}{N}\frac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'})^{-1}\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}},$$

where $\boldsymbol{\delta}_1$ is a point between $\hat{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}_0$. Again, we need to prove that (i) $\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}}$ is asymptotically normal, and (ii) $\frac{1}{N}\frac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'}$ converges in probability to a non-singular matrix. For (i), we have

$$\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} = -\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\tilde{\boldsymbol{u}}_i.$$

Since $\tilde{\boldsymbol{u}}_i$ is uncorrelated with $\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta})$, then the law of iterated expectations gives $\mathbb{E}(\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\tilde{\boldsymbol{u}}_i) = \boldsymbol{0}$ and

$$\begin{aligned}\mathbb{E}(\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\tilde{\boldsymbol{u}}_i\tilde{\boldsymbol{u}}_i{}'\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}) &= \mathbb{E}(\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\Omega}^{-1}\mathbb{E}(\boldsymbol{u}_i\boldsymbol{u}_i' \mid \boldsymbol{X}_i(\boldsymbol{\theta}_0))\boldsymbol{\Omega}^{-1}\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'})\\ &= \mathbb{E}(\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\Omega}^{-1}\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}).\end{aligned}$$

Then, by the CLT, we have

$$\frac{1}{\sqrt{N}}\frac{\partial Q(\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}} \xrightarrow{d} N(\boldsymbol{0}, 4\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\Omega}^{-1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'})).$$

For (ii)

$$\begin{aligned}\tfrac{1}{N}\tfrac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'} &= -\tfrac{2}{N}\sum_{i=1}^{N}(\tfrac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'}(\tilde{\boldsymbol{Y}}_i - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)) - \tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial \boldsymbol{\delta}'}\tfrac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta}})\\ &= \boldsymbol{G}_1 - \boldsymbol{G}_2 - \boldsymbol{G}_3,\end{aligned}$$

where $\boldsymbol{G}_1 = \frac{2}{N}\sum_{i=1}^{N}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial \boldsymbol{\delta}}\frac{\partial f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta}'}$, $\boldsymbol{G}_2 = \frac{2}{N}\sum_{i=1}^{N}\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'}\tilde{\boldsymbol{u}}_i$, and $\boldsymbol{G}_3 = \frac{2}{N}\sum_{i=1}^{N}\frac{\partial^2 f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1)'}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'}(f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_0) - f(\tilde{\boldsymbol{X}}_i;\boldsymbol{\delta}_1))$. Since $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$, $\boldsymbol{\delta}_1$ lies between $\hat{\boldsymbol{\delta}}$ and $\boldsymbol{\delta}_0$, then $\boldsymbol{\delta}_1 \xrightarrow{p} \boldsymbol{\delta}_0$. Then, by the LLN and the CMT, we have

$$-(\frac{1}{N}\frac{\partial^2 Q(\boldsymbol{\delta}_1)}{\partial \boldsymbol{\delta}\partial \boldsymbol{\delta}'})^{-1} \xrightarrow{p} -(2\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\Omega}^{-1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}))^{-1}.$$

Finally, by Slutsky's theorem, we have

$$\sqrt{N}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) \xrightarrow{d} N(\boldsymbol{0}, (\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\Omega}^{-1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}))^{-1}).$$

Again, it is straightforward to show that

$$
\begin{aligned}
(\mathbb{E}(\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\delta}}\boldsymbol{\Omega}^{-1}\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\delta}'}))^{-1} &= \begin{pmatrix} \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}_0)'\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i(\boldsymbol{\theta}_0)) & \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}_0)'\boldsymbol{\Omega}^{-1}\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'}) \\ \mathbb{E}(\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}}\boldsymbol{\Omega}^{-1}\boldsymbol{X}_i(\boldsymbol{\theta}_0)) & \mathbb{E}(\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)'}{\partial \boldsymbol{\theta}}\boldsymbol{\Omega}^{-1}\tfrac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta}_0)}{\partial \boldsymbol{\theta}'}) \end{pmatrix}^{-1} \\
&= \begin{pmatrix} G_{11}(\boldsymbol{\theta}_0) & G_{12}(\boldsymbol{\theta}_0) \\ G_{12}(\boldsymbol{\theta}_0)' & G_{22}(\boldsymbol{\theta}_0) \end{pmatrix}^{-1} \\
&= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},
\end{aligned}
$$

where $A_{11} = (G_{11}(\boldsymbol{\theta}_0) - G_{12}(\boldsymbol{\theta}_0)G_{22}(\boldsymbol{\theta}_0)^{-1}G_{12}(\boldsymbol{\theta}_0)')^{-1}$. Thus, $AVar(\hat{\boldsymbol{\beta}}) = \frac{A_{11}}{N}$.

A.4 Proof of Theorem 1.4

Recall that

$$\hat{\boldsymbol{\beta}}^* = \boldsymbol{\beta}_0 + (\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}(\boldsymbol{\pi}))^{-1}(\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}(\boldsymbol{\theta}_0^*))\boldsymbol{\beta}_0 + (\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}(\boldsymbol{\pi}))^{-1}\tilde{\boldsymbol{X}}(\boldsymbol{\pi})'\tilde{\varepsilon}.$$

The LLN and the CMT give

$$(\frac{\sum_{i=1}^{N}\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})}{N})^{-1} \xrightarrow{p} (\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})))^{-1},$$

$$\frac{\sum_{i=1}^{N}\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)}{N} \xrightarrow{p} \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*)),$$

and

$$\frac{\sum_{i=1}^{N}\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\varepsilon}_i}{N} \xrightarrow{p} \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\varepsilon}_i) = \mathbf{0}.$$

Applying these results, we have

$$ABias(\hat{\boldsymbol{\beta}}^*;\boldsymbol{\beta}_0) = (\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi}))^{-1}\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\theta}_0^*))\boldsymbol{\beta}_0.$$

Besides, the CLT gives

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{\varepsilon}}_i \xrightarrow{d} N(0,\sigma_\varepsilon^2\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi}))),$$

here we use the fact that

$$\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{\varepsilon}}_i\tilde{\boldsymbol{\varepsilon}}_i{}'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})) = \mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})) = \sigma_\varepsilon^2\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})),$$

since $\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{\varepsilon}}_i = \boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{H}_T'\boldsymbol{H}_T\boldsymbol{\varepsilon}_i = \boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{H}_T'\boldsymbol{\varepsilon}_i = \tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\boldsymbol{\varepsilon}_i$. Then the Slutsky's theorem gives $AVar(\hat{\boldsymbol{\beta}}^*) = \frac{\sigma_\varepsilon^2}{N}(\mathbb{E}(\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})'\tilde{\boldsymbol{X}}_i(\boldsymbol{\pi})))^{-1}$.

A.5 Proof of Proposition 1.1

Here, we only focus on the FE model. (i) We start by computing the asymptotic bias and the asymptotic variance of the LS estimator. Recall that, when $K = m$,

$$\begin{cases} X_{it}(\boldsymbol{\pi}) = \pi_1 e_{i,t}^{(m)} + \sum_{j=1}^{m-1}(\rho\pi_j + \pi_{j+1})e_{i,t-j/m}^{(m)} + \rho\pi_m e_{i,t-1}^{(m)}, \\ X_{it}(\boldsymbol{\theta}^*) = w_1(\boldsymbol{\theta}^*)e_{i,t}^{(m)} + \sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*))e_{i,t-j/m}^{(m)} + \rho w_m(\boldsymbol{\theta}^*)e_{i,t-1}^{(m)}, \end{cases}$$

By (1.4.4), to obtain the asymptotic bias of the LS estimator, we need to compute $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})'\boldsymbol{H}_T\boldsymbol{X}_i(\boldsymbol{\pi}))$ and $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})'\boldsymbol{H}_T\boldsymbol{X}_i(\boldsymbol{\theta}^*))$. By the property of the trace and expectation operators, we have

$$\begin{aligned} \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})'\boldsymbol{H}_T\boldsymbol{X}_i(\boldsymbol{\pi})) &= tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')), \\ \mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})'\boldsymbol{H}_T\boldsymbol{X}_i(\boldsymbol{\theta}^*)) &= tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')), \end{aligned}$$

where $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')$ and $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')$ are given by

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} = \begin{cases} \sigma_e^2[\pi_1^2 + \sum_{j=1}^{m-1}(\rho\pi_j + \pi_{j+1})^2 + \rho^2\pi_m^2] & \text{if } r = k \\ \rho\pi_1\pi_m\sigma_e^2 & \text{if } r = k+1 \text{ or } r = k-1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} = \begin{cases} \sigma_e^2[\pi_1 w_1(\boldsymbol{\theta}^*) + \sum_{j=1}^{m-1}(\rho\pi_j + \pi_{j+1})(\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*)) + \rho^2\pi_m w_m(\boldsymbol{\theta}^*)] & \text{if } r = k \\ \sigma_e^2\rho\pi_m w_1(\boldsymbol{\theta}^*) & \text{if } r = k+1 \\ \sigma_e^2\rho\pi_1 w_m(\boldsymbol{\theta}^*) & \text{if } r = k-1 \\ 0 & \text{otherwise} \end{cases}$$

Using the above results, we can obtain the asymptotic bias and the asymptotic variance of the LS estimator.

Similarly, when $K = 2m$,

$$\begin{aligned} X_{it}(\boldsymbol{\pi}) = \quad & \pi_1 e_{i,t}^{(m)} + \sum_{j=1}^{m-1}(\rho\pi_j + \pi_{j+1})e_{i,t-j/m}^{(m)} + (\rho\pi_m + \pi_{m+1})e_{i,t-1}^{(m)} \\ & + \sum_{j=1}^{m-1}(\rho\pi_{m+j} + \pi_{m+j+1})e_{i,t-1-j/m}^{(m)} + \rho\pi_{2m}e_{i,t-2}^{(m)}, \\ X_{it}(\boldsymbol{\theta}^*) = \quad & w_1(\boldsymbol{\theta}^*)e_{i,t}^{(m)} + \sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*))e_{i,t-j/m}^{(m)} + (\rho w_m(\boldsymbol{\theta}^*) + w_{m+1}(\boldsymbol{\theta}^*))e_{i,t-1}^{(m)} \\ & + \sum_{j=1}^{m-1}(\rho w_{m+j}(\boldsymbol{\theta}^*) + w_{m+j+1}(\boldsymbol{\theta}^*))e_{i,t-1-j/m}^{(m)} + \rho w_{2m}(\boldsymbol{\theta}^*)e_{i,t-2}^{(m)}, \end{aligned}$$

then we have

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$
$$\begin{cases} \sigma_e^2[\pi_1^2 + \sum_{j=1}^{2m-1} (\rho\pi_j + \pi_{j+1})^2 + \rho^2\pi_{2m}^2] & \text{if } r = k \\ \sigma_e^2[\pi_1(\rho\pi_m + \pi_{m+1}) + \sum_{j=1}^{m-1} (\rho\pi_j + \pi_{j+1})(\rho\pi_{m+j} + \pi_{m+j+1}) & \\ +\rho\pi_{2m}(\rho\pi_m + \pi_{m+1})] & \text{if } r = k+1 \text{ or } r = k-1 \\ \sigma_e^2\rho\pi_1\pi_{2m} & \text{if } r = k+2 \text{ or } r = k-2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$
$$\begin{cases} \sigma_e^2[\pi_1 w_1(\boldsymbol{\theta}^*) + \sum_{j=1}^{2m-1} (\rho\pi_j + \pi_{j+1})(\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*)) + \rho^2\pi_{2m}w_{2m}(\boldsymbol{\theta}^*)] & \text{if } r = k \\ \sigma_e^2[w_1(\boldsymbol{\theta}^*)(\rho\pi_m + \pi_{m+1}) + \sum_{j=1}^{m-1} (\rho w_j(\boldsymbol{\theta}^*) + w_{j+1}(\boldsymbol{\theta}^*)) & \\ \times(\rho\pi_{m+j} + \pi_{m+j+1}) + \rho\pi_{2m}(\rho w_m(\boldsymbol{\theta}^*) + w_{m+1}(\boldsymbol{\theta}^*))] & \text{if } r = k+1 \\ \sigma_e^2[\pi_1(\rho w_m(\boldsymbol{\theta}^*) + w_{m+1}(\boldsymbol{\theta}^*)) + \sum_{j=1}^{m-1} (\rho w_{m+j}(\boldsymbol{\theta}^*) + w_{m+j+1}(\boldsymbol{\theta}^*)) & \\ \times(\rho\pi_j + \pi_{j+1}) + \rho w_{2m}(\boldsymbol{\theta}^*)(\rho\pi_m + \pi_{m+1})] & \text{if } r = k-1 \\ \sigma_e^2\rho\pi_{2m}w_1(\boldsymbol{\theta}^*) & \text{if } r = k+2 \\ \sigma_e^2\rho\pi_1 w_{2m}(\boldsymbol{\theta}^*) & \text{if } r = k-2 \\ 0 & \text{otherwise} \end{cases}$$

(ii) By (1.3.6), we then compute the asymptotic variance of the MIDAS-NLS estimator. Again, we have

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})'\boldsymbol{H}_T\boldsymbol{X}_i(\boldsymbol{\theta})) = tr(\boldsymbol{H}_T\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')).$$

Using the same way, it follows that

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})'\boldsymbol{H}_T\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\boldsymbol{\theta}'}) = \begin{pmatrix} tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\boldsymbol{X}_i(\boldsymbol{\theta})')) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\boldsymbol{X}_i(\boldsymbol{\theta})')) \end{pmatrix}$$

where

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} = \begin{cases} \sigma_e^2[w_1(\boldsymbol{\theta})^2 + \sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta}) + w_{j+1}(\boldsymbol{\theta}))^2 + \rho^2 w_m(\boldsymbol{\theta})^2] & \text{if } r = k \\ \rho w_m(\boldsymbol{\theta})w_1(\boldsymbol{\theta})\sigma_e^2 & \text{if } r = k+1 \text{ or } r = k-1 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} = \begin{cases} \sigma_e^2[w_1(\boldsymbol{\theta})\frac{\partial w_1(\boldsymbol{\theta})}{\partial\theta_s} + \sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta}) + w_{j+1}(\boldsymbol{\theta}))(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial\theta_s} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial\theta_s}) & \\ +\rho^2 w_m(\boldsymbol{\theta})\frac{\partial w_m(\boldsymbol{\theta})}{\partial\theta_s}]\beta & \text{if } r = k \\ \sigma_e^2\rho w_m(\boldsymbol{\theta})\frac{\partial w_1(\boldsymbol{\theta})}{\partial\theta_s}\beta & \text{if } r = k+1 \\ \sigma_e^2\rho w_1(\boldsymbol{\theta})\frac{\partial w_m(\boldsymbol{\theta})}{\partial\theta_s}\beta & \text{if } r = k-1 \\ 0 & \text{otherwise} \end{cases}$$

Besides, we have

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\boldsymbol{\theta}}'\boldsymbol{H}_T\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\boldsymbol{\theta}'}) = \begin{pmatrix} tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_1})) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_1})) \\ tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_1}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_2})) & tr(\boldsymbol{H}_T\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_2}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_2})) \end{pmatrix},$$

where

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \theta_s})_{kr} =$$
$$\begin{cases} \sigma_e^2[\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_l}\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s} + \sum_{j=1}^{m-1}(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_l} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_l})(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ +\rho^2\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_l}\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s}]\beta^2 & \text{if } r = k \\ \sigma_e^2\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s}\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_l}\beta^2 & \text{if } r = k+1 \\ \sigma_e^2\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_l}\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}\beta^2 & \text{if } r = k-1 \\ 0 & \text{otherwise} \end{cases}$$

Similarly, when $K = 2m$, we have

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$
$$\begin{cases} \sigma_e^2[w_1(\boldsymbol{\theta})^2 + \sum_{j=1}^{2m-1}(\rho w_j(\boldsymbol{\theta}) + w_{j+1}(\boldsymbol{\theta}))^2 + \rho^2 w_{2m}(\boldsymbol{\theta})^2] & \text{if } r = k \\ \sigma_e^2[w_1(\boldsymbol{\theta})(\rho w_m(\boldsymbol{\theta}) + w_{m+1}(\boldsymbol{\theta})) + \sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta}) + w_{j+1}(\boldsymbol{\theta})) & \\ \times(\rho w_{m+j}(\boldsymbol{\theta}) + w_{m+j+1}(\boldsymbol{\theta})) + \rho w_{2m}(\boldsymbol{\theta})(\rho w_m(\boldsymbol{\theta}) + w_{m+1}(\boldsymbol{\theta}))] & \text{if } r = k+1 \text{ or } r = k-1 \\ \sigma_e^2\rho w_1(\boldsymbol{\theta})w_{2m}(\boldsymbol{\theta}) & \text{if } r = k+2 \text{ or } r = k-2 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}\left(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})'\right)_{kr} = \begin{cases} \sigma_e^2[w_1(\boldsymbol{\theta})\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s} + \sum_{j=1}^{2m-1}(\rho w_j(\boldsymbol{\theta}) + w_{j+1}(\boldsymbol{\theta}))(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ +\rho^2 w_{2m}(\boldsymbol{\theta})\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}]\beta & \text{if } r = k \\ \sigma_e^2[(\rho w_m(\boldsymbol{\theta}) + w_{m+1}(\boldsymbol{\theta}))\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s} + \sum_{j=1}^{m-1}(\rho w_{m+j}(\boldsymbol{\theta}) + w_{m+j+1}(\boldsymbol{\theta})) & \\ \times(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s}) + \rho w_{2m}(\boldsymbol{\theta})(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_s})]\beta & \text{if } r = k+1 \\ \sigma_e^2[w_1(\boldsymbol{\theta})(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_s}) + \sum_{j=1}^{m-1}(\rho w_j(\boldsymbol{\theta}) + w_{j+1}(\boldsymbol{\theta})) & \\ \times(\rho\frac{\partial w_{m+j}(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{m+j+1}(\boldsymbol{\theta})}{\partial \theta_s}) + \rho(\rho w_m(\boldsymbol{\theta}) + w_{m+1}(\boldsymbol{\theta}))\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}]\beta & \text{if } r = k-1 \\ \sigma_e^2\rho w_{2m}(\boldsymbol{\theta})\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}\beta & \text{if } r = k+2 \\ \sigma_e^2\rho w_1(\boldsymbol{\theta})\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}\beta & \text{if } r = k-2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\mathbb{E}\left(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial \theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial \theta_s}\right)_{kr} = \begin{cases} \sigma_e^2[\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_l}\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s} + \sum_{j=1}^{2m-1}(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_l} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_l})(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ +\rho^2\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_l}\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}]\beta^2 & \text{if } r = k \\ \sigma_e^2[(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_s})\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_l} + \sum_{j=1}^{m-1}(\rho\frac{\partial w_{m+j}(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{m+j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ \times(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_l} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_l}) + \rho\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_l} + \frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_l})]\beta^2 & \text{if } r = k+1 \\ \sigma_e^2[\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_l} + \frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_l}) + \sum_{j=1}^{m-1}(\rho\frac{\partial w_j(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{j+1}(\boldsymbol{\theta})}{\partial \theta_s}) & \\ \times(\rho\frac{\partial w_{m+j}(\boldsymbol{\theta})}{\partial \theta_l} + \frac{\partial w_{m+j+1}(\boldsymbol{\theta})}{\partial \theta_l}) + \rho(\rho\frac{\partial w_m(\boldsymbol{\theta})}{\partial \theta_s} + \frac{\partial w_{m+1}(\boldsymbol{\theta})}{\partial \theta_s})\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_l}]\beta^2 & \text{if } r = k-1 \\ \sigma_e^2\rho\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_s}\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_l}\beta^2 & \text{if } r = k+2 \\ \sigma_e^2\rho\frac{\partial w_1(\boldsymbol{\theta})}{\partial \theta_s}\frac{\partial w_{2m}(\boldsymbol{\theta})}{\partial \theta_l}\beta^2 & \text{if } r = k-2 \\ 0 & \text{otherwise} \end{cases}$$

for all $s, l = 1, 2$.

A.6 Proof of Proposition 1.3

First, we compute the mean and variance of $x_{it}^{(m)}$. Since $\{x_{it}^{(m)}\}$ follows a stationary AR(1) process, then it follows that

$$\mathbb{E}(x_{it}^{(m)}) = \phi\mathbb{E}(x_{i,t-1/m}^{(m)}) + \mathbb{E}(e_{it}^{(m)}) \Longrightarrow \mathbb{E}(x_{it}^{(m)}) = 0$$

and

$$Var(x_{it}^{(m)}) = \phi^2 Var(x_{i,t-1/m}^{(m)}) + Var(e_{it}^{(m)}) \Longrightarrow Var(x_{it}^{(m)}) = \frac{\sigma_e^2}{1-\phi^2}.$$

Next, since the high frequency regressor $x_{i,t-(j-1)/m}^{(m)}$ can be expressed as a form of $x_{i,t-(m-1)/m}^{(m)}$, where $m > j$,

$$x_{i,t-(j-1)/m}^{(m)} = \phi^{K-j} x_{i,t-(K-1)/m}^{(m)} + \sum_{q=0}^{K-j-1} \phi^q e_{i,t-(q+j-1)/m}^{(m)},$$

then the simple average term and nonlinear term can be expressed as

$$\begin{aligned} X_{it}(\boldsymbol{\pi}) &= \sum_{j=1}^{K-1} \pi_j x_{i,t-(j-1)/m}^{(m)} + \pi_K x_{i,t-(K-1)/m}^{(m)} \\ &= \sum_{j=1}^{K} \pi_j \phi^{K-j} x_{i,t-(K-1)/m}^{(m)} + \sum_{j=1}^{K-1}\sum_{q=1}^{j} \pi_q \phi^{j-q} e_{i,t-(j-1)/m}^{(m)} \end{aligned}$$

and

$$\begin{aligned} X_{it}(\boldsymbol{\theta}^*) &= \sum_{j=1}^{K-1} w_j(\boldsymbol{\theta}^*) x_{i,t-(j-1)/m}^{(m)} + w_K(\boldsymbol{\theta}^*) x_{i,t-(K-1)/m}^{(m)} \\ &= \sum_{j=1}^{K-1} w_j(\boldsymbol{\theta}^*)(\phi^{K-j}-1) x_{i,t-(K-1)/m}^{(m)} + \sum_{j=1}^{K-1}\sum_{q=1}^{j} w_q(\boldsymbol{\theta}^*) \phi^{j-q} e_{i,t-(j-1)/m}^{(m)}. \end{aligned}$$

(i) To obtain the asymptotic bias and the asymptotic variance of the LS estimator, we need to compute $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')$ and $\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')$. It is straightforward to get that for $t > s$

$$x^{(m)}_{i,t-(K-1)/m} = \phi^{m(t-s)} x^{(m)}_{i,s-(K-1)/m} + \sum_{q=0}^{m(t-s)-1} \phi^q e^{(m)}_{i,t-(q+K-1)/m}.$$

When $K = m$,

$$x^{(m)}_{i,t-(m-1)/m} = \phi^{m(t-s)} x^{(m)}_{i,s-(m-1)/m} + \sum_{q=0}^{m(t-s)-m} \phi^q e^{(m)}_{i,t-(q+m-1)/m} + \sum_{q=1}^{m-1} \phi^{m(t-s)-m+q} e^{(m)}_{i,s-(q-1)/m},$$

and therefore

$$\begin{aligned} X_{it}(\boldsymbol{\pi}) &= \sum_{j=1}^{m} \pi_j \phi^{m-j} (\phi^{m(t-s)} x^{(m)}_{i,s-(m-1)/m} + \sum_{q=0}^{m(t-s)-m} \phi^q e^{(m)}_{i,t-(q+m-1)/m} \\ &\quad + \sum_{q=1}^{m-1} \phi^{m(t-s)-m+q} e^{(m)}_{i,s-(q-1)/m}) + \sum_{j=1}^{m-1} \sum_{q=1}^{j} \pi_q \phi^{j-q} e^{(m)}_{i,t-(j-1)/m}, \end{aligned}$$

$$\begin{aligned} X_{it}(\boldsymbol{\theta}^*) &= \sum_{j=1}^{m-1} w_j(\boldsymbol{\theta}^*)(\phi^{m-j} - 1)(\phi^{m(t-s)} x^{(m)}_{i,s-(m-1)/m} + \sum_{q=0}^{m(t-s)-m} \phi^q e^{(m)}_{i,t-(q+m-1)/m} \\ &\quad + \sum_{q=1}^{m-1} \phi^{m(t-s)-m+q} e^{(m)}_{i,s-(q-1)/m}) + \sum_{j=1}^{m-1} \sum_{q=1}^{j} w_q(\boldsymbol{\theta}^*) \phi^{j-q} e^{(m)}_{i,t-(j-1)/m}. \end{aligned}$$

Then, we have

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$
$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m} \pi_j \phi^{m-j})^2 + \sigma_e^2 \sum_{j=1}^{m-1} (\sum_{q=1}^{j} \pi_q \phi^{j-q})^2 & \text{if } r = k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m} \pi_j \phi^{m-j})^2 \phi^{m|r-k|} + \sigma_e^2(\sum_{j=1}^{m} \pi_j \phi^{m-j})(\sum_{j=1}^{m-1} \sum_{q=1}^{j} \pi_q \phi^{m|r-k|-m+2j-q}) & \text{otherwise} \end{cases}$$

and

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$

$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m} \pi_j\phi^{m-j})[\sum_{j=1}^{m-1} w_j(\boldsymbol{\theta}^*)(\phi^{m-j}-1)] \\ +\sigma_e^2 \sum_{j=1}^{m-1} (\sum_{q=1}^{j} \pi_q\phi^{j-q})(\sum_{q=1}^{j} w_q(\boldsymbol{\theta}^*)\phi^{j-q}) & \text{if } r=k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m} \pi_j\phi^{m-j})[\sum_{j=1}^{m-1} w_j(\boldsymbol{\theta}^*)(\phi^{m-j}-1)]\phi^{m(k-r)} \\ +\sigma_e^2[\sum_{j=1}^{m-1} w_j(\boldsymbol{\theta}^*)(\phi^{m-j}-1)](\sum_{j=1}^{m-1}\sum_{q=1}^{j} \pi_q\phi^{m(k-r)-m+2j-q}) & \text{if } r<k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m} \pi_j\phi^{m-j})[\sum_{j=1}^{m-1} w_j(\boldsymbol{\theta}^*)(\phi^{m-j}-1)]\phi^{m(r-k)} \\ +\sigma_e^2(\sum_{j=1}^{m} \pi_j\phi^{m-j})(\sum_{j=1}^{m-1}\sum_{q=1}^{j} w_q(\boldsymbol{\theta}^*)\phi^{m(r-k)-m+2j-q}) & \text{if } r>k \end{cases}$$

When $K=2m$, for $t=s+1$,

$$\begin{aligned} X_{it}(\boldsymbol{\pi}) &= \sum_{j=1}^{2m} \pi_j\phi^{3m-j}x^{(m)}_{i,s-(2m-1)/m} + (\sum_{j=1}^{m} + \sum_{j=m+1}^{2m-1}) \sum_{q=1}^{j} \pi_q\phi^{j-q}e^{(m)}_{i,t-(j-1)/m} \\ &\quad + \sum_{j=1}^{2m} \pi_j\phi^{2m-j} \sum_{q=0}^{m-1} \phi^q e^{(m)}_{i,t-(q+2m-1)/m}, \end{aligned}$$

$$X_{is}(\boldsymbol{\pi}) = \sum_{j=1}^{2m} \pi_j\phi^{2m-j}x^{(m)}_{i,s-(2m-1)/m} + (\sum_{j=1}^{m-1} + \sum_{j=m}^{2m-1}) \sum_{q=1}^{j} \pi_q\phi^{j-q}e^{(m)}_{i,s-(j-1)/m},$$

$$\begin{aligned} X_{it}(\boldsymbol{\theta}^*) &= \sum_{j=1}^{2m-1} w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)\phi^m x^{(m)}_{i,s-(2m-1)/m} + (\sum_{j=1}^{m} + \sum_{j=m+1}^{2m-1}) \sum_{q=1}^{j} w_q(\boldsymbol{\theta}^*)\phi^{j-q}e^{(m)}_{i,t-(j-1)/m} \\ &\quad + \sum_{j=1}^{2m-1} w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1) \sum_{q=0}^{m-1} \phi^q e^{(m)}_{i,t-(q+2m-1)/m}, \end{aligned}$$

$$X_{is}(\boldsymbol{\theta}^*) = \sum_{j=1}^{2m-1} w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)x^{(m)}_{i,s-(2m-1)/m} + (\sum_{j=1}^{m-1} + \sum_{j=m}^{2m-1}) \sum_{q=1}^{j} w_q(\boldsymbol{\theta}^*)\phi^{j-q}e^{(m)}_{i,s-(j-1)/m}.$$

For $t>s+1$,

$$\begin{aligned} X_{it}(\boldsymbol{\pi}) &= \sum_{j=1}^{2m} \pi_j\phi^{2m-j}(\phi^{m(t-s)}x^{(m)}_{i,s-(2m-1)/m} + \sum_{q=0}^{m(t-s)-2m} \phi^q e^{(m)}_{i,t-(q+2m-1)/m} \\ &\quad + \sum_{q=1}^{2m-1} \phi^{m(t-s)-2m+q}e^{(m)}_{i,s-(q-1)/m}) + \sum_{j=1}^{2m-1}\sum_{q=1}^{j} \pi_q\phi^{j-q}e^{(m)}_{i,t-(j-1)/m}, \end{aligned}$$

$$
\begin{aligned}
X_{it}(\boldsymbol{\theta}^*) \;\; = \;\; & \sum_{j=1}^{2m-1} w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)(\phi^{m(t-s)}x^{(m)}_{i,s-(2m-1)/m} + \sum_{q=0}^{m(t-s)-2m} \phi^q e^{(m)}_{i,t-(q+2m-1)/m} \\
& + \sum_{q=1}^{2m-1} \phi^{m(t-s)-2m+q} e^{(m)}_{i,s-(q-1)/m}) + \sum_{j=1}^{2m-1}\sum_{q=1}^{j} w_q(\boldsymbol{\theta}^*)\phi^{j-q} e^{(m)}_{i,t-(j-1)/m}.
\end{aligned}
$$

Then, we have

$$
\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\pi})\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =
$$

$$
\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} \pi_j \phi^{2m-j})^2 + \sigma_e^2 \sum_{j=1}^{2m-1} (\sum_{q=1}^{j} \pi_q \phi^{j-q})^2 & \text{if } r = k \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} \pi_j \phi^{2m-j})^2 \phi + \sigma_e^2 \sum_{j=1}^{m-1} (\sum_{q=1}^{m+j} \pi_q \phi^{m+j-q}) & \\
\times(\sum_{q=1}^{j} \pi_q \phi^{j-q}) + \sigma_e^2(\sum_{j=1}^{2m} \pi_j \phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j} \pi_q \phi^{m+2j-q}) & \text{if } r = k+1 \text{ or } r = k-1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} \pi_j \phi^{2m-j})^2 \phi^{m|r-k|} & \\
+\sigma_e^2(\sum_{j=1}^{2m} \pi_j \phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j} \pi_q \phi^{m|r-k|-2m+2j-q}) & \text{otherwise}
\end{cases}
$$

and

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta}^*)\boldsymbol{X}_i(\boldsymbol{\pi})')_{kr} =$$
$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)] \\ +\sigma_e^2\sum_{j=1}^{2m-1}(\sum_{q=1}^{j}\pi_q\phi^{j-q})(\sum_{q=1}^{j}w_q(\boldsymbol{\theta}^*)\phi^{j-q}) & \text{if } r=k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)]\phi+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{m+j}\pi_q\phi^{m+j-q}) \\ \times(\sum_{q=1}^{j}w_q(\boldsymbol{\theta}^*)\phi^{j-q})+\sigma_e^2(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j}w_q(\boldsymbol{\theta}^*)\phi^{m+2j-q}) & \text{if } r=k+1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)]\phi+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{j}\pi_q\phi^{j-q}) \\ \times(\sum_{q=1}^{m+j}w_q(\boldsymbol{\theta}^*)\phi^{m+j-q})+\sigma_e^2[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)](\sum_{j=0}^{m-1}\sum_{q=1}^{m+j}\pi_q\phi^{m+2j-q}) & \text{if } r=k-1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)]\phi^{m(r-k)} \\ +\sigma_e^2(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j}w_q(\boldsymbol{\theta}^*)\phi^{m(r-k)-2m+2j-q}) & \text{if } r>k+1 \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m}\pi_j\phi^{2m-j})[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)]\phi^{m(k-r)} \\ +\sigma_e^2[\sum_{j=1}^{2m-1}w_j(\boldsymbol{\theta}^*)(\phi^{2m-j}-1)](\sum_{j=1}^{2m-1}\sum_{q=1}^{j}\pi_q\phi^{m(k-r)-2m+2j-q}) & \text{if } r<k-1 \end{cases}$$

(ii) Then we compute the asymptotic variance of the MIDAS-NLS estimator. When $K=m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$
$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}w_j(\boldsymbol{\theta})\phi^{m-j})^2+\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{j-q})^2 & \text{if } r=k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}w_j(\boldsymbol{\theta})\phi^{m-j})^2\phi^{m|r-k|} \\ +\sigma_e^2(\sum_{j=1}^{m}w_j(\boldsymbol{\theta})\phi^{m-j})(\sum_{j=1}^{m-1}\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{m|r-k|-m+2j-q}) & \text{otherwise} \end{cases}$$

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =$$

$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum_{j=1}^{m}w_j(\boldsymbol{\theta})\phi^{m-j})\beta & \\ +\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{j-q})(\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta & \text{if } r=k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum_{j=1}^{m}w_j(\boldsymbol{\theta})\phi^{m-j})\phi^{m(k-r)}\beta & \\ +\sigma_e^2(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum_{j=1}^{m-1}\sum_{q=1}^{j}w_q(\boldsymbol{\theta})\phi^{m(k-r)-m+2j-q})\beta & \text{if } r<k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum_{j=1}^{m}w_j(\boldsymbol{\theta})\phi^{m-j})\phi^{m(r-k)}\beta & \\ +\sigma_e^2(\sum_{j=1}^{m}w_j(\boldsymbol{\theta})\phi^{m-j})(\sum_{j=1}^{m-1}\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m(r-k)-m+2j-q})\beta & \text{if } r>k \end{cases}$$

and

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_s})_{kr} =$$

$$\begin{cases} \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m-j})(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})\beta^2 & \\ +\sigma_e^2\sum_{j=1}^{m-1}(\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{j-q})(\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta^2 & \text{if } r=k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m-j})(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})\phi^{m(k-r)}\beta^2 & \\ +\sigma_e^2(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m-j})(\sum_{j=1}^{m-1}\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m(k-r)-m+2j-q})\beta^2 & \text{if } r<k \\ \frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m-j})(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})\phi^{m(r-k)}\beta^2 & \\ +\sigma_e^2(\sum_{j=1}^{m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m-j})(\sum_{j=1}^{m-1}\sum_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m(k-r)-m+2j-q})\beta^2 & \text{if } r>k \end{cases}$$

When $K = 2m$,

$$\mathbb{E}(\boldsymbol{X}_i(\boldsymbol{\theta})\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =
\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})^2 + \sigma_e^2 \sum_{j=1}^{2m-1} (\sum_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{j-q})^2 & \text{if } r = k \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})^2\phi + \sigma_e^2 \sum_{j=1}^{m-1} (\sum_{q=1}^{m+j} w_q(\boldsymbol{\theta})\phi^{m+j-q}) & \\
\times(\sum_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{j-q}) + \sigma_e^2(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j} w_q(\boldsymbol{\theta})\phi^{m+2j-q}) & \text{if } r = k+1 \text{ or } r = k-1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})^2\phi^{m|r-k|} & \\
+\sigma_e^2(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{m|r-k|-2m+2j-q}) & \text{otherwise}
\end{cases}$$

$$\mathbb{E}(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_s}\boldsymbol{X}_i(\boldsymbol{\theta})')_{kr} =
\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} \frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})\beta & \\
+\sigma_e^2 \sum_{j=1}^{2m-1} (\sum_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{j-q})(\sum_{q=1}^{j} \frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta & \text{if } r = k \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} \frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})\phi\beta + \sigma_e^2 \sum_{j=1}^{m-1} (\sum_{q=1}^{m+j} w_q(\boldsymbol{\theta})\phi^{m+j-q}) & \\
\times(\sum_{q=1}^{j} \frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta + \sigma_e^2(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j} \frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m+2j-q})\beta & \text{if } r = k+1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} \frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})\phi\beta + \sigma_e^2 \sum_{j=1}^{m-1} (\sum_{q=1}^{m+j} \frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m+j-q}) & \\
\times(\sum_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{j-q})\beta + \sigma_e^2(\sum_{j=1}^{2m} \frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=0}^{m-1}\sum_{q=1}^{m+j} w_q(\boldsymbol{\theta})\phi^{m+2j-q})\beta & \text{if } r = k-1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} \frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})\phi^{m(r-k)}\beta & \\
+\sigma_e^2(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j} \frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m(r-k)-2m+2j-q})\beta & \text{if } r > k+1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum_{j=1}^{2m} \frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m} w_j(\boldsymbol{\theta})\phi^{2m-j})\phi^{m(k-r)}\beta & \\
+\sigma_e^2(\sum_{j=1}^{2m} \frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum_{j=1}^{2m-1}\sum_{q=1}^{j} w_q(\boldsymbol{\theta})\phi^{m(k-r)-2m+2j-q})\beta & \text{if } r < k-1
\end{cases}$$

and

$$\mathbb{E}\Big(\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})}{\partial\theta_l}\frac{\partial f(\boldsymbol{X}_i;\boldsymbol{\delta})'}{\partial\theta_s}\Big)_{kr} =$$

$$\begin{cases}
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\beta^2 & \\
+\sigma_e^2\sum\limits_{j=1}^{2m-1}(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{j-q})(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta^2 & \text{if } r=k \\
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\phi\beta^2+\sigma_e^2\sum\limits_{j=1}^{m-1}(\sum\limits_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m+j-q}) & \\
\times(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{j-q})\beta^2+\sigma_e^2(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum\limits_{j=0}^{m-1}\sum\limits_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m+2j-q})\beta^2 & \text{if } r=k+1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\phi\beta^2+\sigma_e^2\sum\limits_{j=1}^{m-1}(\sum\limits_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m+j-q}) & \\
\times(\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{j-q})\beta^2+\sigma_e^2(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=0}^{m-1}\sum\limits_{q=1}^{m+j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m+2j-q})\beta^2 & \text{if } r=k-1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\phi^{m(r-k)}\beta^2 & \\
+\sigma_e^2(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})(\sum\limits_{j=1}^{2m-1}\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_l}\phi^{m(r-k)-2m+2j-q})\beta^2 & \text{if } r>k+1 \\
\frac{\sigma_e^2}{1-\phi^2}(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_s}\phi^{2m-j})\phi^{m(k-r)}\beta^2 & \\
+\sigma_e^2(\sum\limits_{j=1}^{2m}\frac{w_j(\boldsymbol{\theta})}{\partial\theta_l}\phi^{2m-j})(\sum\limits_{j=1}^{2m-1}\sum\limits_{q=1}^{j}\frac{w_q(\boldsymbol{\theta})}{\partial\theta_s}\phi^{m(k-r)-2m+2j-q})\beta^2 & \text{if } r<k-1
\end{cases}$$

for all $s, l = 1, 2$.

Appendix B

Appendix to Chapter 3

B.1 Deriving SQLIML Estimators

Given $\boldsymbol{\delta}$ and ρ, the estimators are

$$\begin{cases} \hat{\sigma}_e^2(\boldsymbol{\delta},\rho) & = & \frac{1}{N}[\boldsymbol{A}(\rho)\boldsymbol{y} - \boldsymbol{M}(\boldsymbol{\delta})\hat{\boldsymbol{\alpha}}(\boldsymbol{\delta},\rho)]'[\boldsymbol{A}(\rho)\boldsymbol{y} - \boldsymbol{M}(\boldsymbol{\delta})\hat{\boldsymbol{\alpha}}(\boldsymbol{\delta},\rho)], \\ \hat{\boldsymbol{\Sigma}} & = & \frac{1}{N}(\boldsymbol{X}_1 - \boldsymbol{Z}\boldsymbol{\delta})'(\boldsymbol{X}_1 - \boldsymbol{Z}\boldsymbol{\delta}), \\ \hat{\boldsymbol{\alpha}}(\boldsymbol{\delta},\rho) & = & (\boldsymbol{M}(\boldsymbol{\delta})'\boldsymbol{M}(\boldsymbol{\delta}))^{-1}\boldsymbol{M}(\boldsymbol{\delta})'\boldsymbol{A}(\rho)\boldsymbol{y}. \end{cases} \tag{B.1.1}$$

Recall that $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{X}_2)$, with $\boldsymbol{Z}_1 = (\boldsymbol{Z}^1, \ldots, \boldsymbol{Z}^{L_1})$. Equation $\frac{\partial lnL(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha}} = \boldsymbol{0}$, therefore, implies that

$$\begin{pmatrix} \delta_1^1 & \cdots & \delta_{L_1}^1 \\ \vdots & \ddots & \vdots \\ \delta_1^{K_1} & \cdots & \delta_{L_1}^{K_1} \end{pmatrix} \begin{pmatrix} \boldsymbol{Z}^{1\prime}\hat{\boldsymbol{e}} \\ \vdots \\ \boldsymbol{Z}^{L_1\prime}\hat{\boldsymbol{e}} \end{pmatrix} = \boldsymbol{0}_{K_1 \times 1}. \tag{B.1.2}$$

To obtain the solution for $\hat{\boldsymbol{\delta}}$, we consider two cases.

Case 1. When the model (3.2.1) is just-identified ($L_1 = K_1$), the above linear system (B.1.2) has a unique solution that is $\boldsymbol{Z}_1'\hat{\boldsymbol{e}} = \boldsymbol{0}$. Plus $\boldsymbol{X}_2'\hat{\boldsymbol{e}} = \boldsymbol{0}$, we have $\boldsymbol{Z}'\hat{\boldsymbol{e}} = \boldsymbol{0}$.

Consequently, $\frac{\partial lnL(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\delta}} = \mathbf{0}$ implies that $\boldsymbol{Z}'\boldsymbol{v}\boldsymbol{\Sigma}^{-1} = \mathbf{0}$, which can be written as

$$\begin{pmatrix} zv_{11} & \cdots & zv_{1K_1} \\ \vdots & \ddots & \vdots \\ zv_{L1} & \cdots & zv_{LK_1} \end{pmatrix} \begin{pmatrix} \Pi_{11} & \cdots & \Pi_{1K_1} \\ \vdots & \ddots & \vdots \\ \Pi_{K_11} & \cdots & \Pi_{K_1K_1} \end{pmatrix} = \mathbf{0}_{L\times K_1},$$

where zv_{ij} is the (i,j)th element of $\boldsymbol{Z}'\boldsymbol{v}$ and Π_{ij} is the (i,j)th element of $\boldsymbol{\Sigma}^{-1}$. It is straightforward to get a linear system of homogeneous equations

$$\begin{cases} \Pi_{11}zv_{11} + \cdots + \Pi_{K_11}zv_{1K_1} = 0, \\ \Pi_{12}zv_{11} + \cdots + \Pi_{K_12}zv_{1K_1} = 0, \\ \vdots \\ \Pi_{1K_1}zv_{11} + \cdots + \Pi_{K_1K_1}zv_{1K_1} = 0. \end{cases}$$

Since $rank(\boldsymbol{\Sigma}^{-1}) = K_1$, $zv_{11} = \cdots = zv_{1K_1} = 0$ is a unique solution to this system. Similarly, we have $\boldsymbol{Z}'\boldsymbol{v} = \mathbf{0}$ and so

$$\hat{\boldsymbol{\delta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}_1.$$

Eventually, combined with (B.1.1), the estimators of σ_e^2, $\boldsymbol{\Sigma}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\delta}$ are unique given by

$$\begin{cases} \hat{\sigma}_e^2(\rho) &= \frac{1}{N}\boldsymbol{y}'\boldsymbol{A}(\rho)'\hat{\boldsymbol{H}}\boldsymbol{A}(\rho)\boldsymbol{y} \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{N}\boldsymbol{X}_1'\boldsymbol{P}\boldsymbol{X}_1 \\ \hat{\boldsymbol{\alpha}}(\rho) &= (\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{M}(\hat{\boldsymbol{\delta}}))^{-1}\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{A}(\rho)\boldsymbol{y} \\ \hat{\boldsymbol{\delta}} &= (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}_1, \end{cases} \tag{B.1.3}$$

where $\hat{\boldsymbol{H}} = \boldsymbol{I}_N - \boldsymbol{M}(\hat{\boldsymbol{\delta}})(\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{M}(\hat{\boldsymbol{\delta}}))^{-1}\boldsymbol{M}(\hat{\boldsymbol{\delta}})'$, $\boldsymbol{M}(\hat{\boldsymbol{\delta}}) = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{P}\boldsymbol{X}_1)$, and $\boldsymbol{P} = \boldsymbol{I}_N - \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$.

Case 2. When the model (3.2.1) is over-identified ($L_1 > K_1$), the linear system (B.1.2) has at least one solution that is $\boldsymbol{Z}_1'\hat{\boldsymbol{e}} = \boldsymbol{0}$. Hence, $\hat{\boldsymbol{\theta}}$ given by (B.1.3) is one estimator of $\boldsymbol{\theta}$, but it is not unique.

B.2 Proof of Theorem 3.1

The way to demonstrate the consistency of $\hat{\boldsymbol{\theta}}$ follows from the uniform convergence and identifiable uniqueness (White (1996, Theorem 3.4)). Recall that $Q(\rho) = max_{\sigma_e^2,\boldsymbol{\Sigma},\boldsymbol{\alpha},\boldsymbol{\delta}}\mathbb{E}(lnL(\boldsymbol{\theta}))$, which is given by

$$Q(\rho) = -\frac{N(K_1+1)}{2}(ln(2\pi)+1) - \frac{N}{2}ln\sigma_e^{*2}(\rho) - \frac{N}{2}ln\left|\boldsymbol{\Sigma}_0\right| + ln\left|\boldsymbol{A}(\rho)\right|.$$

Uniform Convergence

We begin by showing that $sup_{\rho\in\boldsymbol{\Lambda}}\left|\frac{1}{N}lnL(\rho) - \frac{1}{N}Q(\rho)\right| = o_p(1)$, where

$$\frac{1}{N}lnL(\rho) - \frac{1}{N}Q(\rho) = -\frac{1}{2}(ln\hat{\sigma}_e^2(\rho) - ln\sigma_e^{*2}(\rho)) - \frac{1}{2}(ln\left|\hat{\boldsymbol{\Sigma}}\right| - ln\left|\boldsymbol{\Sigma}_0\right|).$$

First, it is easy to show that

$$\hat{\boldsymbol{\delta}} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}_1 = \boldsymbol{\delta}_0 + (\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N})^{-1}(\frac{\boldsymbol{Z}'\boldsymbol{v}_0}{N}) = \boldsymbol{\delta}_0 + o_p(1),$$

and therefore

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}} &= \tfrac{1}{N}[\boldsymbol{X}_1 - \boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{Z}(\boldsymbol{\delta}_0 - \hat{\boldsymbol{\delta}})]'[\boldsymbol{X}_1 - \boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{Z}(\boldsymbol{\delta}_0 - \hat{\boldsymbol{\delta}})] \\
&= \tfrac{1}{N}[\boldsymbol{v}_0'\boldsymbol{v}_0 + 2(\boldsymbol{\delta}_0 - \hat{\boldsymbol{\delta}})'(\boldsymbol{Z}'\boldsymbol{v}_0) + (\boldsymbol{\delta}_0 - \hat{\boldsymbol{\delta}})'(\boldsymbol{Z}'\boldsymbol{Z})(\boldsymbol{\delta}_0 - \hat{\boldsymbol{\delta}})] \\
&= \boldsymbol{\Sigma}_0 + o_p(1),
\end{aligned}$$

then by the continuous mapping theorem (CMT), we have $ln\left|\hat{\boldsymbol{\Sigma}}\right| - ln\left|\boldsymbol{\Sigma}_0\right| = o_p(1)$.

We then show that $\hat{\sigma}_e^2(\rho) - \sigma_e^{*2}(\rho) = o_p(1)$ uniformly on $\mathbf{\Lambda}$. Recall that

$$\begin{aligned}\hat{\sigma}_e^2(\rho) &= \tfrac{1}{N}[\boldsymbol{A}(\rho)\boldsymbol{y} - \boldsymbol{M}(\hat{\boldsymbol{\delta}})\hat{\boldsymbol{\alpha}}(\rho)]'[\boldsymbol{A}(\rho)\boldsymbol{y} - \boldsymbol{M}(\hat{\boldsymbol{\delta}})\hat{\boldsymbol{\alpha}}(\rho)] \\ &= \tfrac{1}{N}[\boldsymbol{J}_1 + \boldsymbol{J}_2 - \boldsymbol{M}(\hat{\boldsymbol{\delta}})(\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{M}(\hat{\boldsymbol{\delta}}))^{-1}\boldsymbol{M}(\hat{\boldsymbol{\delta}})'(\boldsymbol{J}_1 + \boldsymbol{J}_2)]' \\ &\quad \times[\boldsymbol{J}_1 + \boldsymbol{J}_2 - \boldsymbol{M}(\hat{\boldsymbol{\delta}})(\boldsymbol{M}(\hat{\boldsymbol{\delta}})'\boldsymbol{M}(\hat{\boldsymbol{\delta}}))^{-1}\boldsymbol{M}(\hat{\boldsymbol{\delta}})'(\boldsymbol{J}_1 + \boldsymbol{J}_2)]\end{aligned}$$

and

$$\begin{aligned}\sigma_e^{*2}(\rho) = &\tfrac{1}{N}[\mathbb{E}(\boldsymbol{J}_1'\boldsymbol{J}_1) - \mathbb{E}(\boldsymbol{J}_1'\boldsymbol{M}(\boldsymbol{\delta}_0))(\mathbb{E}[\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0)])^{-1}\mathbb{E}(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{J}_1) \\ &+\sigma_{e0}^2 tr(\boldsymbol{A}(\rho_0)^{-1\prime}\boldsymbol{A}(\rho)'\boldsymbol{A}(\rho)\boldsymbol{A}(\rho_0)^{-1})],\end{aligned}$$

where $\boldsymbol{J}_1 = \boldsymbol{A}(\rho)\boldsymbol{A}(\rho_0)^{-1}\boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0$ and $\boldsymbol{J}_2 = \boldsymbol{A}(\rho)\boldsymbol{A}(\rho_0)^{-1}\boldsymbol{e}_0$. By $\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0 = o_p(1)$, the CMT gives $\boldsymbol{M}(\hat{\boldsymbol{\delta}}) - \boldsymbol{M}(\boldsymbol{\delta}_0) = o_p(1)$. It follows that

$$\begin{aligned}\hat{\sigma}_e^2(\rho) &= \tfrac{1}{N}[\boldsymbol{J}_1 + \boldsymbol{J}_2 - \boldsymbol{M}(\boldsymbol{\delta}_0)(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0))^{-1}\boldsymbol{M}(\boldsymbol{\delta}_0)'(\boldsymbol{J}_1 + \boldsymbol{J}_2)]' \\ &\quad \times[\boldsymbol{J}_1 + \boldsymbol{J}_2 - \boldsymbol{M}(\boldsymbol{\delta}_0)(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0))^{-1}\boldsymbol{M}(\boldsymbol{\delta}_0)'(\boldsymbol{J}_1 + \boldsymbol{J}_2)] + o_p(1) \\ &= \tfrac{1}{N}[\boldsymbol{J}_1'\boldsymbol{J}_1 - (\boldsymbol{J}_1'\boldsymbol{M}(\boldsymbol{\delta}_0))(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0))^{-1}(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{J}_1) \\ &\quad +\boldsymbol{J}_2'\boldsymbol{J}_2 + 2\boldsymbol{J}_1'\boldsymbol{J}_2 - 2\boldsymbol{J}_1'\boldsymbol{M}(\boldsymbol{\delta}_0)(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0))^{-1}\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{J}_2 \\ &\quad -(\boldsymbol{J}_2'\boldsymbol{M}(\boldsymbol{\delta}_0))(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{M}(\boldsymbol{\delta}_0))^{-1}(\boldsymbol{M}(\boldsymbol{\delta}_0)'\boldsymbol{J}_2)] + o_p(1) \\ &= \sigma_e^{*2}(\rho) + o_p(1).\end{aligned}$$

Finally, with the above results, the uniform convergence follows.

Uniquely Identifiable

Second, we need to show that $\boldsymbol{\theta}_0$ is uniquely identifiable. We have

$$\frac{1}{N}[Q(\rho) - Q(\rho_0)] = \frac{1}{N}[Q_p(\rho) - Q_p(\rho_0)] - \frac{1}{2}(ln\sigma_e^{*2}(\rho) - ln\sigma_e^2(\rho)),$$

where $\sigma_e^2(\rho) = \frac{1}{N}\sigma_{e0}^2 tr(\boldsymbol{A}(\rho_0)^{-1\prime}\boldsymbol{A}(\rho)'\boldsymbol{A}(\rho)\boldsymbol{A}(\rho_0)^{-1})$ and $Q_p(\rho) = -\frac{N}{2}(ln(2\pi) + 1) - \frac{N}{2}ln\sigma_e^2(\rho) + ln\,|\boldsymbol{A}(\rho)|$. We can describe $Q_p(\rho) = max_{\sigma_e^2}\mathbb{E}(lnL_p(\rho, \sigma_e^2))$, where $lnL_p(\rho, \sigma_e^2)$ is the log-likelihood function of a SLM model $\boldsymbol{y} = \rho_0\boldsymbol{W}\boldsymbol{y} + \boldsymbol{e}_0$, $\boldsymbol{e}_0 \sim NID(0, \sigma_{e0}^2\boldsymbol{I}_N)$. By

Jensen's inequality, we have $Q_p(\rho) - Q_p(\rho_0) \leq 0$ for all ρ. Besides, Assumption 7 that ensures $ln\sigma_e^{*2}(\rho) - ln\sigma_e^2(\rho) > 0$. Hence, the identifiable uniqueness holds.

We have shown that $\frac{1}{N}lnL(\rho) - \frac{1}{N}Q(\rho)$ converges in probability to zero uniformly on $\rho \in \boldsymbol{\Lambda}$ and the identifiable uniqueness holds. Consequently, the consistency of $\hat{\rho}$, and thus, $\hat{\boldsymbol{\theta}}$ follows.

B.3 Proof of Theorem 3.2

Given that $\frac{\partial lnL(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \mathbf{0}$. Then, by the mean value theorem, we have that

$$\mathbf{0} = \frac{\partial lnL(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_1$ is a point between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$, therefore

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'})^{-1}\frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}.$$

(i) First, we show that $\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} - \frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xrightarrow{p} 0$. The second-order derivatives of $lnL(\boldsymbol{\theta})$ in (3.2.8), which are assumed to exist and be continuous in the neighborhood of $\boldsymbol{\theta}_0$, are as follows. For parameter σ_e^2,

$$\begin{aligned}
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \sigma_e^2 \partial \sigma_e^2} &= \frac{N}{2\sigma_e^4} - \frac{\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{e}(\boldsymbol{\phi})}{\sigma_e^6}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \sigma_e^2 \partial \text{vech}(\boldsymbol{\Sigma})'} &= \mathbf{0}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \sigma_e^2 \partial \boldsymbol{\beta}_1'} &= -\frac{\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{X}_1}{\sigma_e^4}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \sigma_e^2 \partial \boldsymbol{\beta}_2'} &= -\frac{\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{X}_2}{\sigma_e^4}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \sigma_e^2 \partial \boldsymbol{\lambda}'} &= -\frac{\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{v}(\boldsymbol{\delta})}{\sigma_e^4}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \sigma_e^2 \partial \text{vec}(\boldsymbol{\delta})'} &= \text{vec}(\frac{\boldsymbol{Z}'\boldsymbol{e}(\boldsymbol{\phi})\boldsymbol{\lambda}'}{\sigma_e^4})', \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \sigma_e^2 \partial \rho} &= -\frac{\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{W}\boldsymbol{y}}{\sigma_e^4}.
\end{aligned}$$

For parameter $\boldsymbol{\Sigma}$,

$$
\begin{aligned}
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})\partial \mathrm{vech}(\boldsymbol{\Sigma})'} &= \boldsymbol{D}'[\boldsymbol{\Sigma}^{-1} \otimes (\tfrac{N}{2}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\boldsymbol{v}(\boldsymbol{\delta})'\boldsymbol{v}(\boldsymbol{\delta})\boldsymbol{\Sigma}^{-1})]\boldsymbol{D}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})\partial \boldsymbol{\beta}_1'} &= \mathbf{0}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})\partial \boldsymbol{\beta}_2'} &= \mathbf{0}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})\partial \boldsymbol{\lambda}'} &= \mathbf{0}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\delta})'} &= -\boldsymbol{D}'(\boldsymbol{\Sigma}^{-1} \otimes (\boldsymbol{\Sigma}^{-1}\boldsymbol{v}(\boldsymbol{\delta})'\boldsymbol{Z})), \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \mathrm{vech}(\boldsymbol{\Sigma})\partial \rho} &= \mathbf{0},
\end{aligned}
$$

where $\boldsymbol{D}$ is the duplication matrix of $\boldsymbol{\Sigma}$. For parameter $\boldsymbol{\beta}_1$,

$$
\begin{aligned}
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_1'} &= -\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2'} &= -\frac{\boldsymbol{X}_1'\boldsymbol{X}_2}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\lambda}'} &= -\frac{\boldsymbol{X}_1'\boldsymbol{v}(\boldsymbol{\delta})}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1 \partial \mathrm{vec}(\boldsymbol{\delta})'} &= \frac{\boldsymbol{\lambda}' \otimes (\boldsymbol{X}_1'\boldsymbol{Z})}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1 \partial \rho} &= -\frac{\boldsymbol{X}_1'\boldsymbol{W}\boldsymbol{y}}{\sigma_e^2}.
\end{aligned}
$$

For parameter $\boldsymbol{\beta}_2$,

$$
\begin{aligned}
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_2'} &= -\frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\lambda}'} &= -\frac{\boldsymbol{X}_2'\boldsymbol{v}(\boldsymbol{\delta})}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_2 \partial \mathrm{vec}(\boldsymbol{\delta})'} &= \frac{\boldsymbol{\lambda}' \otimes (\boldsymbol{X}_2'\boldsymbol{Z})}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_2 \partial \rho} &= -\frac{\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{y}}{\sigma_e^2}.
\end{aligned}
$$

For parameter $\boldsymbol{\lambda}$,

$$
\begin{aligned}
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}'} &= -\frac{\boldsymbol{v}(\boldsymbol{\delta})'\boldsymbol{v}(\boldsymbol{\delta})}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \mathrm{vec}(\boldsymbol{\delta})'} &= \frac{\boldsymbol{\lambda}' \otimes (\boldsymbol{v}(\boldsymbol{\delta})'\boldsymbol{Z}) - \mathbb{I}_{K_1} \otimes (\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{Z})}{\sigma_e^2}, \\
\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \rho} &= -\frac{\boldsymbol{v}(\boldsymbol{\delta})'\boldsymbol{W}\boldsymbol{y}}{\sigma_e^2},
\end{aligned}
$$

where $\mathbb{I}_{K_1}$ is a $K_1 \times K_1$ identity matrix. For parameter $\boldsymbol{\delta}$,

$$\begin{aligned}\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \text{vec}(\boldsymbol{\delta})\partial \text{vec}(\boldsymbol{\delta})'} &= -\frac{(\boldsymbol{\lambda}\boldsymbol{\lambda}')\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{\sigma_e^2} - \boldsymbol{\Sigma}^{-1}\otimes(\boldsymbol{Z}'\boldsymbol{Z}),\\ \frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial \text{vec}(\boldsymbol{\delta})\partial\rho} &= \frac{\boldsymbol{\lambda}\otimes(\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y})}{\sigma_e^2}.\end{aligned}$$

For parameter ρ,

$$\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial\rho\partial\rho} = -\frac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{y}}{\sigma_e^2} - tr(\boldsymbol{G}^2(\rho)).$$

We now show that the convergence of the difference between each of the above second-order derivatives at $\boldsymbol{\theta}_1$ and their counterparts at $\boldsymbol{\theta}_0$ to zero in probability. Denote $\boldsymbol{\theta}_1 = (\sigma_{e1}^2, \text{vech}(\boldsymbol{\Sigma}_1)', \boldsymbol{\beta}_{11}', \boldsymbol{\beta}_{21}', \boldsymbol{\lambda}_1', \text{vec}(\boldsymbol{\delta}_1)', \rho_1)'$, where $\boldsymbol{\lambda}_1 = (\lambda_{11}, \lambda_{21}, \ldots, \lambda_{K_1 1})'$, $\boldsymbol{\delta}_1 = (\boldsymbol{\delta}_1^1, \ldots, \boldsymbol{\delta}_1^{K_1})$. Let $\boldsymbol{\phi}_1 = (\boldsymbol{\alpha}_1', \boldsymbol{\delta}_1', \rho_1)'$, $\boldsymbol{X} \equiv (\boldsymbol{X}_1, \boldsymbol{X}_2)$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ with $\boldsymbol{\alpha}_1 = (\boldsymbol{\beta}_{11}', \boldsymbol{\beta}_{21}', \boldsymbol{\lambda}_1')'$. We have $\boldsymbol{v}(\boldsymbol{\delta}_1) = \boldsymbol{v}_0 + \boldsymbol{Z}(\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1)$ and

$$\boldsymbol{e}(\boldsymbol{\phi}_1) = \boldsymbol{e}_0 + (\rho_0-\rho_1)\boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_1) + \boldsymbol{X}_1(\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}_1) - \boldsymbol{Z}\boldsymbol{\delta}_0(\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}_1) - \boldsymbol{Z}(\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1)\boldsymbol{\lambda}_1.$$

Under Assumptions 5-6 and 9, $\frac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{X}}{N} = O_p(1)$, $\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{N} = O_p(1)$, $\frac{\boldsymbol{Z}'\boldsymbol{X}_1}{N} = O_p(1)$, and $\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N} = O(1)$. For parameter σ_e^2, we have

$$\begin{aligned}&\tfrac{1}{N}\Big(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\sigma_e^2\partial\sigma_e^2} - \tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\sigma_e^2}\Big)\\ &= \tfrac{1}{2}\big(\tfrac{1}{\sigma_{e1}^4} - \tfrac{1}{\sigma_{e0}^4}\big) + \big(\tfrac{1}{\sigma_{e0}^6} - \tfrac{1}{\sigma_{e1}^6}\big)\tfrac{\boldsymbol{e}_0'\boldsymbol{e}_0}{N} - 2(\rho_0-\rho_1)\tfrac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{e}_0}{N\sigma_{e1}^6} - 2(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)'\tfrac{\boldsymbol{X}'\boldsymbol{e}_0}{N\sigma_{e1}^6} - 2(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\tfrac{\boldsymbol{X}_1'\boldsymbol{e}_0}{N\sigma_{e1}^6}\\ &+2(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\tfrac{\boldsymbol{Z}'\boldsymbol{e}_0}{N\sigma_{e1}^6} + 2\boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\tfrac{\boldsymbol{Z}'\boldsymbol{e}_0}{N\sigma_{e1}^6} - (\rho_0-\rho_1)^2\tfrac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^6} - 2(\rho_0-\rho_1)(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)'\tfrac{\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^6}\\ &-2(\rho_0-\rho_1)(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\tfrac{\boldsymbol{X}_1'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^6} + 2(\rho_0-\rho_1)(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\tfrac{\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^6} + 2(\rho_0-\rho_1)\boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\tfrac{\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^6}\\ &-(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)'\tfrac{\boldsymbol{X}'\boldsymbol{X}}{N\sigma_{e1}^6}(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1) - 2(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\tfrac{\boldsymbol{X}_1'\boldsymbol{X}}{N\sigma_{e1}^6}(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1) + 2(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\tfrac{\boldsymbol{Z}'\boldsymbol{X}}{N\sigma_{e1}^6}(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)\\ &+2\boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\tfrac{\boldsymbol{Z}'\boldsymbol{X}}{N\sigma_{e1}^6}(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1) - (\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\tfrac{\boldsymbol{X}_1'\boldsymbol{X}_1}{N\sigma_{e1}^6}(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1) + 2(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\tfrac{\boldsymbol{Z}'\boldsymbol{X}_1}{N\sigma_{e1}^6}(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)\\ &+2\boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\tfrac{\boldsymbol{Z}'\boldsymbol{X}_1}{N\sigma_{e1}^6}(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1) - (\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\tfrac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^6}\boldsymbol{\delta}_0(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)\\ &+2\boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\tfrac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^6}\boldsymbol{\delta}_0(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1) - \boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\tfrac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^6}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)\boldsymbol{\lambda}_1 = o_p(1),\end{aligned}$$

$$\begin{aligned}\tfrac{1}{N}\Big(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\sigma_e^2\partial\boldsymbol{\beta}'} - \tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\boldsymbol{\beta}'}\Big) &= \big(\tfrac{1}{\sigma_{e0}^4} - \tfrac{1}{\sigma_{e1}^4}\big)\tfrac{\boldsymbol{e}_0'\boldsymbol{X}}{N} - (\rho_0-\rho_1)\tfrac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{X}}{N\sigma_{e1}^4} - (\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)'\tfrac{\boldsymbol{X}'\boldsymbol{X}}{N\sigma_{e1}^4}\\ &-(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\tfrac{\boldsymbol{X}_1'\boldsymbol{X}}{N\sigma_{e1}^4} + (\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\tfrac{\boldsymbol{Z}'\boldsymbol{X}}{N\sigma_{e1}^4} + \boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\tfrac{\boldsymbol{Z}'\boldsymbol{X}}{N\sigma_{e1}^4} = o_p(1),\end{aligned}$$

$$\frac{1}{N}(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\sigma_e^2\partial\boldsymbol{\lambda}'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\boldsymbol{\lambda}'})$$
$$= (\frac{1}{\sigma_{e0}^4} - \frac{1}{\sigma_{e1}^4})\frac{\boldsymbol{e}_0'\boldsymbol{v}_0}{N} - (\rho_0-\rho_1)\frac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{v}_0}{N\sigma_{e1}^4} - (\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)'\frac{\boldsymbol{X}'\boldsymbol{v}_0}{N\sigma_{e1}^4} - (\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\frac{\boldsymbol{X}_1'\boldsymbol{v}_0}{N\sigma_{e1}^4} + (\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\frac{\boldsymbol{Z}'\boldsymbol{v}_0}{N\sigma_{e1}^4}$$
$$+\boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{v}_0}{N\sigma_{e1}^4} - \frac{\boldsymbol{e}_0'\boldsymbol{Z}}{N\sigma_{e1}^4}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1) - (\rho_0-\rho_1)\frac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{Z}}{N\sigma_{e1}^4}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1) - (\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)'\frac{\boldsymbol{X}'\boldsymbol{Z}}{N\sigma_{e1}^4}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)$$
$$-(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\frac{\boldsymbol{X}_1'\boldsymbol{Z}}{N\sigma_{e1}^4}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1) + (\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^4}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1) + \boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^4}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1) = o_p(1),$$

$$\begin{aligned}\frac{1}{N}(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\sigma_e^2\partial\mathrm{vec}(\boldsymbol{\delta})'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\mathrm{vec}(\boldsymbol{\delta})'}) &= \mathrm{vec}[(\frac{1}{\sigma_{e1}^4} - \frac{1}{\sigma_{e0}^4})\frac{\boldsymbol{Z}'\boldsymbol{e}_0\boldsymbol{\lambda}_0'}{N} - \frac{\boldsymbol{Z}'\boldsymbol{e}_0}{N\sigma_{e1}^4}(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)' + (\rho_0-\rho_1)\frac{\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^4}\boldsymbol{\lambda}_1\\
&+\frac{\boldsymbol{Z}'\boldsymbol{X}}{N\sigma_{e1}^4}(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)\boldsymbol{\lambda}_1' + \frac{\boldsymbol{Z}'\boldsymbol{X}_1}{N\sigma_{e1}^4}(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)\boldsymbol{\lambda}_1' - \frac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^4}\boldsymbol{\delta}_0(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)\boldsymbol{\lambda}_1'\\
&-\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^4}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1'] = o_p(1),\\
\frac{1}{N}(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\sigma_e^2\partial\rho} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\rho}) &= (\frac{1}{\sigma_{e0}^4} - \frac{1}{\sigma_{e1}^4})\frac{\boldsymbol{e}_0'\boldsymbol{W}\boldsymbol{y}}{N} - (\rho_0-\rho_1)\frac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^4} - (\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)'\frac{\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^4}\\
&-(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\frac{\boldsymbol{X}_1'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^4} + (\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\frac{\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^4} + \boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^4}\\
&= o_p(1),\end{aligned}$$

here we use the fact that $\boldsymbol{\theta}_1 \xrightarrow{p} \boldsymbol{\theta}_0$. Because σ_{e1}^2 lies between $\hat{\sigma_e^2}$ and σ_{e0}^2, σ_{e1}^2 can be represented as $\sigma_{e1}^2 = \lambda\hat{\sigma_e^2} + (1-\lambda)\sigma_{e0}^2$, $\exists\lambda \in (0,1)$. We have shown that $\hat{\sigma_e^2} \xrightarrow{p} \sigma_{e0}^2$, then we have for any $\delta > 0$

$$\begin{aligned}\mathbb{P}(\mid \sigma_{e1}^2 - \sigma_{e0}^2 \mid > \delta) &= \mathbb{P}(\mid \lambda\hat{\sigma_e^2} + (1-\lambda)\sigma_{e0}^2 - \sigma_{e0}^2 \mid > \delta)\\
&= \lambda\mathbb{P}(\mid \hat{\sigma_e^2} - \sigma_{e0}^2 \mid > \delta) \to 0,\end{aligned}$$

which implies that $\sigma_{e1}^2 \xrightarrow{p} \sigma_{e0}^2$. Using the same way, we have $\boldsymbol{\theta}_1 \xrightarrow{p} \boldsymbol{\theta}_0$. Similarly, for parameter $\boldsymbol{\Sigma}$,

$$\frac{1}{N}(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\mathrm{vech}(\boldsymbol{\Sigma})\partial\mathrm{vech}(\boldsymbol{\Sigma})'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\mathrm{vech}(\boldsymbol{\Sigma})\partial\mathrm{vech}(\boldsymbol{\Sigma})'})$$
$$= \frac{1}{2}\boldsymbol{D}'(\boldsymbol{\Sigma}_1^{-1}\otimes\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\Sigma}_0^{-1})\boldsymbol{D} - \boldsymbol{D}'[(\boldsymbol{\Sigma}_1^{-1}\otimes(\boldsymbol{\Sigma}_1^{-1}\frac{\boldsymbol{v}_0'\boldsymbol{v}_0}{N}\boldsymbol{\Sigma}_1^{-1}) - \boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{\Sigma}_0^{-1}\frac{\boldsymbol{v}_0'\boldsymbol{v}_0}{N}\boldsymbol{\Sigma}_0^{-1}))$$
$$+\boldsymbol{\Sigma}_1^{-1}\otimes(\boldsymbol{\Sigma}_1^{-1}\frac{\boldsymbol{v}_0'\boldsymbol{Z}}{N}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)\boldsymbol{\Sigma}_1^{-1}) + \boldsymbol{\Sigma}_1^{-1}\otimes(\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{v}_0}{N}\boldsymbol{\Sigma}_1^{-1})$$
$$+\boldsymbol{\Sigma}_1^{-1}\otimes(\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)\boldsymbol{\Sigma}_1^{-1})]\boldsymbol{D} = o_p(1),$$

$$\frac{1}{N}(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\mathrm{vech}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\delta})'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\mathrm{vech}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\delta})'})$$
$$= -\boldsymbol{D}'[(\boldsymbol{\Sigma}_1^{-1}\otimes(\boldsymbol{\Sigma}_1^{-1}\frac{\boldsymbol{v}_0'\boldsymbol{Z}}{N}) - \boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{\Sigma}_0^{-1}\frac{\boldsymbol{v}_0'\boldsymbol{Z}}{N})) + \boldsymbol{\Sigma}_1^{-1}\otimes(\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N}) = o_p(1).$$

For parameter $\boldsymbol{\beta}_1$,

$$\begin{aligned}
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1'}\right) &= \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{N} = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2'}\right) &= \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{X}_1'\boldsymbol{X}_2}{N} = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\lambda}'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\lambda}'}\right) &= \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{X}_1'\boldsymbol{v}_0}{N} - \frac{\boldsymbol{X}_1'\boldsymbol{Z}}{N\sigma_{e1}^2}(\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1) = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_1\partial\text{vec}(\boldsymbol{\delta})'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\text{vec}(\boldsymbol{\delta})'}\right) &= \left(\frac{1}{\sigma_{e1}^2} - \frac{1}{\sigma_{e0}^2}\right)\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_1'\boldsymbol{Z})}{N} - \frac{(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\otimes(\boldsymbol{X}_1'\boldsymbol{Z})}{N\sigma_{e1}^2} = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_1\partial\rho} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\rho}\right) &= \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{X}_1'\boldsymbol{W}\boldsymbol{y}}{N} = o_p(1).
\end{aligned}$$

For parameter $\boldsymbol{\beta}_2$,

$$\begin{aligned}
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_2'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_2'}\right) &= \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{N} = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\lambda}'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\lambda}'}\right) &= \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{X}_2'\boldsymbol{v}_0}{N} - \frac{\boldsymbol{X}_2'\boldsymbol{Z}}{N\sigma_{e1}^2}(\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1) = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_2\partial\text{vec}(\boldsymbol{\delta})'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2\partial\text{vec}(\boldsymbol{\delta})'}\right) &= \left(\frac{1}{\sigma_{e1}^2} - \frac{1}{\sigma_{e0}^2}\right)\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_2'\boldsymbol{Z})}{N} - \frac{(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)'\otimes(\boldsymbol{X}_2'\boldsymbol{Z})}{N\sigma_{e1}^2} = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\beta}_2\partial\rho} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2\partial\rho}\right) &= \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{y}}{N} = o_p(1).
\end{aligned}$$

For parameter $\boldsymbol{\lambda}$,

$$\begin{aligned}
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}\right) =& \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{v}_0'\boldsymbol{v}_0}{N} - (\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{v}_0}{N\sigma_{e1}^2} - \frac{\boldsymbol{v}_0'\boldsymbol{Z}}{N\sigma_{e1}^2}(\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1)\\
&-(\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^2}(\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1) = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\lambda}\partial\text{vec}(\boldsymbol{\delta})'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}\partial\text{vec}(\boldsymbol{\delta})'}\right) =& \left(\frac{1}{\sigma_{e1}^2} - \frac{1}{\sigma_{e0}^2}\right)\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{v}_0'\boldsymbol{Z})}{N} - (\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}_1)'\otimes\frac{\boldsymbol{v}_0'\boldsymbol{Z}}{N\sigma_{e1}^2} + \boldsymbol{\lambda}_1'\otimes\frac{(\boldsymbol{\delta}_0-\boldsymbol{\delta}_1)'\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^2}\\
&+\left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\mathbb{I}_{K_1}\otimes\frac{\boldsymbol{e}_0'\boldsymbol{Z}}{N} - (\rho_0 - \rho_1)\mathbb{I}_{K_1}\otimes\frac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{Z}}{N\sigma_{e1}^2}\\
&-\mathbb{I}_{K_1}\otimes\left((\boldsymbol{\beta}_0 - \boldsymbol{\beta}_1)'\frac{\boldsymbol{X}'\boldsymbol{Z}}{N\sigma_{e1}^2}\right) - \mathbb{I}_{K_1}\otimes\left((\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}_1)'\frac{\boldsymbol{X}_1'\boldsymbol{Z}}{N\sigma_{e1}^2}\right)\\
&+\mathbb{I}_{K_1}\otimes\left((\boldsymbol{\lambda}_0 - \boldsymbol{\lambda}_1)'\boldsymbol{\delta}_0'\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^2}\right) + \mathbb{I}_{K_1}\otimes\left(\boldsymbol{\lambda}_1'(\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e1}^2}\right) = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\boldsymbol{\lambda}\partial\rho} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}\partial\rho}\right) =& \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{\boldsymbol{v}_0'\boldsymbol{W}\boldsymbol{y}}{N} - (\boldsymbol{\delta}_0 - \boldsymbol{\delta}_1)'\frac{\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y}}{N\sigma_{e1}^2} = o_p(1).
\end{aligned}$$

For parameter $\boldsymbol{\delta}$,

$$\begin{aligned}
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\text{vec}(\boldsymbol{\delta})\partial\text{vec}(\boldsymbol{\delta})'} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\text{vec}(\boldsymbol{\delta})\partial\text{vec}(\boldsymbol{\delta})'}\right) &= \left(\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2}\right)\frac{(\boldsymbol{\lambda}_0\boldsymbol{\lambda}_0')\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N} - \frac{(\boldsymbol{\lambda}_1\boldsymbol{\lambda}_1'-\boldsymbol{\lambda}_0\boldsymbol{\lambda}_0')\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e1}^2}\\
&\quad -(\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\Sigma}_0^{-1})\otimes\frac{\boldsymbol{Z}'\boldsymbol{Z}}{N} = o_p(1),\\
\frac{1}{N}\left(\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\text{vec}(\boldsymbol{\delta})\partial\rho} - \frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\text{vec}(\boldsymbol{\delta})\partial\rho}\right) &= \left(\frac{1}{\sigma_{e1}^2} - \frac{1}{\sigma_{e0}^2}\right)\frac{\boldsymbol{\lambda}_0\otimes(\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y})}{N} - \frac{(\boldsymbol{\lambda}_0-\boldsymbol{\lambda}_1)\otimes(\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y})}{N\sigma_{e1}^2} = o_p(1).
\end{aligned}$$

For parameter ρ,

$$\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_1)}{\partial\rho\partial\rho} - \frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\rho\partial\rho} = (\frac{1}{\sigma_{e0}^2} - \frac{1}{\sigma_{e1}^2})\frac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{y}}{N} - \frac{tr(\boldsymbol{G}^2(\rho_1)) - tr(\boldsymbol{G}^2(\rho_0))}{N} = o_p(1),$$

here we use the mean value theorem that $tr(\boldsymbol{G}^2(\rho_1)) - tr(\boldsymbol{G}^2(\rho_0)) = 2tr(\boldsymbol{G}^3(\rho_2))(\rho_1 - \rho_0) = o_p(1)$, where ρ_2 is a point between ρ_1 and ρ_0, and we use the fact that $tr(\boldsymbol{G}^3(\rho_2)) = O(\frac{N}{h_N})$ and $\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{y} = O_p(\frac{N}{h_N})$. Under Assumption 9, $\boldsymbol{G}(\rho)$ is uniformly bounded in both row and column sums in the neighborhood of ρ_0. Thus, $tr(\boldsymbol{G}^3(\rho_2)) = O(\frac{N}{h_N})$ (Lemma A.8 in Lee (2004)). Therefore, we can conclude that the differences between the second-order derivatives of $\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ at $\boldsymbol{\theta}_1$ and those at $\boldsymbol{\theta}_0$ in probability to zero.

(ii) We show that $\frac{1}{N}[\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \frac{1}{N}\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})] \xrightarrow{p} 0$. The expectation of $\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ for each parameter at $\boldsymbol{\theta}_0$ are as follows. For parameter σ_e^2,

$$\begin{aligned}
\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\sigma_e^2}) &= -\frac{N}{2\sigma_{e0}^4}, \\
\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\boldsymbol{\beta}_1'}) &= \mathbf{0}, \\
\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\boldsymbol{\beta}_2'}) &= \mathbf{0}, \\
\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\boldsymbol{\lambda}'}) &= \mathbf{0}, \\
\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\mathrm{vec}(\boldsymbol{\delta})'}) &= \mathbf{0}, \\
\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\sigma_e^2\partial\rho}) &= -\frac{1}{\sigma_{e0}^4}\mathbb{E}(\boldsymbol{e}_0'\boldsymbol{G}(\boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0 + \boldsymbol{e}_0)) = -\frac{1}{\sigma_{e0}^4}\mathbb{E}(\boldsymbol{e}_0'\boldsymbol{G}\boldsymbol{e}_0) = -\frac{1}{\sigma_{e0}^2}tr(\boldsymbol{G}).
\end{aligned}$$

For parameter $\boldsymbol{\Sigma}$,

$$\begin{aligned}
\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\mathrm{vech}(\boldsymbol{\Sigma})\partial\mathrm{vech}(\boldsymbol{\Sigma})'}) &= -\frac{N}{2}\boldsymbol{D}'(\boldsymbol{\Sigma}_0^{-1} \otimes \boldsymbol{\Sigma}_0^{-1})\boldsymbol{D}, \\
\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\mathrm{vech}(\boldsymbol{\Sigma})\partial\boldsymbol{\delta}'}) &= \mathbf{0}.
\end{aligned}$$

For parameter $\boldsymbol{\beta}_1$,

$$\begin{aligned}
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_1'}) &= -\tfrac{1}{\sigma_{e0}^2}\mathbb{E}[(\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)'(\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)] = -\tfrac{1}{\sigma_{e0}^2}(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta}_0+N\boldsymbol{\Sigma}_0),\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\beta}_2'}) &= -\tfrac{1}{\sigma_{e0}^2}\mathbb{E}[(\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)'\boldsymbol{X}_2] = -\tfrac{1}{\sigma_{e0}^2}\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{X}_2,\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\boldsymbol{\lambda}'}) &= -\tfrac{1}{\sigma_{e0}^2}\mathbb{E}[(\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)'\boldsymbol{v}_0] = -\tfrac{N}{\sigma_{e0}^2}\boldsymbol{\Sigma}_0,\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\mathrm{vec}(\boldsymbol{\delta})'}) &= \tfrac{1}{\sigma_{e0}^2}\boldsymbol{\lambda}_0'\otimes\mathbb{E}[(\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)'\boldsymbol{Z}] = \tfrac{1}{\sigma_{e0}^2}\boldsymbol{\lambda}_0'\otimes(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}),\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_1\partial\rho}) &= -\tfrac{1}{\sigma_{e0}^2}\mathbb{E}[(\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)'\boldsymbol{G}((\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}+\boldsymbol{v}_0\boldsymbol{\lambda}_0+\boldsymbol{e}_0)]\\
&= -\tfrac{1}{\sigma_{e0}^2}(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}) - \tfrac{1}{\sigma_{e0}^2}\mathbb{E}(\boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{v}_0)(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)\\
&= -\tfrac{1}{\sigma_{e0}^2}[(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})+tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)].
\end{aligned}$$

For parameter $\boldsymbol{\beta}_2$,

$$\begin{aligned}
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\beta}_2'}) &= -\tfrac{\boldsymbol{X}_2'\boldsymbol{X}_2}{\sigma_{e0}^2},\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2\partial\boldsymbol{\lambda}'}) &= \boldsymbol{0},\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2\partial\partial\mathrm{vec}(\boldsymbol{\delta})'}) &= \tfrac{1}{\sigma_{e0}^2}\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_2'\boldsymbol{Z}),\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2\partial\rho}) &= -\tfrac{1}{\sigma_{e0}^2}\mathbb{E}[\boldsymbol{X}_2'\boldsymbol{G}((\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}+\boldsymbol{v}_0\boldsymbol{\lambda}_0+\boldsymbol{e}_0)]\\
&= -\tfrac{1}{\sigma_{e0}^2}\boldsymbol{X}_2'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}).
\end{aligned}$$

For parameter $\boldsymbol{\lambda}$,

$$\begin{aligned}
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}\partial\boldsymbol{\lambda}'}) &= -\tfrac{N}{\sigma_{e0}^2}\boldsymbol{\Sigma}_0,\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}\partial\mathrm{vec}(\boldsymbol{\delta})'}) &= \boldsymbol{0},\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}\partial\rho}) &= -\tfrac{1}{\sigma_{e0}^2}\mathbb{E}[\boldsymbol{v}_0'\boldsymbol{G}((\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}+\boldsymbol{v}_0\boldsymbol{\lambda}_0+\boldsymbol{e}_0)]\\
&= -\tfrac{1}{\sigma_{e0}^2}\mathbb{E}(\boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{v}_0)(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0) = -\tfrac{1}{\sigma_{e0}^2}tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0).
\end{aligned}$$

For parameter $\boldsymbol{\delta}$,

$$\begin{aligned}
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\mathrm{vec}(\boldsymbol{\delta})\partial\mathrm{vec}(\boldsymbol{\delta})'}) &= -\tfrac{1'}{\sigma_{e0}^2}(\boldsymbol{\lambda}_0\boldsymbol{\lambda}_0')\otimes(\boldsymbol{Z}'\boldsymbol{Z}) - \boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{Z}'\boldsymbol{Z}),\\
\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\mathrm{vec}(\boldsymbol{\delta})\partial\rho}) &= \tfrac{1}{\sigma_{e0}^2}\boldsymbol{\lambda}_0\otimes\mathbb{E}[\boldsymbol{Z}'\boldsymbol{G}((\boldsymbol{Z}\boldsymbol{\delta}_0+\boldsymbol{v}_0)\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}+\boldsymbol{v}_0\boldsymbol{\lambda}_0+\boldsymbol{e}_0)]\\
&= \tfrac{1}{\sigma_{e0}^2}\boldsymbol{\lambda}_0\otimes(\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})).
\end{aligned}$$

For parameter ρ, we have

$\mathbb{E}(\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{y})$

$= \mathbb{E}[((\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0\boldsymbol{\lambda}_0 + \boldsymbol{e}_0)'\boldsymbol{G}'\boldsymbol{G}((\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0\boldsymbol{\lambda}_0 + \boldsymbol{e}_0)]$

$= \mathbb{E}[(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20}) + (\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0)'\boldsymbol{v}_0'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{e}_0]$

$= (\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20}) + (\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0)'tr(\boldsymbol{G}'\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \sigma_{e0}^2 tr(\boldsymbol{G}'\boldsymbol{G}),$

and hence,

$$\begin{aligned}\mathbb{E}(\tfrac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\rho\partial\rho}) &= -\tfrac{1}{\sigma_{e0}^2}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20}) \\ &\quad -\tfrac{1}{\sigma_{e0}^2}(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0)'tr(\boldsymbol{G}'\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) - tr(\boldsymbol{G}\boldsymbol{G} + \boldsymbol{G}'\boldsymbol{G}).\end{aligned}$$

Now, we show that the convergence of the difference between $\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}$ and $\frac{1}{N}\mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})$ at $\boldsymbol{\theta}_0$ to zero in probability. By Lemma A.10 in Lee (2004), we have $\frac{1}{N}\boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{e}_0 = o_p(1)$, $\frac{1}{N}\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{e}_0 = o_p(1)$, $\frac{1}{N}\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{v}_0 = o_p(1)$, $\frac{1}{N}\boldsymbol{Z}'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{v}_0 = o_p(1)$, $\frac{1}{N}\boldsymbol{Z}'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{e}_0 = o_p(1)$, and

$\frac{1}{N}\boldsymbol{v}_0'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{e}_0 = o_p(1)$. It follows that

$$
\begin{aligned}
\frac{\boldsymbol{e}_0'\boldsymbol{W}\boldsymbol{y}}{N} &= \frac{1}{N}[\boldsymbol{e}_0'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0)] = \frac{1}{N}\boldsymbol{e}_0'\boldsymbol{G}\boldsymbol{e}_0 + o_p(1), \\
\frac{\boldsymbol{X}_1'\boldsymbol{W}\boldsymbol{y}}{N} &= \frac{1}{N}[(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0)] \\
&= \frac{1}{N}[(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20}) + \boldsymbol{\delta}_0'(\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{v}_0)(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{\delta}_0'(\boldsymbol{Z}'\boldsymbol{G}\boldsymbol{e}_0) \\
&\quad + (\boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{Z})\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + (\boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{X}_2)\boldsymbol{\beta}_{20} + \boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{e}_0] \\
&= \frac{1}{N}[(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20}) + \boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0)] + o_p(1), \\
\frac{\boldsymbol{X}_2'\boldsymbol{W}\boldsymbol{y}}{N} &= \frac{1}{N}[\boldsymbol{X}_2'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0)] \\
&= \frac{1}{N}\boldsymbol{X}_2'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20}) + o_p(1), \\
\frac{\boldsymbol{Z}'\boldsymbol{W}\boldsymbol{y}}{N} &= \frac{1}{N}[\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0)] \\
&= \frac{1}{N}\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20}) + o_p(1), \\
\frac{\boldsymbol{v}_0'\boldsymbol{W}\boldsymbol{y}}{N} &= \frac{1}{N}[\boldsymbol{v}_0'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0)] \\
&= \frac{1}{N}\boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + o_p(1), \\
\frac{\boldsymbol{y}'\boldsymbol{W}'\boldsymbol{W}\boldsymbol{y}}{N} &= \frac{1}{N}[(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0)'\boldsymbol{G}' \\
&\quad \times\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20} + \boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0)] \\
&= \frac{1}{N}[(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{X}_2\boldsymbol{\beta}_{20}) \\
&\quad + (\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0)'\boldsymbol{v}_0'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{v}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{e}_0'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{e}_0] + o_p(1),
\end{aligned}
$$

where by Lemmas A.8 and A.11 in Lee (2004), $\mathbb{E}(\boldsymbol{e}_0'\boldsymbol{G}\boldsymbol{e}_0) = \sigma_{e0}^2 tr(\boldsymbol{G})$, $\mathbb{E}(\boldsymbol{e}_0'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{e}_0) = \sigma_{e0}^2 tr(\boldsymbol{G}'\boldsymbol{G})$, and

$$
\begin{aligned}
var(\tfrac{1}{N}\boldsymbol{e}_0'\boldsymbol{G}\boldsymbol{e}_0) &= (\tfrac{\mu_{e0}^4 - 3\sigma_{e0}^4}{N^2})\sum_{i=1}^{N}\boldsymbol{G}_{ii}^2 + \tfrac{\sigma_{e0}^4}{N^2}[tr(\boldsymbol{G}\boldsymbol{G}') + tr(\boldsymbol{G}\boldsymbol{G})] = O(\tfrac{1}{Nh_N}), \\
var(\tfrac{1}{N}\boldsymbol{e}_0'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{e}_0) &= (\tfrac{\mu_{e0}^4 - 3\sigma_{e0}^4}{N^2})\sum_{i=1}^{N}(\boldsymbol{G}'\boldsymbol{G})_{ii}^2 + \tfrac{2\sigma_{e0}^4}{N^2} tr(\boldsymbol{G}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{G}') = O(\tfrac{1}{Nh_N}).
\end{aligned}
$$

Given that

$$
\mathbb{E}(\boldsymbol{v}_0'\boldsymbol{G}\boldsymbol{v}_0) = \mathbb{E}\begin{pmatrix} \boldsymbol{v}_0^{1\prime}\boldsymbol{G}\boldsymbol{v}_0^1 & \cdots & \boldsymbol{v}_0^{1\prime}\boldsymbol{G}\boldsymbol{v}_0^{K_1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{v}_0^{K_1\prime}\boldsymbol{G}\boldsymbol{v}_0^1 & \cdots & \boldsymbol{v}_0^{K_1\prime}\boldsymbol{G}\boldsymbol{v}_0^{K_1} \end{pmatrix} = tr(\boldsymbol{G})\boldsymbol{\Sigma}_0,
$$

where $\mathbb{E}(\boldsymbol{v}_0^{j\prime}\boldsymbol{G}\boldsymbol{v}_0^j) = \sigma_{vj0}^2 tr(\boldsymbol{G})$, $\mathbb{E}(\boldsymbol{v}_0^{i\prime}\boldsymbol{G}\boldsymbol{v}_0^j) = \sigma_{vij0} tr(\boldsymbol{G})$, and for the second moment

$$\mathbb{E}[(\boldsymbol{v}_0^{i\prime}\boldsymbol{G}\boldsymbol{v}_0^j)^2] = \mathbb{E}[(\sum_{l=1}^N\sum_{k=1}^N \boldsymbol{G}_{kl} v_{k0}^i v_{l0}^j)^2] = \mathbb{E}(\sum_{n=1}^N\sum_{m=1}^N\sum_{l=1}^N\sum_{k=1}^N \boldsymbol{G}_{nm}\boldsymbol{G}_{kl} v_{n0}^i v_{m0}^j v_{k0}^i v_{l0}^j).$$

If $i = j$

$$\begin{aligned}
&\mathbb{E}[(\boldsymbol{v}_0^{i\prime}\boldsymbol{G}\boldsymbol{v}_0^i)^2] \\
&= \mu_{vn0}^4 \sum_{n=1}^N \boldsymbol{G}_{nn}^2 + \mathbb{E}(v_{n0}^i v_{n0}^i v_{k0}^i v_{k0}^i)(\sum_{n=1}^N\sum_{k\neq n}^N \boldsymbol{G}_{nn}\boldsymbol{G}_{kk} + \sum_{n=1}^N\sum_{m\neq n}^N \boldsymbol{G}_{nm}\boldsymbol{G}_{mn} + \sum_{n=1}^N\sum_{m\neq n}^N \boldsymbol{G}_{nm}^2) \\
&= (\mu_{vi0}^4 - 3\sigma_{vi0}^4) \sum_{n=1}^N \boldsymbol{G}_{nn}^2 + \sigma_{vi0}^4[tr(\boldsymbol{G})tr(\boldsymbol{G}) + tr(\boldsymbol{G}\boldsymbol{G}) + tr(\boldsymbol{G}\boldsymbol{G}')],
\end{aligned}$$

if $i \neq j$

$$\begin{aligned}
&\mathbb{E}[(\boldsymbol{v}_0^{i\prime}\boldsymbol{G}\boldsymbol{v}_0^j)^2] \\
&= \sum_{n=1}^N \boldsymbol{G}_{nn}^2 \mathbb{E}(v_{n0}^i v_{n0}^j v_{n0}^i v_{n0}^j) + \mathbb{E}(v_{n0}^i v_{n0}^j v_{k0}^i v_{k0}^j)(\sum_{n=1}^N\sum_{k\neq n}^N \boldsymbol{G}_{nn}\boldsymbol{G}_{kk} + \boldsymbol{G}_{nm}\boldsymbol{G}_{mn} \sum_{n=1}^N\sum_{m\neq n}^N \boldsymbol{G}_{nm}^2) \\
&= (\frac{\sigma_{vij0}^2}{\sigma_{vi0}^4}\mu_{vi0}^4 + \sigma_{vi0}^2\sigma_{vj0}^2 - 4\sigma_{vij0}^2) \sum_{n=1}^N \boldsymbol{G}_{nn}^2 + \sigma_{vij0}^2[tr(\boldsymbol{G})tr(\boldsymbol{G}) + tr(\boldsymbol{G}\boldsymbol{G}) + tr(\boldsymbol{G}\boldsymbol{G}')].
\end{aligned}$$

Therefore, $var(\frac{1}{N}\boldsymbol{v}_0^{i\prime}\boldsymbol{G}\boldsymbol{v}_0^j) = O(\frac{1}{Nh_N})$ for all i and j. Similarly, $\mathbb{E}(\boldsymbol{v}_0'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{v}_0) = tr(\boldsymbol{G}'\boldsymbol{G})\boldsymbol{\Sigma}_0$ and $var(\frac{1}{N}\boldsymbol{v}_0^{i\prime}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{v}_0^j) = O(\frac{1}{Nh_N})$ for all i and j.

Besides, we have

$$\begin{aligned}
\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{N} &= \frac{1}{N}[(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)'(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)] = \frac{1}{N}(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0'\boldsymbol{v}_0) + o_p(1), \\
\frac{\boldsymbol{X}_1'\boldsymbol{X}_2}{N} &= \frac{1}{N}(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)'\boldsymbol{X}_2 = \frac{1}{N}\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{X}_2 + o_p(1), \\
\frac{\boldsymbol{X}_1'\boldsymbol{v}_0}{N} &= \frac{1}{N}(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)'\boldsymbol{v}_0 = \frac{1}{N}\boldsymbol{v}_0'\boldsymbol{v}_0 + o_p(1), \\
\frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_1'\boldsymbol{Z})}{N} &= \frac{1}{N}\boldsymbol{\lambda}_0' \otimes ((\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)'\boldsymbol{Z}) = \frac{1}{N}\boldsymbol{\lambda}_0' \otimes (\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}) + o_p(1).
\end{aligned}$$

By the law of large numbers, we have $\frac{\boldsymbol{e}_0'\boldsymbol{e}_0}{N} \xrightarrow{p} \sigma_{e0}^2$, $\frac{\boldsymbol{e}_0'\boldsymbol{X}}{N} \xrightarrow{p} \boldsymbol{0}$, $\frac{\boldsymbol{e}_0'\boldsymbol{v}_0}{N} \xrightarrow{p} \boldsymbol{0}$, $\frac{\boldsymbol{Z}'\boldsymbol{e}_0}{N} \xrightarrow{p} \boldsymbol{0}$, $\frac{\boldsymbol{v}_0'\boldsymbol{v}_0}{N} \xrightarrow{p} \boldsymbol{\Sigma}_0$, $\frac{\boldsymbol{v}_0'\boldsymbol{Z}}{N} \xrightarrow{p} \boldsymbol{0}$. Therefore, with the above results,

$$\frac{1}{N}[\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} - \mathbb{E}(\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})] \xrightarrow{p} 0,$$

and at $\boldsymbol{\theta}_0 = (\sigma_{e0}^2, \mathrm{vech}(\boldsymbol{\Sigma}_0)', \boldsymbol{\beta}_{10}', \boldsymbol{\beta}_{20}', \boldsymbol{\lambda}_0', \mathrm{vec}(\boldsymbol{\delta}_0)', \rho_0)'$

$$\mathbb{E}\Big(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big) =$$

$$\begin{bmatrix}
-\frac{1}{2\sigma_{e0}^4} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -\frac{1}{N\sigma_{e0}^2}tr(\boldsymbol{G}) \\
* & -\frac{\boldsymbol{D}'(\boldsymbol{\Sigma}_0^{-1}\otimes\boldsymbol{\Sigma}_0^{-1})\boldsymbol{D}}{2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
* & * & -\frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta}_0+N\boldsymbol{\Sigma}_0}{N\sigma_{e0}^2} & -\frac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{X}_2}{N\sigma_{e0}^2} & -\frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & \frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2} & -\frac{(\boldsymbol{Z}\boldsymbol{\delta}_0)'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})+tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} \\
* & * & * & -\frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{N\sigma_{e0}^2} & \mathbf{0} & \frac{\boldsymbol{\lambda}_0'\otimes(\boldsymbol{X}_2'\boldsymbol{Z})}{N\sigma_{e0}^2} & -\frac{\boldsymbol{X}_2'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})}{N\sigma_{e0}^2} \\
* & * & * & * & -\frac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2} & \mathbf{0} & -\frac{tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} \\
* & * & * & * & * & -\frac{(\boldsymbol{\lambda}_0\boldsymbol{\lambda}_0')\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N\sigma_{e0}^2} - \frac{\boldsymbol{\Sigma}_0^{-1}\otimes(\boldsymbol{Z}'\boldsymbol{Z})}{N} & \frac{\boldsymbol{\lambda}_0\otimes(\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20}))}{N\sigma_{e0}^2} \\
* & * & * & * & * & * & -\frac{(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})'\boldsymbol{G}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})+(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)'tr(\boldsymbol{G}'\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)}{N\sigma_{e0}^2} - \frac{1}{N}tr(\boldsymbol{G}\boldsymbol{G}+\boldsymbol{G}'\boldsymbol{G})
\end{bmatrix}.$$

B.4 Proof of Theorem 3.3

The Central Limit Theorem for linear-quadratic forms in Kelejian and Prucha (2001) gives

$$\frac{1}{\sqrt{N}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}} + \boldsymbol{\Omega}_{\boldsymbol{\theta}}),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = -\lim_{N\to\infty}\mathbb{E}(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'})$, $\boldsymbol{\Omega}_{\boldsymbol{\theta}} = \lim_{N\to\infty}\boldsymbol{\Omega}_{\boldsymbol{\theta},N}$. Given that

$$\boldsymbol{\Omega}_{\boldsymbol{\theta},N} = \mathbb{E}\Big(\frac{1}{N}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\Big) + \mathbb{E}\Big(\frac{1}{N}\frac{\partial^2 lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\Big).$$

If the errors $\boldsymbol{v}_0^1, \boldsymbol{v}_0^2, \ldots \boldsymbol{v}_0^{K_1}$ are pairwise uncorrelated, the log-likelihood function of (3.2.8) becomes

$$lnL(\boldsymbol{\theta}) = -\frac{N(K_1+1)}{2} ln(2\pi) - \frac{N}{2} ln\sigma_e^2 - \sum_{j=1}^{K_1} \frac{N}{2} ln\sigma_{vj}^2 - \frac{\boldsymbol{e}(\boldsymbol{\phi})'\boldsymbol{e}(\boldsymbol{\phi})}{2\sigma_e^2} - \sum_{j=1}^{K_1} \frac{\boldsymbol{v}^j(\boldsymbol{\delta}^j)'\boldsymbol{v}^j(\boldsymbol{\delta}^j)}{2\sigma_{vj}^2} + ln\left|\boldsymbol{A}(\rho)\right|,$$

where $\boldsymbol{v}^j(\boldsymbol{\delta}^j) = \boldsymbol{X}_1^j - \boldsymbol{Z}\boldsymbol{\delta}^j$. The first order conditions of the log-likelihood function with respect to $\boldsymbol{\theta}$ become

$$\begin{cases}
\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2} &=& \frac{1}{2\sigma_{e0}^4}(\boldsymbol{e}_0'\boldsymbol{e}_0 - N\sigma_{e0}^2) \\
\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vj}^2} &=& \frac{1}{2\sigma_{vj0}^4}(\boldsymbol{v}_0^{j\prime}\boldsymbol{v}_0^j - N\sigma_{vj0}^2),\ j = 1, \ldots, K_1 \\
\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1} &=& \frac{1}{\sigma_{e0}^2}\boldsymbol{X'}_1\boldsymbol{e}_0 \\
\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_2} &=& \frac{1}{\sigma_{e0}^2}\boldsymbol{X'}_2\boldsymbol{e}_0 \\
\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}} &=& \frac{1}{\sigma_{e0}^2}\boldsymbol{v'}_0\boldsymbol{e}_0 \\
\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\delta}^j} &=& -\frac{\lambda_{j0}\boldsymbol{Z}'\boldsymbol{e}_0}{\sigma_{e0}^2} + \frac{\boldsymbol{Z}'\boldsymbol{v}_0^j}{\sigma_{vj0}^2},\ j = 1, \ldots, K_1 \\
\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \rho} &=& \frac{1}{\sigma_{e0}^2}(\boldsymbol{G}\boldsymbol{M}(\boldsymbol{\delta}_0)\boldsymbol{\alpha}_0)'\boldsymbol{e}_0 + \frac{1}{\sigma_{e0}^2}(\boldsymbol{e}_0'\boldsymbol{G}\boldsymbol{e}_0 - \sigma_{e0}^2 tr(\boldsymbol{G}))
\end{cases}$$

The expectation of $\frac{1}{N}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}$ are as follows. For parameter σ_e^2,

$$\begin{aligned}
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2}) &= \tfrac{1}{4N\sigma_{e0}^8}[\mathbb{E}(\boldsymbol{e}_0'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{e}_0) - 2N\sigma_{e0}^2\mathbb{E}(\boldsymbol{e}_0'\boldsymbol{e}_0) + N^2\sigma_{e0}^4] \\
&= \tfrac{1}{4N\sigma_{e0}^8}[Var(\boldsymbol{e}_0'\boldsymbol{e}_0) + \mathbb{E}(\boldsymbol{e}_0'\boldsymbol{e}_0)^2 - N^2\sigma_{e0}^4] = \tfrac{\mu_{e0}^4 - 3\sigma_{e0}^4}{4\sigma_{e0}^8} + \tfrac{1}{2\sigma_{e0}^4}, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vj}^2}) &= 0, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1'}) &= \tfrac{1}{2N\sigma_{e0}^6}\mathbb{E}[(\boldsymbol{e}_0'\boldsymbol{e}_0 - N\sigma_{e0}^2)\boldsymbol{e}_0'(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)] \\
&= \tfrac{1}{2N\sigma_{e0}^6}\mathbb{E}(\boldsymbol{e}_0'\boldsymbol{e}_0\boldsymbol{e}_0')\boldsymbol{Z}\boldsymbol{\delta}_0 = \tfrac{\mu_{e0}^3}{2N\sigma_{e0}^6}\boldsymbol{l}_N'\boldsymbol{Z}\boldsymbol{\delta}_0, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_2'}) &= \tfrac{1}{2N\sigma_{e0}^6}\mathbb{E}[(\boldsymbol{e}_0'\boldsymbol{e}_0 - N\sigma_{e0}^2)\boldsymbol{e}_0'\boldsymbol{X}_2] = \tfrac{\mu_{e0}^3}{2N\sigma_{e0}^6}\boldsymbol{l}_N'\boldsymbol{X}_2, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}'}) &= \tfrac{1}{2N\sigma_{e0}^6}\mathbb{E}[(\boldsymbol{e}_0'\boldsymbol{e}_0 - N\sigma_{e0}^2)\boldsymbol{e}_0'\boldsymbol{v}_0] = \boldsymbol{0}, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\delta}^{j\prime}}) &= -\tfrac{\lambda_{j0}}{2N\sigma_{e0}^6}\mathbb{E}[(\boldsymbol{e}_0'\boldsymbol{e}_0 - N\sigma_{e0}^2)\boldsymbol{e}_0'\boldsymbol{Z}] = -\tfrac{\lambda_{j0}\mu_{e0}^3}{2N\sigma_{e0}^6}\boldsymbol{l}_N'\boldsymbol{Z},
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_e^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \rho}) \\
&= \tfrac{1}{2N\sigma_{e0}^6}\mathbb{E}[(\boldsymbol{e}_0'\boldsymbol{e}_0 - N\sigma_{e0}^2)\boldsymbol{e}_0'(\boldsymbol{GZ}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{GX}_2\boldsymbol{\beta}_{20} + \boldsymbol{Gv}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{Ge}_0)] \\
&= \tfrac{1}{2N\sigma_{e0}^6}\mathbb{E}[\boldsymbol{e}_0'\boldsymbol{e}_0\boldsymbol{e}_0'(\boldsymbol{GZ}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{GX}_2\boldsymbol{\beta}_{20}) + \boldsymbol{e}_0'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{Ge}_0 - N\sigma_{e0}^2\boldsymbol{e}_0'\boldsymbol{Ge}_0] \\
&= \tfrac{1}{2N\sigma_{e0}^6}[\mu_{e0}^3\boldsymbol{l}_N'(\boldsymbol{GZ}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{GX}_2\boldsymbol{\beta}_{20}) + (\mu_{e0}^4 + (N-1)\sigma_{e0}^4)tr(\boldsymbol{G}) - N\sigma_{e0}^4 tr(\boldsymbol{G})].
\end{aligned}$$

For parameter σ_{vj}^2,

$$\begin{aligned}
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vj}^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vi}^2}) &= \tfrac{\mu_{vj0}^4 - 3\sigma_{vj0}^4}{4\sigma_{vj0}^8} + \tfrac{1}{2\sigma_{vj0}^4}, \text{ if } j = i; \text{ 0 otherwise,} \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vj}^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1'}) &= \mathbf{0}, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vj}^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_2'}) &= \mathbf{0}, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vj}^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}'}) &= \mathbf{0}, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vj}^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\delta}^{i\prime}}) &= \tfrac{1}{2N\sigma_{vj0}^6}\mathbb{E}(\boldsymbol{v}_0^{j\prime}\boldsymbol{v}_0^j\boldsymbol{v}_0^{j\prime}\boldsymbol{Z}) = \tfrac{\mu_{vj0}^3}{2N\sigma_{vj0}^6}\boldsymbol{l}_N'\boldsymbol{Z}, \text{ if } j = i; \ \mathbf{0} \text{ otherwise,} \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \sigma_{vj}^2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \rho}) &= 0.
\end{aligned}$$

For parameter $\boldsymbol{\beta}_1$,

$$\begin{aligned}
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1'}) &= \tfrac{1}{N\sigma_{e0}^4}\mathbb{E}[(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)'\boldsymbol{e}_0\boldsymbol{e}_0'(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)] = \tfrac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z}\boldsymbol{\delta}_0 + N\boldsymbol{\Sigma}_0}{N\sigma_{e0}^2}, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_2'}) &= \tfrac{1}{N\sigma_{e0}^4}\mathbb{E}(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{X}_2) = \tfrac{\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{X}_2}{N\sigma_{e0}^2}, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\lambda}'}) &= \tfrac{1}{N\sigma_{e0}^4}\mathbb{E}(\boldsymbol{v}_0'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{v}_0) = \tfrac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2}, \\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\delta}^{j\prime}}) &= -\tfrac{\lambda_{j0}}{N\sigma_{e0}^4}\mathbb{E}(\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{Z}) = -\tfrac{\lambda_{j0}}{N\sigma_{e0}^2}\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{Z},
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_1}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial \rho}) \\
&= \tfrac{1}{N\sigma_{e0}^4}\mathbb{E}[(\boldsymbol{Z}\boldsymbol{\delta}_0 + \boldsymbol{v}_0)'\boldsymbol{e}_0\boldsymbol{e}_0'(\boldsymbol{GZ}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{GX}_2\boldsymbol{\beta}_{20} + \boldsymbol{Gv}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0) + \boldsymbol{Ge}_0)] \\
&= \tfrac{1}{N\sigma_{e0}^4}\mathbb{E}[\boldsymbol{\delta}_0'\boldsymbol{Z}'\boldsymbol{e}_0\boldsymbol{e}_0'(\boldsymbol{GZ}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{GX}_2\boldsymbol{\beta}_{20} + \boldsymbol{Ge}_0) + \boldsymbol{v}_0'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{Gv}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0)] \\
&= \tfrac{1}{N\sigma_{e0}^4}[\sigma_{e0}^2\boldsymbol{\delta}_0'\boldsymbol{Z}'(\boldsymbol{GZ}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10} + \boldsymbol{GX}_2\boldsymbol{\beta}_{20}) + \mu_{e0}^3\sum_{i=1}^{N}\boldsymbol{G}_{ii}(\boldsymbol{\delta}_0'\boldsymbol{Z}')_i + tr(\boldsymbol{G})\sigma_{e0}^2\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10} + \boldsymbol{\lambda}_0)],
\end{aligned}$$

where $(\boldsymbol{\delta}_0'\boldsymbol{Z}')_i$ is the ith column of $\boldsymbol{\delta}_0'\boldsymbol{Z}'$ and $\boldsymbol{G}_{ij}$ is the (i,j)th element of $\boldsymbol{G}$. For parameter $\boldsymbol{\beta}_2$,

$$\begin{aligned}
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2'}) &= \tfrac{1}{N\sigma_{e0}^2}\boldsymbol{X}_2'\boldsymbol{X}_2,\\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}'}) &= \boldsymbol{0},\\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\delta}^{j\prime}}) &= -\tfrac{\lambda_{j0}}{N\sigma_{e0}^2}\boldsymbol{X}_2'\boldsymbol{Z},\\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}_2}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\rho}) &= \tfrac{1}{N\sigma_{e0}^4}\mathbb{E}[\boldsymbol{X}_2'\boldsymbol{e}_0\boldsymbol{e}_0'(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{G}\boldsymbol{X}_2\boldsymbol{\beta}_{20}+\boldsymbol{G}\boldsymbol{v}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)+\boldsymbol{G}\boldsymbol{e}_0)]\\
&= \tfrac{1}{N\sigma_{e0}^4}\mathbb{E}[\boldsymbol{X}_2'\boldsymbol{e}_0\boldsymbol{e}_0'(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{G}\boldsymbol{X}_2\boldsymbol{\beta}_{20})+\boldsymbol{X}_2'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{G}\boldsymbol{e}_0)\\
&= \tfrac{1}{N\sigma_{e0}^4}[\sigma_{e0}^2\boldsymbol{X}_2'(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{G}\boldsymbol{X}_2\boldsymbol{\beta}_{20})+\mu_{e0}^3\sum_{i=1}^{N}\boldsymbol{G}_{ii}(\boldsymbol{X}_2')_i].
\end{aligned}$$

For parameter $\boldsymbol{\lambda}$,

$$\begin{aligned}
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}'}) &= \tfrac{\boldsymbol{\Sigma}_0}{\sigma_{e0}^2},\\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\delta}^{j\prime}}) &= \boldsymbol{0},\\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\lambda}}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\rho}) &= \tfrac{1}{N\sigma_{e0}^2}tr(\boldsymbol{G})\boldsymbol{\Sigma}_0(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0).
\end{aligned}$$

For parameter $\boldsymbol{\delta}$,

$$\begin{aligned}
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\delta}^j}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\delta}^{i\prime}}) &= \tfrac{\lambda_{j0}^2\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e0}^2}+\tfrac{\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{vj0}^2},\text{ if } j=i;\ \tfrac{\lambda_{j0}\lambda_{i0}\boldsymbol{Z}'\boldsymbol{Z}}{N\sigma_{e0}^2}\text{ otherwise},\\
\mathbb{E}(\tfrac{1}{N}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\delta}^j}\tfrac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\rho}) &= -\tfrac{\lambda_{j0}}{N\sigma_{e0}^4}\mathbb{E}(\boldsymbol{Z}'\boldsymbol{e}_0\boldsymbol{e}_0'(\boldsymbol{G}\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{G}\boldsymbol{X}_2\boldsymbol{\beta}_{20})+\boldsymbol{Z}'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{G}\boldsymbol{e}_0)\\
&= -\tfrac{\lambda_{j0}}{N\sigma_{e0}^4}(\sigma_{e0}^2\boldsymbol{Z}'\boldsymbol{G}(\boldsymbol{Z}\boldsymbol{\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{X}_2\boldsymbol{\beta}_{20})+\mu_{e0}^3\sum_{i=1}^{N}\boldsymbol{G}_{ii}(\boldsymbol{Z}')_i).
\end{aligned}$$

For parameter ρ,

$$
\begin{aligned}
&\mathbb{E}\left(\frac{1}{N}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\rho}\frac{\partial lnL(\boldsymbol{\theta}_0)}{\partial\rho}\right)\\
&=\frac{1}{N\sigma_{e0}^4}\mathbb{E}[(\boldsymbol{GZ\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{GX}_2\boldsymbol{\beta}_{20})'\boldsymbol{e}_0\boldsymbol{e}_0'(\boldsymbol{GZ\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{GX}_2\boldsymbol{\beta}_{20})\\
&+2(\boldsymbol{GZ\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{GX}_2\boldsymbol{\beta}_{20})'(\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{Ge}_0)+(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)'(\boldsymbol{v}_0'\boldsymbol{G}'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{Gv}_0)(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)\\
&+\boldsymbol{e}_0'\boldsymbol{G}'\boldsymbol{e}_0\boldsymbol{e}_0'\boldsymbol{Ge}_0-2\sigma_{e0}^2tr(\boldsymbol{G})\boldsymbol{e}_0'\boldsymbol{G}'\boldsymbol{e}_0+\sigma_{e0}^4tr(\boldsymbol{G})tr(\boldsymbol{G})]\\
&=\frac{1}{N\sigma_{e0}^4}[\sigma_{e0}^2(\boldsymbol{GZ\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{GX}_2\boldsymbol{\beta}_{20})'(\boldsymbol{GZ\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{GX}_2\boldsymbol{\beta}_{20})\\
&+2\mu_{e0}^3\sum_{i=1}^N\boldsymbol{G}_{ii}((\boldsymbol{GZ\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{GX}_2\boldsymbol{\beta}_{20})')_i+(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)'\sigma_{e0}^2\boldsymbol{\Sigma}_0tr(\boldsymbol{G}'\boldsymbol{G})(\boldsymbol{\beta}_{10}+\boldsymbol{\lambda}_0)\\
&+(\mu_{e0}^4-3\sigma_{e0}^4)\sum_{i=1}^N\boldsymbol{G}_{ii}^2+\sigma_{e0}^4tr(\boldsymbol{GG}+\boldsymbol{G}'\boldsymbol{G})].
\end{aligned}
$$

Consequently, at $\boldsymbol{\theta}_0=(\sigma_{e0}^2,\boldsymbol{\sigma}_{v0}^2{}',\boldsymbol{\beta}_{10}',\boldsymbol{\beta}_{20}',\boldsymbol{\lambda}_0',\text{vec}(\boldsymbol{\delta}_0)',\rho_0)'$, we have

$$
\boldsymbol{\Omega}_{\boldsymbol{\theta},N}=
$$

$$
\left[\begin{array}{ccccc}
\frac{\mu_{e0}^4-3\sigma_{e0}^4}{4\sigma_{e0}^8} & \mathbf{0} & \frac{\mu_{e0}^3}{2N\sigma_{e0}^6}\boldsymbol{l}_N'\boldsymbol{Z\delta}_0 & \frac{\mu_{e0}^3}{2N\sigma_{e0}^6}\boldsymbol{l}_N'\boldsymbol{X}_2 & \mathbf{0}\\
* & \text{diag}\left(\frac{\mu_{v10}^4-3\sigma_{v10}^4}{4\sigma_{v10}^8},\dots,\frac{\mu_{vK_10}^4-3\sigma_{vK_10}^4}{4\sigma_{vK_10}^8}\right) & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
* & * & \mathbf{0} & \mathbf{0} & \mathbf{0}\\
* & * & * & \mathbf{0} & \mathbf{0}\\
* & * & * & * & \mathbf{0}\\
* & * & * & * & *\\
* & * & * & * & *
\end{array}\right.
$$

$$
\left.\begin{array}{cc}
-\frac{\mu_{e0}^3\boldsymbol{\lambda}_0'\otimes(\boldsymbol{l}_N'\boldsymbol{Z})}{2N\sigma_{e0}^6} & \frac{1}{2N\sigma_{e0}^6}[\mu_{e0}^3\boldsymbol{l}_N'(\boldsymbol{GZ\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{GX}_2\boldsymbol{\beta}_{20})+(\mu_{e0}^4-3\sigma_{e0}^4)tr(\boldsymbol{G})]\\
\frac{\text{diag}(\boldsymbol{\zeta}_{v0})\otimes\boldsymbol{l}_N'\boldsymbol{Z}}{2N} & \mathbf{0}\\
\mathbf{0} & \frac{\mu_{e0}^3}{N\sigma_{e0}^4}\sum_{i=1}^N\boldsymbol{G}_{ii}(\boldsymbol{\delta}_0'\boldsymbol{Z}')_i\\
\mathbf{0} & \frac{\mu_{e0}^3}{N\sigma_{e0}^4}\sum_{i=1}^N\boldsymbol{G}_{ii}(\boldsymbol{X}_2')_i\\
\mathbf{0} & \mathbf{0}\\
\mathbf{0} & -\frac{\mu_{e0}^3}{N\sigma_{e0}^4}\boldsymbol{\lambda}_0\otimes\sum_{i=1}^N\boldsymbol{G}_{ii}(\boldsymbol{Z}')_i\\
* & \frac{2\mu_{e0}^3}{N\sigma_{e0}^4}\sum_{i=1}^N\boldsymbol{G}_{ii}((\boldsymbol{GZ\delta}_0\boldsymbol{\beta}_{10}+\boldsymbol{GX}_2\boldsymbol{\beta}_{20})')_i+\frac{(\mu_{e0}^4-3\sigma_{e0}^4)}{N\sigma_{e0}^4}\sum_{i=1}^N\boldsymbol{G}_{ii}^2
\end{array}\right],
$$

where $\boldsymbol{\zeta}_{v0} = (\zeta_{v10}, \ldots, \zeta_{vK_10})$ with $\zeta_{vj0} = \mu^3_{vj0}/\sigma^6_{vj0}$. Importantly, if $(e_i, \boldsymbol{v}_i)$ are normal distributed, $\mu^3_{e0} = \mu^3_{vj0} = 0$, $\mu^4_{e0} = 3\sigma^4_{e0}$, and $\mu^4_{vj0} = 3\sigma^4_{vj0}$, and hence $\boldsymbol{\Omega}_{\boldsymbol{\theta},N} = \mathbf{0}$.

B.5 Supplementary Materials: Tables

In this supplement, we provide Monte Carlo results for the case where u_i is given by

$$u_i = \lambda v_i + 0.5(v_i^2 - 5/3) + e_i \tag{B.5.1}$$

and other DGPs are the same as those in Tables 3.1-3.5. In this setup, Assumption 2, which requires a linear relationship between $\boldsymbol{u}$ and $\boldsymbol{v}$, is not satisfied. To assess the consequences of violating this linearity assumption, we report the following statistics in Tables B.1-B.3:

- Average bias of $\hat{\beta}_{SQLIML\_nonlinear}$ and $\hat{\rho}_{SQLIML\_nonlinear}$, where $\hat{\beta}_{SQLIML\_nonlinear}$ and $\hat{\rho}_{SQLIML\_nonlinear}$ are the SQLIML estimators of β and ρ in models where $\boldsymbol{u}$ is given by (B.5.1);
- Relative efficiency of β: $RE(\hat{\beta}) = RMSE(\hat{\beta}_{SQLIML})/RMSE(\hat{\beta}_{SQLIML\_nonlinear})$, where $\hat{\beta}_{SQLIML}$ is the SQLIML estimator of β in models where $\boldsymbol{u}$ is given by (3.4.2);
- Relative efficiency of ρ: $RE(\hat{\rho}) = RMSE(\hat{\rho}_{SQLIML})/RMSE(\hat{\rho}_{SQLIML\_nonlinear})$, where $\hat{\rho}_{SQLIML}$ is the SQLIML estimator of ρ in models where $\boldsymbol{u}$ is given by (3.4.2);
- Size and Power of exogeneity test, $H_0 : \lambda = 0$, in models where $\boldsymbol{u}$ is given by (B.5.1).

Table B.1: Bias of SQLIML Estimates in Models Violating the Linearity Assumption

	Average Bias of $\hat{\beta}_{SQLIML\_nonlinear}$						Average Bias of $\hat{\rho}_{SQLIML\_nonlinear}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$L = 1$												
$\lambda = 0.1$	0.0001	0.0014	0.0003	-0.0126	-0.0058	0.0031	-0.0084	-0.0083	-0.0050	-0.0022	-0.0033	-0.0027
$\lambda = 0.5$	-0.0201	0.0058	0.0006	-0.0225	-0.0135	-0.0056	-0.0061	-0.0026	-0.0045	-0.0025	-0.0034	-0.0022
$\lambda = 1$	-0.0259	-0.0158	-0.0006	-0.0229	-0.0143	0.0014	-0.0041	-0.0044	-0.0051	-0.0031	-0.0010	-0.0031
$\lambda = 1.5$	-0.0981	-0.0345	-0.0107	-0.0388	-0.0144	0.0009	-0.0060	-0.0046	-0.0044	-0.0032	-0.0034	-0.0012
$L = 4$												
$\lambda = 0.1$	0.0104	0.0054	0.0031	0.0010	-0.0020	0.0003	-0.0048	-0.0047	-0.0044	-0.0017	-0.0023	-0.0031
$\lambda = 0.5$	0.0106	0.0010	0.0050	0.0175	0.0093	0.0054	-0.0069	-0.0047	-0.0059	-0.0022	-0.0027	-0.0021
$\lambda = 1$	0.0197	0.0078	0.0034	0.0228	0.0114	0.0063	-0.0038	-0.0043	-0.0048	-0.0025	-0.0023	-0.0019
$\lambda = 1.5$	0.0296	0.0056	0.0010	0.0280	0.0107	0.0055	-0.0046	-0.0037	-0.0035	-0.0020	2.45E-05	2.89E-05
$L = 8$												
$\lambda = 0.1$	0.0051	-0.0017	-0.0031	0.0040	0.0019	0.0014	-0.0056	-0.0051	-0.0073	-0.0025	-0.0026	-0.0023
$\lambda = 0.5$	0.0061	0.0043	0.0014	0.0016	0.0056	0.0009	-0.0076	-0.0036	-0.0026	-0.0039	-0.0007	-0.0021
$\lambda = 1$	0.0272	0.0100	0.0039	0.0122	0.0042	0.0015	-0.0035	-0.0032	-0.0025	-0.0029	-0.0027	-0.0020
$\lambda = 1.5$	0.0418	0.0157	0.0034	0.0254	0.0052	0.0024	-0.0031	-0.0026	-0.0025	-0.0015	-0.0024	-0.0014
$L = 16$												
$\lambda = 0.1$	0.0019	0.0012	0.0009	0.0008	0.0011	-0.0003	-0.0054	-0.0074	-0.0058	-0.0054	-0.0019	-0.0011
$\lambda = 0.5$	0.0188	0.0069	0.0029	0.0030	0.0047	0.0023	-0.0034	-0.0026	-0.0053	-0.0017	-0.0040	-0.0040
$\lambda = 1$	0.0336	0.0128	0.0038	0.0186	0.0075	0.0041	-0.0057	-0.0057	-0.0036	-0.0031	-0.0033	-0.0014
$\lambda = 1.5$	0.0458	0.0177	0.0069	0.0199	0.0089	0.0041	-0.0046	-0.0022	-0.0020	-0.0027	-0.0020	-0.0014

Table B.2: Relative Efficiency of SQLIML Estimates in Models Meeting and Violating the Linearity Assumption

	Relative Efficiency of $\hat{\beta}_{SQLIML}/\hat{\beta}_{SQLIML\_nonlinear}$						Relative Efficiency of $\hat{\rho}_{SQLIML}/\hat{\rho}_{SQLIML\_nonlinear}$					
	$N = 500$			$N = 1000$			$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$L = 1$												
$\lambda = 0.1$	0.5197	0.5215	0.5093	0.3952	0.3874	0.5073	0.9063	0.9016	0.9263	0.9307	0.9031	0.9210
$\lambda = 0.5$	0.5350	0.5924	0.5273	0.4549	0.4494	0.5333	0.8685	0.8347	0.8554	0.8309	0.8160	0.8213
$\lambda = 1$	0.6793	0.6381	0.6081	0.6568	0.6439	0.6360	0.8169	0.8254	0.7798	0.7865	0.7525	0.7364
$\lambda = 1.5$	0.7344	0.7369	0.7164	0.7405	0.6420	0.7232	0.7288	0.7387	0.8009	0.7412	0.7515	0.7379
$L = 4$												
$\lambda = 0.1$	0.5081	0.5072	0.5071	0.5313	0.5176	0.5296	0.8679	0.8477	0.8150	0.8917	0.8783	0.8099
$\lambda = 0.5$	0.5321	0.5553	0.5682	0.4826	0.4828	0.4834	0.8241	0.8226	0.8156	0.8282	0.7972	0.7609
$\lambda = 1$	0.6277	0.6310	0.6345	0.5772	0.5784	0.5795	0.7766	0.7706	0.7498	0.7567	0.7441	0.7178
$\lambda = 1.5$	0.6516	0.7442	0.7457	0.6741	0.7262	0.7279	0.7291	0.6912	0.6748	0.7137	0.6970	0.6840
$L = 8$												
$\lambda = 0.1$	0.4799	0.5121	0.5007	0.4626	0.4600	0.4838	0.8890	0.8603	0.7535	0.8916	0.8482	0.7608
$\lambda = 0.5$	0.5798	0.5308	0.5358	0.5300	0.5635	0.5635	0.8273	0.8174	0.7712	0.8083	0.8040	0.7365
$\lambda = 1$	0.6199	0.6204	0.6224	0.6495	0.6493	0.6505	0.7902	0.7717	0.7407	0.7777	0.7452	0.7111
$\lambda = 1.5$	0.7125	0.7099	0.6920	0.7424	0.7563	0.7542	0.7569	0.7477	0.6975	0.7337	0.7015	0.6672
$L = 16$												
$\lambda = 0.1$	0.4968	0.4846	0.5244	0.4091	0.4928	0.4844	0.8679	0.7376	0.7257	0.8592	0.7733	0.7282
$\lambda = 0.5$	0.5146	0.5255	0.5695	0.5803	0.5350	0.5368	0.8093	0.7762	0.7147	0.8375	0.7156	0.6664
$\lambda = 1$	0.6891	0.6721	0.6413	0.6275	0.6236	0.6396	0.7494	0.7357	0.6847	0.7123	0.6879	0.7007
$\lambda = 1.5$	0.7584	0.7509	0.7465	0.7804	0.7585	0.7269	0.7310	0.6740	0.6484	0.6928	0.6698	0.6573

Table B.3: Size and Power of Exogeneity Test in Models Violating the Linearity Assumption

	$N = 500$			$N = 1000$		
	$c = 0.3$	$c = 0.5$	$c = 0.8$	$c = 0.3$	$c = 0.5$	$c = 0.8$
$L = 1$						
$\lambda = 0$	0.107	0.248	0.376	0.134	0.277	0.399
$\lambda = 0.1$	0.109	0.255	0.371	0.157	0.296	0.450
$\lambda = 0.5$	0.325	0.605	0.789	0.523	0.789	0.893
$\lambda = 1$	0.728	0.912	0.957	0.894	0.968	0.973
$\lambda = 1.5$	0.937	0.979	0.989	0.976	0.986	0.990
$L = 4$						
$\lambda = 0$	0.274	0.412	0.504	0.314	0.480	0.558
$\lambda = 0.1$	0.302	0.441	0.535	0.339	0.508	0.572
$\lambda = 0.5$	0.681	0.796	0.871	0.793	0.888	0.917
$\lambda = 1$	0.936	0.958	0.971	0.966	0.981	0.985
$\lambda = 1.5$	0.984	0.992	0.993	0.986	0.990	0.994
$L = 8$						
$\lambda = 0$	0.375	0.487	0.523	0.434	0.552	0.562
$\lambda = 0.1$	0.384	0.508	0.566	0.449	0.571	0.614
$\lambda = 0.5$	0.791	0.865	0.897	0.901	0.930	0.945
$\lambda = 1$	0.961	0.971	0.979	0.979	0.981	0.982
$\lambda = 1.5$	0.988	0.995	0.992	0.993	0.999	0.997
$L = 16$						
$\lambda = 0$	0.439	0.536	0.569	0.507	0.591	0.581
$\lambda = 0.1$	0.463	0.526	0.593	0.517	0.597	0.643
$\lambda = 0.5$	0.850	0.894	0.892	0.908	0.947	0.951
$\lambda = 1$	0.970	0.979	0.984	0.989	0.991	0.992
$\lambda = 1.5$	0.995	0.994	0.995	0.997	0.999	0.996

Bibliography

Adie, D. K. (1973). Teen-age unemployment and real federal minimum wages. Journal of Political Economy 81(2, Part 1), 435–441.

Alessi, L., E. Ghysels, L. Onorante, R. Peach, and S. Potter (2014). Central bank macroeconomic forecasting during the global financial crisis: The European Central Bank and Federal Reserve Bank of New York experiences. Journal of Business & Economic Statistics 32(4), 483–500.

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. In Handbook of labor economics, Volume 3, pp. 3143–3259. Elsevier B.V.

Andreou, E., E. Ghysels, and A. Kourtellos (2010). Regression models with mixed sampling frequencies. Journal of Econometrics 158(2), 246–261. Specification Analysis in Honor of Phoebus J. Dhrymes.

Andreou, E., E. Ghysels, and A. Kourtellos (2011). Forecasting with mixed-frequency data. In The Oxford handbook of economic forecasting.

Angel-Urdinola, D. (2008). Can the introduction of a minimum wage in FYR Macedonia decrease the gender wage gap? The World Bank, Policy Research Working Paper Series.

Anselin, L. (1988). Spatial econometrics : Methods and models. Kluwer, Dordrecht.

Anselin, L. (2010). Thirty years of spatial econometrics. Papers in Regional Science 89(1), 3–25.

Anselin, L. and A. K. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. Statistics Textbooks and Monographs 155, 237–290.

Anselin, L. and R. J. Florax (1995). New directions in spatial econometrics: Introduction. In New directions in spatial econometrics, pp. 3–18. Springer.

Anselin, L., J. Le Gallo, and H. Jayet (2006). Spatial panel econometrics. In L. Mátyás and P. e. Sevestre (Eds.), The econometrics of panel data: fundamentals and recent developments in theory and practice (3 ed.)., pp. 901–969. Kluwer, Dordrecht.

Anselin, L. and N. Lozano-Gracia (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. Empirical Economics 34(1), 5–34.

Anselin, L. and S. J. Rey (1997). Introduction to the special issue on spatial econometrics. International Regional Science Review 20(1-2), 1–7.

Antonakakis, N. and R. Gupta (2017). Is economic policy uncertainty related to suicide rates? Evidence from the United States. Social Indicators Research 133(2), 543–560.

Autor, D. H., A. Manning, and C. L. Smith (2016). The contribution of the minimum wage to US wage inequality over three decades: A reassessment. American Economic Journal: Applied Economics 8(1), 58–99.

Babii, A., E. Ghysels, and J. Striaukas (2021). Machine learning time series regressions with an application to nowcasting. Journal of Business & Economic Statistics, 1–23.

Bai, J., E. Ghysels, and J. H. Wright (2013). State space models and MIDAS regressions. Econometric Reviews 32(7), 779–813.

Baker, S. R., N. Bloom, and S. J. Davis (2012). Has economic policy uncertainty hampered the recovery? Government Policies and the Delayed Economic Recovery 70.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. The Quarterly Journal of Economics 131(4), 1593–1636.

Behrman, J. R. and A. B. Deolalikar (1988). Health and nutrition. Handbook of Development Economics 1, 631–711.

Besley, T. and A. Case (1995). Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. The American Economic Review 85(1), 25–45.

Blau, F. D. and L. M. Kahn (1997). Swimming upstream: Trends in the gender wage differential in the 1980s. Journal of Labor Economics 15(1, Part 1), 1–42.

Blau, F. D. and L. M. Kahn (2003). Understanding international differences in the gender pay gap. Journal of Labor Economics 21(1), 106–144.

Blau, F. D. and L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. Journal of Economic Literature 55(3), 789–865.

Bloom, N. (2014). Fluctuations in uncertainty. Journal of Economic Perspectives 28(2), 153–76.

BLS (2020). Characteristics of minimum wage workers, 2019. U.S. Bureau of Labor Statistics.

Brainerd, E. (2001). Economic reform and mortality in the former Soviet Union: A study of the suicide epidemic in the 1990s. European Economic Review 45(4), 1007–1019. 15th Annual Congress of the European Economic Association.

Breitung, J. and N. R. Swanson (2002). Temporal aggregation and spurious instantaneous causality in multiple time series models. Journal of Time Series Analysis 23(6), 651–665.

Brenner, M. H. (1973). Mental illness and the economy. Harvard University Press.

Brenner, M. H. (1975). Trends in alcohol consumption and associated illnesses. Some effects of economic changes. American Journal of Public Health 65(12), 1279–1292.

Brenner, M. H. (1979). Mortality and the national economy: A review, and the experience of England and Wales, 1936-76. The Lancet 314(8142), 568–573.

Brenner, M. H. (1987). Economic change, alcohol consumption and heart disease mortality in nine industrialized countries. Social Science & Medicine 25(2), 119–132.

Brenner, M. H. (2005). Commentary: Economic growth is the basis of mortality rate decline in the 20th century–experience of the United States 1901–2000. International Journal of Epidemiology 34(6), 1214–1221.

Breuer, C. (2015). Unemployment and suicide mortality: Evidence from regional panel data in Europe. Health Economics 24(8), 936–950.

Brogaard, J. and A. Detzel (2015). The asset-pricing implications of government economic policy uncertainty. Management Science 61(1), 3–18.

Brown, C., C. Gilroy, and A. Kohen (1982). The effect of the minimum wage on employment and unemployment: A survey. NBER Working Paper Series, 846.

Burgard, S. A., J. E. Brand, and J. S. House (2009). Perceived job insecurity and worker health in the United States. Social Science & Medicine 69(5), 777–785.

Burgoa, M., E. Regidor, C. Rodriguez, and J. Gutierrez-Fisac (1998). Mortality by cause of death and marital status in Spain. European Journal of Public Health 8(1), 37–42.

Cain, G. G. (1986). The economic analysis of labor market discrimination: A survey. In Handbook of labor economics, Volume 1, pp. 693–785. Elsevier B.V.

Carroll, C. D. (1997). Buffer-stock saving and the life cycle-permanent income hypothesis. The Quarterly Journal of Economics 112(1), 1–55.

Casarin, R., C. Foroni, M. Marcellino, and F. Ravazzolo (2018). Uncertainty through the lenses of a mixed-frequency Bayesian panel Markov-switching model. The Annals of Applied Statistics 12(4), 2559–2586.

Case, A. and A. Deaton (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. Proceedings of the National Academy of Sciences 112(49), 15078.

Cavalcanti, T. and J. Tavares (2016). The output cost of gender discrimination: A model-based macroeconomics estimate. The Economic Journal 126(590), 109–134.

Clements, M. P. and A. B. Galvão (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. Journal of Business & Economic Statistics 26, 546–554.

Cliff, A. D. and J. K. Ord (1973). Spatial autocorrelation. Monographs in spatial and environmental systems analysis, 5. London: Pion.

Cliff, A. D. and J. K. Ord (1981). Spatial processes : Models and applications. Monographs in spatial and environmental systems analysis, 5. London: Pion.

Cooper, M. L., M. Russell, J. B. Skinner, M. R. Frone, and P. Mudar (1992). Stress and alcohol use: Moderating effects of gender, coping, and alcohol expectancies. Journal of Abnormal Psychology 101(1), 139–152.

Da, Z., J. Engelberg, and P. Gao (2015). The sum of all FEARS investor sentiment and asset prices. The Review of Financial Studies 28(1), 1–32.

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

Dehejia, R. and A. Lleras-Muney (2004). Booms, busts, and babies' health. The Quarterly Journal of Economics 119(3), 1091–1130.

Dinardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. Econometrica 64(5), 1001–1044.

Dixit, A. K. (1994). Investment under uncertainty. Princeton, N.J. ; Chichester: Princeton University Press.

Drukker, D. M., P. Egger, and I. R. Prucha (2013). On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. Econometric Reviews 32(5-6), 686–733.

Drukker, D. M., I. R. Prucha, and R. Raciborski (2013). A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. The Stata Journal 13(2), 287–301.

Elhorst, J. P. (2010). Spatial panel data models. In M. M. Fischer and A. Getis (Eds.), Handbook of applied spatial analysis, pp. 377–407. Springer, Berlin, Heidelberg.

Elhorst, J. P. (2014). Spatial econometrics: From cross-sectional data to spatial panels, Volume 479. Springer.

England, P. (1992). Comparable worth: Theories and evidence. Transaction Publishers.

Engle, R. and C. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. Econometrica (1986-1998) 55(2), 251.

Engle, R. F. (1969). Biases from time-aggregation of distributed lag models. Cornell University.

Engle, R. F., E. Ghysels, and B. Sohn (2013). Stock market volatility and macroeconomic fundamentals. Review of Economics and Statistics 95(3), 776–797.

Erlandsson, G., M. Lambe, S. Cnattingius, and A. Ekbom (2002). Hyperemesis gravidarum and subsequent breast cancer risk. British Journal of Cancer 87(9), 974.

Ferrie, J. E., M. J. Shipley, K. Newman, S. A. Stansfeld, and M. Marmot (2005). Self-reported job insecurity and health in the Whitehall II study: Potential explanations of the relationship. Social Science & Medicine 60(7), 1593–1602.

Fingleton, B. and J. Le Gallo (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. Papers in Regional Science 87(3), 319–339.

Forbes, J. F. and A. McGregor (1984). Unemployment and mortality in post-war Scotland. Journal of Health Economics 3(3), 239–257.

Foroni, C., P. Guérin, and M. Marcellino (2015). Markov-switching mixed-frequency VAR models. International Journal of Forecasting 31(3), 692–711.

Foroni, C., M. Marcellino, and C. Schumacher (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. Journal of the Royal Statistical Society: Series A (Statistics in Society) 178(1), 57–82.

Galvão, A. B. (2013). Changes in predictive ability with mixed frequency data. International Journal of Forecasting 29(3), 395–410.

Gerdtham, U.-G. and C. J. Ruhm (2006). Deaths rise in good economic times: Evidence from the OECD. Economics & Human Biology 4(3), 298–316.

Ghysels, E. (2013). Matlab toolbox for mixed sampling frequency data analysis using MIDAS regression models. Available on MATLAB Central at http://www. mathworks. com/matlabcentral/fileexchange/45150-midas-regression.

Ghysels, E. and M. Marcellino (2018). Applied economic forecasting using time series methods. Oxford University Press.

Ghysels, E. and J. I. Miller (2015). Testing for cointegration with temporally aggregated and mixed-frequency time series. Journal of Time Series Analysis 36(6), 797–816.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004). The MIDAS touch: Mixed data sampling regression models. IDEAS Working Paper Series from RePEc.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2005). There is a risk-return trade-off after all. Journal of Financial Economics 76(3), 509–548.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. Journal of Econometrics 131(1), 59–95.

Ghysels, E., A. Sinko, and R. Valkanov (2007). MIDAS regressions: Further results and new directions. Econometric Reviews 26(1), 53–90.

Ghysels, E. and J. H. Wright (2009). Forecasting professional forecasters. Journal of Business & Economic Statistics 27(4), 504–516.

Goldin, C. (2014). A grand gender convergence: Its last chapter. American Economic Review 104(4), 1091–1119.

Götz, T. B., A. Hecq, and J. Urbain (2014). Forecasting mixed-frequency time series with ECM-MIDAS models. Journal of Forecasting 33(3), 198–213.

Granados, J. A. T. (2005). Recessions and mortality in Spain, 1980–1997. European Journal of Population/Revue européenne de Démographie 21(4), 393–422.

Granger, C. W. (1987). Implications of aggregation with common factors. Econometric Theory 3(2), 208–222.

Guérin, P. and M. Marcellino (2013). Markov-switching MIDAS models. Journal of Business & Economic Statistics 31(1), 45–56.

Gulen, H. and M. Ion (2016). Policy uncertainty and corporate investment. The Review of Financial Studies 29, 523–564.

Gunderson, M. (1975). Male-female wage differentials and the impact of equal pay legislation. The Review of Economics and Statistics 57(4), 462.

Hallward-Driemeier, M., B. Rijkers, and A. Waxman (2017). Can minimum wages close the gender wage gap? Evidence from Indonesia. Review of Income and Wealth 63(2), 310–334.

Handley, K. and N. Limao (2015). Trade and investment under policy uncertainty: theory and firm evidence. American Economic Journal: Economic Policy 7(4), 189–222.

Hartmann, H., J. Hayes, and J. Clark (2014). How equal pay for working women would reduce poverty and grow the American economy. Washington, DC.: Institute for Women's Policy Research, Briefing paper (IWPR# C411). Retrieved on January 10, 2016.

Heutel, G. and C. J. Ruhm (2016). Air pollution and procyclical mortality. Journal of the Association of Environmental and Resource Economists 3(3), 667–706.

Johnson, K. (2014, February). Crossing borders and changing lives, lured by higher state minimum wages. The New York Times.

Kalemli-Ozcan, S. (2002). Does the mortality decline promote economic growth? Journal of Economic Growth 7(4), 411–439.

Kang, W., K. Lee, and R. Ratti (2014). Economic policy uncertainty and firm-level investment. Journal of Macroeconomics 39, 42.

Kapoor, M., H. H. Kelejian, and I. R. Prucha (2007). Panel data models with spatially correlated error components. Journal of Econometrics 140(1), 97–130.

Karnizova, L. and J. C. Li (2014). Economic policy uncertainty, financial markets and probability of US recessions. Economics Letters 125(2), 261–265.

Kelejian, H. H. and I. R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics 17(1), 99–121.

Kelejian, H. H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40(2), 509–533.

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104(2), 219–257.

Kelejian, H. H. and I. R. Prucha (2004). Estimation of simultaneous systems of spatially interrelated cross sectional equations. Journal of Econometrics 118(1-2), 27–50.

Khalaf, L., M. Kichian, C. J. Saunders, and M. Voia (2021). Dynamic panels with MIDAS covariates: Nonlinearity, estimation and fit. Journal of Econometrics 220(2), 589–605.

Korkeamäki, O. and T. Kyyrä (2006). A gender wage gap decomposition for matched employer-employee data. Labour Economics 13(5), 611–638.

Laporte, A. (2004). Do economic cycles have a permanent effect on population health? Revisiting the Brenner hypothesis. Health Economics 13(8), 767–779.

Leahy, J. V. and T. M. Whited (1996). The effect of uncertainty on investment: Some stylized trends. Journal of Money, Credit & Banking 28(1), 64.

Lee, D. S. (1999). Wage inequality in the United States during the 1980s: Rising dispersion or falling minimum wage? The Quarterly Journal of Economics 114(3), 977–1023.

Lee, L.-F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18(2), 252–277.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72(6), 1899–1925.

LeSage, J. P. (1997). Bayesian estimation of spatial autoregressive models. International Regional Science Review 20(1-2), 113–129.

LeSage, J. P. and M. M. Fischer (2008). Spatial growth regressions: Model specification, estimation and interpretation. Spatial Economic Analysis 3(3), 275–304.

LeSage, J. P. and R. K. Pace (2009). Introduction to spatial econometrics. Statistics, textbooks and monographs. Boca Raton ; London: CRC Press.

Li, K. (2017). Fixed-effects dynamic spatial panel data models and impulse response analysis. Journal of Econometrics 198(1), 102–121.

Liu, X. (2012). On the consistency of the LIML estimator of a spatial autoregressive model with many instruments. Economics Letters 116(3), 472–475.

Liu, X. and L.-F. Lee (2013). Two-stage least squares estimation of spatial autoregressive models with endogenous regressors and many instruments. Econometric Reviews 32(5-6), 734–753.

Liu, X. and P. Saraiva (2015). GMM estimation of SAR models with endogenous regressors. Regional Science and Urban Economics 55, 68–79.

Magnus, J. R. and H. Neudecker (2019). Matrix differential calculus with applications in statistics and econometrics. John Wiley & Sons.

Manesh, S. N., J. O. Choi, B. K. Shrestha, J. Lim, and P. P. Shrestha (2020). Spatial analysis of the gender wage gap in architecture, civil engineering, and construction occupations in the United States. Journal of Management in Engineering 36(4), 04020023.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. Journal of Business & Economic Statistics 17(1), 129–136.

Marshall, J. R. and D. P. Funch (1979). Mental illness and the economy: A critique and partial replication. Journal of Health and Social Behavior, 282–289.

McCall, L. (1998). Spatial routes to gender wage (in) equality: Regional restructuring and wage differentials by gender and education. Economic Geography 74(4), 379–404.

McCall, L. (2001). Complex inequality: Gender, class, and race in the new economy. Psychology Press.

Miller, J. I. (2014). Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. Journal of Financial Econometrics 12(3), 584–614.

Mincer, J. (1976). Unemployment effects of minimum wages. Journal of Political Economy 84(4, Part 2), S87–S104.

Mincer, J. and S. Polachek (1974). Family investments in human capital: Earnings of women. Journal of Political Economy 82(2, Part 2), S76–S108.

Moran, P. A. (1950). Notes on continuous stochastic phenomena. Biometrika 37(1/2), 17–23.

Neumayer, E. (2004). Recessions lower (some) mortality rates: Evidence from Germany. Social Science & Medicine 58(6), 1037–1047.

Newbold, P. and C. W. J. Granger (1974). Experience with forecasting univariate time series and the combination of forecasts. Journal of the Royal Statistical Society: Series A (General) 137(2), 131–146.

Nicolini, E. A. (2007). Was Malthus right? A VAR analysis of economic and demographic interactions in pre-industrial England. European Review of Economic History 11, 99–121.

Nordentoft, M., K. Wahlbeck, J. Hallgren, J. Westman, U. Osby, H. Alinaghizadeh, M. Gissler, and T. M. Laursen (2013). Excess mortality, causes of death and life expectancy in 270,770 patients with recent onset of mental disorders in Denmark, Finland and Sweden. PLoS ONE 8(1), e55176.

Ord, K. (1975). Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70(349), 120–126.

Paelinck, J. H. and L. H. Klaassen (1979). Spatial econometrics, Volume 1. Saxon House.

Parker, K. and C. Funk (2017). Gender discrimination comes in many forms for today's working women. Pew Research Center 14.

Pástor, L. and P. Veronesi (2013). Political uncertainty and risk premia. Journal of Financial Economics 110(3), 520–545.

Pinkse, J. and M. E. Slade (2010). The future of spatial econometrics. Journal of Regional Science 50(1), 103–117.

Robinson, H. (2005). Regional evidence on the effect of the national minimum wage on the gender pay gap. Regional Studies 39(7), 855–872.

Ruhm, C. J. (1995). Economic conditions and alcohol problems. Journal of Health Economics 14(5), 583–603.

Ruhm, C. J. (2000). Are recessions good for your health? The Quarterly Journal of Economics 115(2), 617.

Ruhm, C. J. (2005). Healthy living in hard times. Journal of Health Economics 24(2), 341–363.

Ruhm, C. J. (2015). Recessions, healthy no more? Journal of Health Economics 42, 17–28.

Ruhm, C. J. (2016). Health effects of economic crises. Health Economics 25(S2), 6–24.

Sbarra, D. A., R. W. Law, and R. M. Portley (2011). Divorce and death: A meta-analysis and research agenda for clinical, social, and health psychology. Perspectives on Psychological Science 6(5), 454–474.

Schober, T. and R. Winter-Ebmer (2011). Gender wage inequality and economic growth: Is there really a puzzle? A comment. World Development 39(8), 1476 – 1484.

Shor, E., D. J. Roelfs, P. Bugyi, and J. E. Schwartz (2012). Meta-analysis of marital dissolution and mortality: Reevaluating the intersection of gender and age. Social Science & Medicine 75(1), 46–59.

Stuckler, D., S. Basu, M. Suhrcke, A. Coutts, and M. Mckee (2009). The public health effect of economic crises and alternative policy responses in Europe: An empirical analysis. The Lancet 374(9686), 315–323.

Stuckler, D., S. Basu, M. Suhrcke, A. Coutts, and M. Mckee (2011). Effects of the 2008 recession on health: A first look at European data. The Lancet 378(9786), 124–125.

Tapia Granados, J. A. and E. L. Ionides (2017). Population health and the economy: Mortality and the great recession in Europe. Health Economics 26(12), e219–e235.

Tiao, G. C. (1972). Asymptotic behaviour of temporal aggregates of time series. Biometrika 59(3), 525–531.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. Economic Geography 46(sup1), 234–240.

Upton, G. and B. Fingleton (1985). Spatial data analysis by example. Volume 1: Point pattern and quantitative data. John Wiley & Sons Ltd.

Vandoros, S., M. Avendano, and I. Kawachi (2019). The association between economic uncertainty and suicide in the short-run. Social Science & Medicine 220, 403–410.

Vandoros, S. and I. Kawachi (2021). Economic uncertainty and suicide in the united states. European journal of epidemiology 36(6), 641–647.

Vecchia, C. and S. Franceschi (1991). Reproductive factors and colorectal cancer. Cancer Causes & Control 2(3), 193–200.

Wagstaff, A. (1985). Time series analysis of the relationship between unemployment and mortality: A survey of econometric critiques and replications of Brenner's studies. Social Science & Medicine 21(9), 985–996.

Wang, Y., C. R. Chen, and Y. S. Huang (2014). Economic policy uncertainty and corporate investment: Evidence from China. Pacific-Basin Finance Journal 26, 227–243.

Wei, W. W. (1978). The effect of temporal aggregation on parameter estimation in distributed lag model. Journal of Econometrics 8(2), 237–246.

White, H. (1996). Estimation, inference and specification analysis. Number 22. Cambridge University Press.

Wiseman, T. and N. Dutta (2016). Religion and the gender wage gap: A US state-level study. Available at SSRN 2738523.

Wooldridge, J. M. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. Journal of Econometrics 182(1), 226–234.