Essays on Multivariate and Simultaneous Equations
Spatial Autoregressive Models

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy in the Graduate School of The Ohio State
University

By

Kai Yang, M.A.

Graduate Program in Economics

The Ohio State University

2016

Dissertation Committee:

Lung-fei Lee, Advisor
Robert de Jong
Jason Blevins
Abstract

Databases with cross-sectional interdependent variables have highlighted the need for new data analysis techniques to model interdependence patterns cross-sectional units. Among various models to describe the interdependence, spatial autoregressive models (SAR) have attracted much attention. The theory and practice of single dependent variable SAR have been well established. Although a large number of economic theories may concern about interrelations among several economic variables, econometric studies regarding multivariate and simultaneous equations SAR models are limited. This dissertation is filling in this gap. This dissertation is composed of two chapters, the first chapter focuses on models with cross-sectional data, while the second chapter is on models in panel data which incorporates both intertemporal dynamics and spatial interdependence.

The first chapter investigates a simultaneous equations spatial autoregressive model which incorporates simultaneity effects, own-variable spatial lags and cross-variable spatial lags as explanatory variables, and allows for correlation between disturbances across equations. In exposition, this chapter also discusses a multivariate spatial autoregressive model that can be treated as a reduced form of the simultaneous equations model. For a multivariate model, we provide identification conditions in terms of the existence of instruments for spatial lags and
regularities of the weight matrix structure. Rank conditions and order conditions are provided for identification of structural parameters in the simultaneous equations model. In this chapter we study parameter spaces, the parameter identification, asymptotic properties of the quasi-maximum likelihood estimation, and computational issues. Monte Carlo experiments illustrate the advantages of the QML, broader applicability and efficiency, compared to instrumental variables based estimation methods in the existing literature.

The second chapter introduces multivariate and simultaneous equations dynamic panel spatial autoregressive models in the cases of stability and spatial cointegration. A spatial unit is assumed to depend on its lagged term, and to respond to its neighbours’ or peers’ behaviour in the current period (spatial lags), and in the previous period (space-time lags). The disturbances in the model are specified with time fixed effects and individual fixed effects in addition to idiosyncratic disturbances. This chapter investigates identification for the model with simultaneous effects, time dynamic effects, and spatial effects. In the estimation of stable and spatially cointegrated models, we investigate QMLE and establish asymptotic properties of the estimator. Convergence rates of parameters may change depending on variables being stable or unstable. We analyze asymptotic biases and suggest bias-corrected estimates. We also study a robust estimation method which can be applied to stable case, spatial cointegration case and some spatial explosion cases. We apply the model to study the grain market integration using a unique historical dataset of rice and wheat prices of 65 cities in 49 years in Yangtze River Basin. The empirical result shows that rice and wheat prices are spatially cointegrated across cities. These results provide
evidences of interregional and intertemporal grain market integration and trading network in the eighteenth-century Yangtze River basin.
Dedicated to my beloved parents and wife.
Acknowledgments

First of all, I would like to express my deepest gratitude and respect to my advisor, Professor Lung-fei Lee, for his intellectual inspiration, tremendous support, guidance and encouragement. He is a great teacher and advisor. I would not be able to complete this dissertation without his years of support.

I am grateful to my committee members and econometrics faculty at the Ohio State University, Professor Robert de Jong, Professor Jason Blevins and Professor Stephen Cosslett, for their constructive guidance and inspired suggestions on my work. Their insightful comments have greatly improved my dissertation. I would also like to thank the econometrics group – Jihai Yu, Xi Qu, Xingbai Xu, Xiaoyi Han, Chao Yang, Neslihan Sakarya, Wei Shi, Tuo Liu, Wei Cheng, Shin-Yi Wu and Hanbat Jeong – for their valuable discussions on my dissertation.

My thanks also go to faculty, staff, and graduate students in the Economics Department for their assistance and support throughout my PhD journey.

I owe a lot to my family and friends for their support over the past few years. None of this would have been possible without the love and support of my family and friends. I devote my deepest gratitude to my parents, Zonghe Yang and Qingyu Qu, and my wife Wei Kong, for their love and support in me.
Vita

2011 .............................................. B.A. Economics, Central University of Finance and Economics, China
2012 .............................................. M.A. Economics, The Ohio State University
2011-present ................................. Graduate Teaching Associate, The Ohio State University

Fields of Study

Major Field: Economics
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>v</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>vi</td>
</tr>
<tr>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>1. Identification and QML Estimation of Multivariate and Simultaneous</td>
<td>1</td>
</tr>
<tr>
<td>Equations Spatial Autoregressive Models</td>
<td></td>
</tr>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Model Specifications: SESAR and MSAR Models</td>
<td>5</td>
</tr>
<tr>
<td>1.2.1 Basic Model Structure, Parameter Space and Stability</td>
<td>7</td>
</tr>
<tr>
<td>1.2.2 Computation Algorithm</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Identification</td>
<td>14</td>
</tr>
<tr>
<td>1.3.1 Identification of the General MSAR Models</td>
<td>14</td>
</tr>
<tr>
<td>1.3.2 Identification of Structural Parameters of the SESAR Model</td>
<td>18</td>
</tr>
<tr>
<td>1.4 Consistency and Asymptotic Normality of the QML</td>
<td>24</td>
</tr>
<tr>
<td>1.5 Monte Carlo Experiments</td>
<td>28</td>
</tr>
<tr>
<td>1.5.1 Monte Carlo Experiments for MSAR Models</td>
<td>29</td>
</tr>
<tr>
<td>1.5.2 Monte Carlo Experiments for simultaneous equations Models</td>
<td>35</td>
</tr>
<tr>
<td>1.6 Conclusion and Remarks</td>
<td>36</td>
</tr>
</tbody>
</table>
2. Multivariate and Simultaneous Equations Dynamic Panel Spatial Autoregressive Models ........................................ 41
   2.1 Introduction ...................................................... 41
   2.2 A Multivariate Dynamic Panel Spatial Autoregressive Model . 46
      2.2.1 Stability and Parameter Space ........................... 47
   2.3 Identification and QML Estimation for Stable MDP-SAR Model 55
      2.3.1 Assumptions ................................................. 55
      2.3.2 Identification and Consistency of QMLE ................. 58
      2.3.3 Asymptotic Normality of QMLE .......................... 62
   2.4 Identification and QML Estimation of the Spatially Conintegrated
      MDP-SAR Model .................................................. 66
      2.4.1 Error Correction Model Representation .................... 67
      2.4.2 Alternative Representation: Unit Roots View on Rotations
      across Multivariate Variables and over Space .................. 69
      2.4.3 Assumptions ................................................. 70
      2.4.4 Reparameterization and Concentrated Log Likelihood Func-
      tion .............................................................. 72
      2.4.5 Identification and Consistency ............................ 74
      2.4.6 Asymptotic Distribution and Bias Correction .............. 75
   2.5 Robust Estimation ................................................ 78
      2.5.1 Asymptotic Properties of RQMLE ......................... 81
   2.6 Monte Carlo Experiments ....................................... 85
   2.7 simultaneous equations Dynamic Panel SAR Model .............. 89
   2.8 Application: Market Integration of Yangtze River Basin ..... 92
      2.8.1 Data ......................................................... 94
      2.8.2 Model, Results and Robust Check .......................... 95
   2.9 Conclusion ....................................................... 99

Bibliography .......................................................... 104

Appendices ............................................................ 109

A. Covariance Matrix for QMLE in Chapter 1 ....................... 109
   A.1 SESAR Model .................................................... 109
   A.2 MSAR Model ..................................................... 114

B. 2SLS and 3SLS estimations with optimal IVs for SESAR .......... 118
   B.1 2SLS with Optimal IVs ........................................ 119
B.2 3SLS with Optimal IVs ................................................. 121

C. Proofs of Lemmas, Theorems and Corollaries in Chapter 1 .............. 123

D. Some Detailed Derivation and Useful Lemmas in Chapter 2 .............. 145
   D.1 Detailed Derivations ............................................. 145
   D.2 Useful Lemmas .................................................. 146

E. Details on First Order Derivatives and Second Order Derivatives of Quasi-Maximum Likelihood Functions for MDP-SAR ................. 161
   E.1 The Stable Case ............................................... 161
   E.2 The Spatial Cointegration Case ................................. 172
   E.3 Robust Estimation ............................................. 175

F. Proofs of Lemmas, Theorems and Corollaries in Chapter 2 .............. 179
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Monte Carlo Results for Example 1</td>
<td>30</td>
</tr>
<tr>
<td>1.2</td>
<td>Monte Carlo Results for Example 2</td>
<td>30</td>
</tr>
<tr>
<td>1.3</td>
<td>Monte Carlo Results for Example 3</td>
<td>32</td>
</tr>
<tr>
<td>1.4</td>
<td>Monte Carlo Results for Example 3 Using Optimal IVs</td>
<td>33</td>
</tr>
<tr>
<td>1.5</td>
<td>Monte Carlo for Example 3 with non-normal distributed disturbance</td>
<td>34</td>
</tr>
<tr>
<td>1.6</td>
<td>Monte Carlo Results for Example 4</td>
<td>37</td>
</tr>
<tr>
<td>1.7</td>
<td>Monte Carlo Results for Example 4 with Optimal IVs</td>
<td>38</td>
</tr>
<tr>
<td>2.1</td>
<td>Monte Carlo for MDP-SAR, Stable Case</td>
<td>87</td>
</tr>
<tr>
<td>2.2</td>
<td>Monte Carlo for MDP-SAR, Spatial Cointegration Case</td>
<td>88</td>
</tr>
<tr>
<td>2.3</td>
<td>Monte Carlo for MDP-SAR, Robust Estimation</td>
<td>90</td>
</tr>
<tr>
<td>2.4</td>
<td>Data Summary</td>
<td>95</td>
</tr>
<tr>
<td>2.5</td>
<td>Empirical Application: System Estimation Results</td>
<td>98</td>
</tr>
<tr>
<td>2.6</td>
<td>Empirical Application: Robust Check</td>
<td>101</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Impulse Response to One Unit Shock on Rice Price in Anqing City, Anhui Province</td>
<td>100</td>
</tr>
</tbody>
</table>
Chapter 1: Identification and QML Estimation of Multivariate and Simultaneous Equations Spatial Autoregressive Models

1.1 Introduction

The single equation spatial autoregressive (SAR) model has received much attention in spatial econometrics. Anselin (1988) summarizes early development in estimation and testing for SAR models, while Kelejian and Prucha (1998, 1999) and Lee (2004, 2007) investigate the two stage least squares (2SLS), the three stage least squares (3SLS), the quasi-maximum likelihood (QML) and generalized method of moments (GMM) estimation methods. Though univariate SAR models have been well developed, the identification and estimation of multivariate SAR (MSAR) models, including simultaneous equations, are limited with a few exceptions: Kelejian and Prucha (2004), Baltagi and Bresson (2011), Baltagi and Deng (2012), Cohen-Cole et al. (2013) and Liu (2014).

The studies on MSAR model have been empirically motivated. The MSAR and simultaneous equations SAR (SESAR) models have been employed in regional science, such as studies on migration, employment and population (de Graaff et al., 2012; Gebremariam et al., 2011); housing economics (Baltagi and Bresson, 2011; Jeanty, 2010); fiscal policies analysis, such as fiscal competition
over taxes and public input (Hauptmeier et al., 2012), and interactions between governments expenditures (Aller and Elhorst, 2011); and topics in agricultural economics (e.g. Wu and Lin, 2010). Furthermore, the MSAR and SESAR models can be seen as best response functions derived from network games with multi-choice by capturing own-choice peer effects, cross-choices peer effects, simultaneity effects and correlated effects across disturbances. Cohen-Cole et al. (2013) propose a multi-choice network game framework with a linear-quadratic utility function. Their empirical example studies cross-choices peer effects between time spent watching TV and students’ GPA, by estimating bivariate SAR models using Kelejian and Prucha’s (2004) generalized spatial 2SLS. The empirical result demonstrates the presence of cross-variable spatial effects.

The research on ML estimation of a MSAR model is theoretically and empirically motivated as well. Kelejian and Prucha (2004) consider two stage least square (2SLS) and three stage least square (3SLS) estimation methods for a SESAR model, incorporating spatial lags in dependent variables and allowing for spatial correlation in disturbance terms. Their investigation of estimation and estimators’ large sample properties suggest computational simple estimation methods for empirical studies. Baltagi and Deng (2012) extend the model to fit panel data by deriving a 3SLS estimator for SESAR model with random effects. Motivated by multi-choice network games, Cohen-cole et al (2013) investigate the identification of various bivariate spatial autoregressive models: the “seemingly unrelated” equations model\(^1\) with correlated disturbances that contains endogenous spatial effects and contextual effects; the triangular system model which

\(^1\)This model excludes simultaneity effects and cross-choice peer effects as explanatory variables.
introduces a one-way cross-choice peer effect in a two-equations system; and a square system of simultaneous equations with simultaneity effects and two-way cross-choice peer effects. While the first two models can be identified, the last one requires exclusive restrictions. They also utilize 2SLS and 3SLS empirically to estimate all three models. Instead of the instrumental variables (IV)-based estimation, Baltagi and Bresson (2011) employ the ML method to estimate a spatial seemingly unrelated regression panel model. The disturbances allow spatial error (SE) and spatial error components. They also propose joint and conditional Lagrange multiplier tests for the presence of spatial correlation and random effects.

However, there are several issues remained. First, IV-based estimation methods, such as 2SLS and 3SLS, do not work for models without exogenous variable and regression models with SE disturbances. Also, as we have shown in Monte Carlo experiments, in some cases, conventional IV methods can result in very large standard deviations of estimator. IV-based estimation methods also have efficiency losses even when they are precise. Second, the parameter space of spatial effects need to be studied since it is crucial for the stabilization of the MSAR model related to the interpretation of empirical results. Third, identification conditions of models’ parameters, including the identification of parameters in the MSAR models, and order and rank conditions in the SESAR models, should be developed. Last, computational challenges using QML methods are to be considered.
In this paper, we focus on the identification, QML estimation, parameter spaces, and computational issues. This paper investigates identification and QML estimation of SESAR and MSAR models taking own-variable spatial effects, cross-variable spatial effects and correlated effects into consideration. A SESAR model is specified as: \[ Y_{nm} = W_{n}Y_{nm}\Gamma_{m} + X_{n}C_{m} + U_{nm}, \] which contains \( m \) endogenous variables within a system of \( m \) simultaneous equations. A MSAR model has the specification \[ Y_{nm} = W_{n}Y_{nm}\Psi_{m} + X_{n}\Pi_{m} + V_{nm}, \] which consists of \( m \) equations with \( m \) endogenous variables. The MSAR model can be regarded as the ‘quasi-reduced form’ of a SESAR model. We investigate parameter identification and asymptotic properties of the full information maximum likelihood estimator, which is strictly speaking a QML as normality is not really imposed. We provide identification conditions analogous to the rank identification in the classical linear simultaneous equations (SEM) model. A central limit theorem of an extended linear-quadratic form is developed and employed to characterize large sample distribution of the estimator. Furthermore, we derive the asymptotic distribution and finite sample performances of 2SLS and 3SLS estimation with optimal IVs in order to supplement the existing literature and for comparison (Appendix B and Section 5).

Monte Carlo experiments in Section 5 illustrate two advantages of the QML estimation. First, while some SESAR and MSAR models without exogenous variables cannot be estimated by IV based methods, they may still be estimated by QML. Second, when there are problems with IV, such as imprecise 2/3SLS estimators, the QML estimator (QMLE) may still be efficient and precise. The QMLE
can also gain efficiency by reducing 15%-50% standard deviation over 3SLS estimator (3SLSE) even when IV-based estimators are precise. Section 6 concludes. Proofs, details of derivations, a supplementary discussion of identification and 2SLS/3SLS estimation with optimal IVs are reported either in the appendix or in an online supplementary file.

1.2 Model Specifications: SESAR and MSAR Models

We study a SESAR model:

\[ Y_{nm} \Gamma_m = W_n Y_{nm} \Lambda_m + X_n C_m + U_{nm}. \]  

(1.1)

where \( Y_{nm} \) is a \( n \times m \) matrix, which consists of \( m \) endogenous variables; and \( n \) is the number of observations. Neighborhood relationships are summarized by \( W_n, \) a \( n \times n \) matrix. \( X_n \) represents the \( n \times k_x \) matrix of all observations on \( k_x \) exogenous variables. The \( n \times m \) matrix, \( U_{nm}, \) is the disturbance matrix. There are three parameter matrices, \( \Gamma_m, \) a \( m \times m \) matrix, \( \Lambda_m, \) a \( m \times m \) matrix, and \( C_m \) a \( k_x \times m \) matrix. In addition, we assume that \( u_{n,i}, \) the \( i \)th row of \( U_{nm}, \) are i.i.d. with zero mean and covariance matrix \( \Sigma_{um} \) for \( i = 1, 2, ..., n. \) This model includes simultaneity effects, represented by \( \Gamma_m \) which is invertible with its diagonal elements being normalized to ones, own-variable spatial effects represented by diagonal elements of \( \Lambda_m, \) cross-variable spatial effects represented by off-diagonal elements of \( \Lambda_m, \) and correlated effects across equations incorporated in the covariance matrix of disturbances. The traditional SEM can be viewed as \( Y_{nm} \Gamma_m = X_n C_m + U_{nm}. \) The difference between traditional SEM and SESAR model is that the later contains the additional explanatory regressor \( W_n Y_{nm}, \) which introduces spatial (cross-sectional) dependence among agents. For example, in the traditional
model of markets, each \( i \) would be an isolated market from which equilibrium \( Y_{n,i} \), the \( i \)th row of the matrix \( Y_{nm} \), would be determined by the \( i \)th market alone. But for the SESAR model, the equilibrium matrix \( Y_{nm} \) needs to be determined simultaneously by all markets.

Then, the corresponding MSAR model is

\[
Y_{nm} = W_n Y_{nm} \Psi_m + X_n \Pi_m + V_{nm},
\]

which can be seen as a quasi-reduced form of SESAR (1.1) with
\[
\Psi_m = \Lambda_m \Gamma_m^{-1}, \quad \Pi_m = C_m \Gamma_m^{-1} \quad \text{and} \quad \Sigma_{vm} = \Gamma_m^{-1} \Sigma_{um} \Gamma_m^{-1}
\]

where might be constraints across parameters in \( \Psi_m \) and \( \Pi_m \). On the other hand, it can also be a special case of (1) when \( \Gamma_m = I_m \). In this case, the parameters in (2) might not have cross parameter constraints. By taking vectorization, the model is transformed into

\[
\vec(Y_{nm}) = (\Psi'_m \otimes W_n) Y_{nm} + (I_m \otimes X_n) \vec(\Pi_m) + \vec(V_{nm}).
\]

With i.i.d. \( v_{n,i} \) across \( i = 1, \ldots, n \), the sample average log likelihood function for (1.2) is:

\[
\frac{1}{mn} \ln L_{nm}(\Psi_m, \Pi_m, \Sigma_{vm}) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2m} \ln |\Sigma_{vm}| + \frac{1}{mn} \ln |S_{nm}|
\]

\[
-\frac{1}{2mn} [S_{nm} \vec(Y_{nm}) - (I_m \otimes X_n) \vec(\Pi_m)]' \left( \Sigma_{vm}^{-1} \otimes I_n \right) [S_{nm} \vec(Y_{nm}) - (I_m \otimes X_n) \vec(\Pi_m)],
\]

where \( S_{nm} = I_{nm} - \Psi'_m \otimes W_n \).

For the SESAR model (1.1), the log likelihood function is

\[
\frac{1}{mn} \ln L_{nm}(\Gamma_m, \Lambda_m, C_m, \Sigma_{um}) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2m} \ln |\Sigma_{um}| + \frac{1}{mn} \ln |\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n|
\]

\[
-\frac{1}{2mn} [(\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n) \vec(Y_{nm}) - (I_m \otimes X_n) \vec(C_m)]' \left( \Sigma_{um}^{-1} \otimes I_n \right)
\]

\[
\times [(\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n) \vec(Y_{nm}) - (I_m \otimes X_n) \vec(C_m)].
\]

It is apparent that this likelihood function is almost the same as that of the MSAR model by replacing \( \Psi_m = \Lambda_m \Gamma_m^{-1}, \Pi_m = C_m \Gamma_m^{-1} \) and \( \Sigma_{vm} = \Gamma_m^{-1} \Sigma_{um} \Gamma_m^{-1} \). In order
to derive consistency and asymptotic normality for the QMLE of SESAR and that of a general MSAR model, we make some regularity conditions.

1.2.1 Basic Model Structure, Parameter Space and Stability

The following assumptions on disturbances and regressors are basic for our models.

**Assumption 1.1.** (1) For the SESAR model, the $i$th row $u_{n,i}$ of $U_{nm}$ in (1.1), is a random vector with zero mean and covariance matrix $\Sigma_{um}$ and is i.i.d. for all $i$. The elements of disturbances satisfy the moment condition $\mathbb{E}|u_{n,ik}u_{n,il}u_{n,ip}u_{n,iq}|^{1+\delta} < D_4$ for any $i = 1, 2, ..., n$ and $k, l, p, q = 1, 2, ..., m$ for some positive constants $D_4$ and $\delta > 0$.

(2) For the MSAR model, $v_{n,i}$, the $i$th row of $V_{nm}$ in (1.2), is a random vector of dimension $m$ with zero mean and covariance matrix $\Sigma_{vm}$ and are i.i.d. across rows. The elements of disturbances satisfy the moment condition that $\mathbb{E}|v_{n,ik}v_{n,il}v_{n,ip}v_{n,iq}|^{1+\delta}$ is uniformly bounded for any $i = 1, 2, ..., n$ and $k, l, p, q = 1, 2, ..., m$ for some positive constants $\delta$.

Assumption 1.1 allows correlation between disturbance terms from different structural equations for each spatial unit. This assumption means that different
endogenous variables for an individual unit can be influenced by some correlated unobservables, and/or some common unobservables entering the disturbance process. Higher than the fourth moments are bounded in order to derive asymptotic properties of QMLEs in following sections. Under the i.i.d. random vector of disturbances, the absolute moment $\mathbb{E} |u_{n,ik}u_{n,it}u_{n,iq}u_{n,ip}|^{1+\delta}$ is a constant, which will be denoted by $|\mu|_{1+\delta,klpq}$. Assumption 1.1(2) is similar to Assumption 1.1(1) but for the disturbances of the MSAR model. For the case, the MSAR is a quasi-reduced form of the SESAR, and the i.i.d. and moment properties of $V_{nm} = U_{nm}\Gamma_m$ will be inherited from $U_{nm}$ in Assumption 1.1(1).

**Assumption 1.2.** Elements of $X_n$ are exogenous constants and are uniformly bounded for all $n$.

The above assumption states exogeneity of $X_n$. The uniform boundedness will simplify our asymptotic analysis and the use of a central limit theorem\(^2\). Additional regularity will be assumed for $X_n$ for the identification purpose in Assumption 1.5 below.

The parameter space of $\Psi_m$ in (1.2), which can be the quasi-reduced form of (1.1), is important to guarantee that the SAR equations system is stable and the Jacobian matrix has a positive determinant. Since $|I_{nm} - \Psi_m' \otimes W_n| = \prod_{i=1}^p \prod_{j=1}^m (1 - \rho_j \lambda_i)$, where $\rho_j$ is an eigenvalue of $\Psi_m$ and $\lambda_i$ is an eigenvalue of $W_n$, in order to give a positive determinant, for each real $\rho_j$, where $j = 1, \cdots, m$, $1/\min_i (\lambda_i) < \rho_j < 1/\max_i (\lambda_i)$ where $\lambda_i$’s are real eigenvalues of $W_n$. For complex $\rho_j$ and real $\lambda_i$, $(1 - \rho_j \lambda_i)$ will be a complex number and its appearance will associate with

\(^2\)It is possible to relax this boundedness assumption to stochastic $X_n$ with some moment conditions.
a complex conjugate \((1 - \bar{\rho}_j \lambda_i)\). As \((1 - \rho_j \lambda_i)(1 - \bar{\rho}_j \lambda_i) = |1 - \rho_j \lambda_i|^2 > 0\) because \(1 - \rho_j \lambda_i\) can not be zero, the cases with complex \(\rho\) and real \(\lambda\) will have positive contribution to the determinant and they will not matter (LeSage and Pace (2010)). So for the case that \(W_n\) has all real eigenvalues, the parameter space can be completely determined by all the real eigenvalues \(\rho_j\) of \(\Psi_m\) such that \(1/\min_i(\lambda_i) \leq \rho_j < 1/\max_i(\lambda_i)\). For the case that both \(\rho_j\) and \(\lambda_i\) are complex, even \((1 - \rho_j \lambda_i)(1 - \bar{\rho}_j \lambda_i) = |1 - \rho_j \lambda_i|^2\) is nonnegative, it can be zero in the event \(\rho_j = 1/\lambda_i = \bar{\lambda}_i/|\lambda_i|^2\). In order to rule out the latter, an additional sufficient condition can be \(|\rho_j| < 1/|\lambda_i|\) for all complex \(\lambda_i\) for each complex \(\rho_j\). Since all the eigenvalues of \(\Psi_m\) should lie in intervals indicated, a sufficient condition is that \(\rho_{sr}(\Psi_m) \rho_{sr}(W_n) < 1\), where \(\rho_{sr}(A)\) is the spectral radius of matrix \(A\). \(\rho_{sr}(W_n)\) can be calculated directly since it is the largest absolute value of eigenvalues of \(W_n\). The spectral radius of \(\Psi_m\), is less than any of its induced matrix norm according to the spectral radius theorem. For instance, a strong sufficient condition is that \(||\Psi_m||_\infty < 1/\rho_{sr}(W_n)\). Intuitively, it means that, for any dependent variable, the sum of own variable and cross variable spatial effects in absolute value is bounded. Another sufficient condition can be that \(||\Psi_m||_1 < 1/\rho_{sr}(W_n)\), which means the sum of own variable and cross variable spatial effects from each variable in absolute value is bounded. We make the following assumption:

**Assumption 1.3.** (1) For the SESAR model, the parameter spaces of \(\Gamma_m, \Lambda_m, C_m\) are compact sets. For any \(\Gamma_m\) in its compact parameter space, \(\Gamma_m\) is invertible. The covariance matrix \(\Sigma_{um}\) lies in a compact space and is nonsingular. The true parameters \(\Gamma_{m0}, \Lambda_{m0}, C_{m0},\) and \(\Sigma_{um0}\) are in the interior of the product parameter space.
(2) For the MSAR model, the parameter spaces of $\Pi_m, \Psi_m,$ and the covariance matrix $\Sigma_{vm}$ are compact. Any $\Sigma_{vm}$ in its parameter space is nonsingular. Any $\Psi_m$ in its parameter space satisfies property that $|I_{nm} - \Psi_m' \otimes W_n| > 0.$ The true parameters $\Pi_{m0}, \Psi_{m0}$ and $\Sigma_{v0}$ are in the interior of their product parameter space.

Compact parameter spaces are employed in proofs of consistency and asymptotic properties. The compact parameter space is desirable in proving the uniform convergence of the log likelihood function. Typically, the normalized covariance matrices $\Sigma_{um}$ and $\Sigma_{vm}$ are positive definite, which is so unless some disturbances have multicollinearity. For the SESAR model (1.1) with compact parameter spaces for its structural coefficients, the parameters of the implied quasi-reduced MSAR model will be compact because the inverse and multiplication operations are continuous, so there are no contradictions among them. The requirement of compactness is for $\Psi_m$ since it rules out its boundary at $|I_{nm} - \Psi_m' \otimes W_n| = 0.$ However, as long as the true interaction parameters are bounded away from boundaries of parameters space,$^3$ this does not influence the estimator in practice since the maximizing algorithm when searching for QMLE will avoid, with large probability, any estimator of $\Psi_m$ close to its boundary where the log likelihood would take negative infinity as $\ln |I_{nm} - \Psi_m' \otimes W_n| = ln0 = -\infty.$ This is the advantage for using ML approach for SAR models as pointed out in LeSage and Pace (2009).

$^3$In the event that the true parameters form a sequence (depending on $n$) which converges to the boundary, one might face an estimation issue related to near unit roots. Some limited studies on SAR models are in Baltagi et al. (2013) and Lee and Yu (2013). For such situations, the rate of convergence of estimates would not be the usual $\sqrt{n}$-rate, and the asymptotic analysis will be quite different from the analysis in this paper.
Assumption 1.4. Row and column sums of $W_n$ in absolute value are uniformly bounded, uniformly in $n$ (UB). For any possible $\Psi_m$ in its parameter space, $S_{nm}$ is nonsingular, and $S_{nm}^{-1}$ is UB, uniformly in $\Psi_m$.

Assumption 1.4 is essential in order for a SAR model to be stable across space. It is a typical assumption for stable spatial econometric models (Kelejian and Prucha, 1998; Lee, 2004). As sequences, $S_{nm}$ in $n$ will be UB by UB of $W_n$ and the compact parameter space of $\Psi_m$. The sequence $S_{nm}^{-1}$ is also assumed to be UB. Lemma 1.1 below indicates how these properties can be derived from properties of parameters and the weight matrix. This assumption, along with Assumption 1.3, are used to establish the uniform convergence of the QMLE. They can be used to show that $\ln |S_{nm}|$ as a function of $\Psi_m$ is a Lipschitz function. For example, if we let $m = 1$, $e = [1, ..., 1]'$, and $W_n = (n - 1)^{-1}[ee' - I_n]$. Under the compact assumption of $\psi \in [-1 + \epsilon, 1 - \epsilon]$, the row and column sum norms are bounded by $1/\epsilon$ and $\ln |S_{nm}|$ has Lipschitz bound $1/\epsilon^2$. The assumption is limited since theoretically a natural constraint for $\Psi_m$ is a bounded open set. However, empirically, when optimizing the log-likelihood function and $\Psi_m$ is approaching its boundary, the log-likelihood function would approach negative infinity with probability close to 1 when $\Psi_m \theta$ is in the interior of its parameter space (LeSage and Page 2009).

Lemma 1.1. The sequences $S_{nm}$ and $S_{nm}^{-1}$ are uniformly bounded in column sum norm, uniformly in $\Psi_m$, if $\sup_{\Psi_m} \|\Psi_m' \otimes W_n\|_1 < 1$. They are uniformly bounded in row sum norm, uniformly in $\Psi_m$ if $\sup_{\Psi_m} \|\Psi_m' \otimes W_n\|_\infty < 1$.

4We appreciate a referee points out this example.
The above lemma gives sufficient conditions for the UB of $S_{nm}$ and $S_{nm}^{-1}$, uniformly in $\Psi_m$. This means that their UB property requires that the coefficients of spatial effects and cross-variable spatial effects cannot be very large. For example, when $W_n$ is row normalized, $\| \Psi_m \|_1 < 1$ and $\| \Psi_m \|_\infty < 1 / \sup_n \| W_n \|_1$ will be sufficient to guarantee the above lemma.

### 1.2.2 Computation Algorithm

In this subsection, we pay attention on the computational aspect of QML estimation of the SESAR model. With $m$ equations, there are at most $m(m+1)/2 + 2m^2 + k_x m - m$ parameters to be estimated, which may be subject to “curse of dimension” if $m$ is large.\(^5\) In the log likelihood function, there is a determinant $|\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n|$ containing $2m^2 - m$ parameters. This section briefly discusses some computational techniques that can ease computational burden.

For the determinant of the Jacobian transformation in the objective function, the following result suggests a way to analyze it. For a square matrix $A_m$, its characteristic polynomial is

$$p_{A_m}(\rho) = c_m \rho^m + c_{m-1} \rho^{m-1} + c_{m-2} \rho^{m-2} + ... + c_1 \rho + c_0,$$

where $c_0 = |A_m|$, $c_m = (-1)^m$, and $c_{m-j} = (-1)^{m-j} M_{j,A_m}$ with $M_{j,A_m}$ being the sum of all $j \times j$ principle minors of $A_m$ for $j = 1, 2, ..., m - 1$. $|\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n| = |\Gamma_m| n \prod_{i=1}^n \prod_{j=1}^m (1 - \rho_j \lambda_i)$, where $\rho_j$ is eigenvalue of $\Lambda_m \Gamma_m^{-1}$ and $\lambda_i$ is eigenvalue of $W_n$ for $j = 1, \cdots, m$ and $i = 1, \cdots, n$. There are $m$ terms in $\prod_{j=1}^m (1 - \rho_j \lambda_i)$

\(^5\)For the implementation of the QMLE with an optimization subroutine, the number of parameters to be evaluated is crucial for convergence speed. Therefore, we propose to estimate the model using concentrated likelihood function after substantial reduction of parameters to save time.
containing $\lambda_i$. As we know, $|\Lambda_m \Gamma^{-1}_m - \rho I_m| = \prod_{j=1}^{m} (\rho_j - \rho) = p_{\Lambda_m \Gamma^{-1}_m}(\rho)$. If we multiply $(-1)^m \lambda^m_i$ to both sides of $\prod_{j=1}^{m} (\rho_j - \rho) = p_{\Lambda_m \Gamma^{-1}_m}(\rho)$ and let $\rho = 1/\lambda_i$, where $\lambda_i \neq 0$, it becomes $\prod_{j=1}^{m} (1 - \rho_j \lambda_i) = p_{\Lambda_m \Gamma^{-1}_m}(1/\lambda_i)(-1)^m \lambda^m_i$. With the characteristic polynomial in the above, the log determinant term in the likelihood function becomes $\ln |I_m - (\Lambda_m \Gamma^{-1}_m) \otimes W_n| = \sum_{i=1}^{n} \ln \left[ 1 + \sum_{j=1}^{m} (-1)^j M_{j, \Lambda_m \Gamma^{-1}_m} \lambda^2_j \right]$. Compute the eigenvalues of $W_n$, and have them fixed once for all iterations. During iterations only all the principle minors of $\Lambda_m \Gamma^{-1}_m$ need to be updated, and so is $|\Gamma_m|$. So in general, the determinant of the Jacobian transformation can be effectively evaluated during iterations. This device is motivated by that of Ord (1975). Comparing with the evaluation of such a determinant for a single equation SAR model, the additional computation is to evaluate all the principal minors of $\Lambda_m \Gamma^{-1}_m$ and the determinant of $\Gamma_m$ during iterations of a maximization algorithm. But $m$ has usually a much smaller dimension than $n$. For $m = 2$, the trace and the determinant of $\Lambda_2 \Gamma^{-1}_2$ need to be evaluated during each iteration, as $\ln |I_{2n} - (\Lambda_2 \Gamma^{-1}_2) \otimes W_n| = \sum_{i=1}^{n} \ln \left[ 1 - \text{Tr}(\Lambda_2 \Gamma^{-1}_2) \lambda_i + \text{det}(\Lambda_2 \Gamma^{-1}_2) \lambda^2_i \right]$.

A 3-step method via a concentrated likelihood which has a reduced number of parameters can be as follows:

- **Step 1**: Given $\Lambda_m$, $\Gamma_m$ and $\Sigma_{um}$, we estimate $C_m$ by a generalized least square estimation:

$$\text{vec}(C_m) = \left[(I_m \otimes X_n')(\Sigma^{-1}_{um} \otimes I_n)(I_m \otimes X_n)\right]^{-1} \times (I_m \otimes X_n')(\Sigma^{-1}_{um} \otimes I_n)(\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n)\text{vec}(Y_{nm}).$$

Substituting $\text{vec}(C_m)$ into the log likelihood function, we have

$$\ln L_{nm} = \text{constant} - \frac{n}{2} \ln |\Sigma_{um}| + \ln |\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n|$$

$$- \frac{1}{2} \text{vec}(Y_{nm})'(\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n)'[\Sigma^{-1}_{um} \otimes (I_n - P_{X_n})](\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n)\text{vec}(Y_{nm}),$$
where $P_{X_n} = X_n(X_n'X_n)^{-1}X_n'$. Let $\tilde{y}_{nm}(\Gamma_m, \Lambda_m) = [I_m \otimes (I_n - P_{X_n})](\Gamma_m' \otimes I_n - \Lambda_m' \otimes W_n)\text{vec}(Y_{nm})$, which can be denoted by $(\tilde{y}_{n1}(\Gamma_m, \Lambda_m), \ldots, \tilde{y}_{nm}(\Gamma_m, \Lambda_m))'$, where $\tilde{y}_{n,j}(\Gamma_m, \Lambda_m)$ is an $n$ dimensional column vector for $j = 1, \ldots, m$. The derivative with respect to $\sigma_m^{(kl)}$, the $(k, l)$th elements of $\Sigma_{um}^{-1}$, is

$$\frac{\partial \ln L_{nm}}{\partial \sigma_m^{(kl)}} = \frac{n}{2} \Sigma_{um,kl} - \frac{1}{2} \tilde{y}_{n,k}(\Gamma_m, \Lambda_m)\tilde{y}_{n,l}(\Gamma_m, \Lambda_m),$$

because $\frac{\partial \ln |\Sigma_{um}^{-1}|}{\partial \Sigma_{um}} = \Sigma_{um}$.

- Step 2: $\Sigma_{um,kl}(\Gamma_m, \Lambda_m) = \frac{1}{n} \tilde{y}_{n,k}(\Gamma_m, \Lambda_m)\tilde{y}_{n,l}(\Gamma_m, \Lambda_m)$ for $k, l = 1, \ldots, m$.

- Step 3: We maximize the concentrated log likelihood function with respect to $\Gamma_m, \Lambda_m$:

$$\ln L_{nm} = \text{constant} - \frac{n}{2} \ln |\Sigma_{um}(\Gamma_m, \Lambda_m)| + \sum_{i=1}^{n} \ln \left[1 + \sum_{j=1}^{m} (-1)^j M_{j,\Lambda_m r_m^{-1}} \lambda_j^i\right]$$

$$+ n \ln |\Gamma_m| - \frac{1}{2} \text{vec}(Y_{nm})' (\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n) [\Sigma_{um}(\Gamma_m, \Lambda_m) \otimes (I_n - P_{X_n})]$$

$$\times (\Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n)\text{vec}(Y_{nm}).$$

### 1.3 Identification

In this section, we first discuss identification of the general MSAR model. Then, we provide rank conditions that guarantee the structural parameters in the SESAR model can be identified from the quasi-reduced form parameters of its implied MSAR model.

#### 1.3.1 Identification of the General MSAR Models

The MSAR model contains own-variable spatial lags and cross-variable spatial lags as explanatory variables, which result in endogenous regressors. In addition,
the present of endogenous variables on the right-hand side of an equation makes identification an issue, which needs detailed analysis. The $\Psi_{m0}$, $\Pi_{m0}$, $\Sigma_{vm0}$ and $S_{nm0}$ denote the true parameter values corresponding to $\Psi_m$, $\Pi_m$, $\Sigma_{vm}$ and $S_{nm}$.

Our identification analysis will be based on the information inequality approach in Rothenberg (1971). It employs the expected likelihood function of the MSAR model:

$$
\frac{1}{mn} \mathbb{E} \ln L_{nm}(\Psi_m, \Pi_m, \Sigma_{vm}) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2m} \ln |\Sigma_{vm}| + \frac{1}{mn} \ln |S_{nm}|
$$

$$
-\frac{1}{2mn} \left[ S_{nm} S_{nm0}^{-1}(I_m \otimes X_n) \text{vec}(\Pi_{m0}) - (I_m \otimes X_n) \text{vec}(\Pi_m) \right]' \left( \Sigma_{vm}^{-1} \otimes I_n \right)
$$

$$
\times \left[ S_{nm} S_{nm0}^{-1}(I_m \otimes X_n) \text{vec}(\Pi_{m0}) - (I_m \otimes X_n) \text{vec}(\Pi_m) \right]
$$

$$
-\frac{1}{2mn} \text{Tr} \left[ \left( \Sigma_{vm}^{-1} \otimes I_n \right) S_{nm} S_{nm0}^{-1} \left( \Sigma_{vm0}^{-1} \otimes I_n \right) S_{nm0}^{-1}' S_{nm} \right].
$$

(1.5)

Let $J_i$ ($i = 1, 2, \ldots, m$) be a $1 \times m$ row vector with all zero elements except for the $i$th entry, which is 1. Let $\tilde{X}_{1,n} = (J_1 \otimes I_n) S_{nm0}^{-1}(I_m \otimes (W_n X_n)) \text{vec}(\Pi_{m0}), \ldots, \tilde{X}_{m,n} = (J_m \otimes I_n) S_{nm0}^{-1}(I_m \otimes (W_n X_n)) \text{vec}(\Pi_{m0})$ and $A_{n \times (k_x + m)} = [X_n, \tilde{X}_{1,n}, \tilde{X}_{2,n}, \ldots, \tilde{X}_{m,n}]$.

**Assumption 1.5.** The limiting matrix $\lim_{n \to \infty} \frac{1}{n} \left( A_{n \times (k_x + m)}' A_{n \times (k_x + m)} \right)$ exists and is nonsingular.

Assumption 1.5 guarantees that $\Pi_{m0}$ and $\Psi_{m0}$ can be identified$^6$ by maximizing the log likelihood function. For a single equation SAR model, Assumption 5 reduces to $(X_n, W_n S_{n10}^{-1} X_n, \Pi_{m10})$ having full column rank, which is the same in Lee (2004).$^7$ This sufficient condition guarantees the existence of best IV estimator.

$^6$The identification adopts Definition 3.3 “identifiable uniqueness” in its parameter space in White (1994).

$^7$When $m = 1$, we do not need $W_n$ and $S_{n10}^{-1}$ to be symmetric, because as $S_{n10} = I_n - \Psi_{10} W_n$, where $\Psi_{10}$ is a scalar, it follows that $W_n S_{n10} = W_n - \Psi_{10} W_n^2 = S_{n10} W_n$. Therefore, $S_{n10}^{-1} W_n = W_n S_{n10}^{-1}$ and $(X_n, W_n S_{n10}^{-1} X_n, \Pi_{m10})$ is equal to $(X_n, S_{n10}^{-1} W_n X_n, \Pi_{m10})$. 

15
for each structural equation. This can be seen as follows. The original system (1.2) has its $l$th equation being

\[ y_{nl} = W_n Y_{nm} \psi_{,l} + X_n \pi_{,l} + v_{nl} = \sum_{r=1}^{m} W_n y_{nr} \psi_{,rl} + X_n \pi_{,l} + v_{nl}. \]

For $W_n y_{nr} = W_n (J_r \otimes I_n) \text{vec}(Y_{nm}) = (1 \otimes W_n) (J_r \otimes I_n) \text{vec}(Y_{nm}) = (J_r \otimes I_n) (I_m \otimes W_n) \text{vec}(Y_{nm})$, as \( \text{vec}(Y_{nm}) = S_{nm0}^{-1} \text{vec}(X_n \Pi_{m0}) + \text{vec}(V_{nm}) \), it is apparent that $(J_r \otimes I_n) (I_m \otimes W_n) S_{nm0}^{-1} \text{vec}(X_n \Pi_{m0})$ is a proper best IV for $W_n y_{nr}$ for each $r, r = 1, \cdots, m$.

The above assumption concerns about the implied reduced form regression equation but not the implied variance structure. In order to identify the covariance matrix $\Sigma_{vm}$, we need the following assumption. Let $\sigma_m^2 = \text{Tr}(\Sigma_{vm})$ and $\Sigma_{vm}^* = \Sigma_{vm} / \sigma_m^2$.

**Assumption 1.6.**

\[
\lim_{n \to +\infty} \left[ \frac{1}{mn} \text{Tr} \left[ \left( \Sigma_{vm}^{*-1} \otimes I_n \right) S_{nm} S_{nm0}^{-1} \left( \Sigma_{vm0}^{*-1} \otimes I_n \right) S_{nm0}' S_{nm}' \right] \right. 
- \left. \left| \left( \Sigma_{vm}^{*-1} \otimes I_n \right) S_{nm} S_{nm0}^{-1} \left( \Sigma_{vm0}^{*-1} \otimes I_n \right) S_{nm0}' S_{nm}' \right| \right] > 0
\]

unless $\Sigma_{vm}^* = \Sigma_{vm0}^*$ and $\Psi_m = \Psi_{m0}$.\(^8\)

The intuition of Assumption 1.6 can be illustrated by some simpler sufficient conditions in a finite sample setting.

**Proposition 1.1.** (A) No matter whether $W_n$ is symmetric or asymmetric, suppose no linear combination of $W_n, W_n'$ and $W_n' W_n$ can be proportional to $I_n$, then,

\[
\frac{1}{mn} \text{Tr} \left[ \left( \Sigma_{vm}^{*-1} \otimes I_n \right) S_{nm} S_{nm0}^{-1} \left( \Sigma_{vm0}^{*-1} \otimes I_n \right) S_{nm0}' S_{nm}' \right] 
> \left| \left( \Sigma_{vm}^{*-1} \otimes I_n \right) S_{nm} S_{nm0}^{-1} \left( \Sigma_{vm0}^{*-1} \otimes I_n \right) S_{nm0}' S_{nm}' \right| \frac{1}{mn}
\]

\(^8\)Assumption 6 is essentially an inequality which is a sufficient condition for information inequality to hold. In principle, the numerical evaluation of the log likelihood function for a unique global maximum is implicitly an indirect check on Assumptions 6 with finite sample.
unless $\Sigma^*_{vm} = \Sigma^*_{vm0}$.  

(B) Suppose $W_n$ is asymmetric, $I_n$, $W_n$, $W_n'$ and $W_nW_n'$ are linearly independent, then,

$$
\frac{1}{mn} \text{Tr} \left[ \left( \Sigma_{vm}^{* -1} \otimes I_n \right) S_{nm}S_{nm0}^{-1} \left( \Sigma_{vm0}^{*} \otimes I_n \right) S_{nm0}'S_{nm}' \right] > \frac{1}{mn} \left| \left( \Sigma_{vm}^{* -1} \otimes I_n \right) S_{nm}S_{nm0}^{-1} \left( \Sigma_{vm0}^{*} \otimes I_n \right) S_{nm0}'S_{nm}' \right|
$$

unless $\Sigma^*_{vm} = \Sigma^*_{vm0}$ and $\Psi_m = \Psi_{m0}$.

**Proposition 1.2.** In the limiting case, under Assumption 1.1(2), 1.2, 1.3(2), 1.4 and 1.5, $\Psi_{m0}$ and $\Pi_{m0}$ can be identified; and the covariance matrix $\Sigma_{vm0}$ can be identified when Assumption 1.6 is satisfied.

We will prove that Assumption 1.5 with a finite $n$ and conditions in Proposition 1.1 (A) are sufficient to identify all parameters via the likelihood function in a finite sample.

**Corollary 1.1.** (I) Suppose Assumptions 1.1(2), 1.2, 1.3(2) and 1.4 are satisfied, and $A_n' \times (k_x + m)$ is nonsingular, then $\Psi_{m0}$ and $\Pi_{m0}$ can be identified. The covariance matrix $\Sigma_{vm0}$ can be identified when conditions in Proposition 1.1 (A) hold.

(II) Suppose Assumptions 1.1(2), 1.2, 1.3(2) and 1.4 are satisfied, conditions in Proposition 1.1 (B) hold, and $X_n$ has full column rank, then all the true parameters can be identified.

---

9 Note that when $W_n$ is symmetric, $\Psi_m$ might not be identified. The conditions to identify $\Psi_m$, when $W_n$ is symmetric, are $\Psi_m\Sigma_{vm0}^{* -1}\Psi_m' = \Psi_{m0}\Sigma_{vm0}^{* -1}\Psi_{m0}'$ and $\Psi_m\Sigma_{vm0}^{* -1} + \Sigma_{vm0}^{* -1}\Psi_m = \Psi_{m0}\Sigma_{vm0}^{* -1} + \Sigma_{vm0}^{* -1}\Psi_{m0}'$. Consider a case when $m = 2$. Suppose $\Sigma_{vm0}^{* -1} = I_2$, $\Psi_{m0} = \begin{pmatrix} 0.1 & -0.3 \\ 0.3 & 0.1 \end{pmatrix}$, both $\Psi_m$ and $\Psi_m = \begin{pmatrix} 0.1 & 0.3 \\ -0.3 & 0.1 \end{pmatrix}$ satisfy the conditions. Hence, we can not identify $\Psi_m$ in some special cases when $m > 1$.

10 There are several sufficient conditions for conditions in Proposition 1.1 (A), for example, “suppose $W_n$ is symmetric, $I_n$, $W_n$, and $W_n^2$ are linearly independent”, or, “suppose $W_n$ is asymmetric, $I_n$, $W_n$, $W_n'$ and $W_nW_n'$ are linearly independent”.

17
Some special MSAR models may be regarded as MSAR models with restrictions on parameters. Linear restrictions on coefficients of some explanatory variables are common among empirical models. Consider the restrictions in the form, vec \([(\Pi_m', \Psi'_m)'] = R_m\theta_l\), where \(\theta_l\) is a \(l \times 1\) vector representing “free” parameters in a MSAR model with \(l < m(k_x + m)\), and \(R_m\) is a \((m^2 + mk_x) \times l\) known matrix with full rank. The following corollary provides a formal statement of identification with linear restrictions on coefficients.

Corollary 1.2. Under Assumptions 1.1(2), 1.2, 1.3(2) and 1.4, with linear restrictions on \(\Pi_m\) and \(\Psi_m\), such that vec \([(\Pi_m', \Psi'_m)'] = R_m\theta_l\), an identification condition for \(\theta_l\) is that the limiting matrix \(\lim_{n \to +\infty} \frac{1}{n} R_m' \left[ \Sigma_{vm}^{-1} \otimes (A_n'(k_x+m), A_n'(k_x+m)) \right] R_m\) exists and is nonsingular.

As an example, consider \(Y_{n1} = W_n Y_{n2} \Psi_{21} + X_n \Pi_{m1} + V_{n1}\), and \(Y_{n2} = W_n Y_{n2} \Psi_{22} + X_n \Pi_{m2} + V_{n2}\), where \(m_1\) and \(m_2\) are, respectively, numbers of columns in \(Y_{n1}\) and \(Y_{n2}\), which represent the numbers of endogenous variables in the two subsystems. One interesting point is that the system does not include \(W_n Y_{n1}\). For this model, \(R_m = I_m \otimes \begin{pmatrix} I_{k_x} & 0_{k_x \times m_2} \\ 0_{m_1 \times k_x} & 0_{m_1 \times m_2} \\ 0_{m_2 \times k_x} & I_{m_2} \end{pmatrix}\). Then, \((I_m \otimes A_{m \times (m+k_x)}) R_m = I_m \otimes [X_n, \tilde{X}_{m1+1,n}, \ldots, \tilde{X}_{m,n}]\). Therefore, the identification condition is the existence and nonsingularity of \(\lim_{n \to \infty} \frac{1}{n} [X_n, \tilde{X}_{m1+1,n}, \ldots, \tilde{X}_{m,n}]' [X_n, \tilde{X}_{m1+1,n}, \ldots, \tilde{X}_{m,n}]\), which is a weaker condition than that in Assumption 1.5.

1.3.2 Identification of Structural Parameters of the SESAR Model

Let \(Z_{nm} = [Y_{nm}, W_n Y_{nm}, X_n]\) and \(\alpha_m = (\Gamma'_m, -A'_m, -C'_m)'\), the model (1.1) becomes \(Z_{nm} \alpha_m = U_{nm}\). For identification, we assume that \(\Psi_m\) and \(\Pi_m\) in its MSAR quasi-reduced form are identifiable. So it remains to consider the identification
of the structural parameters $\Gamma_m, \Lambda_m$ and $C_m$. By the linear relation, $\Psi_m \Gamma_m = \Lambda_m$ and $\Pi_m \Gamma_m = C_m$, the identification issue is similar algebraically to the classic identification of structural parameters from reduced-form parameters, e.g. in Schmidt (1976). For the SESAR model, the statement that structural parameters can be deducted from reduced form parameters means that the structural parameters vector is a unique solution to $(\Lambda_m', C_m') \Gamma_m^{-1} = (\Psi_m', \Pi_m')$ given $\Psi_m$ and $\Pi_m$, which is equivalent to $(\Psi_m', \Pi_m') \Gamma_m = (\Lambda_m', C_m')$, where $\Gamma_m$ is invertible with scale normalization that its diagonal elements are all ones. Let $(\Psi_m', \Pi_m')' = \theta_{\Psi, \Pi}$, the equation system for solution becomes $(\theta_{\Psi, \Pi}, I_m + k_x) \alpha_m = 0$.

We first consider the identification of parameters subject to exclusive restrictions.

Lemma 1.2. Suppose $\Phi$ is a $R \times (2m + k_x)$ matrix representing exclusive restrictions on the coefficients of the first structural equation such that $\Phi \alpha_{1,m} = 0$, and there is no restriction on the disturbance variance, then the sufficient and necessary rank condition for the identification of $\alpha_{1,m}$ is $\text{rank}(\Phi_{\alpha,m}) = m - 1$. And the necessary order condition for identification is $R \geq m - 1$.

Similar to that for the classic linear SEM, the order condition is in particular useful for counting exclusion restrictions of variables in a structural equation. The issue for the SESAR model is on the spatial lagged variables in the system. They are endogenous variables in statistical sense. However, as we demonstrate below by an example, without exclusion of exogenous regressors $X$’s but with exclusion of spatial lags (which are relevant for the whole system) in a structural equation, that structural equation may still be identifiable. For the purpose of counting excluded variables in a structural equation, it is appropriate to
regard spatial lags as ‘exogenous’ variables. This interpretation is possible because spatial lags bring in neighboring characteristics as extra exogenous variables for identification and estimation as in a single SAR model with relevant exogenous variables. However, without exogenous variables in the system, i.e., pure SESAR model, due to the structure of disturbances, parameter identification may still be possible even though spatial lags would no longer bring in any neighboring exogenous characteristics but neighboring disturbances. With exclusive restrictions, one may count the number of included endogenous variables (not including spatial lags) and the number of excluded spatial lags and exogenous regressors. The order condition is equivalent to the condition that the number of excluded exogenous variables and spatial lags is at least as large as the number of endogenous variables excluding the one with normalized coefficient being a unity in the structural equation.

Some empirical models utilize not only exclusive restrictions but also other linear restrictions including normalization to identify true parameters. The following lemma summarizes identification with rank conditions:

**Lemma 1.3.** Suppose $\Phi$ is a $R \times (2m+k_x)$ matrix representing linear restrictions (including normalization) on the coefficients of the first equation such that $\bar{\Phi}_{\alpha,1,m} = c$, where $c$ is a vector of constants, and there is no restriction on the disturbance variance, then the sufficient and necessary rank condition for the identification of $\alpha_{1,m}$ is $\text{rank}(\bar{\Phi}_{\alpha,m}) = m$.

Restrictions on disturbances may also help identification. Discussions on identification with restrictions on disturbances are collected in the supplementary file.
Since QML estimation is of interest, we should consider the identification of the whole system when linear constraints may also include across-equations restrictions in the system. We may summarize linear restrictions including cross equation ones (including normalization) into a matrix $\Psi$ such that $\Psi \vec{\alpha_m} = c$ where $c$ is a $R \times 1$ constant vector and $\Psi$ is a $R \times (2m^2 + k_x m)$ matrix.

**Lemma 1.4.** A necessary and sufficient condition to identify the whole system is $\text{rank}\{I_m \otimes \alpha_m\} = m^2$.

**Examples**

Two special groups of SESAR models are relatively simpler but useful in empirical studies. Their identification issues are investigated here.

**Example 1.** Here, $Y_{n1}$ consist of a dependent variable which can be explained by explanatory regressors $X$ and $Y_2$, where $Y_2$ is a vector of endogenous explanatory regressors subject to spatial interactions. Explicitly, the system consists of $Y_{n1} + Y_{n2} \Gamma_{21} = X_n C_{m1} + U_{n1}$ and $Y_{n2} \Gamma_{22} = W_n Y_{n2} \Lambda_{22} + X_n C_{m2} + U_{n2}$, where $Y_{n1}$ is an $n \times 1$ matrix, $Y_{n2}$ is an $n \times m_2$ matrix, $X_n$ is an $n \times k_x$ matrix, $\Gamma_{21}$ is an $m_2 \times 1$ matrix, $\Gamma_{22}$ is an invertible $m_2 \times m_2$ matrix with normalized diagonals of ones, $\Lambda_{22}$ is an $m_2 \times m_2$ matrix, $C_{m1}$ is a $k_x \times 1$ matrix and $C_{m2}$ is a $k_x \times m_2$ matrix. Here $m = 1 + m_2$. These equations can be put into a matrix form:

$$(Y_{n1}, Y_{n2}) \begin{pmatrix} 1 & 0_{1 \times m_2} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} = (W_n Y_{n1}, W_n Y_{n2}, X_n) \begin{pmatrix} 0 & 0_{1 \times m_2} \\ 0_{m_2 \times 1} & \Lambda_{22} \\ C_{m1} & C_{m2} \end{pmatrix} + (U_{n1}, U_{n2}).$$

Define $\alpha_m = \begin{pmatrix} 1 & 0_{1 \times m_2} \\ \Gamma_{21} & \Gamma_{22} \\ 0 & 0_{1 \times m_2} \\ 0_{m_2 \times 1} & -\Lambda_{22} \\ -C_{m1} & -C_{m2} \end{pmatrix}$. $\Phi_1 = \begin{pmatrix} 0 & 0_{1 \times m_2} & 1 & 0_{1 \times m_2} & 0_{1 \times k_x} \\ 0_{m_2 \times 1} & 0_{m_2 \times m_2} & 0_{m_2 \times 1} & I_{m_2} & 0_{m_2 \times k_x} \end{pmatrix}$ represents exclusive restrictions for the first equation. For the rank conditions,
with $\Phi_1 \alpha_m = \begin{pmatrix} 0_{m_1 \times 1} & 0_{m_1 \times m_2} \\ 0_{m_2 \times 1} & -\Lambda_{22} \end{pmatrix}$, if $\text{rank}(\Lambda_{22}) = m_2$, $\text{rank}(\Phi_1 \alpha_m) = m - 1$, so the rank condition is satisfied and the structural parameters in the first equation can be identified. The possible identification of the equation in $Y_{n1}$ is intuitively appealing because the second set of equations for $Y_{n2}$ in its reduced form is a MSAR, which can be identifiable, so there are valid IVs for $Y_{n2}$, which can be used for an IV estimation of the equation $Y_{n1}$.

Next, for the first equation from the second group,

$$\Phi_2 = \begin{pmatrix} 1 & 0_{1 \times m_2} & 0_{1 \times m_2} & 0_{1 \times k_x} \\ 0 & 1 & 0_{1 \times m_2} & 0_{1 \times k_x} \end{pmatrix}$$

with $\Phi_2 \alpha_m = \begin{pmatrix} 1 & 0_{1 \times m_2} \\ 0 & 0_{1 \times m_2} \end{pmatrix}$. This equation cannot be identified whenever $m_2 > 1$, since the rank condition is not satisfied. When $m_2 = 1$, $\text{rank}(\Phi_2 \alpha_m) = 1$ which equals $m - 1$ because $m - 1 = m_2 = 1$, so the rank condition would be satisfied. These situations are quite easy to be understood. If $m_2 = 1$, we have a univariate SAR model $Y_{n2} = \lambda_{22} W_n Y_{n2} + X_n C_{m_2} + U_{n2}$, which is identifiable. However, if $m_2 > 1$, we have the MSAR $Y_{n2} = W_n Y_{n2} \Lambda_{22} \Gamma_{22}^{-1} + X_n C_{m_2} \Gamma_{22}^{-1} + U_{n2} \Gamma_{22}^{-1}$, of which only the products of coefficient matrices are identifiable, but not the structural parameters separately.

For a similar equation system but without the presence of any exogenous $X_n$ in the equations, i.e., $Y_{n1} + Y_{n2} \Gamma_{21} = U_{n1}$, and $Y_{n2} \Gamma_{22} = W_n Y_{n2} \Lambda_{22} + U_{n2}$, the preceding identification analysis will also be valid such that $\Gamma_{21}$ can be identified in the first equation, but $\Gamma_{22}$ and $\Lambda_{22}$ in the second equation would be underidentified.

**Example 2.** Consider the system consisting of $Y_{n1} \Gamma_{11} + Y_{n2} \Gamma_{21} = W_n Y_{n1} \Lambda_{11} + X_{n1} C_{k_x,1 m_1} + U_{n1}$ and $Y_{n2} = X_{n1} C_{k_x,1 m_2} + X_{n2} C_{k_x,2 m_2} + U_{n2}$, where $Y_{n1}$ is an $n \times m_1$ matrix, $Y_{n2}$ is an $n \times m_2$ matrix, $X_{n1}$ is an $n \times k_{x,1}$ matrix, $X_{n2}$ is an $n \times k_{x,2}$ matrix,
\( \Gamma_{21} \) is an \( m_2 \times m_1 \) matrix, \( \Gamma_{11} \) is an \( m_1 \times m_1 \) invertible matrix with diagonal elements normalized to be ones, \( \Lambda_{11} \) is an \( m_1 \times m_1 \) matrix, \( C_{k_x,1m_1} \) is a \( k_x \times m_1 \) matrix, \( C_{k_x,1m_2} \) is a \( k_x \times m_2 \) matrix, and \( C_{k_x,2m_2} \) is a \( k_x \times m_2 \) matrix. Here, \( m = m_1 + m_2 \) and \( Y_1 \) is a SESAR system of dependent variables explained by regressors \( X_1 \) and endogenous regressors \( Y_2 \).

Each equation of \( Y_{n2} \) in the second group is a regression equation and hence can be identified as long as regressors in \( X_1 \) and \( X_2 \) are not linearly dependent, which is assumed. Consider the first equation in the first group. As \( Z_{nm} \) denotes [\( Y_{n1}, Y_{n2}, W_n Y_{n1}, W_n Y_{n2}, X_{n1}, X_{n2} \)] and, \( \alpha_m = \begin{pmatrix} \Gamma_{11} & 0_{m_1 \times m_2} \\ \Gamma_{21} & I_{m_2} \\ -\Lambda_{11} & 0_{m_1 \times m_2} \\ 0_{m_1 \times m_1} & 0_{m_1 \times m_2} \\ -C_{k_x,1m_1} & -C_{k_x,2m_2} \\ 0_{k_x,2 \times m_1} & -C_{k_x,2m_2} \end{pmatrix} \), the \( \Phi_1 = \begin{pmatrix} 0_{m_2 \times m_1} & 0_{m_2 \times m_2} & 0_{m_2 \times m_1} & I_{m_2} & 0_{m_2 \times k_x,1} & 0_{m_2 \times k_x,2} \\ 0_{k_x,2 \times m_1} & 0_{k_x,2 \times m_2} & 0_{k_x,2 \times m_1} & 0_{k_x,2 \times m_2} & 0_{k_x,2 \times k_x,1} & I_{k_x,2} \end{pmatrix} \) represents the exclusive restriction for that equation.

With \( \Phi_1 \alpha_m = \begin{pmatrix} 0_{m_2 \times 1} & 0_{m_2 \times (m_1 - 1)} & 0_{m_2 \times m_2} \\ 0_{k_x,2 \times 1} & 0_{k_x,2 \times (m_1 - 1)} & -C_{k_x,2m_2} \end{pmatrix} \). The rank condition requires that \( \text{rank}(C_{k_x,2m_2}) = m_1 + m_2 - 1 \), however, \( \text{rank}(C_{k_x,2m_2}) \leq m_2 \), so the structural parameters in the first equation cannot be identified unless \( m_1 = 1 \) and \( \text{rank}(C_{k_x,2m_2}) = m_2 \). When \( m_1 = 1 \), the first group consists of a single equation \( Y_{n1} = W_n Y_{n1} \Lambda_{11} - Y_{n2} \Gamma_{21} + X_n C_{k_x,1m_1} + U_{n1} \), which, with the second set of equations, implies \( Y_{n1} = W_n Y_{n1} \Lambda_{11} + X_n (C_{k_x,1m_1} - C_{k_x,1m_2} \Gamma_{21}) - X_n C_{k_x,2m_2} \Gamma_{21} + U_{n1} - U_{n2} \Gamma_{21} \). This is a univariate SAR equation, so \( \Lambda_{11} \) and the combined parameters \( (C_{k_x,1m_1} - C_{k_x,1m_2} \Gamma_{21}) \) and \( C_{k_x,2m_2} \Gamma_{21} \) are identifiable. As \( C_{k_x,1m_2} \) and \( C_{k_x,2m_2} \) are identifiable from the regression equations of \( Y_{n2} \), with full rank of \( C_{k_x,2m_2} \), the parameters \( C_{k_x,1m_1} \) and \( \Gamma_{21} \) are identifiable. The rank condition \( \text{rank}(C_{k_x,2m_2}) = m_2 \)
means that there are enough exclusion restrictions in \(Y_{n1}\) equation to find IV’s for \(Y_{n2}\). For \(m_1 > 1\), identification is not possible because the reduced form MSAR equations in the best can only identify, for example, the product \(\Lambda_{11}\Gamma_{11}^{-1}\).

1.4 Consistency and Asymptotic Normality of the QML

We propose a QML method for estimation of the SESAR model as well as a MSAR model. With constraints in the rank identification condition for the SESAR model, \(\theta\) will consist of free parameter in \(\text{vec}(\Gamma_m), \text{vec}(C_m), \text{vec}(\Lambda_m), \) and distinctive parameters of \(\Sigma_{um}\). For a general MSAR model, \(\theta\) will consist of free parameters in \(\text{vec}(\Psi_m), \text{vec}(\Phi_{km})\) and \(\Sigma_{vm}\). Without loss of generality, we shall focus on the estimation without explicitly stating constraints by setting. For example, \(\theta\) being a subvector of \((\text{vec}^*(\Gamma_m), \text{vec}(C_m)', \text{vec}(\Lambda_m)', \text{vec}^*(\Sigma_{vm})')'\), where \(\text{vec}^*(\Sigma_{vm})\) selects only free parameters from \(\text{vec}(\Sigma_{um})\), and has only \(m(m + 1)/2\) free parameters because the variance matrix \(\Sigma_{vm}\) must be symmetric. For \(\text{vec}^*(\Gamma_m)\), the diagonal entries are not included. The \(\theta\) is a at most \((k_x m + 2m^2 + m(m + 1)/2 - m)\) dimensional column vector. The \(\theta\) may also capture any other restrictions on the MSAR or SESAR model. The parameter space of \(\theta\) will be denoted by \(\Theta\).

In this section, we discuss and prove consistency and asymptotic normality of the QMLE. In order to derive consistency and asymptotic normality for the QMLE of the structural parameters, we make explicitly the identification condition.

Assumption 1.7. All structural parameters can be identified from the parameters in the quasi-reduced form. The quasi-reduced form of this SESAR model satisfies Assumptions 1.1(2), 1.2-1.6.
The following lemma shows that the difference of the average concentrated log likelihood function and the expected concentrated log likelihood function converges in probability to zero, uniformly in its parameter space.

**Lemma 1.5.** Under Assumptions 1.1 and 1.7, as \( n \to \infty \),

\[
\frac{1}{mn} \left[ \ln L_{nm}(\theta) - \mathbb{E} \ln L_{nm}(\theta) \right] \overset{p}{\to} 0 \text{ uniformly in } \theta \text{ in their compact parameter spaces } \Theta.
\]

The consistency of the QMLE is summarized in the following theorem.

**Theorem 1.1.** Under Assumption 1.1 and 1.7, the structural parameters in \( \Gamma_m \), \( \Lambda_m \), \( C_m \), and \( \Sigma_{um} \) with all the rank identification restrictions imposed, can be consistently estimated by the QML.

Asymptotic distribution of the QMLE will be useful for statistical inference. The asymptotic distribution of \( \hat{\theta} \) can be derived from the Taylor expansion of (1.4), given by

\[
\hat{\theta} - \theta_0 = - \left[ \frac{\partial^2 \ln L_{nm}(\theta)}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \theta}.
\]

The first step is to derive the asymptotic distribution of \( \frac{1}{\sqrt{n}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \theta} \). Let \( F_{m,ij} \) represent the \( m \times m \) matrix with zero entries except the \((i, j)\)th and \((j, i)\)th elements, which are one, so \( F_{m,ij} \) is symmetric. Also let \( E_{m,ij} \) represent a \( m \times m \) matrix with zero entries except the \((i, j)\)th element which is one. Specifically, let \( e_m \) represent the \( i \)th unit column vector of dimension \( m \), which is zero except the \( i \)th element which is one. Then,

\[
E_{m,ij} = e_{mi}e'_{mj}, \quad F_{m,ij} = E_{m,ij} + E'_{m,ij} = e_{mi}e'_{mj} + e_{mj}e'_{mi}, \quad \text{when } i \neq j, \text{ and } F_{m,ii} = E_{m,ii}.
\]

First order derivatives of the likelihood function with respect to the structural parameters are:

\[
\frac{\partial \ln L_{nm}(\theta_{m0})}{\partial C_{m,ij}} = e'_{(i-1)m+j} \left( \Sigma_{um0}^{-1} \otimes X'_n \right) \text{vec}(U_{nm}), \quad \text{for } i = 1, ..., k \text{ and } j = 1, ..., m
\]

\[
\frac{\partial \ln L_{nm}(\theta_{m0})}{\partial \Sigma_{um,ij}} = \frac{1}{2} \left[ \text{vec}(U_{nm})' \left[ \left( \Sigma_{um0}^{-1} F_{m,ij} \Sigma_{um0}^{-1} \right) \otimes I_n \right] \text{vec}(U_{nm}) \right]
\]
\[
\frac{\partial \ln L_{nm}(\theta_{m0})}{\partial \Lambda_{m,ij}} = -\text{vec}(C_{m0})'(I_m \otimes X_n')S^{-1}_{nm0,s}[(E_{m,ij}\Sigma^{-1}_{um0}) \otimes W_n'] \text{vec}(U_{nm}) \\
\quad + \left[\text{vec}(U_{nm})' \left[(\Sigma^{-1}_{um0}E'_{n,ij}) \otimes W_n \right] S^{-1}_{nm0,s} \text{vec}(U_{nm}) \right] - \\
\quad \text{Tr} \left[S^{-1}_{nm0,s} \left(E'_{n,ij} \otimes W_n \right) \right], \quad \text{and} \\
\frac{\partial \ln L_{nm}(\theta_{m0})}{\partial \Gamma_{m,ij}} = -\text{vec}(C_{m0})'(I_m \otimes X_n')S^{-1}_{nm0,s}[(E_{m,ij}\Sigma^{-1}_{um0}) \otimes I_n] \text{vec}(U_{nm}) \\
\quad -\left[\text{vec}(U_{nm})' \left[(\Sigma^{-1}_{um0}E'_{n,ij}) \otimes I_n \right] S^{-1}_{nm0,s} \text{vec}(U_{nm}) \right] - \text{Tr} \left[S^{-1}_{nm0,s} \left(E'_{n,ij} \otimes I_n \right) \right],
\]

for \(i, j = 1, \ldots, m\), where \(e_{(i-1)m+j}\) is a \(k_xm \times 1\) vector with all zeros except the \((i - 1)m + j\)th entry, which is one, and \(S_{nm,s} = \Gamma'_m \otimes I_n - \Lambda'_m \otimes W_n\). For a general MSAR model, \(\Gamma_m = I_m\), \(\Lambda_m = \Psi_m\) and \(C_m = \Pi_m\); so the derivatives with respect to \(\Gamma_m\) are irrelevant.

One observes that these first order derivatives have linear-quadratic forms. Characterized by the general form:

\[
B'_{nm} \text{vec}(V_{nm}) + \left[\text{vec}(V_{nm})'A_{nm} \text{vec}(V_{nm}) - \mathbb{E}(\text{vec}(V_{nm})'A_{nm} \text{vec}(V_{nm})) \right],
\]

where \(B_{nm}\) is a \(nm \times 1\) vector with uniformly bounded constants, and \(\{A_{nm}\}\) is a sequence of \(nm \times nm\) constant UB matrices. It differs from the linear-quadratic form in Kelejian and Prucha (1998) in that the elements of \(\text{vec}(V_{nm})\) are not independently distributed. Therefore, the central limit theorem with i.i.d. random variables for the linear-quadratic form in Kelejian and Prucha (1998) is not directly applicable. We need to extend the linear-quadratic central limit theorem to the multivariate case.\(^{11}\)

\(^{11}\)Qu and Lee (2002) provides an extension of linear-quadratic central limit theorem to a bivariate case.
Lemma 1.6. Suppose \( Q_n = \sum_{k=1}^{m} b'_{nk} V_{nk} + \sum_{k=1}^{m} \sum_{l=1}^{m} [V'_{nk} A_{nkl} V_{nl} - \mathbb{E}(V'_{nk} A_{nkl} V_{nl})] \),

where \( b_{nk} \) is a \( n \times 1 \) constant vector with uniformly bounded elements, \( A_{nkl} \) is a \( n \times n \) constant UB matrix, and \( V_{nk} \) and \( V_{nl} \) are \( n \times 1 \) random vectors with \( \mathbb{E}(V_{nk} V'_{nl}) = \sigma_{kl} I_n \) for any \( k,l = 1,2,\ldots,m \), which satisfy Assumption 1. Furthermore, suppose that the variance of \( Q_n \) is \( \sigma^2_{Q_n} \), which is \( O(n) \), and \( \frac{1}{n} \sigma^2_{Q_n} \) is bounded away from zero.

Then \( \frac{Q_n}{\sigma_{Q_n}} \xrightarrow{d} N(0,1) \).

This CLT shows that any linear combination of these first order derivatives that has a linear-quadratic form is asymptotically normal. Thus, the score vector is jointly normally distributed according to the Cramer-Wold device. The next step is to calculate the variance and covariance of the score vector. Lemma 1.7 provides the covariances of linear-quadratic forms, where the disturbances across equations can be correlated.

Lemma 1.7. For the two linear-quadratic forms

\[
Q_{n}^{(a,b)} = B'_{nm} \text{vec}(V_{nm}) + [\text{vec}(V_{nm})' A_{nm} \text{vec}(V_{nm}) - \mathbb{E}(\text{vec}(V_{nm})' A_{nm} \text{vec}(V_{nm}))]
\]

and \( Q_{n}^{(c,d)} = D'_{nm} \text{vec}(V_{nm}) + [\text{vec}(V_{nm})' C_{nm} \text{vec}(V_{nm}) - \mathbb{E}(\text{vec}(V_{nm})' C_{nm} \text{vec}(V_{nm}))] \),

where \( V_{nm} \), a generic \( n \times m \) matrix, satisfies Assumption 1(1), then

\[
\mathbb{E}[Q_{n}^{(a,b)} Q_{n}^{(c,d)}] = B'_{nm} (\Sigma_{vm} \otimes I_n) D_{nm} + \text{Tr}[A_{nm} (\Sigma_{vm} \otimes I_n) (C_{nm} + C'_{nm}) (\Sigma_{vm} \otimes I_n)]
\]

\[
+ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{i=1}^{n} \mu_{klp} \sum_{i=1}^{n} b_{nk,i} C_{nlp,ii} + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{i=1}^{n} \mu_{klp} \sum_{i=1}^{n} d_{nk,i} A_{nlp,ii}
\]

\[
+ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ (\mu_{klpq} - \sigma_{kq} \sigma_{lp} - \sigma_{kl} \sigma_{pq}) \left( \sum_{i=1}^{n} A_{nkl,ii} C_{npq,ii} \right) \right]
\]

where \( \mathbb{E}[\text{vec}(V_{nm}) \text{vec}(V_{nm})'] = \Sigma_{vm} \otimes I_n, \mu_{klp} = \mathbb{E}[v_{n,ik}v_{n,il}v_{n,ip}], \mu_{klpq} = \mathbb{E}[v_{n,ik}v_{n,il}v_{n,ip}v_{n,iq}], \)

\( A_{nkl,ii} \) denotes the \( i \)th element of \( A_{nkl} \) which is the \( (k,l) \)th block of \( A_{nm} \) and \( b_{nk,i} \) denotes the \( i \)th entry of vector \( b_{nk} \) which is the \( k \)th block of \( B_{nm} \), and similarly \( C_{nlp,ii} \) and \( d_{nk,i} \) are defined.
For our SESAR model, 
\[ E \left[ \left( \frac{1}{\sqrt{n}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \theta} \right) \left( \frac{1}{\sqrt{n}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \theta} \right)' \right] = \Omega_\theta + \Xi_\theta, \]
where the first part \( \Omega_\theta \) represents a typical information matrix under normality and the second part \( \Xi_\theta \) may contain high order moments of disturbances, when disturbances are not necessarily normal. The information matrix \( \Omega_\theta \) can be derived from the second order derivatives of log likelihood function via \( \Omega_\theta = -E \left( \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \theta \partial \theta'} \right) \).

Detailed expressions of these derivatives, \( \Omega_\theta \) and \( \Xi_\theta \) are in the appendix.

**Lemma 1.8.** Under Assumptions 1.1 and 1.7, for any consistent estimate \( \hat{\theta} \) of \( \theta_0 \),
\[
\frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\hat{\theta})}{\partial \theta \partial \theta'} \right] - E \left[ \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \theta \partial \theta'} \right] \xrightarrow{P} 0.
\]

Utilizing the limiting distribution of the score vector and the limiting matrix of the second derivatives, we arrive at the theorem on the asymptotic distribution of the QMLE.

**Theorem 1.2.** Under Assumptions 1.1 and 1.7, suppose the information matrix \( \Omega_\theta \) is nonsingular, \( \sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega^{-1}_\theta + \Omega^{-1}_\theta \Xi_\theta \Omega^{-1}_\theta) \)

### 1.5 Monte Carlo Experiments

Monte Carlo experiments are designed to investigate finite sample properties of the QMLE. We first look at MSAR models and then move on to SESAR models because the latter can be regarded as a MSAR with parameter constraints. In all our experiments, the spatial weight matrix \( W \) is generated from 760 counties in great plains states of USA. When two counties, \( i \) and \( j \), share border, \( w_{ij} = w_{ji} = 1 \). When sample size is \( n \), we randomly pick an \( n \times n \) block along the diagonal and row normalize the matrix.
1.5.1 Monte Carlo Experiments for MSAR Models

Three groups of experiments are conducted. The first group focus on the situation when 2SLS/3SLS cannot be applied. An example is a MSAR process without exogenous regressors.

Example 1: \( y_1 = \psi_{11}Wy_1 + \psi_{21}Wy_2 + u_1 \) and \( y_2 = \psi_{12}Wy_1 + \psi_{22}Wy_2 + u_2 \), where no exogenous regressors are present. This is a pure bivariate SAR system so they cannot be estimated by 2SLS or 3SLS. For this system, the disturbances \( u_1 \) and \( u_2 \) are normally distributed with mean 0 and variances \( \sigma_1^2 = \sigma_2^2 = 1 \) and their covariance \( \sigma_{12} = 0.5 \). The spatial effects matrix \( \Psi_m \) is set as \( \psi_{11} = \psi_{22} = 0.5 \) and \( \psi_{12} = \psi_{21} = 0.3 \).

Table 1 presents the Monte Carlo results when the sample sizes are 100, 300 and 500. We conduct 500 repetitions for each design that return empirical mean and standard deviation of the QMLEs. All the estimates are significant and the biases for \( \psi' \)'s are about 4% or less, even when the sample size is just 300. When the sample size increases, estimator’ bias decreases. Standard deviations decrease in the rate of \( \sqrt{n} \), which is consistent with the implication of Theorem 1.2.

The second group of Monte Carlo experiments is conducted for the following model.

Example 2: \( y_1 = \psi_{11}Wy_1 + \psi_{21}Wy_2 + \beta_1 x + u_1 \) and \( y_2 = \psi_{12}Wy_1 + \psi_{22}Wy_2 + \beta_2 x + u_2 \), where the true \( \beta_1 = \beta_2 = 1 \) and \( x \) is uniform distributed \( U(0, 2) \). As suggested by the existing literature\(^{12}\), we use \( Wx, W^2x \) and \( W^3x \) as instrument variables to do 2SLS. Results for 2SLS and QML are reported in table 2.

Table 1.1: Monte Carlo Results for Example 1

<table>
<thead>
<tr>
<th>Method</th>
<th>TRUE Size</th>
<th>QMLE</th>
<th>QMLE</th>
<th>QMLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
<td>300</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>$\psi_{11}$ 0.5</td>
<td>Mean</td>
<td>0.4917</td>
<td>0.4780</td>
<td>0.4892</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.2855)</td>
<td>(0.1813)</td>
<td>(0.1606)</td>
</tr>
<tr>
<td>$\psi_{21}$ 0.3</td>
<td>Mean</td>
<td>0.3130</td>
<td>0.2983</td>
<td>0.2928</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.2820)</td>
<td>(0.1808)</td>
<td>(0.1516)</td>
</tr>
<tr>
<td>$\psi_{12}$ 0.3</td>
<td>Mean</td>
<td>0.3299</td>
<td>0.2950</td>
<td>0.3004</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.3011)</td>
<td>(0.1745)</td>
<td>(0.1645)</td>
</tr>
<tr>
<td>$\psi_{22}$ 0.5</td>
<td>Mean</td>
<td>0.4782</td>
<td>0.4848</td>
<td>0.4862</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.2787)</td>
<td>(0.1707)</td>
<td>(0.1585)</td>
</tr>
<tr>
<td>$\sigma_{11}$ 1</td>
<td>Mean</td>
<td>0.9690</td>
<td>0.9914</td>
<td>0.9992</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1480)</td>
<td>(0.0813)</td>
<td>(0.0619)</td>
</tr>
<tr>
<td>$\sigma_{12}$ 0.5</td>
<td>Mean</td>
<td>0.4804</td>
<td>0.4947</td>
<td>0.4973</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1181)</td>
<td>(0.0611)</td>
<td>(0.0515)</td>
</tr>
<tr>
<td>$\sigma_{22}$ 1</td>
<td>Mean</td>
<td>0.9672</td>
<td>0.9904</td>
<td>0.9901</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1395)</td>
<td>(0.0779)</td>
<td>(0.0614)</td>
</tr>
</tbody>
</table>

Table 1.2: Monte Carlo Results for Example 2

<table>
<thead>
<tr>
<th>Method</th>
<th>QMLE</th>
<th>QMLE</th>
<th>QMLE</th>
<th>2SLS</th>
<th>2SLS</th>
<th>2SLS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>Size</td>
<td>100</td>
<td>300</td>
<td>500</td>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>$\psi_{11}$ 0.5</td>
<td>Mean</td>
<td>0.4861</td>
<td>0.4939</td>
<td>0.4942</td>
<td>0.4946</td>
<td>0.4942</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1566)</td>
<td>(0.1037)</td>
<td>(0.0895)</td>
<td>(2.5522)</td>
<td>(2.6320)</td>
</tr>
<tr>
<td>$\psi_{21}$ 0.3</td>
<td>Mean</td>
<td>0.3102</td>
<td>0.3044</td>
<td>0.3043</td>
<td>0.3168</td>
<td>0.3049</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1646)</td>
<td>(0.1092)</td>
<td>(0.0929)</td>
<td>(2.5164)</td>
<td>(2.5791)</td>
</tr>
<tr>
<td>$\psi_{12}$ 0.3</td>
<td>Mean</td>
<td>0.3155</td>
<td>0.3040</td>
<td>0.3020</td>
<td>0.4174</td>
<td>0.4852</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1680)</td>
<td>(0.1055)</td>
<td>(0.0924)</td>
<td>(2.2030)</td>
<td>(2.4126)</td>
</tr>
<tr>
<td>$\psi_{22}$ 0.5</td>
<td>Mean</td>
<td>0.4806</td>
<td>0.4945</td>
<td>0.4955</td>
<td>0.3639</td>
<td>0.3151</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1613)</td>
<td>(0.1010)</td>
<td>(0.0904)</td>
<td>(2.1293)</td>
<td>(2.3996)</td>
</tr>
<tr>
<td>$\beta_{1}$ 1</td>
<td>Mean</td>
<td>1.0164</td>
<td>1.0089</td>
<td>1.0074</td>
<td>0.9990</td>
<td>1.0057</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1552)</td>
<td>(0.0910)</td>
<td>(0.0771)</td>
<td>(0.1421)</td>
<td>(0.0847)</td>
</tr>
<tr>
<td>$\beta_{2}$ 1</td>
<td>Mean</td>
<td>1.0225</td>
<td>1.0048</td>
<td>1.0105</td>
<td>1.0042</td>
<td>1.0020</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1508)</td>
<td>(0.0987)</td>
<td>(0.0767)</td>
<td>(0.1390)</td>
<td>(0.0844)</td>
</tr>
<tr>
<td>$\sigma_{11}$ 1</td>
<td>Mean</td>
<td>0.9652</td>
<td>0.9895</td>
<td>0.9981</td>
<td>1.9413</td>
<td>1.8615</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1515)</td>
<td>(0.0829)</td>
<td>(0.0624)</td>
<td>(4.6898)</td>
<td>(5.4109)</td>
</tr>
<tr>
<td>$\sigma_{12}$ 0.5</td>
<td>Mean</td>
<td>0.4814</td>
<td>0.4939</td>
<td>0.4975</td>
<td>0.8084</td>
<td>0.9621</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1210)</td>
<td>(0.0624)</td>
<td>(0.0523)</td>
<td>(1.5044)</td>
<td>(3.2092)</td>
</tr>
<tr>
<td>$\sigma_{22}$ 1</td>
<td>Mean</td>
<td>0.9634</td>
<td>0.9878</td>
<td>0.9894</td>
<td>1.6716</td>
<td>1.7343</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1412)</td>
<td>(0.0779)</td>
<td>(0.0617)</td>
<td>(2.3701)</td>
<td>(3.5218)</td>
</tr>
</tbody>
</table>
The result suggests that the conventional IV (2SLS) method has large standard deviations for the 2SLSE of the spatial effects parameters though the exogenous effects can be estimated much more precisely and with small biases. On the other hand, the QML has much smaller biases and standard deviations for the estimates of the spatial effects parameters, compared to those of the 2SLS. The standard deviations of the 2SLSEs of the spatial effects can be 15 to 30 times larger than those of the QMLEs.

The next example considers two functions with different exogenous variables. 

Example 3: \( y_1 = \psi_{11} W y_1 + \psi_{21} W y_2 + \beta_1 x_1 + u_1 \) and \( y_2 = \psi_{12} W y_1 + \psi_{22} W y_2 + \beta_2 x_2 + u_2 \), where \( \beta_1 = \beta_2 = 1 \). \( x_1 \) and \( x_2 \) are generated from uniform distribution \( U(0, 2) \). Here we use \( W x, W^2 x \) and \( W^3 x \) for both \( X_1 \) and \( X_2 \) as instrument variables in the 2SLS estimation.

The results in Table 3 suggest that the biases of 2SLSEs, 3SLSEs and QMLEs are small and the estimates are significant for all spatial effects and exogenous effects parameters. Comparing the precision of the estimates, while there are not much differences in the estimates of the \( \beta \)'s, the standard deviations of the QMLEs of the spatial effects are much smaller than those of the 2SLSE and 3SLSE. QMLEs show substantial efficiency improvement compared to 2SLSEs and 3SLSE as their standard deviations are only about 60% those of the 2SLSEs and 3SLSEs.

In Appendix B we derive the optimal IVs for endogenous variables. Table 4 demonstrates the results. In finite sample cases, the 2SLSEs and 3SLSEs with optimal IVs do not improve the efficiency much: the standard deviations are mostly similar to those of 2SLSEs and 3SLSEs in Table 3.
### Table 1.3: Monte Carlo Results for Example 3

<table>
<thead>
<tr>
<th>Method</th>
<th>QMLE 100</th>
<th>QMLE 300</th>
<th>QMLE 500</th>
<th>2SLS 100</th>
<th>2SLS 300</th>
<th>2SLS 500</th>
<th>3SLS 100</th>
<th>3SLS 300</th>
<th>3SLS 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi_{11}$ 0.5</td>
<td>Mean</td>
<td>0.4891</td>
<td>0.4986</td>
<td>0.4966</td>
<td>0.4972</td>
<td>0.5045</td>
<td>0.5007</td>
<td>0.4965</td>
<td>0.5034</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1236)</td>
<td>(0.0763)</td>
<td>(0.0690)</td>
<td>(0.1926)</td>
<td>(0.1205)</td>
<td>(0.1079)</td>
<td>(0.1910)</td>
<td>(0.1202)</td>
</tr>
<tr>
<td>$\psi_{21}$ 0.3</td>
<td>Mean</td>
<td>0.3085</td>
<td>0.2996</td>
<td>0.3026</td>
<td>0.3048</td>
<td>0.2967</td>
<td>0.3000</td>
<td>0.3060</td>
<td>0.2973</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1340)</td>
<td>(0.0811)</td>
<td>(0.0717)</td>
<td>(0.1987)</td>
<td>(0.1195)</td>
<td>(0.1066)</td>
<td>(0.1966)</td>
<td>(0.1198)</td>
</tr>
<tr>
<td>$\psi_{12}$ 0.3</td>
<td>Mean</td>
<td>0.3152</td>
<td>0.3030</td>
<td>0.3019</td>
<td>0.3093</td>
<td>0.3021</td>
<td>0.2967</td>
<td>0.3103</td>
<td>0.3031</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1291)</td>
<td>(0.0813)</td>
<td>(0.0686)</td>
<td>(0.1921)</td>
<td>(0.1305)</td>
<td>(0.1054)</td>
<td>(0.1918)</td>
<td>(0.1303)</td>
</tr>
<tr>
<td>$\psi_{22}$ 0.5</td>
<td>Mean</td>
<td>0.4831</td>
<td>0.4960</td>
<td>0.4975</td>
<td>0.4946</td>
<td>0.4980</td>
<td>0.5047</td>
<td>0.4936</td>
<td>0.4977</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1271)</td>
<td>(0.0774)</td>
<td>(0.0664)</td>
<td>(0.2028)</td>
<td>(0.1300)</td>
<td>(0.1068)</td>
<td>(0.2015)</td>
<td>(0.1297)</td>
</tr>
<tr>
<td>$\beta_1$ 1</td>
<td>Mean</td>
<td>1.0076</td>
<td>1.0080</td>
<td>1.0028</td>
<td>0.9914</td>
<td>0.9953</td>
<td>0.9976</td>
<td>0.9894</td>
<td>0.9976</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1380)</td>
<td>(0.0798)</td>
<td>(0.0659)</td>
<td>(0.1717)</td>
<td>(0.0987)</td>
<td>(0.0774)</td>
<td>(0.1563)</td>
<td>(0.0869)</td>
</tr>
<tr>
<td>$\beta_2$ 1</td>
<td>Mean</td>
<td>1.0047</td>
<td>1.0021</td>
<td>0.9999</td>
<td>0.9816</td>
<td>0.9986</td>
<td>0.9918</td>
<td>0.9827</td>
<td>0.9954</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1346)</td>
<td>(0.0831)</td>
<td>(0.0686)</td>
<td>(0.1721)</td>
<td>(0.1038)</td>
<td>(0.0824)</td>
<td>(0.1465)</td>
<td>(0.0901)</td>
</tr>
<tr>
<td>$\sigma_{11}$ 1</td>
<td>Mean</td>
<td>0.9649</td>
<td>0.9905</td>
<td>0.9882</td>
<td>0.9734</td>
<td>0.9931</td>
<td>0.9998</td>
<td>0.9751</td>
<td>0.9943</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1511)</td>
<td>(0.0828)</td>
<td>(0.0625)</td>
<td>(0.1531)</td>
<td>(0.0839)</td>
<td>(0.0634)</td>
<td>(0.1536)</td>
<td>(0.0842)</td>
</tr>
<tr>
<td>$\sigma_{12}$ 0.5</td>
<td>Mean</td>
<td>0.4830</td>
<td>0.4956</td>
<td>0.4978</td>
<td>0.4761</td>
<td>0.4934</td>
<td>0.4966</td>
<td>0.4857</td>
<td>0.4968</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1214)</td>
<td>(0.0621)</td>
<td>(0.0519)</td>
<td>(0.1229)</td>
<td>(0.0623)</td>
<td>(0.0522)</td>
<td>(0.1237)</td>
<td>(0.0626)</td>
</tr>
<tr>
<td>$\sigma_{22}$ 1</td>
<td>Mean</td>
<td>0.9621</td>
<td>0.9884</td>
<td>0.9891</td>
<td>0.9701</td>
<td>0.9927</td>
<td>0.9907</td>
<td>0.9719</td>
<td>0.9930</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1408)</td>
<td>(0.0779)</td>
<td>(0.0619)</td>
<td>(0.1470)</td>
<td>(0.0805)</td>
<td>(0.0624)</td>
<td>(0.1463)</td>
<td>(0.0797)</td>
</tr>
<tr>
<td>Method</td>
<td>B2SLS</td>
<td>B2SLS</td>
<td>B2SLS</td>
<td>B3SLS</td>
<td>B3SLS</td>
<td>B3SLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRUE</td>
<td>Size</td>
<td>100</td>
<td>300</td>
<td>500</td>
<td>100</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.5</td>
<td>Mean</td>
<td>0.4958</td>
<td>0.5084</td>
<td>0.4964</td>
<td>0.4955</td>
<td>0.5073</td>
<td>0.4971</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.2011)</td>
<td>(0.1239)</td>
<td>(0.1100)</td>
<td>(0.2000)</td>
<td>(0.1238)</td>
<td>(0.1094)</td>
<td></td>
</tr>
<tr>
<td>$\psi_{21}$</td>
<td>0.3</td>
<td>Mean</td>
<td>0.3035</td>
<td>0.2922</td>
<td>0.3036</td>
<td>0.3054</td>
<td>0.2930</td>
<td>0.3037</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.2098)</td>
<td>(0.1229)</td>
<td>(0.1087)</td>
<td>(0.2078)</td>
<td>(0.1233)</td>
<td>(0.1087)</td>
<td></td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0.3</td>
<td>Mean</td>
<td>0.3130</td>
<td>0.2975</td>
<td>0.2988</td>
<td>0.3141</td>
<td>0.2988</td>
<td>0.2988</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.1944)</td>
<td>(0.1338)</td>
<td>(0.1060)</td>
<td>(0.1946)</td>
<td>(0.1335)</td>
<td>(0.1059)</td>
<td></td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>0.5</td>
<td>Mean</td>
<td>0.4883</td>
<td>0.5022</td>
<td>0.5019</td>
<td>0.4881</td>
<td>0.5018</td>
<td>0.5019</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.2044)</td>
<td>(0.1334)</td>
<td>(0.1073)</td>
<td>(0.2041)</td>
<td>(0.1329)</td>
<td>(0.1073)</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1</td>
<td>Mean</td>
<td>1.0007</td>
<td>0.9974</td>
<td>1.0000</td>
<td>0.9931</td>
<td>0.9987</td>
<td>0.9965</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.1719)</td>
<td>(0.0980)</td>
<td>(0.0777)</td>
<td>(0.1569)</td>
<td>(0.0865)</td>
<td>(0.0683)</td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>Mean</td>
<td>0.9907</td>
<td>1.0001</td>
<td>0.9943</td>
<td>0.9870</td>
<td>0.9959</td>
<td>0.9945</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.1702)</td>
<td>(0.1037)</td>
<td>(0.0822)</td>
<td>(0.1458)</td>
<td>(0.0901)</td>
<td>(0.0727)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>1</td>
<td>Mean</td>
<td>0.9777</td>
<td>0.9938</td>
<td>1.0004</td>
<td>0.9784</td>
<td>0.9949</td>
<td>1.0005</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.1544)</td>
<td>(0.0838)</td>
<td>(0.0636)</td>
<td>(0.1545)</td>
<td>(0.0840)</td>
<td>(0.0634)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.5</td>
<td>Mean</td>
<td>0.4789</td>
<td>0.4943</td>
<td>0.4968</td>
<td>0.4876</td>
<td>0.4976</td>
<td>0.4987</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.1247)</td>
<td>(0.0625)</td>
<td>(0.0522)</td>
<td>(0.1252)</td>
<td>(0.0628)</td>
<td>(0.0524)</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>1</td>
<td>Mean</td>
<td>0.9743</td>
<td>0.9934</td>
<td>0.9911</td>
<td>0.9757</td>
<td>0.9936</td>
<td>0.9916</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>(0.1480)</td>
<td>(0.0817)</td>
<td>(0.0624)</td>
<td>(0.1472)</td>
<td>(0.0807)</td>
<td>(0.0625)</td>
<td></td>
</tr>
</tbody>
</table>
To investigate the robustness of estimators with non-normally distributed disturbances, we generate disturbances using non-normal distributions in the current example. We first generate two independent uniformly distributed random variables with mean zero and variance one, then multiply the vector of the two uniform random variables by a constant matrix $C$, where $CC' = \Sigma_u = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, so that the resulted disturbances $u_1$ and $u_2$ have the same variance-covariance matrix of the previous designs. The estimated results are presented in Table 5. The results are similar to those when the disturbances are normally distributed.

### Table 1.5: Monte Carlo for Example 3 with non-normal distributed disturbance

<table>
<thead>
<tr>
<th>Method</th>
<th>Size</th>
<th>QMLE Mean</th>
<th>QMLE Mean</th>
<th>QMLE Mean</th>
<th>2SLS Mean</th>
<th>2SLS Mean</th>
<th>2SLS Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>100</td>
<td>0.4911</td>
<td>0.4920</td>
<td>0.4957</td>
<td>0.5010</td>
<td>0.4836</td>
<td>0.4983</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.1257</td>
<td>0.0816</td>
<td>0.0661</td>
<td>0.1873</td>
<td>0.1311</td>
<td>0.1070</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.5</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>$\psi_{21}$</td>
<td>0.3</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0.3</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>0.5</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>1</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.5</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>1</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
</tbody>
</table>

34
1.5.2 Monte Carlo Experiments for simultaneous equations Models

We also conduct Monte Carlo experiments for SESAR models to investigate the finite sample properties of the QMLE.

**Example 4:**

\[ y_1 = \gamma_1 y_2 + \lambda_{11} W y_1 + \lambda_{21} W y_2 + \beta_1 x_1 + u_1 \quad \text{and} \quad y_2 = \gamma_2 y_1 + \lambda_{12} W y_1 + \lambda_{22} W y_2 + \beta_2 x_2 + u_2. \]

For this system, the disturbances \( u_1 \) and \( u_2 \) are normally distributed with mean 0 and their variances \( \sigma^2_1 = \sigma^2_2 = 1 \) and covariance \( \sigma_{12} = 0.5 \). The simultaneity effects are set as \( \gamma_1 = \gamma_2 = -0.5 \). The spatial effects matrix \( \Lambda_m \) has \( \lambda_{11} = \lambda_{22} = 0.5 \) and \( \lambda_{12} = \lambda_{21} = 0.3 \). Exogenous effects are \( \beta_1 = \beta_2 = 1 \), where both \( x_1 \) and \( x_2 \) are uniform distributed \( U(0, 3) \) as before. Due to exclusion restrictions of exogenous variables, both structural equations satisfy their rank conditions for identification. Again, as suggested by the existed literature\(^{13}\), we also use \( W x_i, W^2 x_i \) and \( W^3 x_i \), for \( i = 1, 2 \), as instrument variables in 2SLS. Results for 2SLSEs, 3SLSEs and QMLEs are reported in table 6.

Table 6 presents the Monte Carlo results when the sample sizes are 100, 300 and 500. We conduct 500 repetitions for each design which return empirical mean and standard deviations of the QMLEs. The 2SLSEs, 3SLSEs and QMLEs return significant simultaneity effects, spatial effects and exogenous effects parameters, while QMLEs present efficiency improvement compared to 2SLSEs and 3SLSEs. The reduction of standard deviations in the QMLEs for the spatial effects is about one third relative to those of the 2SLSEs and 3SLSEs, and the reduction of standard deviations of the QMLEs of the simultaneity effect parameters can be

15%. Again, utilizing optimal IVs cannot increase efficiency for those estimators in finite samples (see Table 7).

In sum, Monte Carlo experiments illustrate that the QML estimation has two advantages. First, QML estimation can be applied to MSAR models without exogenous variables, while such models cannot be estimated by IV based methods. The second advantage is when there are exogenous variables in the model so that IV approach is feasible, the QMLEs can be much more efficient than the 2SLSEs and 3SLSEs.

1.6 Conclusion and Remarks

In this paper, we consider the identification and estimation of MSAR and SESAR models. The MSAR model consists of several equations, which incorporate own-variable spatial lags and cross-variable spatial lags as explanatory variables, and allow for correlation between disturbances cross equations. The SESAR model, in addition to incorporating spatial lags, includes also endogenous dependent variables as explanatory variables.

We first investigate identification conditions for SESAR and MSAR models. For the SESAR model, we derive rank identification conditions by treating an implied MSAR model as its quasi-reduced form equations. The QML is proposed with special attention on the computation of the determinant of the Jacobian transformation for the model. The consistency of the estimator is derived. We also study the asymptotic distribution of the QMLE. Monte Carlo experiments illustrate the advantages of QML estimation with respect to IV methods. For models
<table>
<thead>
<tr>
<th>Method</th>
<th>QMLE 100</th>
<th>QMLE 300</th>
<th>QMLE 500</th>
<th>2SLS 100</th>
<th>2SLS 300</th>
<th>2SLS 500</th>
<th>3SLS 100</th>
<th>3SLS 300</th>
<th>3SLS 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRUE</td>
<td>Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>-0.5</td>
<td>Mean</td>
<td>-0.4869</td>
<td>-0.4985</td>
<td>-0.4923</td>
<td>-0.5102</td>
<td>-0.4956</td>
<td>-0.5007</td>
<td>-0.4979</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1121)</td>
<td>(0.0736)</td>
<td>(0.0546)</td>
<td>(0.1344)</td>
<td>(0.0774)</td>
<td>(0.0588)</td>
<td>(0.1370)</td>
<td>(0.0781)</td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>-0.5</td>
<td>Mean</td>
<td>-0.4975</td>
<td>-0.4989</td>
<td>-0.4962</td>
<td>-0.4991</td>
<td>-0.5015</td>
<td>-0.4962</td>
<td>-0.4848</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1188)</td>
<td>(0.0703)</td>
<td>(0.0552)</td>
<td>(0.1247)</td>
<td>(0.0792)</td>
<td>(0.0606)</td>
<td>(0.1292)</td>
<td>(0.0801)</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.5</td>
<td>Mean</td>
<td>0.4884</td>
<td>0.4917</td>
<td>0.4933</td>
<td>0.5061</td>
<td>0.5006</td>
<td>0.5020</td>
<td>0.5025</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.0775)</td>
<td>(0.0468)</td>
<td>(0.0422)</td>
<td>(0.1237)</td>
<td>(0.0789)</td>
<td>(0.0637)</td>
<td>(0.1256)</td>
<td>(0.0796)</td>
</tr>
<tr>
<td>$\psi_{21}$</td>
<td>0.3</td>
<td>Mean</td>
<td>0.2930</td>
<td>0.2978</td>
<td>0.2926</td>
<td>0.3080</td>
<td>0.3025</td>
<td>0.2972</td>
<td>0.2943</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1282)</td>
<td>(0.0809)</td>
<td>(0.0659)</td>
<td>(0.1552)</td>
<td>(0.1008)</td>
<td>(0.0811)</td>
<td>(0.1587)</td>
<td>(0.1014)</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0.3</td>
<td>Mean</td>
<td>0.2995</td>
<td>0.2879</td>
<td>0.2954</td>
<td>0.3139</td>
<td>0.2958</td>
<td>0.3010</td>
<td>0.3027</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1395)</td>
<td>(0.0842)</td>
<td>(0.0692)</td>
<td>(0.1906)</td>
<td>(0.0992)</td>
<td>(0.0814)</td>
<td>(0.1933)</td>
<td>(0.0997)</td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>0.5</td>
<td>Mean</td>
<td>0.4834</td>
<td>0.4928</td>
<td>0.4935</td>
<td>0.5056</td>
<td>0.4989</td>
<td>0.5017</td>
<td>0.5004</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.0815)</td>
<td>(0.0468)</td>
<td>(0.0420)</td>
<td>(0.1518)</td>
<td>(0.0786)</td>
<td>(0.0661)</td>
<td>(0.1549)</td>
<td>(0.0793)</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1</td>
<td>Mean</td>
<td>1.0155</td>
<td>1.0415</td>
<td>1.0105</td>
<td>0.9856</td>
<td>0.9998</td>
<td>0.9973</td>
<td>0.9899</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1450)</td>
<td>(0.0887)</td>
<td>(0.0757)</td>
<td>(0.2004)</td>
<td>(0.1116)</td>
<td>(0.0879)</td>
<td>(0.2040)</td>
<td>(0.1124)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>Mean</td>
<td>1.0169</td>
<td>1.0100</td>
<td>1.0066</td>
<td>0.9824</td>
<td>0.9993</td>
<td>0.9946</td>
<td>0.9884</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1527)</td>
<td>(0.0867)</td>
<td>(0.0749)</td>
<td>(0.1997)</td>
<td>(0.1110)</td>
<td>(0.0911)</td>
<td>(0.2019)</td>
<td>(0.1112)</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>1</td>
<td>Mean</td>
<td>0.9683</td>
<td>0.9935</td>
<td>1.0010</td>
<td>2.0536</td>
<td>2.0661</td>
<td>2.0918</td>
<td>2.0307</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1522)</td>
<td>(0.0847)</td>
<td>(0.0623)</td>
<td>(0.5100)</td>
<td>(0.2719)</td>
<td>(0.2109)</td>
<td>(0.5059)</td>
<td>(0.2707)</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.5</td>
<td>Mean</td>
<td>0.4603</td>
<td>0.4822</td>
<td>0.4854</td>
<td>-0.4504</td>
<td>-0.4051</td>
<td>-0.3734</td>
<td>-0.4648</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1752)</td>
<td>(0.1020)</td>
<td>(0.0882)</td>
<td>(0.2050)</td>
<td>(0.1144)</td>
<td>(0.0929)</td>
<td>(0.2104)</td>
<td>(0.1161)</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>1</td>
<td>Mean</td>
<td>0.9689</td>
<td>0.9920</td>
<td>0.9911</td>
<td>2.3522</td>
<td>2.0959</td>
<td>2.0870</td>
<td>2.3097</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>(0.1436)</td>
<td>(0.0783)</td>
<td>(0.0622)</td>
<td>(0.5699)</td>
<td>(0.2895)</td>
<td>(0.2307)</td>
<td>(0.5560)</td>
<td>(0.2877)</td>
</tr>
<tr>
<td>Method</td>
<td>TRUE Size</td>
<td>B2SLS 100</td>
<td>B2SLS 300</td>
<td>B2SLS 500</td>
<td>B3SLS 100</td>
<td>B3SLS 300</td>
<td>B3SLS 500</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma_1 )</td>
<td>-0.5 Mean</td>
<td>-0.5002</td>
<td>-0.4944</td>
<td>-0.5001</td>
<td>-0.5044</td>
<td>-0.4999</td>
<td>-0.4981</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.1393)</td>
<td>(0.0776)</td>
<td>(0.0597)</td>
<td>(0.1346)</td>
<td>(0.0742)</td>
<td>(0.0623)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \gamma_2 )</td>
<td>-0.5 Mean</td>
<td>-0.4895</td>
<td>-0.5022</td>
<td>-0.4950</td>
<td>-0.5122</td>
<td>-0.4977</td>
<td>-0.4974</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.1275)</td>
<td>(0.0814)</td>
<td>(0.0608)</td>
<td>(0.1303)</td>
<td>(0.0827)</td>
<td>(0.0608)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \psi_{11} )</td>
<td>0.5 Mean</td>
<td>0.4951</td>
<td>0.5000</td>
<td>0.4984</td>
<td>0.5043</td>
<td>0.4987</td>
<td>0.4945</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.1234)</td>
<td>(0.0795)</td>
<td>(0.0648)</td>
<td>(0.1241)</td>
<td>(0.0801)</td>
<td>(0.0614)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \psi_{21} )</td>
<td>0.3 Mean</td>
<td>0.3003</td>
<td>0.3026</td>
<td>0.2965</td>
<td>0.3198</td>
<td>0.3058</td>
<td>0.2984</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.1559)</td>
<td>(0.1043)</td>
<td>(0.0819)</td>
<td>(0.1616)</td>
<td>(0.1054)</td>
<td>(0.0849)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \psi_{12} )</td>
<td>0.3 Mean</td>
<td>0.3041</td>
<td>0.2941</td>
<td>0.3015</td>
<td>0.3096</td>
<td>0.3030</td>
<td>0.3003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.1987)</td>
<td>(0.1003)</td>
<td>(0.0829)</td>
<td>(0.1955)</td>
<td>(0.1054)</td>
<td>(0.0859)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \psi_{22} )</td>
<td>0.5 Mean</td>
<td>0.4885</td>
<td>0.4996</td>
<td>0.4983</td>
<td>0.4973</td>
<td>0.4917</td>
<td>0.4941</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.1542)</td>
<td>(0.0792)</td>
<td>(0.0660)</td>
<td>(0.1559)</td>
<td>(0.0848)</td>
<td>(0.0696)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>1 Mean</td>
<td>1.0002</td>
<td>1.0010</td>
<td>1.0002</td>
<td>0.9825</td>
<td>0.9995</td>
<td>1.0007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.1982)</td>
<td>(0.1111)</td>
<td>(0.0886)</td>
<td>(0.1940)</td>
<td>(0.1185)</td>
<td>(0.0847)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>1 Mean</td>
<td>0.9980</td>
<td>0.9988</td>
<td>0.9979</td>
<td>0.9817</td>
<td>0.9997</td>
<td>1.0047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.1972)</td>
<td>(0.1116)</td>
<td>(0.0912)</td>
<td>(0.1951)</td>
<td>(0.1129)</td>
<td>(0.0892)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_{11} )</td>
<td>1 Mean</td>
<td>2.0369</td>
<td>2.0630</td>
<td>2.0915</td>
<td>2.0593</td>
<td>2.0835</td>
<td>2.0859</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.5115)</td>
<td>(0.2709)</td>
<td>(0.2132)</td>
<td>(0.4837)</td>
<td>(0.2644)</td>
<td>(0.2211)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_{12} )</td>
<td>0.5 Mean</td>
<td>-0.4706</td>
<td>-0.4049</td>
<td>-0.3764</td>
<td>-0.4668</td>
<td>-0.4057</td>
<td>-0.3789</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.2075)</td>
<td>(0.1161)</td>
<td>(0.0929)</td>
<td>(0.2289)</td>
<td>(0.1253)</td>
<td>(0.0926)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \sigma_{22} )</td>
<td>1 Mean</td>
<td>2.3277</td>
<td>2.0989</td>
<td>2.0841</td>
<td>2.4450</td>
<td>2.0945</td>
<td>2.0946</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.5685)</td>
<td>(0.2964)</td>
<td>(0.2307)</td>
<td>(0.6358)</td>
<td>(0.2956)</td>
<td>(0.2222)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
without exogenous variables, only the QML estimation is feasible but not IV approach. Relative to the 2SLS and 3SLS estimators, the QMLE has smaller biases and is much more efficient than the 2SLS and 3SLS estimators. The Monte Carlo experiments present efficiency improvement compared to 2SLS and 3SLS estimates in finite samples.

The consistency of the QML for a SAR model is based on homoskedastic variance of disturbances in the model. In the presence of unknown heteroskedastic variances, the QML would be inconsistent due to its use of the quadratic moment condition which explores homoskedastic variances of disturbances. On the other hand, the 2SLS uses only linear moment conditions and so can be consistent even variances of disturbances are heteroskedastic. In that case, it would be of interest to consider an extension of the 2SLS estimation by a GMM estimation method which explores the use of proper quadratic moments in addition to linear moments. The GMM estimation might also be computationally simpler as it does not involve the evaluation of the determinant of a Jacobian transformation. For the estimation of a univariate SAR model, such a GMM method has been considered in Lee (2007). For future research, it will be of interest to extend it for the estimation of SESAR model under unknown heteroskedasticity.

SESAR and MSAR models help in understanding own-variable spatial effects and cross-variable spatial effects\textsuperscript{14} in addition to direct simultaneous interrelated effects of dependent variables. Extended models could include spatial lags in disturbance terms. It is also interesting to study nonlinear MSAR models such

\textsuperscript{14}In Cohen-cole (2013), cross-variable spatial effects are referred to cross-activity peer effects.
as models with truncated dependent variables and semiparametric or nonparametric estimation for them. In addition, dynamic models with both spatial and time lags are of interest to be investigated in future research.
Chapter 2: Multivariate and Simultaneous Equations Dynamic Panel Spatial Autoregressive Models

2.1 Introduction

Panel databases with cross-sectional dependent variables have highlighted the need for new data analysis techniques to model dependence patterns within these databases. The vector autoregressive model has been proven to be useful for describing dynamic behaviors of economic variables. However, the number of dependent variables must be relatively small, which limits its application to analyze data with large cross-sectional units. Regional databases often contain much larger cross-sectional units (counties, metropolitan areas, cities or states). So to overcome the cross sectional dimension problem, researchers assume relatively prior strengths of connections via the specification of a spatial weight matrix. Among various models to describe regional dependence, spatial autoregressive models (SAR) have attracted much attention. Early works include Anselin (1988), Kelejian and Prucha (1998, 1999) and Lee (2004, 2007). Panel data models incorporating spatial autoregression are studied in Elhorst (2003), Baltagi (2006) and Yu et al. (2008, 2012) among others. Asymptotic properties of estimators for single equation dynamic panel spatial autoregressive models are established in Yu et al. (2008, 2012). Although a large number of economic theory may concern
about interrelations among several economic variables, econometric studies regarding multivariate and simultaneous equations SAR models are limited with a few exceptions: Kelejian and Prucha (2004), Baltagi and Deng (2012), Cohen-Cole et al (2013), Liu (2014) and Chapter 1.

This paper introduces multivariate and simultaneous equations dynamic panel SAR models. Such models incorporate more than one dependent variables. In addition to time lags, a cross-sectional unit may respond to its neighbours’ or peers’ behaviour in the current period (spatial lags), and in previous periods (spatial-time lags; diffusion). Therefore, we could expect that a shock to one variable in one region could prorogate to other variables, across time dimension and spatial dimension. In order to fit more empirical specifications, overall disturbances in such a model may include time fixed effects and individual fixed effects in addition to idiosyncratic shocks. Therefore, such models can be applied to analyze equation systems with time dynamics and spatial spillover effects in regional science (de Graaff et al., 2012; Gebremariam et al., 2011), fiscal policy with government competition (Hauptmeier et al 2012, Aller and Elhorst, 2011), macroeconomic or financial analysis with internal and external habits in consumption (Korniotis, 2007), spatial propagation of macroeconomic shocks (Dewachter, 2012) and spillover effects in environmental economics, housing prices, and social network.

For a study of dynamic panel multivariate and simultaneous equations, we shall focus on several issues. The first issue is model identification. In a simultaneous equations spatial model, there are simultaneous and spatial effects present in the equations. The presence of dependent variables as endogenous regressors
raises the problem of identification of simultaneous effect, while the presence of spatial lags raises the question of identification of spatial interactions. Second, in a panel model, one needs to account for time and individual fixed effects which raise the incidental parameters problem for estimation, as well as time dynamics. Third, some time series in the model might be non-stable. Since there are \( n \times m \) different time series (\( n \) is the number of cross-sectional units and \( m \) is the number of dependent variables/equations for each unit), the possible implication of non-stability becomes more complex over time and space as well as cross variables. For the latter, we might have cointegration across variables and spatial cointegration across spatial units. One needs to investigate how estimates would behave when non-stability and cointegration are present.

For a simultaneous equations system, one may investigate relations of structural parameters with reduced form parameters. This method can achieve interesting results for a two equations SAR model (Cohen-Cole, et al, 2013, Liu 2014) but tends to be difficult when \( m > 2 \). For cross sectional data, Chapter 1 treat the implied multivariate model as a “quasi-reduced” form of a simultaneous equations model. The identification from quasi-reduced form parameters to structural parameters can then be similar to a traditional linear simultaneous equations model. However, the identification of multivariate model is not trivial. Chapter 1 provide two sets of identification conditions: the existence of instruments for spatial lags without multicollinearity, and regularities of the weight matrix to obtain information from disturbance terms. For a dynamic spatial simultaneous equations model, the identification requirement via instrument variables
become relatively easier in the presence of valid predetermined time lagged variables.

For estimation, we consider the QMLE under the sample scenario that both \( n \) and \( T \) tend to infinity. We first eliminate the time fixed effects and concentrate out the individual fixed effects in estimation. The concentrated log likelihood function is maximized and we establish the asymptotic properties of the estimator. The asymptotic bias of estimates may occur depending on the ratio of \( T \) and \( n \). When \( T \) is asymptotically larger than \( n \), the QMLE is asymptotically normal and is centered around zero. When \( T \) is asymptotically proportional to \( n \), estimator is asymptotically normal but is not centered around zero. When \( T \) is asymptotically smaller than \( n \), the QMLE is \( T \)-consistent and converges in probability to a constant. Asymptotic properties of such a nature are revealed in the QMLE for a single equation SDPD model in Yu et al. (2008). So asymptotic bias issue occurs as expected for dynamic multivariate and simultaneous equations SAR models.

For possible non-stability, we analyze spatial cointegration as well as possible cointegration across multivariate dependent variables, which means that even there are unstable components in time series, there may exist linear combinations among variables and across spatial units with stable errors. We analyze conditions under which cointegration may present. Cointegration can be revealed in our model via an error correction representation and unit roots processes formed by rotation of variables. The latter helps also the determination of the cointegration rank. For estimation, the QMLEs may have different rates of convergence rather than the usual \( \sqrt{nT} \)-rate. The higher rate of convergence occurs for some linear combinations of parameters, which can be revealed by reparameterizing
the model so that all unstable components are combined into a single variable component for estimation. While individual coefficient estimates are still $\sqrt{nT}$ consistent, estimate of the linear combination of coefficients can be $\sqrt{nT^3}$ super-consistent. This feature is similar to the estimates for the single equation SDPD model in Yu et al. (2012).

We also study a robust estimation method which can be applied to stable case, spatial cointegration case and possibly some spatial explosion cases.

In application, we apply the model to study the grain market integration in historical China using a unique historical dataset of rice and wheat prices of 65 cities in 49 years in Yangtze River Basin in 18th century. Previous researches consider rice prices solely. However, rice was not the only food in grain market, and wheat was account for around 10% (Huang, 2009). In addition, the change of rice prices cannot be completely attributed to markets’ incentive or integration since local governments have policies to stabilize rice prices. Therefore, we add the multivariate feature since wheat is believed to be a substitute for rice and price changes could be due to market behaviour. This is because that it was not the primary food and local governments would pay less attention to it as to rice prices. The empirical results show that the rice prices and wheat prices are spatially cointegrated across cities. These results provide evidence of inter-regional and intertemporal grain market integration and trading network in the eighteenth-century Yangtze River basin.

Section 2 specifies the dynamic panel multivariate SAR model. Section 3 studies identification and QML estimation of the stable model while Section 4 studies the cointegration model. Asymptotic properties of the QMLEs for both stable and
Section 5 introduces the robust estimation. Section 6 presents Monte Carlo experiments. Section 7 analyzes the identification and estimation of structural parameters in a simultaneous equations SAR model. Section 8 studies an empirical application. All proofs of main results are collected in Appendices.

2.2 A Multivariate Dynamic Panel Spatial Autoregressive Model

We specify a multivariate dynamic panel spatial autoregressive (MDP-SAR) model as:

\[ Y_{nm,t} = W_n Y_{nm,t-1} \Psi_m + Y_{nm,t-1} P_m + W_n Y_{nm,t-1} \Phi_m + X_{nk,t} \Pi_{km} + C_{nm} + D_{m,t} + V_{nm,t}, \quad (2.1) \]

for \( t = 1, 2, ..., T \), where \( D_{m,t} = \alpha_m' t \otimes l_n \) (\( \alpha_m,t \) is \( m \times 1 \)). Elements of idiosyncratic disturbance term \( V_{nm,t} \) are assumed to be i.i.d. \((0, \Sigma_v)\), across spatial units and over time, but correlations between \( v_{nk,t,i} \) and \( v_{nl,t,i} \) for different equations \( k \) and \( l \) are allowed to be correlated, so off diagonal entries of \( \Sigma_{vm} \) may be non-zero. \( Y_{nm,t} \) is a \( n \times m \) matrix, which consists of \( m \) endogenous variables (one for each column); \( t \) represents the time index and \( n \) is the number of observations. Neighborhood relationships are summarized by \( W_n \), a \( n \times n \) matrix. \( Y_{nm,t-1} \) is the time lag and \( W_n Y_{nm,t-1} \) is the space-time lag, which captures diffusion. \( X_{nk,t} \) is a \( n \times k \) matrix which contains all \( k \) exogenous variables of the \( n \) spatial units at time \( t \). This model includes own-variable spatial effects represented by diagonal elements of \( \Psi_m \), cross-variable spatial effects represented by off-diagonal elements of \( \Psi_m \), and correlated effects across equations incorporated in the covariance matrix \( \Sigma_{vm} \) of disturbances. Furthermore, \( P_m \) and \( \Phi_m \) are \( m \times m \) matrices, representing the dynamic time effects and space-time diffusion effects. \( \Pi_{km} \)
a $k \times m$ coefficient matrix for regressors. The $C_{nm}$ and $\alpha_{m,t}$ capture, respectively, individuals effects and time effects for the $m$ variables.

To get the reduced form of the MDP-SAR model, we may first transform the model,

$$\begin{align*}
Y'_{nm,t} &= \Psi_m' Y'_{nm,t-1} + P_m' Y'_{nm,t-1} W_n' + \Pi_m' X_{nt} + C_{nm}' + D_{m,t}' + V_{nm,t}',
\end{align*}$$

and then take the vectorization

$$\begin{align*}
\text{vec}(Y'_{nm,t}) &= (W_n \otimes \Psi_m') \text{vec}(Y'_{nm,t}) + (I_n \otimes P_m') \text{vec}(Y'_{nm,t-1}) \\
&+ (W_n \otimes \Phi_m') \text{vec}(Y'_{nm,t-1}) + (X_{nk,t} \otimes I_m') \text{vec}(\Pi_m) + \text{vec}(C_{nm}') \\
&+ I_n \otimes \alpha_{m,t} + \text{vec}(V_{nm,t}').
\end{align*}$$

For this arrangement, we have organized the $nm$ vector at time $t$ by first packing $m$ equations together for each individual and then order the individuals. This arrangement is convenient for analyzing possible cointegration across variables.

Let $S_{nm} = I_{nm} - W_n \otimes \Psi_m$ and $H_{nm} = S_{nm}^{-1}(I_n \otimes P_m + W_n \otimes \Phi_m')$. The reduced form system is

$$\begin{align*}
\text{vec}(Y'_{nm,t}) &= \sum_{h=0}^{+\infty} H_{nm}^h S_{nm}^{-1} [(X_{nk,t-h} \otimes I_m) \text{vec}(\Pi_{km}) + \text{vec}(C_{nm}')] + l_n \otimes \alpha_{m,t-h} + \text{vec}(V_{nm,t-h}').
\end{align*}$$

### 2.2.1 Stability and Parameter Space

The dynamic panel spatial system has “dynamics” in two dimensions – one in space and the other in time. Therefore we analyze whether the model is stable for the two dimensions. In spatial dimension, the analysis is analog to Chapter 1. The Jacobian matrix need to have a positive determinant. A sufficient condition is that $\rho(\Psi_m)\rho(W_n) < 1$, where $\rho(A)$ is the spectral radius of matrix $A$.  

47
In time dimension, we will analyze two cases: stability and (spatial) cointegration. Stability requires all the eigenvalues $\lambda_H$ in absolute value of $H_{nm}$, $|\lambda_H| < 1$. Because $S_{nm}$ is invertible, the stable condition becomes that all solutions $\lambda$ to $|(I_n \otimes P'_m + W_n \otimes \Phi'_m) - \lambda(I_{nm} - W_n \otimes \Psi'_m)| = 0$ are inside the unit circle because

$$|(I_n \otimes P'_m + W_n \otimes \Phi'_m)S_{nm}^{-1} - \lambda I_{nm}| = 0$$

$$\Leftrightarrow |(I_n \otimes P'_m + W_n \otimes \Phi'_m) - \lambda(I_{nm} - W_n \otimes \Psi'_m)| = 0.$$ 

A possible sufficient condition can be derived:

$$||H_{nm}|| \leq ||I_n \otimes P'_m|| + ||W_n \otimes \Phi'_m|| ||S_{nm}^{-1}||$$

$$\leq (||I_n \otimes P'_m|| + ||W_n \otimes \Phi'_m||) \frac{1}{1 - ||W_n \otimes \Psi'_m||} < 1$$

By using row sum norm, a stronger sufficient condition from the preceding inequality is $||P'_m||_\infty + ||\Phi'_m||_\infty ||W_n||_\infty + ||\Psi'_m||_\infty ||W_n||_\infty < 1$. Furthermore, for the case with a row normalized weight matrix, $||P'_m||_\infty + ||\Phi'_m||_\infty + ||\Psi'_m||_\infty < 1$.

In order to analyze stability in details, we assume that the weight matrix $W_n$ is row normalized and diagonalizable with real eigenvalues in our model. In the single variable case ($m = 1$), Yu, de Jong and Lee (2012) analyze an unstable scenario which indicates that dependent variables across space may be cointegrated. However, the multivariate SAR model is more complicated, since there may exist cointegration among variables as well as cointegration of variables across spaces.

This assumption is stronger than the subsequent Assumption 2.2.1, however, we maintain this assumption in cointegration part for easiness of analysis. In addition, a typical undirected weight matrix is diagonalizable with real eigenvalues because of its symmetry. If $W_n$ is row-normalized from an original symmetric model as in spatial econometric models, $W_n$ is diagonalizable with all eigenvalues being real. And these weight matrices are widely employed in empirical studies of regional economics and social network analysis.
For the spatial network under consideration, the weight matrix, $W_n$, can be diagonalized as $\Gamma_n \omega_n \Gamma_n^{-1}$ with a diagonal real eigenvalue matrix $\omega_n$. Since the weight matrix is row normalized, the largest eigenvalue is 1. Let the eigenvalue matrix have $n_1$ eigenvalues to be one while others are less than one in absolute value, i.e.,

$$\omega_n = \text{diag}\{1, ..., 1, \omega_{n,n_1+1}, ..., \omega_{n,n}\} = \mathbb{1}_{n,n_1} + \text{diag}\{0, ..., 0, \omega_{n,n_1+1}, ..., \omega_{n,n}\},$$

where $\mathbb{1}_{n,n_1}$ represents the $n \times n$ diagonal matrix with the first $n_1$ diagonal elements being one and remaining ones being zero. Therefore,

$$S_{nm}^{-1} = (I_{nm} - W_n \otimes \Psi_m')^{-1} = (\Gamma_n \otimes I_m)(I_{nm} - \omega_n \otimes \Psi_m')^{-1}(\Gamma_n^{-1} \otimes I_m),$$

$$H_{nm} = S_{nm}^{-1}(I_n \otimes P_m' + W_n \otimes \Phi_m')$$

$$= (\Gamma_n \otimes I_m)(I_{nm} - \omega_n \otimes \Psi_m')^{-1}(I_n \otimes P_m' + \omega_n \otimes \Phi_m')(\Gamma_n^{-1} \otimes I_m).$$

Define a diagonal block $nm \times nm$ matrix $\tilde{B}_{nm} = \text{diag}\{0, \cdots, 0, (I_m - \omega_{n,n_1+1} \Psi_m')^{-1}(P_m + \omega_{n,n_1+1} \Phi_m)', \cdots, (I_m - \omega_{n,n} \Psi_m')^{-1}(P_m + \omega_{n,n} \Phi_m)'\}$, where each of the diagonal block is a $m \times m$ submatrix. We have

$$(I_n \otimes I_m - \omega_n \otimes \Psi_m')^{-1}(I_n \otimes P_m' + \omega_n \otimes \Phi_m')$$

$$= \text{diag}\{(I_m - \Psi_m')^{-1}(P_m + \Phi_m)', \cdots, (I_m - \Psi_m')^{-1}(P_m + \Phi_m)\},$$

$$= (I_m - \omega_{n,n_1+1} \Psi_m')^{-1}(P_m + \omega_{n,n_1+1} \Phi_m)', \cdots, (I_m - \omega_{n,n} \Psi_m')^{-1}(P_m + \omega_{n,n} \Phi_m)'.$$

$$= \mathbb{1}_{n,n_1} \otimes (I_m - \Psi_m')^{-1}(P_m + \Phi_m) + \tilde{B}_{nm}.$$

Further define $B_{nm} = (\Gamma_n \otimes I_m)\tilde{B}_{nm}(\Gamma_n^{-1} \otimes I_m)$, we have

$$H_{nm}^h = \Gamma_n \mathbb{1}_{n,n_1} \Gamma_n^{-1} \otimes [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)']^h + B_{nm}^h,$$

$$H_{nm}^h S_{nm}^{-1} = \Gamma_n \mathbb{1}_{n,n_1} \Gamma_n^{-1} \otimes [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)']^h (I_m - \Psi_m')^{-1} + B_{nm}^h S_{nm}^{-1}.$$
Finally, for simpler exposition, we collect terms excluding spatial lags, time lags and time effect components into

\[ Q_{nm,t} = (X_{nt} \otimes I_m)\text{vec}(\Pi'_m) + \text{vec}(C'_{nm}) + \text{vec}(V'_{nm,t}) \]

With this abbreviation, the reduced form by backward substitution to the period \( t = -1 \) is

\[ \text{vec}(Y'_{nm,t}) = H_{nm}\text{vec}(Y'_{nm,t-1}) + S_{nm}^{-1}[Q_{nm,t} + \text{vec}(D'_{m,t})] \]

\[ = H_{nm}^t\text{vec}(Y'_{nm,-1}) + \sum_{h=0}^{t} H_{nm}^h S_{nm}^{-1}[Q_{nm,t-h} + \text{vec}(D'_{m,t-h})]. \]

Therefore, we can decompose the dependent variable into several components as (see details in appendix),

\[ \text{vec}(Y'_{nm,t}) = \text{vec}^{(u)}(Y'_{nm,t}) + \text{vec}^{(s)}(Y'_{nm,t}) + \text{vec}^{(a)}(Y'_{nm,t}), \]

where

\[ \text{vec}^{(u)}(Y'_{nm,t}) = (\Gamma_n 1_{n,n}1_{n,n}^{-1} \otimes [(I_m - \Psi'_m)^{-1}(P_m + \Phi_m)]^{t+1}) \text{vec}(Y'_{nm,-1}) \]

\[ + \sum_{h=0}^{t} \{ \Gamma_n 1_{n,n}1_{n,n}^{-1} \otimes [(I_m - \Psi'_m)^{-1}(P_m + \Phi_m)]^h(I_m - \Psi'_m)^{-1} \} Q_{nm,t-h} \]

\[ \text{vec}^{(s)}(Y'_{nm,t}) = \sum_{h=0}^{\infty} B_{nm}^h S_{nm}^{-1} Q_{nm,t-h} \]

and \( \text{vec}^{(a)}(Y'_{nm,t}) = \sum_{h=0}^{t} \{ I_n \otimes [(I_m - \Psi'_m)^{-1}(P_m + \Phi_m)]^h(I_m - \Psi'_m)^{-1} \alpha_{m,t-h} \}. \)

(2.4)

In the stable case, according to the following assumption, all three components are stable.

**Assumption 2.2.1.** \( W_n \) is a nonstochastic row normalized weight matrix. Row and column sums of \( W_n \) in absolute value are uniformly bounded, uniformly in \( n \). For
any possible parameters in their parameter spaces, $S_{nm}$ is nonsingular, and $S_{nm}^{-1}$ and $\sum_{h=1}^{+\infty} \text{abs}(H_{nm}^h)$ are bounded in row and column sum norms.

The parameter space of coefficients $\Pi_{km}$ is compact; the parameter space of coefficients $\Psi_m, P_m$ and $\Phi_m$ is compact such that $\rho(\Psi_m)\rho(W_n) < 1$ and all solutions $\lambda$ to $|S_{nm}^{-1}(I_n \otimes P'_m + W_n \otimes \Phi'_m) - \lambda I_{nm}| = 0$ are inside the unit circle; and the covariance matrix $\Sigma_{vm}$ lies on a compact space and is nonsingular. The true parameters are located in the interior of their parameter space.

A spatial weight matrix is the key factor in a SAR model. In this study, we focus on the scenarios of a nonstochastic and row-normalized spatial weights matrix as these features are prevalent in regional economics and social network analysis (as a social norm effect, see Liu, Patacchinni and Zenou, 2013). As usual, we assume that the row and column sum norms of $\{W_n\}$ are bounded, which could be justified by: the sparsity of the weight matrix or decreasing interactions of spatial units with far away neighbors.

Assumption 2.2.1 doesn’t assume that the weight matrix is diagonalizable and all eigenvalues are real, which is not necessary for consistency and asymptotic normality of the QMLE. The absolute summability condition for $H_{nm}$ and its powers, and also its uniform boundedness in spatial dimension is to justify stability of the process through space and time. It is ultimately related to the requirement that $(I_m - \Psi'_m)^{-1}(P_m + \Phi_m)'$, $(I_m - \omega_i \Psi'_m)^{-1}(P_m + \omega_i \Phi_m)'$ and $\tilde{B}_{nm}$ have all their eigenvalues less than 1 in absolute value. And, hence, all eigenvalues of $S_{nm}^{-1}(I_n \otimes P'_m + W_n \otimes \Phi'_m)$ are inside the unit circle. The assumption that all eigenvalues are inside the unit circle could be justified by some stronger conditions,
one of them mentioned earlier is \( ||P'_m||_\infty + ||\Phi'_m||_\infty + ||\Psi'_m||_\infty < 1 \). We maintain the relative higher level assumption in Assumption 2.2.1 as a general case.

For the cointegration case, according to the assumption below, the \( \tilde{B}_{nm} \) still has all its eigenvalues strictly less than one in absolute value, however, \( (I_m - \Psi'_m)^{-1}(P_m + \Phi_m)' \) has \( m_1 \) unit eigenvalues, and other eigenvalues \( (m_2 = m - m_1 \) where \( m_2 \) is possibly zero) are strictly less than one in absolute value.

**Assumption 2.2.2.** \( W_n \) is a nonstochastic row normalized and diagonalizable weight matrix with real eigenvalues. Row and column sums of \( W_n \) in absolute value are uniformly bounded, uniformly in \( n \). For any possible parameters in their parameter spaces, the corresponding \( S_{nm} \) is nonsingular, and \( S_{nm}^{-1} \) and \( \sum_{h=1}^{+\infty} \text{abs}(B_{nm}^h) \) are bounded in row and column sums norms.

The parameter spaces of regression coefficients \( \Pi_{km} \) and the covariance matrix \( \Sigma_{vm} \), which is nonsingular, are compact. The parameter space of coefficients of spatial interactions and dynamics \( \Psi_m, P_m \) and \( \Phi_m \) is compact such that \( \rho(\Psi_m)\rho(W_n) < 1 \), the largest eigenvalues of \( (I_m - \Psi'_m)^{-1}(P_m + \Phi_m)' \) equal to one, any remaining eigenvalues of \( (I_m - \Psi'_m)^{-1}(P_m + \Phi_m)' \) are less than one in absolute value, and \( \rho((I_m - \omega_{n,j}\Psi_m)^{-1}(P_m + \omega_{n,j}\Phi_m)') < 1 \) for all those eigenvalues \( w_{n,j} \) of \( W_n \) less than one in absolute value, where such eigenvalues are assumed to be bounded away from 1, when \( n \) tends to infinity. The true parameters are located in the interior of their parameter spaces.

The assumption on the weight matrix that it is diagonalizable with real eigenvalues is stronger than that in Assumption 2.2.1. But is similar to that in Yu, de Jong and Lee (2012). In this case, \( \text{vec}^{(u)}(Y'_{nm,t}) \) may have unstable components, \( \text{vec}^{(a)}(Y'_{nm,t}) \) represents the time effects component and all components of
vec(s)(Y′_{nm,t}) are stable. The spectral radius \( \rho((I_m - \omega_{n,j}\Psi_m)^{-1}(P_m + \omega_{n,j}\Phi_m)) < 1 \) can be justified by some stronger conditions, a possible sufficient condition is:

\[
\|(\omega_{n,j}\Phi_m + P_m)(I_m - \omega_{n,j}\Psi_m)^{-1}\| < 1 \text{ for any } |\omega_{n,j}| < 1,
\]

where the matrix norm can be row sum norm or column sum norm. A stronger condition is \( \|P_m\| + \max_{j=n_1,\ldots,n}\omega_{n,j}(\|\Phi_m\| + \|\Psi_m\|) < 1 \). The \( (I_m - \Psi_m)^{-1}(P_m + \Phi_m) \) has its largest eigenvalue equal to one and remaining eigenvalues, if any, are less than one in absolute value, which means that \( 1 \leq \|(I_m - \Psi_m)^{-1}(P_m + \Phi_m)\| \leq \|P_m + \Phi_m\|(1 - \|\Psi_m\|)^{-1} \). In turn, this implies \( \|P_m\| + \|\Phi_m\| + \|\Psi_m\| \geq 1 \). Combined with the former assumption that \( \|P_m\| + \max_{j=n_1,\ldots,n}\omega_{n,j}(\|\Phi_m\| + \|\Psi_m\|) < 1 \), we need to assume that the limit of \( \max_{j=n_1,\ldots,n}\omega_{n,j} \) is strictly less than one, i.e.,

\[
\lim_{n \to \infty} \max_{j=n_1,\ldots,n}\omega_{n,j} < 1 - \epsilon \text{ for some positive number } \epsilon
\]

if we maintain the stronger assumption in this case.

The following example provides a special case of stable analysis for triangular system. As we introduce the triangular system in the empirical application, it is informative to illustrate and sharpen our assumptions for this system.

**Example: A Triangular System**

Consider a triangular system,

\[
Y'_{nm,t} = \Psi_m Y'_{nm,t} W_n + P_m Y'_{nm,t-1} + \Phi_m Y'_{nm,t-1} W_n + \Pi_m X'_{nt} + C'_{nm} + D'_m + V'_{nm,t},
\]

where the matrices \( \Psi_m', \Phi_m' \) and \( P_m' \) are upper (or lower) triangular matrices. Take \( m = 2 \) for instance,

\[
Y_{1n,t} = \psi_{11} W_n Y_{1n,t} + p_{11} Y_{1n,t-1} + \phi_{11} W_n Y_{1n,t-1} + U_{1n,t},
\]

\[
Y_{2n,t} = \psi_{12} W_n Y_{2n,t} + \psi_{22} W_n Y_{2n,t} + p_{12} Y_{1n,t-1} + p_{22} Y_{2n,t-1} + \phi_{12} W_n Y_{1n,t-1} + \phi_{22} W_n Y_{2n,t-1} + U_{2n,t},
\]

53
where $U_{jn,t}$ represents remaining components consisting of time fixed effects, individual fixed effects and idiosyncratic disturbance terms.

The system has an advantage that concerned eigenvalues for stability can be expressed explicitly as polynomials of the parameters in the model. In the general triangular $m > 1$ system, the eigenvalues for $(I_m - \omega_n, i \Psi'_m)^{-1}(\omega_n, i \Phi'_m + P'_m)$, where $i = n_1 + 1, \ldots, n$, are $\{p_{jj} + w_{n,i} \phi_{jj} \over 1 - w_{n,i} \psi_{jj}\}_{i=1}^m$ for $j = 1, \ldots, m$. Therefore, for an exact region of parameters which characterizes the stability of the system can be derived with reference to Lee and Yu (2015). The derivative with respect to $w_{n,i}$ is

$$\frac{\partial}{\partial w_{n,i}} p_{jj} + w_{n,i} \phi_{jj} = \phi_{jj} + \psi_{jj} p_{jj}$$

Therefore, the monotonicity of eigenvalues of $(I_m - \omega_n, i \Psi'_m)^{-1}(\omega_n, i \Phi'_m + P'_m)$ respect to $w_{n,i}$ depends on the value of $\phi_{jj} + \psi_{jj} p_{jj}$ for each $j = 1, 2, \ldots, m$. Defining $w_{\text{min}}$ as the minimum eigenvalue of $W_n$, the stable case is characterized by

$$\begin{cases} p_{jj} + \phi_{jj} + \psi_{jj} < 1, & \text{when } \phi_{jj} + \psi_{jj} p_{jj} > 0; \\ p_{jj} + \phi_{jj} + \psi_{jj} > 1, & \text{when } \phi_{jj} + \psi_{jj} p_{jj} = 0; \\ p_{jj} + \phi_{jj} + \psi_{jj} > -1, & \text{when } \phi_{jj} + \psi_{jj} p_{jj} < 0. \end{cases}$$

And the cointegration case can be characterized as

$$\begin{cases} p_{jj} + \phi_{jj} + \psi_{jj} \leq 1, & \text{when } \phi_{jj} + \psi_{jj} p_{jj} > 0; \\ p_{jj} + \phi_{jj} + \psi_{jj} \leq 1, \text{ and some } j \in 1, \ldots, m, \text{ such that } \\ p_{jj} + \phi_{jj} + \psi_{jj} > -1, & \text{when } \phi_{jj} + \psi_{jj} p_{jj} = 0; \\ p_{jj} + \phi_{jj} + \psi_{jj} > 1, & \text{when } \phi_{jj} + \psi_{jj} p_{jj} < 0. \end{cases}$$

Note that the above conditions generalize those in Yu and Lee (2015) where $m=1$. Our conditions require all $p_{jj}$, $\psi_{jj}$ and $\phi_{jj}$ for $j = 1, \ldots, m$ are in the same geometric figures provided in their chapter.

16 They provide a geometric illustration in their paper.
2.3 Identification and QML Estimation for Stable MDP-SAR Model

We first discuss the identification and QML estimation of the multivariate dynamic panel SAR model when all dependent variables are stable.

2.3.1 Assumptions

Assumption 2.3.1. \( v_{n,t,i} \), the \( i \)th row of \( V_{nm,t} \), is independent and identically distributed random vector of dimension \( m \) with zero mean and covariance matrix \( \Sigma_{vm} \). The elements of disturbances satisfy that \( \mathbb{E}|v_{n,t,ik}v_{n,t,il}v_{n,t,ip}v_{n,t,iq}|^{1+\delta} \) is bounded for any \( i = 1, 2, ..., n \) and \( k, l, p, q = 1, 2, ..., m \) for some constant \( \delta > 0 \).

According to Assumption 2.3.1, though the disturbances are i.i.d across time and individuals, they are allowed to be correlated among different variables for the same individual at the same time period.

Assumption 2.3.2. Elements of \( X_{nk,t} \) are exogenous constants and are uniformly bounded for all \( n \) and \( t \).

Assumption 2.3.3. \( n \) is a nondecreasing function of \( T \). Both \( T \) and \( n \) tends to infinity.

We consider the model with large \( T \). Short \( T \) case is also interesting for some dataset, which would be estimated by other methods such as IVs or GMM instead of QML. As we are focusing on QML estimation without extra initial value specification, we need large \( T \).

Note that we can accommodate the finite \( n \) and large \( T \) case, but the estimation and asymptotic properties of estimates for finite \( n \) have been discussed in classic vector autoregression models (see Hamilton (1994)). So we focus on
the case with both large \( n \) and \( T \). When \( n \) and \( T \) tend to infinity, the number of individual and time fixed effects goes to infinity. We can eliminate the time fixed effects and concentrate out the individual fixed effects so to focus our asymptotic analysis on the QML estimates of common parameters of interest. The elimination of time effects can avoid the incidental parameter problem due to the many time dummies\(^{17}\).

Define \( J_n = I_n - \frac{1}{n} l_n l_n' \). \((F_{n,n-1}, \frac{ln}{\sqrt{n}})\) is the orthonormal eigenvector matrix of \( J_n \) where columns of \( F_{n,n-1} \) are vectors corresponding to eigenvalue ones, i.e., \( J_n F_{n,n-1} = F_{n,n-1} \). Therefore, \( F_{n,n-1} F_{n,n-1}' = J_n \) and \( F_{n,n-1}' F_{n,n-1} = I_{n-1} \). Since, \( J_n (\alpha_{m,t} \otimes l_n) = 0 \), multiplying \( J_n \) to the model from left and then \( F_{n,n-1}' \) or simply multiplying \( F_{n,n-1}' \), we get

\[
F_{n,n-1}' Y_{nm,t} = V_{nm,t}^{*} \Psi_{m} + P_{m} + X_{nk,t} \Pi_{km} + C_{nm} + V_{nm,t}.
\]

Define \( W_{n}^{*} = F_{n,n-1}' W_{n} F_{n,n-1} \), which is a \((n - 1) \times (n - 1)\) matrix. The resulted equation has the time dummy eliminated but is now of dimension \((n - 1)\) with the loss of one degree of freedom in each time period. By taking a transpose, the equation can be vectorized to solve for \( vec(Y_{nm,t} F_{n,n-1}) \) at \( t \) via \( S_{nnm}^{-1} \), where \( S_{nnm} = (F_{n,n-1} \otimes I_m)(I_{nm} - W_n \otimes \Psi_m')(F_{n,n-1} \otimes I_m) \). The determinant of \( S_{nnm}^{-1} \) can be evaluated via \( S_{nm} \):

\[
|S_{nm}| = |(F_{n,n-1}, \frac{ln}{\sqrt{n}})' \otimes I_m| \cdot |S_{nm}| \cdot |(F_{n,n-1}, \frac{ln}{\sqrt{n}}) \otimes I_m|
\]

\[
= |[(F_{n,n-1}, \frac{ln}{\sqrt{n}})' \otimes I_m](I_{nm} - W_n \otimes \Psi_m')[(F_{n,n-1}, \frac{ln}{\sqrt{n}}) \otimes I_m]|^{17}
\]

The likelihood constructed after the elimination of the time effects is strictly speaking a partial likelihood.
Ψ

ables after the elimination of time dummies is

\[ I_{nm} - \left( \begin{array}{c c c}
F'_{n,n-1} W_n F_{n,n-1} & F'_{n,n-1} W_n (l_n / \sqrt{n}) \\
(l'_n / \sqrt{n}) W_n F_{n,n-1} & (l'_n / \sqrt{n}) W_n l_n / \sqrt{n}
\end{array} \right) \otimes \Psi'_m \]

\[ = I_{nm} - \left( \begin{array}{c c}
F'_{n,n-1} W_n F_{n,n-1} & 0 \\
0 & 1
\end{array} \right) \otimes \Psi'_m \]

\[ = |I_m - \Psi'_m||I_{(n-1)m} - W_n \otimes \Psi'_m| = |I_m - \Psi'_m||S_{nm}^*|. \]

Furthermore, define \( V_{nm,t}^*(\theta) = F'_{n,n-1} V_{nm,t}(\theta) \). Hence,

\[ \text{vec}(V_{nm,t}^*(\theta)) = [F'_{n,n-1} \otimes I_m - (W_n F'_{n,n-1}) \otimes \Psi'_m] \text{vec}(Y'_{nm,t}) - [F'_{n,n-1} \otimes P_m - (W_n F'_{n,n-1}) \otimes \Psi'_m] \text{vec}(Y'_{nm,t}) - [(F'_{n,n-1} X_{nk,t}) \otimes I_m] \text{vec}(\Pi'_{km}) - (F'_{n,n-1} \otimes I_m) \text{vec}(C'_{nm}), \]

where \( \theta = (\text{vec}(\Pi'_{km}), \text{vec}(P_m'), \text{vec}(\Psi'_m), \text{vec}(\Psi'_m'), \text{vec}(\Sigma_{vm})')' \) with vec being vectorization with all distinct elements for a symmetric matrix.

\[ F'_{n,n-1} \otimes I_m - (W_n F'_{n,n-1}) \otimes \Psi'_m = (F'_{n,n-1} \otimes I_m)(I_{nm} - W_n \otimes \Psi'_m)(F_{n,n-1} F'_{n,n-1} \otimes I_m) \]

\[ = (F'_{n,n-1} \otimes I_m)(I_{nm} - W_n \otimes \Psi'_m)(I_{nm} - l_n l_n' n) \otimes I_m \]

\[ = (F'_{n,n-1} \otimes I_m)(I_{nm} - W_n \otimes \Psi'_m), \]

where the last equality holds because \((I_{nm} - W_n \otimes \Psi'_m) \cdot (l_n l_n' n) = l_n l_n' n \otimes (I_m - \Psi'_m) \) and \( F'_{n,n-1} l_n = 0 \). Similarly, \( F'_{n,n-1} \otimes P_m + (W_n F'_{n,n-1}) \otimes \Phi'_m = (F'_{n,n-1} \otimes I_m)(I_n \otimes P_m + W_n \otimes \Phi'_m). \) Therefore, \( \text{vec}(V_{nm,t}^*(\theta)) = (F'_{n,n-1} \otimes I_m)[(I_{nm} - W_n \otimes \Psi'_m) \text{vec}(Y'_{nm,t}) - (I_n \otimes P_m + W_n \otimes \Phi'_m) \text{vec}(Y'_{nm,t}) - (X_{nk,t} \otimes I_m) \text{vec}(\Pi_{km}) - \text{vec}(C'_{nm})] = (F'_{n,n-1} \otimes I_m) \text{vec}(V_{nm,t}(\theta)). \) The partial likelihood function for the transformed variables after the elimination of time dummies is

\[ \frac{1}{n - 1} \ln I_{nT,m} = - \frac{m}{2} \ln(2\pi) + \frac{1}{n - 1} \ln |S_{nm}| - \frac{1}{n - 1} \ln |I_m - \Psi'_m| - \frac{1}{2} \ln |\Sigma_{vm}| \]

\[ - \frac{1}{2(n - 1)T} \sum_{t=1}^{T} \text{vec}(V_{nm,t}(\theta))' (J_n \otimes \Sigma_{vm}^{-1}) \text{vec}(V_{nm,t}(\theta)). \]

(2.5)
The first order condition with respect to \( \text{vec}(C'_{nm}) \) gives an estimate of individual effects in terms of other parameters,

\[
\text{vec}(C'_{nm}) = \frac{1}{T} \sum_{t=1}^{T} [S_{nm} \text{vec}(Y'_{nm,t}) - (I_n \otimes P'_m + W_n \otimes \Phi'_m) \text{vec}(Y'_{nm,t-1}) - (X_{nk,t} \otimes I_m) \text{vec}(\Pi'_{km})].
\]

Define \( \text{vec}(\bar{Y}'_{nm,T}) = \frac{1}{T} \sum_{t=1}^{T} \text{vec}(Y'_{nm,t}) \), as time means, and \( \text{vec}(\tilde{Y}'_{nm,t}) = \text{vec}(Y'_{nm,t}) - \text{vec}(\bar{Y}'_{nm,T}) \), vec(\( \bar{Y}'_{nm,T} \)) = \frac{1}{T} \sum_{t=1}^{T} \text{vec}(Y'_{nm,t}-1) \) and \( \text{vec}(\tilde{Y}'_{nm,t}) = \text{vec}(Y'_{nm,t}) - \text{vec}(\bar{Y}'_{nm,T}) \), as deviations from time means of the relevant dependent variables. The concentrated likelihood function with individual effects concentrated out is

\[
\frac{1}{(n-1)V} \ln L^*_{nT,m} = -\frac{m}{2} \ln(2\pi) + \frac{1}{n-1} \ln |S_{nm}| - \frac{1}{n-1} \ln |I_m - \Psi'_m| - \frac{1}{2} \ln |\Sigma_{vm}|
\]

\[
- \frac{1}{2(n-1)V} \sum_{t=1}^{T} \text{vec}(\tilde{V}'_{nm,t}(\theta))' (J_n \otimes \Sigma_{vm}^{-1}) \text{vec}(\tilde{V}'_{nm,t}(\theta)),
\]

where

\[
\text{vec}(\tilde{V}'_{nm,t}(\theta)) = S_{nm} \text{vec}(\tilde{Y}'_{nm,t}) - (I_n \otimes P'_m + W_n \otimes \Phi'_m) \text{vec}(\tilde{Y}'_{nm,t-1}) - (X_{nk,t} \otimes I_m) \text{vec}(\Pi'_{km}).
\]

With individual (time) effects concentrated out (eliminated), we can focus asymptotic analysis on the common parameters’ estimates via the concentrated likelihood without facing explicitly an infinite number of parameters.

### 2.3.2 Identification and Consistency of QMLE

The concentrated sample average log likelihood function is

\[
\frac{1}{(n-1)V} \ln L_{nT,m} = -\frac{m}{2} \ln(2\pi) + \frac{1}{(n-1)} \ln |S_{nm}| - \frac{1}{(n-1)} \ln |I_m - \Psi'_m| - \frac{1}{2} \ln |\Sigma_{vm}|
\]

\[
- \frac{1}{2(n-1)V} \sum_{t=1}^{T} \text{vec}(\tilde{V}'_{nm,t}(\theta))' (J_n \otimes \Sigma_{vm}^{-1}) \text{vec}(\tilde{V}'_{nm,t}(\theta)),
\]

where

\[
\text{vec}(\tilde{V}'_{nm,t}(\theta)) = S_{nm} \text{vec}(\tilde{Y}'_{nm,t}) - (I_n \otimes P'_m + W_n \otimes \Phi'_m) \text{vec}(\tilde{Y}'_{nm,t-1}) - (X_{nk,t} \otimes I_m) \text{vec}(\Pi'_{km}).
\]
Let $e'_{m,i}$ be a $1 \times m$ row vector with all zero elements except for its $i$th entry which has the value one. Define $\tilde{A}_z = \left[ (W_n \otimes e'_{m,1})S^{-1}_{nm0}((I_n \otimes P_{m0} + W_n \otimes \Phi_{m0})vec(\tilde{Y}'_{nm,t-1}) + (\tilde{X}_{nk,t} \otimes I_m)vec(\Pi'_{km0}), \ldots, (W_n \otimes e'_{m,m})S^{-1}_{nm0}((I_n \otimes P'_{m0} + W_n \otimes \Phi_{m0})vec(\tilde{Y}'_{nm,t-1}) + (\tilde{X}_{nk,t} \otimes I_m)vec(\Pi'_{km0})) \right]$, a $n \times m$ matrix, and $\tilde{Z}_{nm,t} = [\tilde{Y}_{nm,t-1}, W_n \tilde{Y}_{nm,t-1}, \tilde{X}_{nk,t}]$ a $n \times (2m + k)$ matrix. The following two assumptions provide ways to identify the parameters. Assumption 2.3.4 guarantees the existence of instruments for every endogenous regressor, while Assumption 2.3.5 employs the disturbance structure to identify parameters.

**Assumption 2.3.4.** The limiting matrix

$$
\lim_{T \to +\infty} \frac{1}{(n-1)T} \sum_{t=1}^{T} E \left[ \begin{pmatrix} 
\tilde{Z}'_{nm,t} J_n \tilde{Z}'_{nm,t} & \tilde{Z}'_{nm,t} J_n \tilde{A}_z,t \\
\tilde{A}_z,t J_n \tilde{Z}_{nm,t} & \tilde{A}_z,t J_n \tilde{A}_z,t 
\end{pmatrix} \right]
$$

exists and is nonsingular.

Assumption 2.3.4 indicates that the parameters can be identified when (best) instruments for endogenous variables exist and they are not linearly dependent with each other. The following assumption, however, employs information from the disturbances to help identification. Let $\sigma^2_m = \text{Tr}(\Sigma_{vm})$ and $\Sigma^*_{vm} = \Sigma_{vm}/\sigma^2_m$.

**Assumption 2.3.5.**

$$
\lim_{n \to +\infty} \left| S'_{nm0} S_{nm}(J_n \otimes \Sigma_{vm}) S^{-1}_{nm0} S'_{nm0}(J_n \otimes \Sigma^*_{vm0}) \right|^{1/(m(n-1))} 
- \frac{1}{m(n-1)} \text{Tr} \left[ S'_{nm0} S_{nm}(J_n \otimes \Sigma^*_{vm}) S^{-1}_{nm0} S'_{nm0}(J_n \otimes \Sigma^*_{vm0}) \right] < 0
$$

unless $\Sigma^*_{vm} = \Sigma^*_{vm0}$ and $\Psi_m = \Psi_{m0}$. 

59
Assumption 2.3.5 holds in finite sample for any finite \( n \) due to the geometric-arithmetic mean inequality. We need the relationship being preserved in limiting case to establish identification. The identification of parameters is established in the following proposition.

**Proposition 2.1.** (1) Under Assumption 2.2.1, 2.3.1-2.3.3 and 2.3.4, \( \Psi_m, P_m, \Phi_m \) and \( \Pi_{km} \) can be identified;

(2) Under Assumption 2.2.1, 2.3.1-2.3.3 and 2.3.5, \( \Psi_m \) and \( \Sigma_{vm} \) can be identified.

Identification of parameters in a finite sample is empirically interesting. Corollary 2.1 suggests a simple sufficient condition based on the structure of \( W_n \) to verify the finite sample counterpart of Assumption 2.3.5.

**Corollary 2.1.** In finite sample, when \( I_{n-1}, W_n^*, W_n^* W_n^* \) are linearly independent, then

\[
|S_{nn0}^{-1} S_{nm} (J_n \otimes \Sigma_{vm}^{-1}) S_{nm0} S_{nm0}^{-1} (J_n \otimes \Sigma_{vm0}^{-1}) | \frac{1}{m(n-1)}
\]

\[
- \frac{1}{m(n-1)} \text{Tr} \left[ S_{nm0}^{-1} S_{nm} (J_n \otimes \Sigma_{vm}^{-1}) S_{nm} S_{nm0}^{-1} (J_n \otimes \Sigma_{vm0}^{-1}) \right] < 0,
\]

unless \( \Sigma_{vm}^* = \Sigma_{vm0}^* \) and \( \Psi_m = \Psi_{m0} \).

Lemmas 2.1 and 2.2 are implied results by some basic lemmas in Appendix D. They are used to establish the uniform convergence of the sample average concentrated log likelihood function to its expectation.

**Lemma 2.1.** Under Assumptions 2.2.1, 2.3.1-2.3.3, suppose a generic matrix \( B_{nm} \)

is uniformly bounded in row and column sum norms, then

\[
\frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m) \text{vec}(\tilde{Y}_{nm,t-1}^\prime)]^\prime B_{nm} \text{vec}(\tilde{V}_{nm,t}^\prime)
\]
Proposition 2.2. Under Assumption 2.2.1, 2.3.1-2.3.3, as T goes to infinity, \( \frac{1}{(n-1)^T} \ln L_{nm}(\theta) - Q_{nT,m} \xrightarrow{p} 0 \) uniformly in \( \theta \) in any compact parameter set, where \( Q_{nT,m} = E[\frac{1}{(n-1)^T} \ln L_{nm}(\theta)] \).

Furthermore, \( \frac{1}{n^T} \sum_{t=1}^{T} [(J_n \otimes I_m) \text{vec}(\tilde{Y}_{nm,t-1})]' B_{nm} [(J_n \otimes I_m) \text{vec}(\tilde{Y}_{nm,t-1})] \) and \( \frac{1}{n^T} \sum_{t=1}^{T} [(J_n \otimes I_m) \text{vec}(\tilde{Y}_{nm,t-1})]' B_{nm} [(J_n \otimes I_m) \text{vec}(\tilde{Y}_{nm,t-1})] \) both go to infinity, \( T \rightarrow \infty \).

Lemma 2.2. Under Assumption 2.3.1, if a generic matrix \( B_{nm} \) is bounded in row and column sum norms, then

\[
E \left\{ \frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m) \text{vec}(\tilde{Y}_{nm,t-1})]' B_{nm} \text{vec}(\tilde{Y}_{nm,t}) \right\} = O\left( \frac{1}{T} \right);
\]

\[
E \left\{ \frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m) \text{vec}(\tilde{Y}_{nm,t-1})]' B_{nm} [(J_n \otimes I_m) \text{vec}(\tilde{Y}_{nm,t-1})] \right\} = O(1).
\]

(1)

\[
E \left[ \frac{1}{nT} \sum_{t=1}^{T} \text{vec}(V_{nm,t}'B_{nm} \text{vec}(V_{nm,t}) \right] = \frac{1}{n} \text{Tr}[B_{nm} (I_n \otimes \Sigma_{nm0})] = O(1);
\]

\[
\frac{1}{nT} \sum_{t=1}^{T} \text{vec}(V_{nm,t}'B_{nm} \text{vec}(V_{nm,t}) - E \left[ \frac{1}{nT} \sum_{t=1}^{T} \text{vec}(V_{nm,t}'B_{nm} \text{vec}(V_{nm,t}) \right] = O_p \left( \frac{1}{\sqrt{nT}} \right).
\]

(2)

\[
E \left[ \frac{1}{n} \text{vec}(\tilde{V}_{nm,T}'B_{nm} \text{vec}(\tilde{V}_{nm,T}) \right] = \frac{1}{nT} \text{Tr}[B_{nm} (I_n \otimes \Sigma_{nm0})] = O\left( \frac{1}{T} \right);
\]

\[
\frac{1}{n} \text{vec}(\tilde{V}_{nm,T}'B_{nm} \text{vec}(\tilde{V}_{nm,T}) - E \left[ \frac{1}{n} \text{vec}(\tilde{V}_{nm,T}'B_{nm} \text{vec}(\tilde{V}_{nm,T}) \right] = O_p \left( \frac{1}{\sqrt{nT^2}} \right).
\]

(3)

\[
\frac{1}{nT} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}'B_{nm} \text{vec}(\tilde{V}_{nm,t}) - E \left[ \frac{1}{nT} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}'B_{nm} \text{vec}(\tilde{V}_{nm,t}) \right] = O_p \left( \frac{1}{\sqrt{nT}} \right).
\]
With the identification in and the uniform convergence, consistency of the QML estimator can be established in the following theorem.

**Theorem 2.1.** Under assumptions 2.2.1, 2.3.1-2.3.3, 2.3.4 and 2.3.5, the QMLE for the MDP-SAR model is consistent.

### 2.3.3 Asymptotic Normality of QMLE

Asymptotic distribution of the QML estimator $\hat{\theta}$ will be useful for statistical inference. The asymptotic distribution of $\hat{\theta}$ can be derived from the Taylor expansion of the log likelihood function, given by the mean value theorem

$$\hat{\theta} - \theta_0 = -\left[\frac{\partial^2 \ln L_{nT,m}(\bar{\theta})}{\partial \theta \partial \theta'}\right]^{-1} \frac{\partial \ln L_{nT,m}(\theta_0)}{\partial \theta}.$$

In appendix we provide detailed formula of first order derivatives and second order derivatives. When evaluated at the true value $\theta_0$, we can have the decomposition

$$\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} = \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^1}{\partial \theta} + \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \theta}$$

into two components, where the first component captures the score with zero mean, and the remaining component is mainly for possible asymptotic bias. By inspection,

$$\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^1}{\partial \theta}$$

is a statistic of the general form:

$$\sum_{t=1}^T [\mathbb{U}_{nm,t-1} \text{vec}(V_{nm,t}') + \text{vec}(V_{nm,t}') B_{nm} \text{vec}(V_{nm,t}') + D_{nm,t} \text{vec}(V_{nm,t}')] - \text{Tr}(B_{nm}(I_n \otimes \Sigma_{nm0}))],$$

where $\mathbb{U}_{nm,t-1}$ is defined in Appendix D, as $\mathbb{U}_{nm,t} = \sum_{h=1}^{+\infty} G_{nm,h} \text{vec}(V_{nm,t+1-h}') = \sum_{h=1}^{+\infty} G_{nm,c} G_{nm,d}^h \text{vec}(V_{nm,t+1-h}')$ where $G_{nm,h}$ $G_{nm,c}$ and $G_{nm,d}$ are $nm \times nm$ matrices, and $G_{nm,c}$ and $\sum_{h=1}^{+\infty} \text{abs}(G_{nm,d}^h)$ are bounded in row and column sum norms. $D_{nm,t}$ is a $nm \times 1$ constant vector with uniformly bounded elements, and $B_{nm}$ is a $nm \times nm$ constant matrix with uniformly bounded row and column sum norms.
It differs from the linear-quadratic form in Yu, de Jong and Lee (2008) in that the elements of $vec(V'_{nm,t})$ are not independently distributed. Therefore, the central limit theorem with i.i.d. random variables for their linear-quadratic form needs to be extended to the multivariate case.

**Lemma 2.3.** Under assumptions 2.3.1 and 2.3.3, the fourth moment of the any entry of any $U_{nm,t−1}$ is uniformly bounded, i.e., for any $i = 1,..,n$ and $k = 1, ..., m$, $E|u_{nk,t−1,i}|^4 = O(1)$ uniformly in $k$ and $i$.

**Lemma 2.4.** Suppose $Q_{nT,m} = \sum_{t=1}^{T} [U'_{nm,t−1} vec(V'_{nm,t}) + vec(V'_{nm,t})' B_{nm} vec(V'_{nm,t}) + D'_{nm,t} vec(V'_{nm,t})] = Tr(B_{nm} \Sigma_{vm})$, where $U_{nm,t−1} = \sum_{h=1}^{+\infty} G_{nm,h} vec(V'_{nm,t−h})$, $D_{nm,t}$ is a constant vector with uniformly bounded elements, and $B_{nm}$ is a constant matrix with uniformly bounded row and column sums. $V_{nm,t} = (V_{n1,t}, ..., V_{nm,t})$ where $V_{nk,t}$ and $V_{nl,t}$ are $n \times 1$ random vectors with $E(V_{nk,t} V'_{nl,t}) = \sigma_{kl}^2 I_n$ for any $k, l = 1, 2, ..., m$, which satisfy Assumption 1. Furthermore, suppose that the variance $\sigma^2_{Q_{nT,m}}$ of $Q_{nT,m}$ is $O(nT)$, and $\frac{1}{nT^2} \sigma^2_{Q_{nT,m}}$ is bounded away from zero. Then, $\frac{Q_{nT,m}}{\sigma_{Q_{nT,m}}} \xrightarrow{d} N(0, 1)$.

The variance matrix of the normalized score vector is (details in Appendix)

$$var \left( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L^1_{nT,m0}}{\partial \theta} \right) = \Omega_{\theta_0,nT} + \Xi_{\theta_0,nT} + O \left( \frac{1}{T} \right).$$

Denote $\Omega_{\theta_0} = \lim_{T \to +\infty} \Omega_{\theta_0,nT}$ and $\Xi_{\theta_0} = \lim_{T \to +\infty} \Xi_{\theta_0,nT}$. Thus, as

$$\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L^1_{nT,m0}}{\partial \theta} \xrightarrow{d} N(0, \Omega_{\theta_0} + \Xi_{\theta_0}).$$

For the case that the disturbances are normally distributed, $\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L^1_{nT,m0}}{\partial \theta} \xrightarrow{d} N(0, \Omega_{\theta_0}).$
Excluding the scalar factor \( \sqrt{\frac{T}{n-1}} \), terms of \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} \) have the expressions in the forms (1) \( \text{vec}(\bar{V}_{nm,T}')\bar{B}_{nm} \text{vec}(\bar{V}_{nm,T}) \); and/or (2) \( \bar{U}_{nm,T} \text{vec}(\bar{V}_{nm,T}) \), where \( \bar{B}_{nm} \) is an \( n \times m \) matrix with bounded row and column sum norms, and \( \bar{U}_{nm,T} = \frac{1}{T} \sum_{t=0}^{T-1} \sum_{h=1}^{+\infty} G_{nm,h} \text{vec}(V_{nm,t-h}) \).

According to Lemma D.8,

\[
\bar{U}_{nm,T} \text{vec}(\bar{V}_{nm,T}) = E(\bar{U}_{nm,T} \text{vec}(\bar{V}_{nm,T})) + O_p \left( \frac{\sqrt{n}}{T^2} \right)
\]

Collecting \( \text{Tr}(\Sigma_{vm0}) \) and \( \frac{1}{n} \text{Tr}((I_n \otimes \Sigma_{vm0}) G_{nm,c} \sum_{g=1}^{+\infty} G_{nm,d}) \) as entries in \( \Delta_R \), we can thus summarize the second component as \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} - \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT.m}}{\partial \theta \partial \theta'} = \sqrt{\frac{n-1}{T}} \Delta_R + O_p \left( \frac{1}{T^2} \right) \), where the explicit expressions for entries in \( \Delta_R \) corresponds to those in Appendix E.

For the normalized second order derivatives \( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m}}{\partial \theta \partial \theta'} \), it has the following regular convergence properties which are needed for the derivation of asymptotic distribution of the QMLE.

**Lemma 2.5.** Under Assumptions 2.2.1, 2.3.1-2.3.3,

(1) For any consistent estimate \( \tilde{\theta} \) of \( \theta_0 \), \( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m}(\tilde{\theta})}{\partial \theta \partial \theta'} - \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m0}}{\partial \theta \partial \theta'} = o_p(1) \);

(2) \( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m0}}{\partial \theta \partial \theta'} - \Omega_{\theta_0} = o_p(1) \).

Now define \( \Delta_\theta = \Omega_{\theta_0}^{-1} \Delta_R \), the QMLE \( \hat{\theta} \) has the following asymptotic expansion.
Theorem 2.2. Under Assumptions 2.2.1, 2.3.1-2.3.3, when \( T \to +\infty \),

\[
\sqrt{(n-1)T}(\hat{\theta} - \theta_0) - \sqrt{n-1} \frac{T}{T^3} \Delta_\theta + O_p \left( \max \left( \sqrt{\frac{n-1}{T^3}}, \sqrt{\frac{1}{T}} \right) \right) \\
\overset{d}{\to} N(0, \Omega_{\theta_0}^{-1}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega_{\theta_0}^{-1}).
\]

Consequently, (1) if \( n/T \to 0 \), then the bias \( \sqrt{n-1} \Delta_\theta \) disappears, and \( \sqrt{(n-1)T}(\hat{\theta} - \theta_0) \overset{d}{\to} N(0, \Omega_{\theta_0}^{-1}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega_{\theta_0}^{-1}) \);

(2) if \( n/T \to M \), which is a finite positive constant, \( \sqrt{(n-1)T}(\hat{\theta} - \theta_0) - \sqrt{M} \Delta_\theta \overset{d}{\to} N(0, \Omega_{\theta_0}^{-1}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega_{\theta_0}^{-1}) \);

(3) if \( n/T \to \infty \), \( T(\hat{\theta} - \theta_0) - \Delta_\theta \overset{p}{\to} 0 \).

For the case that the disturbances are normally distributed, \( \Xi_{\theta_0} = 0 \).

Bias Corrected Estimator

When \( \frac{n}{T} \to M > 0 \), the QMLE is not asymptotically centered at zero. We can reduce the asymptotic bias by bias reduction. The asymptotic bias is \( -\frac{1}{T} \Delta_\theta \). We define the bias corrected estimator as \( \hat{\theta}^c = \hat{\theta} + \frac{1}{T} \Delta_\theta(\hat{\theta}) \), where the bias term is \( \Delta_\theta(\hat{\theta}) = \left[ E \left( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{n,T,m}(\hat{\theta})}{\partial \theta \partial \theta'} \right) \right]^{-1} \Delta R(\hat{\theta}) \). In the following theorem, we show that when \( T/n^3 \to \infty \), the bias corrected estimator \( \hat{\theta}^c \) is \( \sqrt{nT} \) consistent and asymptotically normal centered at zero. This result holds when \( n/T \to M \) as \( M \) is a constant and when \( n/T \to \infty \). Additional assumption is needed to guarantee asymptotic normality.

Assumption 2.3.6. The row and column sum norms of \( \sum_{g=1}^{\infty} \sum_{h=0}^{g-1} H_{nm}^{h} S^{-1}(W_n \otimes E_{m,ij})H_{nm}^{g-1-h} \) and \( \sum_{g=1}^{\infty} \sum_{h=0}^{g-1} H_{nm}^{h} S^{-1}(I_n \otimes E_{m,ij}')H_{nm}^{g-1-h} \) are bounded uniformly in absolute value in the parameter space for \( i, j = 1, ..., m \).
The asymptotic normality is achieved when \( n \) grows not too faster than \( T \) under Assumption 2.3.6.

**Theorem 2.3.** *Under Assumptions 2.2.1 and 2.3.1-2.3.6, when \( n/T^3 \to 0 \), the QMLE with bias correction, \( \hat{\theta}_c \), has*

\[
\sqrt{(n-1)T}(\hat{\theta}_c - \theta_0) \xrightarrow{d} N(0, \Omega^{-1}_{\theta_0}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega^{-1}_{\theta_0}).
\]

### 2.4 Identification and QML Estimation of the Spatially Conintegrated MDP-SAR Model

In the presence of unit roots in the equations, there are features of non-stability for the system. As dependent variables are interrelated and there are spatial interactions, cointegration among variables and across space may occur. Traditional cointegration analysis has emphasized on how to detect cointegration. In addition to detect cointegration, our analysis will focus on the model when cointegration is presented and its consequence on asymptotic properties of the QMLE. In this section, we provide a revealing ECM representation and an unit-root view on rotations across multivariate variables and over space. Then, we formally state assumptions, reparameterize the model and establish asymptotic properties of the QML estimator. The reparameterization generalizes that for a univariate model in Yu et al. (2012) to the multivariate case. The reparameterization was motivated by Sims et al. (1990) and provides a more tractable way for asymptotic analysis.
2.4.1 Error Correction Model Representation

To obtain an ECM representation, we subtract a time lagged term from both sides of the equation $\text{vec}(Y_{nm,t}') = H_{nm}\text{vec}(Y_{nm,t-1}') + S_{nm}^{-1}[Q_{nm,t} + \text{vec}(D_{m,t}')],$ which is the MDP-SAR model in (2.2) with abbreviation. It follows $\text{vec}(Y_{nm,t}') - \text{vec}(Y_{nm,t-1}') = (H_{nm} - I_{nm})\text{vec}(Y_{nm,t-1}') + S_{nm}^{-1}[Q_{nm,t} + \text{vec}(D_{m,t}')].$ As

$$H_{nm} - I_{nm} = S_{nm}^{-1}[I_{n} \otimes (P_{m}' - I_{m})' + W_{n} \otimes (\Psi_{m}' + \Phi_{m}')'] = (I_{n} - W_{n}) \otimes (P_{m}' - I_{m})' + W_{n} \otimes (\Psi_{m}' + \Phi_{m}')',$$

where $A_{nm}^{0} = I_{n} \otimes (P_{m}' - I_{m})' + W_{n} \otimes (\Psi_{m}' + \Phi_{m}')' = (I_{n} - W_{n}) \otimes (P_{m}' - I_{m})' + W_{n} \otimes (\Psi_{m}' + \Phi_{m}' + P_{m}' - I_{m})'$. The above expression becomes

$$\text{vec}(Y_{nm,t}') - \text{vec}(Y_{nm,t-1}') = S_{nm}^{-1}A_{nm}^{0}\text{vec}(Y_{nm,t-1}') + S_{nm}^{-1}[Q_{nm,t} + \text{vec}(D_{m,t}')].$$

In the cointegration case, the components $\text{vec}^{(u)}(Y_{nm,t}')$ and $\text{vec}^{(a)}(Y_{nm,t}')$ in (2.4) of $\text{vec}(Y_{nm,t}')$ may be unstable. This is so because some of the eigenvalues of $(I_{m} - \Psi_{m})^{-1}(P_{m} + \Phi_{m})$ are ones. We see that

$$\text{vec}^{(u)}(Y_{nm,t}') = \{I_{n} \otimes [(I_{m} - \Psi_{m})^{-1}(P_{m} + \Phi_{m})]'(I_{m} - \Psi_{m})^{-1}(P_{m} + \Phi_{m})]'[I_{m} - \Psi_{m}]^{-1}\alpha_{m,t-h}\}
+ \sum_{h=0}^{t}\{I_{n} \otimes [(I_{m} - \Psi_{m})^{-1}(P_{m} + \Phi_{m})]'[I_{m} - \Psi_{m}]^{-1}\alpha_{m,t-h}\},$$

may have unstable in terms of time tend due to the summation $\sum_{h=0}^{t}\{I_{n} \otimes [(I_{m} - \Psi_{m})^{-1}(P_{m} + \Phi_{m})]'[I_{m} - \Psi_{m}]^{-1}\alpha_{m,t-h}\},$ in the presence of unit roots in the matrix product $(I_{m} - \Psi_{m})^{-1}(P_{m} + \Phi_{m})'.$ These unstable components of $\text{vec}(Y_{nm,t}')$ can be taken by linear combinations to become stable. To see this possibility, we note that $(\Psi_{m} + \Phi_{m} + P_{m} - I_{m})'[I_{m} -
\[
\Psi_m^{-1}(P_m+\Phi_m)^{\prime \prime} = (I_m-\Psi_m')[(I_m-\Psi_m)^{-1}(P_m+\Phi_m)'-I_m][(I_m-\Psi_m)^{-1}(P_m+\Phi_m)]^{\prime h}.
\]

Suppose \( \lambda_m \) is an eigenvalue of \((I_m-\Psi_m)^{-1}(P_m+\Phi_m)'\), then the eigenvalue of \([(I_m-\Psi_m)^{-1}(P_m+\Phi_m)'-I_m][(I_m-\Psi_m)^{-1}(P_m+\Phi_m)]^{\prime h}\) should be \((\lambda_m-1)^{\lambda_m^h}\), which will be 0 for an eigenvalue \( \lambda_m = 1 \), or less than 1 in absolute value when \(|\lambda_m| < 1\). Indeed, \( \lambda_m^h \) geometrically declines in \( h \). In addition to \((I_n-W_n)I_n = 0\), and \((I_n-W_n)\Gamma_n\mathbb{1}_{n,n_1}\Gamma_n^{-1} = 0\). Thus \( A_{nm,\text{vec}(\alpha)}(Y_{nm,t})' \) and \( A_{nm,\text{vec}(\alpha)}(Y_{nm,t})' \) become stable\(^{18}\). In consequence, \( A_{nm,\text{vec}}(Y'_{nm,t}) \) is stable and we have cointegration.

The cointegration rank of the system is the rank of the matrix \( A_{nm}^h \). Note \( A_{nm}^h = I_n \otimes (P_m-I_m)' + W_n \otimes (\Psi_m+\Phi_m)' \). With \( W_n = \Gamma_n \bar{\omega}_n \Gamma_n^{-1}, A_{nm}^h = (\Gamma_n \otimes I_m)(I_n \otimes (P_m-I_m)' + \bar{\omega}_n \otimes (\Psi_m+\Phi_m)')(\Gamma_n^{-1} \otimes I_m) \). Multiply by \( S_{nm}^{-1} \),

\[
S_{nm}^{-1}[I_n \otimes (P_m-I_m)' + W_n \otimes (\Psi_m+\Phi_m)'] = (\Gamma_n \otimes I_m) \begin{pmatrix} A_{n1,m} & 0 \\ 0 & A_{n2,m} \end{pmatrix} (\Gamma_n^{-1} \otimes I_m),
\]

where

\[
A_{n1,m} = I_{n1} \otimes (I_m-\Psi_m)^{-1}(P_m-I_m+\Psi_m+\Phi_m)',
\]

and

\[
A_{n2,m} = (I_{n2} \otimes I_m - \bar{\omega}_{n2} \otimes \Psi_m)^{-1}[I_{n2} \otimes P_m' + \bar{\omega}_{n2} \otimes (\Psi_m+\Phi_m)'-I_{n2} \otimes I_m],
\]

with \( \bar{\omega}_{n2} = \text{diag}\{0, \ldots, 0, w_{n,n_1+1}, \ldots, w_{n,n}\} \). The \( n_2m \)-dimensional square matrix \( A_{n2,m} \) can have full rank, \( n_2m \), as discussed earlier. As \( (I_m-\Psi_m')^{-1}(P_m-I_m+\Psi_m+\Phi_m)' = [(I_m-\Psi_m)^{-1}(P_m+\Phi_m)'-I_m] \) has eigenvalues \( \lambda_m-1 \) where \( \lambda_m \) are eigenvalues of \((I_m-\Psi_m)^{-1}(P_m+\Phi_m)'\). \( A_{n1,m} \) has rank \( n_1(m-m_1) \) where \( m_1 \) is the number of unit eigenvalues of \((I_m-\Psi_m)^{-1}(P_m+\Phi_m)'\). Therefore, cointegration rank of the system is \( nm-n_1m_1 \).

\(^{18}\)Strictly speaking, for the claim that \( A_{nm,\text{vec}(\alpha)}(Y'_{nm,t}) \) to be stable, we need to exclude possible explosive behavior for \( \alpha_{m,t} \) over time \( t \). For estimation, we have \( \alpha_{m,t} \) eliminated, so its behavior will be inconsequential.
2.4.2 Alternative Representation: Unit Roots View on Rotations across Multivariate Variables and over Space

We can rotate the dependent variables, which are linear combinations of variables, in order to explicitly view how unit root processes are generated due to unit eigenvalues of $W_n$ and $(I_m - \Psi_m)^{-1}(P_m + \Phi_m)\Psi_m$. This viewpoint is in spirit of the single variable counterpart of Yu, de Jong and Lee (2012).19

The diagonalizability of $W_n$ gives $W_n = \Gamma_n \bar{\omega}_n \Gamma_n^{-1}$, where $\bar{\omega}_n$ has $n_1$ unit eigenvalues and the remaining $n_2 = (n - n_1)$ eigenvalues are strictly less than one in absolute value. The original model system

$$Y_{nm,t} = W_n Y_{nm,t} Y_{nm,t} + Y_{nm,t-1} P_m + W_n Y_{nm,t-1} \Phi_m + U_{nm,t},$$

where $U_{nm,t}$ represents the remaining model components (regressors, time fixed effects, individual fixed effects and idiosyncratic disturbances) for simplicity. Applying the transformation $\Gamma_n^{-1}$ to $Y_{nm,t}$, we arrive at $Y_{nm,t}^+ = \bar{\omega}_n Y_{nm,t}^+ + Y_{nm,t-1}^+ P_m + \bar{\omega}_n Y_{nm,t-1}^+ \Phi_m + \Gamma_n^{-1} U_{nm,t}$, where $Y_{nm,s}^+ = \Gamma_n^{-1} Y_{nm,s}$ for any $s$. With $\bar{\omega}_n = \begin{pmatrix} I_{n_1} & 0 \\ 0 & \bar{\omega}_{n_2} \end{pmatrix}$, decompose $\Gamma_n^{-1} = \begin{pmatrix} \Gamma_{n_1} \\ \Gamma_{n_2} \end{pmatrix}$ and $Y_{nm,s}^+ = \begin{pmatrix} Y_{n_{1m,s}}^+ \\ Y_{n_{2m,s}}^+ \end{pmatrix}$, so we have

$$Y_{n_{1m,t}}^+ = Y_{n_{1m,t}}^+ \Psi_m + Y_{n_{1m,t-1}}^+ P_m + Y_{n_{1m,t-1}}^+ \Phi_m + \Gamma_{n_1} U_{nm,t},$$
$$Y_{n_{2m,t}}^+ = \bar{\omega}_{n_2} Y_{n_{2m,t}}^+ \Psi_m + Y_{n_{2m,t-1}}^+ P_m + \bar{\omega}_{n_2} Y_{n_{2m,t-1}}^+ \Phi_m + \Gamma_{n_2} U_{nm,t}. \quad (2.8)$$

The first $n_1$ equations of (2.8) imply that

$$Y_{n_{1m,t}}^+ = Y_{n_{1m,t-1}}^+ (P_m + \Phi_m) (I_m - \Psi_m)^{-1} + \Gamma_{n_1} U_{nm,t} (I_m - \Psi_m)^{-1}.$$

As $(P_m + \Phi_m) (I_m - \Psi_m)^{-1} = S_m \Lambda_m S_m^{-1}$ for some $S_m$ nonsingular matrix, where eigenvalue matrix $\Lambda_m$ has $m_1$ unit eigenvalues, the unit values can be arranged

19That was originally pointed out by Professor Peter Phillips in conversation.
such that \( \Lambda_m = \begin{pmatrix} I_{m_1} & 0 \\ 0 & \Lambda_{m_2} \end{pmatrix} \), i.e., the first \( m_1 \) eigenvalues are 1 and the remaining ones are smaller than 1 in absolute values. The above subsystem can then be transformed into

\[
Y^+_{n_1 m, t} S_m = Y^+_{n_1 m, t-1} S_m \Lambda_m + \Gamma_{n_1} U_{n m, t} (I_m - \Psi_m)^{-1} S_m.
\]

Denote \( Y^*_{n_1 m, t} = Y^+_{n_1 m, t} S_m \) and, conformably \( Y^*_{n_1 m, t} = (Y^*_{n_1 m_1, t}, Y^*_{n_1 m_2, t}) \) and \( S_m = [S_{m_1}, S_{m_2}] \). Then we have the following unit root subsystem,

\[
Y^*_{n_1 m_1, t} = Y^*_{n_1 m_1, t-1} + \Gamma_{n_1} U_{n m, t} (I_m - \Psi_m)^{-1} S_{m_1},
\]

and the remaining stable components,

\[
Y^*_{n_1 m_2, t} = Y^*_{n_1 m_2, t-1} \Lambda_{m_2} + \Gamma_{n_1} U_{n m, t} (I_m - \Psi_m)^{-1} S_{m_2}.
\]

Recall that

\[
Y^+_{n_2 m, t} = \omega_{n_2} Y^+_{n_2 m, t} \Psi_m + Y^+_{n_2 m, t-1} P_m + \omega_{n_2} Y^+_{n_2 m, t-1} \Phi_m + \Gamma_{n_2} U_{n m, t}
\]

is also a stable subsystem. In sum, the transformed system can be divided into three subsystem, the first one consists of unit root processes while the remaining two are stable. The unit root subsystem contains \( n_1 m_1 \) processes while the stable subsystem has \( nm - n_1 m_1 \) equations, where the number of equations equals to the cointegration rank.

### 2.4.3 Assumptions

We list some regularity conditions for the cointegration case.

**Assumption 2.4.1.** \( v_{n, t, i} \), the \( i \)th row of \( V_{n,m,t} \), of dimension \( m \), for \( i = 1, \ldots, n \) are i.i.d. \((0, \Sigma_{vm})\). The \( E|v_{n, t, i k} v_{n, t, il} v_{n, t, ip} v_{n, t, iq}|^{1+\delta} \) is uniformly bounded for \( i = 1, 2, \ldots, n \) and \( k, l, p, q = 1, 2, \ldots, m \) for some positive constants \( \delta \).
Assumption 2.4.2. Elements of $X_{nk,t}$ are exogenous constants and are uniformly bounded for all $n$ and $t$. Elements of individual effects in $C_{nm}$ are uniformly bounded for all $n$.

Assumption 2.4.1 is on stochastic properties of disturbances which are the same of stable case in Assumption 2.3.1. Assumption 2.4.2 is on exogenous regressors which are the same of stable case in Assumption 2.3.2.

Define $E_{m,ij}$ represent zero $m \times m$ matrix but the $(i,j)$th element equals to one,

$X_{nm,t} = \sum_{h=0}^{t-1} (I_n \otimes [(I_m - \Psi_{m0}^{-1}(\Phi_{m0} + P_{m0})')h(I_m - \Psi_{m0}')^{-1}])(X_{nk,t-h} \otimes I_m)vec(\Pi_{km0})$

and $C_{nm,t} = \sum_{h=1}^{t-1} (I_n \otimes [((I_m - \Psi_{m0}')^{-1}(\Phi_{m0} + P_{m0})')h(I_m - \Psi_{m0}')^{-1}]vec(C_{nm})$. Define $\bar{X}_{nm,T} = \frac{1}{T} \sum_{t=1}^{T} X_{nm,t}$ and $\bar{C}_{nm,T} = \frac{1}{T} \sum_{t=1}^{T} C_{nm,t}$. Also, $\tilde{X}_{nm,t} = X_{nm,t} - \bar{X}_{nm,T}$ and $\tilde{C}_{nm,t} = C_{nm,t} - \bar{C}_{nm,T}$. Define a $m^2 \times m^2$ matrix $H_{X,C,m^2}$ with the element being $[\tilde{X}_{nm,t} + \tilde{C}_{nm,t}]'[(\Gamma_n \mathbb{I}_{n,n_1} \Gamma_n^{-1})'E_{ij,m}(E_{kl,m}^{-1})'][(\Gamma_n \mathbb{I}_{n,n_1} \Gamma_n^{-1})][\tilde{X}_{nm,t} + \tilde{C}_{nm,t}]$.

Assumption 2.4.3. $n$ is a nondecreasing function of $T$, where $T$ tends to infinity. The limit $m^2 \times m^2$ matrix, $\lim_{T \to \infty} \frac{1}{m^2T} \sum_{t=1}^{T} H_{X,C,m^2}$, exists and is nonsingular.

In the presence of unit roots, individual effects and non-zero means of regressors would generate deterministic trend in $vec^{(n)}(Y_{nm,t})$, while disturbances generate a stochastic trend. It is known that a deterministic time trend will dominate the stochastic trend in terms of probability order (see, e.g., Hamilton (1994)). Thus, the deterministic trend is the dominate source of non-stability. Assumption 2.4.3 provides a proper scale normalization on parts of the asymptotic variance of the QMLE. The assumption is essential for asymptotic normal distribution of the QMLE because if there were no deterministic trend, the stochastic
trend in \( \text{vec}^{(u)}(Y_{nm,t}) \) would become the sole force of non-stability, which would change the convergence rate as well as the asymptotic distribution of estimators. \(^{20}\)

2.4.4 Reparameterization and Concentrated Log Likelihood Function

We use the QML method to estimate parameters of the model. For analytical purpose on asymptotic distribution of estimators, we first reparameterize the model so that all unstable components can be combined into one variable. Then, we maximize the log likelihood function with time fixed effects eliminated and individual fixed effects concentrated. The QMLEs of parameters in front of stable variables should be \( \sqrt{\frac{n}{T}} \) consistent while those in front of unstable variables can be superconsistent. The advantage of the reparameterization procedure is that asymptotic properties of the QMLE can be established similarly to that of stable case in Section 3 once scale normalization for each of the unstable and stable variables can be properly done so that the corresponding limited information matrix is of full rank.

By reparameterization, \( P^*_m = P_m + \Psi_m + \Phi_m \), the model (2) with abbreviation becomes

\[
\text{vec}(Y'_{nm,t}) = (I_n \otimes \Psi'_m)[(W_n \otimes I_m)\text{vec}(Y'_{nm,t}) - \text{vec}(Y'_{nm,t-1})]
+ (I_n \otimes \Phi'_m)[(W_n \otimes I_m)\text{vec}(Y'_{nm,t-1}) - \text{vec}(Y'_{nm,t-1})] \\
+ (I_n \otimes P^*_m)\text{vec}(Y'_{nm,t-1}) + \text{vec}(D'_{nm,t-1}) + Q_{nm,t}. \tag{2.9}
\]

\(^{20}\)This case may exist when there were no fixed effects and no exogenous regressors. This scenario is not included in this paper.
First, the transformed variable \((W_n \otimes I_m)vec(Y'_{nm,t}) - vec(Y'_{nm,t-1})\) is stable because
\[
(W_n \otimes I_m)vec(Y'_{nm,t-1}) - vec(Y'_{nm,t-1}) = (W_n \otimes I_m)vec(\alpha'(Y'_{nm,t-1})) - vec(\alpha'(Y'_{nm,t-1})).
\]
This follows, because \((W_n \otimes I_m)(\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes I_m) = \Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes I_m\), we have
\[
(W_n \otimes I_m)vec^{(u)}(Y'_{nm,t}) = vec^{(u)}(Y'_{nm,t}), \text{ and } (W_n \otimes I_m)vec^{(a)}(Y'_{nm,t}) = vec^{(a)}(Y'_{nm,t}).
\]
Second, \([(W_n \otimes I_m)vec(Y'_{nm,t}) - vec(Y'_{nm,t-1})]\) is also stable. This can be justified because
\[
vec^{(u)}(Y'_{nm,t}) - vec^{(u)}(Y'_{nm,t-1})
\]
\[
= (\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)]^t[(I_m - \Psi_m')^{-1}(P_m + \Phi_m)' - I_m])vec(Y'_{nm,-1})
\]
\[
+ \sum_{h=0}^{t-1} \{\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)]^h(I_m - \Psi_m')^{-1}\}Q_{nm,t-h}
\]
\[
- \sum_{h=0}^{t-1} \{\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)]^h(I_m - \Psi_m')^{-1}\}Q_{nm,t-1-h}
\]
\[
= (\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)]^t[(I_m - \Psi_m')^{-1}(P_m + \Phi_m)' - I_m])vec(Y'_{nm,-1})
\]
\[
+ \{\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes (I_m - \Psi_m')^{-1}\}Q_{nm,t}
\]
\[
+ \sum_{h=0}^{t-1} \{\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)]^h
\]
\[
\times [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)' - I_m](I_m - \Psi_m')^{-1}\}Q_{nm,t-1-h};
\]
and
\[
vec^{(a)}(Y'_{nm,t}) - vec^{(a)}(Y'_{nm,t-1}) = l_n \otimes (I_m - \Psi_m')^{-1}\alpha_m'
\]
\[
+ l_n \otimes \sum_{h=0}^{t-1} [(I_m - \Psi_m')^{-1}(P_m + \Phi_m)]^h[(I_m - \Psi_m')^{-1}(P_m + \Phi_m)' - I_m](I_m - \Psi_m')^{-1}\alpha_m't_{-1-h}.
\]
In the above two equations, \([(I_m - \Psi_m')^{-1}(P_m + \Phi_m)]^h[(I_m - \Psi_m')^{-1}(P_m + \Phi_m)' - I_m]\)
has eigenvalues \(\lambda_m^h(\lambda_m - 1)\), which eliminates the unit eigenvalues. So the only
unstable term in (2.9) is \(vec(Y_{nm,t-1})\).
The concentrated log likelihood in (2.6) can be expressed as

$$\frac{1}{n-1} \ln \frac{1}{T} L_{nT,m} = -\frac{m}{2} \ln (2\pi) + \frac{1}{n-1} \ln |S_{nm}| - \frac{1}{n-1} \ln |I_m - \Psi_m'|$$

$$- \frac{1}{2} \ln |\Sigma_m| - \frac{1}{2(n-1)T} \sum_{t=1}^{T} \text{vec}(\tilde{V}_t(\theta))'(J_n \otimes \Sigma_{vm}^{-1}) \text{vec}(\tilde{V}_t(\theta)),$$

where \( \text{vec}(\tilde{V}_t(\theta)) = \text{vec}(\tilde{Y}'_{nm,t}) - (I_n \otimes \Psi_m')([W_n \otimes I_m] \text{vec}(\tilde{Y}'_{nm,t}) - \text{vec}(\tilde{Y}'_{nm,t-1})]$$

$$- (I_n \otimes \Phi_m')([W_n \otimes I_m] \text{vec}(\tilde{Y}'_{nm,t-1}) - \text{vec}(\tilde{Y}'_{nm,t-1}))$$

$$- (I_n \otimes P_m') \text{vec}(\tilde{Y}'_{nm,t-1}) - (\tilde{X}_{nk,t} \otimes I_m) \text{vec}(\Pi_{km}).$$

(2.10)

### 2.4.5 Identification and Consistency

The identification of parameters can be similar to that in stable case with a few modifications. The key difference is that \( \tilde{Y}'_{nm,t-1} \) needs to be normalized by \( T \) since it has a time trend.

Define \( \tilde{Z}_{nm,t}^{(0)} = [W_n \tilde{Y}'_{nm,t-1} - \tilde{Y}'_{nm,t-1}, \frac{1}{T} \tilde{Y}'_{nm,t-1}, \tilde{X}_{nk,t}] \), and \( \tilde{A}_{z,t} = [\tilde{a}_1, ..., \tilde{a}_m] \) where

$$\tilde{a}_j = [I_n \otimes e'_{m,j}] \left\{ \left( W_n \otimes I_m \right) S_{nm0}^{-1} \left[ (I_n \otimes \Psi_{m0}') [W_n \otimes I_m] \text{vec}^{(s)}(\tilde{Y}'_{nm,t-1}) \right. \right.$$

$$\left. - \text{vec}^{(s)}(\tilde{Y}'_{nm,t-1}) \right] + (I_n \otimes P_{m0}') \text{vec}^{(s)}(\tilde{Y}'_{nm,t-1}) + (\tilde{X}_{nk,t} \otimes I_m) \text{vec}(\Pi_{km0}) \right\}$$

$$- (I_n \otimes \Psi_{m0}') \text{vec}^{(s)}(\tilde{Y}'_{nm,t-1})$$

$$+ I_n \otimes [(I_m - \Psi'_{m0})^{-1}(\Phi_{m0}' + P_{m0}') - I_m] \text{vec}^{(u)}(\tilde{Y}'_{nm,t-1}).$$

The following assumptions are for identification. The first assumption states that the best instruments for all endogenous variables exist and there is no multicollinearity across them. The second assumption is similar to Assumption 2.3.5, which explores information from the structure of disturbance terms.
Assumption 2.4.4. The limit matrix

\[
\lim_{T \to \infty} \frac{1}{(n-1)T} \sum_{t=1}^{T} E \left[ \begin{pmatrix} \tilde{Z}_{nm,t}^{(0)}' J_{n} \tilde{Z}_{nm,t}^{(0)} & \tilde{Z}_{nm,t}^{(0)}' J_{n} \tilde{A}_{z,t} \\ \tilde{A}_{z,t}' J_{n} \tilde{Z}_{nm,t}^{(0)} & \tilde{A}_{z,t}' J_{n} \tilde{A}_{z,t} \end{pmatrix} \right]
\]

exists and is nonsingular.

Assumption 2.4.5.

\[
\lim_{n \to +\infty} \left[ \left| S_{nm0}' S_{nm} (J_n \otimes \Sigma^{*-1}) S_{nm} S_{nm0}' (J_n \otimes \Sigma^{*}) \right| \right] < 0
\]

unless \( \Sigma_{vm} = \Sigma_{vm0} \) and \( \Psi = \Psi_{m0} \).

Proposition 2.3. (1) Under Assumption 2.4.1-2.4.3 and 2.4.4, \( \Psi_{m}, P_{m}^{*}, \Phi_{m} \) and \( \Pi_{km} \) can be identified;

(2) Under Assumption 2.4.1-2.4.3 and 2.4.5, \( \Psi_{m} \) and \( \Sigma_{vm} \) can be identified.

The uniform convergence of the objective functions can similarly be established with procedures in Section 3. This is because the model with transformed variable \( \text{vec}(\tilde{Y}_{nm,t-1}')/T \) has similar features for uniform convergence. In sum, we can establish consistency of the QMLEs

Theorem 2.4. Under assumptions 2.4.1-2.4.3, 2.4.4 and 2.4.5, the QML estimators

\( (\text{vec}(\hat{\Psi}_{m})' - \text{vec}(\Psi_{m0})', \text{vec}(\hat{\Phi}_{m})' - \text{vec}(\Phi_{m0})', T(\text{vec}(\hat{P}_{m})' - \text{vec}(P_{m0})')', \text{vec}^* (\hat{\Sigma}_{vm})' - \text{vec}^* (\Sigma_{vm0})')' \xrightarrow{p} 0 \), where \( \text{vec}^* \) means vectorization with distinct elements.

2.4.6 Asymptotic Distribution and Bias Correction

The asymptotic distribution of the QMLE \( \hat{\theta}^{*} \) of the transformed parameter

\( \theta^{*} = (\text{vec}(\Pi_{km})', \text{vec}(P_{m}')', \text{vec}(\Phi_{m}')', \text{vec}(\Psi_{m}')', \text{vec}^* (\Sigma_{vm})')' \) can be derived via the
expansion
\[
\hat{\theta}^* - \theta^*_0 = - \left[ T^{-1} \frac{\partial^2 \ln L_{nT,m}(\hat{\theta}^*)}{\partial \theta \partial \theta^*} T^{-1} \right]^{-1} T^{-1} \frac{\partial \ln L_{nT,m}(\theta^*_0)}{\partial \theta},
\]
where \( \theta^* = (\text{vec}(\Pi_{km}')', \text{vec}(P_m')', \text{vec}(\Phi_m')', \text{vec}(\Psi_m')', \text{vec}(\Sigma_{vm})')' \) and normalization matrix \( T = \text{diag}\{l_{km}', Tl_m', l_{2m^2+m(m+1)/2}'\} \). The \( \theta^* \) can be transformed back to \( \theta \), via
\[
P = \begin{pmatrix}
I_{km} & 0_{km \times m^2} & 0_{km \times m^2} & 0_{km \times m^2} & 0_{km \times m(m+1)/2} \\
0_{m^2 \times km} & I_{m^2} & -I_{m^2} & -I_{m^2} & 0_{m^2 \times m(m+1)/2} \\
0_{m^2 \times km} & 0_{m^2 \times m^2} & I_{m^2} & 0_{m^2 \times m^2} & 0_{m^2 \times m(m+1)/2} \\
0_{m^2 \times km} & 0_{m^2 \times m^2} & 0_{m^2 \times m^2} & I_{m^2} & 0_{m^2 \times m(m+1)/2} \\
0_{m(m+1)/2 \times km} & 0_{m(m+1)/2 \times m^2} & 0_{m(m+1)/2 \times m^2} & 0_{m(m+1)/2 \times m^2} & I_{m(m+1)/2} \\
\end{pmatrix}
\]

When evaluated at the true value \( \theta_0 \), we can decompose the first derivative \( T^{-1} \frac{\partial \ln L_{nT,m0}}{\partial \theta} = T^{-1} \frac{\partial \ln L_{nT,m0}^L}{\partial \theta} + T^{-1} \frac{\partial \ln L_{nT,m0}^R}{\partial \theta} \) to separate the possible asymptotic bias in the second component. Similarly to Section 3, we have \( T^{-1} \frac{\partial \ln L_{nT,m0}^R}{\partial \theta} = \sqrt{\frac{n-1}{T}} \Delta^R_0 + O \left( \sqrt{\frac{n}{T}} \right) + O_p \left( \sqrt{\frac{1}{T}} \right) \). The details are in Appendix E.

Applying Lemma 2.4 when \( T \to \infty \)
\[
T^{-1} \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} \to N(0, \Omega_{\theta_0}^{\text{co}} + \Xi_{\theta_0}^{\text{co}}),
\]
where the detailed expression \( \Omega_{\theta_0}^{\text{co}} \) and \( \Xi_{\theta_0}^{\text{co}} \) are very similar to those in stable case.

Combining the second order conditions of the log likelihood function, since
\[
T^{-1} \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m}}{\partial \theta \partial \theta^*} T^{-1} - T^{-1} \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m0}^L}{\partial \theta \partial \theta^*} T^{-1} = o_p(1), \text{ and } T^{-1} \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m0}^R}{\partial \theta \partial \theta^*} T^{-1} = \Omega_{\theta_0}^{\text{co}} = o_p(1),
\]
we arrive at the asymptotic distribution of the QMLE of model with multivariate spatial cointegration in the following theorem.

**Theorem 2.5.** Under Assumptions 2.2.2, 2.4.1-2.4.5, let \( \Delta_{\theta}^{\text{co}} = \Omega_{\theta_0}^{\text{co}} \Delta_0^{\text{co}} \) where \( \Omega_{\theta_0}^{\text{co}} \) is nonsingular and when \( T \to \infty \).
\[
\sqrt{(n-1)T} T (\hat{\theta}^* - \theta_0^*) - \sqrt{\frac{n-1}{T}} \Delta_{\theta}^{\text{co}} + O_p \left( \max \left( \sqrt{\frac{n-1}{T}}, \sqrt{\frac{1}{T}} \right) \right) \quad d \to N(0, \Omega_{\theta_0}^{\text{co}}^{-1}(\Omega_{\theta_0}^{\text{co}} + \Xi_{\theta_0}^{\text{co}})\Omega_{\theta_0}^{\text{co}}^{-1}).
\]
Consequently, (1) if \( n/T \to 0 \), then 
\[
\sqrt{(n-1)T} (\hat{\theta} - \theta_0^*) \xrightarrow{d} N(0, \Omega_{\theta_0}^{co-1}(\Omega_{\theta_0}^{co} + \Xi_{\theta_0}^{co})\Omega_{\theta_0}^{co-1});
\]
(2) if \( n/T \to M \), 
\[
\sqrt{(n-1)T} (\hat{\theta} - \theta_0^*) \xrightarrow{d} N(\sqrt{M} \Delta_{\theta_0}^{co}, \Omega_{\theta_0}^{co-1}(\Omega_{\theta_0}^{co} + \Xi_{\theta_0}^{co})\Omega_{\theta_0}^{co-1});
\]
(3) if \( n/T \to \infty \), 
\[
TT (\hat{\theta} - \theta_0^*) \xrightarrow{p} \Delta_{\theta_0}^{co}.
\]

The asymptotic distribution of \( \hat{\theta} \) is
\[
\sqrt{(n-1)T} (\hat{\theta} - \theta_0) - \mathbb{P}^{-1} \sqrt{\frac{n-1}{T}} \Delta_{\hat{\theta}}^{co} + O_p \left( \max \left( \sqrt{\frac{n-1}{T^3}}, \sqrt{\frac{1}{T}} \right) \right) \xrightarrow{d} N(0, \lim_{T \to \infty} \mathbb{P}^{-1} \Omega_{\theta_0}^{co-1}(\Omega_{\theta_0}^{co} + \Xi_{\theta_0}^{co})\Omega_{\theta_0}^{co-1} \mathbb{P}^{-1}).
\]

For the case that the disturbance terms are normally distributed, \( \Xi_{\theta_0}^{co} = 0 \).

The limiting covariance matrix in (2.11) is nonsingular while the limiting covariance matrix in (2.11) is singular. Similar to the stable case, the QMLEs may not asymptotically centered at zero due to bias term \( \Delta_{\theta_0}^{co} \) when \( n/T \to M > 0 \). But we can correct the biases using bias-corrected estimators. The asymptotic bias for \( \hat{\theta} \) is \( \mathbb{P}^{-1} \frac{1}{T} \Delta_{\hat{\theta}}^{co} \) from (2.11). We define the bias corrected estimator as
\[
\hat{\theta}^c = \hat{\theta} + \frac{1}{T} \Delta_{\hat{\theta}},
\]
where \( \Delta_{\hat{\theta}} = \left[ E \left( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{T,m}(\theta)}{\partial \theta \partial \theta'} \right) \right]^{-1} \mathbb{P} \Delta_{\hat{\theta}}^{co}(\hat{\theta}) \). In the following theorem, we show that when \( T/n^3 \to \infty \), the QMLE with bias correction \( \hat{\theta}^c \) is \( \sqrt{nT} \) consistent and asymptotically normal centered at zero. This result holds when \( n/T \to M \) as \( M \) is a constant number and when \( n/T \to \infty \). Additional assumption is needed to guarantee the asymptotical normality.

**Assumption 2.4.6.** The row and column sum norms of
\[
\sum_{g=1}^{\infty} \sum_{h=0}^{g-1} B_{nm}^{g-1-h} (W_n \otimes E'_{m,ij}) B_{nm}^{g-1-h} \quad \text{and} \quad \sum_{g=1}^{\infty} \sum_{h=0}^{g-1} B_{nm}^{h} S^{-1}(I_n \otimes E'_{m,ij}) B_{nm}^{g-1-h}
\]
are bounded uniformly in absolute value, uniformly in the parameter space for \( i, j = 1, ..., m \).
Theorem 2.6. Under Assumptions 2.4.1-2.4.6, when \( n/T^3 \to 0 \), the QMLE with bias correction, \( \hat{\theta}^c \), has

\[
\sqrt{(n - 1)T}(\hat{\theta}^c - \theta_0) \overset{d}{\to} N(0, \lim_{T \to \infty} P T^{-1}(\Omega_{\theta_0}^{\alpha} + \Xi_{\theta_0}^{\alpha})^{-1} T^{-1} P^\prime).
\]

2.5 Robust Estimation

The model can be decomposed into three components, where \( \text{vec}(u) (Y_{nm,t}) \) and \( \text{vec}(\alpha) (Y_{nm,t}) \) are viewed as possibly unstable components. However, even though \( \text{vec}(s) (Y_{nm,t}) \) is stable, the other two components can be “explosive”, which means some eigenvalues of \( (I_m - \Psi'_m)^{-1}(P_m + \Phi_m)' \) is larger than 1 while all the eigenvalues of \( B_{nm} \) are inside the unit circle. We call this case “explosive” case. When the true data generating process is in explosive case, both asymptotic properties of QMLE in stable case and spatial cointegration case are not correct. Another concern lies on Assumption 2.4.3, which is not satisfied when a subsystem of the model is stable and the other subsystem is spatially cointegrated. For example, consider the recursive system,

\[
Y_{1n,t} = \psi_{11} W_n Y_{1n,t} + \psi_{21} W_n Y_{2n,t} + p_{11} Y_{1n,t-1} + p_{21} Y_{2n,t-1} + \phi_{11} W_n Y_{1n,t-1} + \phi_{21} W_n Y_{2n,t-1} + U_{1n,t},
\]

\[
Y_{2n,t} = \psi_{22} W_n Y_{2n,t} + p_{22} Y_{2n,t-1} + \phi_{22} W_n Y_{2n,t-1} + U_{2n,t}.
\]

Suppose \(-1 < \psi_{22} + p_{22} + \phi_{22} < 1\) and \(\psi_{11} + p_{11} + \phi_{11} = 1\), then the subsystem of \(Y_{1n}\) is not stable while the subsystem of \(Y_{2n}\) can be stable. Therefore, \(\lim_{T \to \infty} \frac{1}{nT^3} \sum_{t=1}^{T} H_{X,C,m^2} \) exists but does not have the full rank. Hence the QMLE does not have the asymptotic properties in Theorem 2.5.
We introduce a transformed quasi-maximum likelihood estimation method, which is robust to some explosive cases and some cases when Assumption 2.4.3 is violated. So we call it robust quasi-maximum likelihood (RQML) estimation method. First, we make assumptions.

**Assumption 2.5.1.** $v_{n,t,i}$, the $i$th row of $V_{nm,t}$, is independent and identically distributed random vector of dimension $m$ with zero mean and covariance matrix $\Sigma_{vm}$. The elements of disturbances satisfy that $\mathbb{E}|v_{n,t,ik}v_{n,t,il}v_{n,t,ip}v_{n,t,iq}|^{1+\delta}$ is bounded for any $i = 1, 2, ..., n$ and $k, l, p, q = 1, 2, ..., m$ for some constant $\delta > 0$.

**Assumption 2.5.2.** Elements of $X_{nk,t}$ are exogenous constants and are uniformly bounded for all $n$ and $t$.

**Assumption 2.5.3.** $n$ is a nondecreasing function of $T$. Both $T$ and $n$ tends to infinity.

**Assumption 2.5.4.** $W_n$ is a nonstochastic row normalized and diagonalizable weight matrix with real eigenvalues. Row and column sums of $W_n$ in absolute value are uniformly bounded, uniformly in $n$. For any possible parameters in their parameter spaces, the corresponding $S_{nm}$ is nonsingular, and $S_{nm}^{-1}$ and $\sum_{h=1}^{+\infty} \text{abs}(B_{nm}^h)$ are bounded in row and column sums norms.

The parameter spaces of regression coefficients $\Pi_{km}$ and the covariance matrix $\Sigma_{vm}$, which is nonsingular, are compact. The parameter space of coefficients of spatial interactions and dynamics $\Psi_m$, $P_m$ and $\Phi_m$ is compact such that $\rho(\Psi_m)\rho(W_n) < 1$, and $\rho((I_m - \omega_{n,j}\Psi_m)^{-1}(P_m + \omega_{n,j}\Phi_m)') < 1$ for all those eigenvalues $w_{n,j}$ of $W_n$ less than one in absolute value, where such eigenvalues are assumed to be bounded.
away from 1, when \( n \) tends to infinity. The true parameters are located in the interior of their parameter spaces.

Assumption 2.5.4 is weaker than Assumption 2.2.2. We assume that \( \rho((I_m - \omega_{n,j} \Psi'_m)^{-1}(P_m + \omega_{n,j} \Phi_m)^') < 1 \) but put no restrictions on \( \rho((I_m - \Psi'_m)^{-1}(P_m + \Phi_m)^') \), which means that the assumption allows for some explosive cases. Also, we do not need Assumption 2.4.3.

In estimation, we’d like to eliminate the \( \text{vec}(u) (Y'_{nm,t}) \) and \( \text{vec}(\alpha) (Y'_{nm,t}) \) at the same time and estimate the parameters using information in \( \text{vec}(s) (Y'_{nm,t}) \). Therefore, this method is robust to some explosive cases and violations of Assumption 2.4.3. In addition, when the model satisfies stale case and spatial cointegration case, we can also apply RQML estimation.

\( I_n - W_n \) has the same eigenvectors \( \Gamma_n \) with \( W_n \) but the corresponding eigenvalues became zero when \( W_n \) has unity eigenvalues. \( [(I_n - W_n) \otimes I_m] \text{vec}(Y'_{nm,t}) = [(I_n - W_n) \otimes I_m] \text{vec}(u)(Y'_{nm,t}) + [(I_n - W_n) \otimes I_m] \text{vec}(s)(Y'_{nm,t}) + [(I_n - W_n) \otimes I_m] \text{vec}(\alpha)(Y'_{nm,t}) \),

where \( [(I_n - W_n) \otimes I_m] \text{vec}(u)(Y'_{nm,t}) = 0 \) since \( (I_n - W_n) \Gamma_n \|_{n,n_1} \Gamma_n^{-1} = 0 \) and \( [(I_n - W_n) \otimes I_m] \text{vec}(\alpha)(Y'_{nm,t}) = 0 \) since \( (I_n - W_n) l_n = 0 \) when \( W_n \) is row normalized.

Therefore, \( [(I_n - W_n) \otimes I_m] \text{vec}(Y'_{nm,t}) = [(I_n - W_n) \otimes I_m] \text{vec}(s)(Y'_{nm,t}) \) is stable under regularity conditions.

\[
E([(I_n - W_n) \otimes I_m] \text{vec}(V'_{nm,t}) \text{vec}(V'_{nm,t})^t [(I_n - W_n) \otimes I_m]) = \Sigma_n \otimes \Sigma_{vm0},
\]

where \( \Sigma_n = (I_n - W_n)(I_n - W_n)^t \). \( \Sigma_n \) is symmetric but rank is \( n - n_1 \). To eliminate the linear dependence among the transformed disturbances, we use eigenvalues and eigenvectors decomposition\(^{21}\). Therefore, decompose \( \Sigma_n = R \Lambda \Sigma_n R' \) where \( R \) are orthonormal eigenvectors. Define \( R = [R_1, R_2] \) where \( R_1 \) is corresponding

\(^{21}\)This method can be found in textbooks, such as Theil (1971)
2.5.1 Asymptotic Properties of RQMLE

Define $W^\dagger = \Lambda_1^{-\frac{1}{2}}R'_1 W_n R_1 \Lambda_1^{-\frac{1}{2}}$, and $Y_{nm,t}^\dagger = \Lambda_1^{-\frac{1}{2}}R'_1 (I_n - W_n) Y_{nm,t}$, we can get

$$Y_{nm,t}^\dagger = W_n^\dagger Y_{nm,t}^\dagger \Psi_m + Y_{nm,t-1}^\dagger P_m + W_n^\dagger Y_{nm,t-1}^\dagger \Phi_m + X_{nk,t}^\dagger \Pi_{km} + C_{nm}^\dagger + Y_{nm,t}^\dagger.$$

The covariance matrix is $E(vec(V_{nm,t}^\dagger)vec(V_{nm,t}^\prime)) = (\Lambda_1^{-\frac{1}{2}}R'_1 \Sigma_1 R_1 \Lambda_1^{-\frac{1}{2}}) \otimes \Sigma_{vm0} = I_{n-n_1} \otimes \Sigma_{vm0}$. $S_{nm}^\dagger = \Lambda_1^{-\frac{1}{2}}(R_1 \otimes I_m)' S_{nm}(R_1 \otimes I_m) \Lambda_1^{-\frac{1}{2}}$, therefore $|S_{nm}^\dagger| = |(R'_1 \otimes I_m)S_{nm}(R_1 \otimes I_m)|$.

$$R' \otimes I_m (I_{nm} - W_n \otimes \Psi_m') R \otimes I_m$$

$$= \left( \begin{array}{cc} (R'_1 R_1) \otimes I_m - (R'_1 W_n R_1) \otimes \Psi_m' & (R'_1 R_2) \otimes I_m - (R'_1 W_n R_2) \otimes \Psi_m' \\ (R'_2 R_1) \otimes I_m - (R'_2 W_n R_1) \otimes \Psi_m' & (R'_2 R_2) \otimes I_m - (R'_2 W_n R_2) \otimes \Psi_m' \end{array} \right).$$

Since $R'_1 R_1 = I_{n-n_1}$, $R'_j R_{-j} = 0$, and $R'_2 W_n = R'_2$ (because $R'_2(I_n - W_n)(I_n - W_n)' R_2 = 0$ thus $R'_2(I_n - W_n) = 0$), $|S_{nm}^\dagger| = |S_{nm}'| |I_m - \Psi_m'|^n_1$.

Using FOC to concentrate out the term vec($C_{nm}'$), we can derive the concentrated log likelihood function:

$$\frac{1}{(n-n_1)T} \ln L_{nT,m,ro} = -\frac{m}{2} \ln (2\pi) + \frac{1}{n-n_1} \ln |S_{nm}|$$

$$- \frac{n_1}{n-n_1} \ln |I_m - \Psi_m'| - \frac{1}{2} \ln |\Sigma_{vm}|$$

$$- \frac{1}{2(n-n_1)T} \sum_{t=1}^T vec(\tilde{V}_{nm,t}(\theta))'(J_n^* \otimes \Sigma_{vm}^{-1}) vec(\tilde{V}_{nm,t}(\theta))$$

where $J_n^* = (I_n - W_n)' R_1 \Lambda_1^{-1} R'_1 (I_n - W_n)$ and vec($\tilde{V}_{nm,t}(\theta)$) = $S_{nm} vec(\tilde{Y}_{nm,t}'(\theta)) = (I_n \otimes P_m' + W_n \otimes \Phi_m' vec(\tilde{X}_{nk,t} \otimes I_m) vec(\Pi_{km}')$.

### 2.5.1 Asymptotic Properties of RQMLE

For consistency, we need boundedness assumptions on $R_1 \Gamma_1^{-1} R'_1$, the generalized inverse of $(I_n - W_n)(I_n - W_n)'$. 

81
Assumption 2.5.5. The row and column sum norms of $R_1 \Gamma_1^{-1} R'_1$ is uniformly bounded, uniformly in $n$.

There are two sets of identification conditions. The first set of conditions requires existence and non-multicollinearity of IV for spatial lags:

\[ \tilde{Z}_{nm,t} = [\tilde{Y}_{nm,t-1}, W_n \tilde{Y}_{nm,t-1}, \tilde{X}_{nk,t}], \]

and remind that $\tilde{A}_z = \left[ (W_n \otimes e_{m,1}) S_{nm0}^{-1} (I_n \otimes P_{m0} + W_n \otimes \Phi_{m0}) \text{vec}(\tilde{Y}_{nm,t-1}) + (\tilde{X}_{nk,t} \otimes I_m) \text{vec}(\Pi_{km0}'), \ldots, (W_n \otimes s_{m,m}) S_{nm0}^{-1} (I_n \otimes P_{m0} + W_n \otimes \Phi_{m0}) \text{vec}(\tilde{Y}_{nm,t-1}) + (\tilde{X}_{nk,t} \otimes I_m) \text{vec}(\Pi_{km0}) \right]$, a $n \times m$ matrix. Assumption 2.5.6 guarantees the existence of instruments for every endogenous regressor. It is similar in the form with Assumption 2.3.4 and Assumption 2.4.4.

Assumption 2.5.6. The limiting matrix

\[
\lim_{T \to +\infty} \frac{1}{(n-n_1)T} \sum_{t=1}^{T} E \left[ \begin{pmatrix} \tilde{Z}_{nm,t}' J_n^* \tilde{Z}_{nm,t} & \tilde{Z}_{nm,t}' J_n^* \tilde{A}_z,t \\ \tilde{A}_z,t' J_n^* \tilde{Z}_{nm,t} & \tilde{A}_z,t' J_n^* \tilde{A}_z,t \end{pmatrix} \right]
\]

exists and is nonsingular.

In finite sample, it needs $[I_T \otimes \Lambda_{1}^{1/2} R'_1 (I_n - W_n)] \left( \begin{pmatrix} \tilde{A}_z,1 & \tilde{Z}_{nm,1} \\ \vdots & \vdots \\ \tilde{A}_z,T & \tilde{Z}_{nm,T} \end{pmatrix} \right)$ has full column rank.

Assumption 2.5.7, similarly with Assumptions 2.3.5 and 2.4.5, employs information from the disturbances to help identification. Let $\sigma_m^2 = \text{Tr}(\Sigma_{vm})$ and $\Sigma_{vm}^* = \Sigma_{vm}/\sigma_m^2$.

Assumption 2.5.7.

\[
\lim_{n \to +\infty} \left[ S_{nm0}^{-1} S_{nm}' (J_n^* \otimes \Sigma_{vm}^{-1}) S_{nm} S_{nm0}^{-1} (J_n^* \otimes \Sigma_{vm0}^{-1}) \right]^{\frac{1}{m(n-1)}} = 0
\]

unless $\Sigma_{vm}^* = \Sigma_{vm0}^*$ and $\Psi_m = \Psi_m^0$.
Assumption 2.5.7 holds in finite sample for any finite \( n \) due to the geometric-arithmetic mean inequality.

**Proposition 2.4.** (1) Under Assumptions 2.5.1-2.5.6 and 2.5.5, \( \Psi_m, P_m, \Phi_m \) and \( \Pi_{km} \) can be identified;

(2) Under Assumptions 2.5.1-2.5.3, 2.5.7 and 2.5.5, \( \Psi_m \) and \( \Sigma_{vm} \) can be identified.

The robust quasi-maximum likelihood estimators are consistency under above assumptions.

**Theorem 2.7.** Under assumptions 2.5.1-2.5.5 the RQMLE for the MDP-SAR model is consistent.

Asymptotic distribution of the RQML estimator \( \hat{\theta}_r \) will be useful for statistical inference. The asymptotic distribution of \( \hat{\theta}_r \) can be derived from the Taylor expansion of the log likelihood function, given by the mean value theorem

\[
\hat{\theta}_r - \theta_0 = - \left[ \frac{\partial^2 \ln L_{nT,m,ro}(\tilde{\theta}_r)}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial \ln L_{nT,m,ro}(\theta_0)}{\partial \theta}.
\]

where \( \tilde{\theta}_r \) is between \( \theta_0 \) and \( \hat{\theta}_r \). Similarly with stable case, we can decompose the first derivatives of log-likelihood function as

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m,0,ro}}{\partial \theta} = \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m,0,ro}}{\partial \theta} + \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m,0,ro}}{\partial \theta} \text{ with details in appendix.}
\]

The first component has zero mean and is proved to be asymptotic normally distributed while the second component results in asymptotic bias. Using the Lemma 2.4 when \( T \to \infty \)

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0,ro}}{\partial \theta} \overset{d}{\to} N(0, \Omega_{\theta_0}^{ro} + \Xi_{\theta_0}^{ro}).
\]

With the convergence of the second derivatives, we arrive at the asymptotic distribution of RQMLE.
Theorem 2.8. Under Assumptions 2.5.1-2.5.7, let \( \Delta_{\theta_0}^\theta = \Omega_{\theta_0}^{-1} \Delta_{\theta_0}^R \) where \( \Omega_{\theta_0} \) is non-singular and when \( T \to \infty \).

\[
\sqrt{(n-1)T}(\hat{\theta}_r - \theta_0) - \sqrt{\frac{n-1}{T}} \Delta_{\theta_0}^\theta + O_p \left( \max\left( \sqrt{\frac{n-1}{T^3}}, \sqrt{\frac{1}{T}} \right) \right)
\]

\( d \to N(0, \Omega_{\theta_0}^{-1}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega_{\theta_0}^{-1}) \) (2.13)

Consequently, (1) if \( n/T \to 0 \), then

\[
\sqrt{(n-1)T}(\hat{\theta}_r - \theta_0) \stackrel{d}{\to} N(0, \Omega_{\theta_0}^{-1}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega_{\theta_0}^{-1})
\]

(2) if \( n/T \to M \),

\[
\sqrt{(n-1)T}(\hat{\theta}_r - \theta_0) \stackrel{d}{\to} N(\sqrt{M} \Delta_\theta^\theta, \Omega_{\theta_0}^{-1}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega_{\theta_0}^{-1})
\]

(3) if \( n/T \to \infty \),

\[
T(\hat{\theta}_r - \theta_0) \xrightarrow{p} \Delta_{\theta_0}^\theta.
\]

For the case that the disturbance terms are normally distributed, \( \Xi_{\theta_0} = 0 \).

When \( \frac{n}{T} \to M > 0 \), the RQMLE is not asymptotically centered at zero. The asymptotic bias is \( -\frac{1}{T} \Delta_{\theta_0}^\theta \). We define the bias corrected estimator as \( \hat{\theta}_c = \hat{\theta}_r + \frac{1}{T} \Delta_{\theta_0}^\theta(\hat{\theta}_r) \), where the bias term is \( \Delta_{\theta_0}^\theta(\hat{\theta}_r) = \left[ E\left( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m}(\hat{\theta}_r)}{\partial \theta \partial \theta'} \right) \right]^{-1} \Delta_{\theta_0}^R(\hat{\theta}_r) \).

In the following theorem, we show that when \( T/n^3 \to \infty \), the bias corrected estimator \( \hat{\theta}_c \) is \( \sqrt{nT} \) consistent and asymptotically normal centered at zero. This result holds when \( n/T \to M \) as \( M \) is a constant and when \( n/T \to \infty \). Additional assumption is needed to guarantee asymptotic normality.

Assumption 2.5.8. The row and column sum norms of \( \sum_{g=1}^\infty \sum_{h=0}^{g-1} B_{nm}^h S^{-1}(W_n \otimes E_{m,ij})B_{nm}^{g-1-h} \) and \( \sum_{g=1}^\infty \sum_{h=0}^{g-1} B_{nm}^h S^{-1}(I_n \otimes E_{m,ij})B_{nm}^{g-1-h} \) are bounded uniformly in absolute value in the parameter space for \( i, j = 1, ..., m \).

The asymptotic normality is achieved when \( n \) grows slower than or proportional to \( T \) under Assumption 2.5.8.

Theorem 2.9. Under Assumptions 2.5.1-2.5.8, when \( n/T^3 \to 0 \), the QMLE with bias correction, \( \hat{\theta}_c \), has

\[
\sqrt{(n-1)T}(\hat{\theta}_c - \theta_0) \stackrel{d}{\to} N(0, \Omega_{\theta_0}^{-1}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega_{\theta_0}^{-1})
\]
2.6 Monte Carlo Experiments

We conduct a monte carlo experiment to investigate small sample properties of the QML estimates for the multivariate SAR model. In the experiment, the spatial weight matrix, \( W \) is generated from 760 counties in great plains states of USA. When two counties, \( i \) and \( j \), share border, \( w_{ij} = w_{ji} = 1 \). When sample size is \( n \), we randomly pick a \( n \times n \) block along the diagonal and row normalize the matrix. The MDP-SAR process we investigated is:

\[
\begin{align*}
    y_{1,t} &= \psi_{11}W_{y1,t} + \psi_{21}W_{y2,t} + p_{11}y_{1,t-1} + p_{21}y_{2,t-1} + \phi_{11}W_{y1,t-1} + \phi_{21}W_{y2,t-1} + \beta_{11}x_{1,t} + u_{1,t}, \\
    y_{2,t} &= \psi_{12}W_{y1,t} + \psi_{22}W_{y2,t} + p_{12}y_{1,t-1} + p_{22}y_{2,t-1} + \phi_{12}W_{y1,t-1} + \phi_{22}W_{y2,t-1} + \beta_{22}x_{1,t} + u_{2,t},
\end{align*}
\]

where \( u_{j,t} = c_j + d_{j,t} + v_{j,t} \) for \( j = 1, 2 \). For this system, the disturbances \( v_{1,t} \) and \( v_{2,t} \) are normally distributed with mean 0 and the covariance matrix: \( \Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix} \). \( c_l \) and \( d_{l,t} \) are individual fixed effects and time fixed effects, generated by uniformly distributed random variable on \([0, 1]\). The coefficients matrices are set as

\[

\Psi_m = P_m = \Phi_m = \begin{pmatrix} 0.4 & -0.3 \\ 0.3 & -0.4 \end{pmatrix}.
\]

Table 2.1 presents the Monte Carlo results when the sample sizes are \((n, T) = (20, 20), (40, 20), (20, 40) \) and \((40, 40)\). We conduct 300 repetitions for each design which return empirical bias and standard deviation of the QML estimates. All the estimates without bias correction are significant and most of the biases are about 5% or less when the sample size is just \((20, 20)\). Two estimators has about 10% bias but they can be corrected using bias correction estimator. When time period is small, the biases for time lagged effects \( P_m \) are relatively large (see the experiment with \((40, 20)\)). When cross-sectional unit is small the biases for \( \Psi_m \) are relatively
large (see (20,40)). Furthermore, the bias corrected estimator can reduce biases of some estimators while it doesn’t work well in some other estimators. When the sample size increases, estimators’ standard deviations decrease in the rate of $\sqrt{nT}$.

For the system with spatial cointegration, the disturbances $v_{1,t}$ and $v_{2,t}$ are normally distributed with mean 0 and the covariance matrix: $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$. $c_t$ and $d_{t,t}$ are individual fixed effects and time fixed effects, generated by uniformly distributed random variable on $[0, 1]$. The coefficients matrices are set as $\text{vec}(\Psi_m) = \text{vec}(P_m) = (0.7, 0, -0.4, 0.7)'$ and $\text{vec}(\Phi_m) = (-0.4, 0, 0.8, -0.4)$.

Table 2.2 presents the Monte Carlo results when the sample sizes are $(n, T) = (20,20), (40,20), (20,40)$ and $(40,40)$. We conduct 300 repetitions for each design which return empirical bias and standard deviation of the QML estimates. All the nonzero estimates without bias correction are significant and most of the biases are about 5% or less when the sample size is just (20,20). The estimators of $\Phi_m$ has relatively large biases but bias correction works well for them. When time period is small, the biases for time lagged effects $P_m$ are relatively large (see the experiment with (40,20)). Furthermore, the bias corrected estimator can reduce biases of some estimators while it doesn’t work well in some other estimators, e.g. $\Psi_m$. When the sample size increases, estimators’ standard deviations decrease in the rate of $\sqrt{nT}$.

We also conduct Monte Carlo experiments for robust estimation; here we use the same parameter in last set of experiments.

Table 2.3 presents the Monte Carlo results when the sample sizes are $(n, T) = (20,20), (40,20), (20,40)$ and $(40,40)$. We conduct 300 repetitions for each design
Table 2.1: Monte Carlo for MDP-SAR, Stable Case

<table>
<thead>
<tr>
<th>TRUE</th>
<th>n,T</th>
<th>20,20</th>
<th>20,40</th>
<th>40,20</th>
<th>40,40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\psi_{11})</td>
<td>Bias</td>
<td>0.0234</td>
<td>0.0343</td>
<td>0.0234</td>
<td>0.0288</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0429</td>
<td>0.0424</td>
<td>0.0289</td>
<td>0.0286</td>
</tr>
<tr>
<td>(\psi_{12})</td>
<td>0.3</td>
<td>Bias</td>
<td>-0.0117</td>
<td>-0.0323</td>
<td>-0.0120</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0763</td>
<td>0.0741</td>
<td>0.0498</td>
<td>0.0489</td>
</tr>
<tr>
<td>(\psi_{21})</td>
<td>-0.3</td>
<td>Bias</td>
<td>0.0111</td>
<td>0.0282</td>
<td>0.0105</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0641</td>
<td>0.0621</td>
<td>0.0429</td>
<td>0.0423</td>
</tr>
<tr>
<td>(\psi_{22})</td>
<td>-0.4</td>
<td>Bias</td>
<td>0.0107</td>
<td>-0.0014</td>
<td>0.0205</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0572</td>
<td>0.0562</td>
<td>0.0382</td>
<td>0.0379</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.5</td>
<td>Bias</td>
<td>0.0068</td>
<td>-0.0103</td>
<td>-0.0097</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.1653</td>
<td>0.1649</td>
<td>0.1146</td>
<td>0.1147</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>0.5</td>
<td>Bias</td>
<td>0.0096</td>
<td>0.0321</td>
<td>0.0016</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.1538</td>
<td>0.1536</td>
<td>0.1237</td>
<td>0.1236</td>
</tr>
<tr>
<td>(p_{11})</td>
<td>0.4</td>
<td>Bias</td>
<td>-0.0687</td>
<td>-0.0128</td>
<td>-0.0394</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0539</td>
<td>0.0560</td>
<td>0.0383</td>
<td>0.0390</td>
</tr>
<tr>
<td>(p_{12})</td>
<td>0.3</td>
<td>Bias</td>
<td>-0.0015</td>
<td>-0.0075</td>
<td>-0.0029</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0443</td>
<td>0.0446</td>
<td>0.0294</td>
<td>0.0295</td>
</tr>
<tr>
<td>(p_{21})</td>
<td>-0.3</td>
<td>Bias</td>
<td>-0.0375</td>
<td>-0.0110</td>
<td>-0.0200</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0542</td>
<td>0.0551</td>
<td>0.0361</td>
<td>0.0365</td>
</tr>
<tr>
<td>(p_{22})</td>
<td>-0.4</td>
<td>Bias</td>
<td>-0.0147</td>
<td>-0.0011</td>
<td>-0.0060</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0447</td>
<td>0.0452</td>
<td>0.0292</td>
<td>0.0294</td>
</tr>
<tr>
<td>(\phi_{11})</td>
<td>0.4</td>
<td>Bias</td>
<td>-0.0054</td>
<td>-0.0447</td>
<td>-0.0083</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0822</td>
<td>0.0845</td>
<td>0.0595</td>
<td>0.0603</td>
</tr>
<tr>
<td>(\phi_{12})</td>
<td>0.3</td>
<td>Bias</td>
<td>-0.0011</td>
<td>-0.0212</td>
<td>-0.0038</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0648</td>
<td>0.0637</td>
<td>0.0432</td>
<td>0.0429</td>
</tr>
<tr>
<td>(\phi_{21})</td>
<td>-0.3</td>
<td>Bias</td>
<td>-0.0097</td>
<td>-0.0211</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0896</td>
<td>0.0896</td>
<td>0.0626</td>
<td>0.0626</td>
</tr>
<tr>
<td>(\phi_{22})</td>
<td>-0.4</td>
<td>Bias</td>
<td>0.0063</td>
<td>-0.0115</td>
<td>0.0125</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0646</td>
<td>0.0647</td>
<td>0.0425</td>
<td>0.0425</td>
</tr>
<tr>
<td>(\sigma_{11})</td>
<td>1</td>
<td>Bias</td>
<td>-0.0312</td>
<td>0.0226</td>
<td>0.0270</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0844</td>
<td>0.0891</td>
<td>0.0609</td>
<td>0.0625</td>
</tr>
<tr>
<td>(\sigma_{12})</td>
<td>0.5</td>
<td>Bias</td>
<td>-0.0292</td>
<td>-0.0030</td>
<td>-0.0274</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0740</td>
<td>0.0781</td>
<td>0.0529</td>
<td>0.0543</td>
</tr>
<tr>
<td>(\sigma_{22})</td>
<td>1</td>
<td>Bias</td>
<td>0.0224</td>
<td>0.0792</td>
<td>0.0670</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0933</td>
<td>0.0985</td>
<td>0.0595</td>
<td>0.0610</td>
</tr>
</tbody>
</table>

Note: \(\theta\) is QMLE and \(\theta^c\) is QMLE with bias correction.
<table>
<thead>
<tr>
<th>$\psi_{11}$</th>
<th>Bias</th>
<th>$\theta$</th>
<th>$\theta^c$</th>
<th>$\theta$</th>
<th>$\theta^c$</th>
<th>$\theta$</th>
<th>$\theta^c$</th>
<th>$\theta$</th>
<th>$\theta^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>Bias</td>
<td>0.0130</td>
<td>-0.0376</td>
<td>0.0094</td>
<td>0.0424</td>
<td>0.0080</td>
<td>0.0337</td>
<td>0.0048</td>
<td>0.0175</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0354</td>
<td>0.0262</td>
<td>0.0312</td>
<td>0.0251</td>
<td>0.0305</td>
<td>0.0183</td>
<td>0.0194</td>
<td></td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>Bias</td>
<td>-0.0071</td>
<td>-0.0292</td>
<td>-0.0555</td>
<td>0.0045</td>
<td>-0.0022</td>
<td>0.0160</td>
<td>-0.0033</td>
<td>0.0052</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0407</td>
<td>0.0283</td>
<td>0.0341</td>
<td>0.0288</td>
<td>0.0351</td>
<td>0.0205</td>
<td>0.0215</td>
<td></td>
</tr>
<tr>
<td>$\psi_{21}$</td>
<td>Bias</td>
<td>0.0176</td>
<td>0.0154</td>
<td>0.0179</td>
<td>0.0083</td>
<td>0.0136</td>
<td>0.0042</td>
<td>0.0071</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0346</td>
<td>0.0236</td>
<td>0.0232</td>
<td>0.0240</td>
<td>0.0224</td>
<td>0.0169</td>
<td>0.0166</td>
<td></td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>Bias</td>
<td>0.0100</td>
<td>0.0103</td>
<td>0.0224</td>
<td>0.0038</td>
<td>0.0262</td>
<td>0.0083</td>
<td>0.0182</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0278</td>
<td>0.0201</td>
<td>0.0224</td>
<td>0.0223</td>
<td>0.0263</td>
<td>0.0139</td>
<td>0.0144</td>
<td></td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>Bias</td>
<td>-0.0533</td>
<td>-0.0202</td>
<td>-0.0197</td>
<td>0.0003</td>
<td>-0.019</td>
<td>0.0013</td>
<td>-0.0021</td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.1616</td>
<td>0.1111</td>
<td>0.1112</td>
<td>0.1094</td>
<td>0.1105</td>
<td>0.0787</td>
<td>0.0783</td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>Bias</td>
<td>-0.0122</td>
<td>-0.0198</td>
<td>-0.0147</td>
<td>-0.0096</td>
<td>-0.0105</td>
<td>0.0119</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.1511</td>
<td>0.1017</td>
<td>0.1017</td>
<td>0.1189</td>
<td>0.1187</td>
<td>0.0748</td>
<td>0.0748</td>
<td></td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>Bias</td>
<td>-0.0976</td>
<td>-0.0438</td>
<td>-0.0949</td>
<td>-0.0194</td>
<td>-0.0421</td>
<td>-0.0050</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0505</td>
<td>0.0308</td>
<td>0.0311</td>
<td>0.0335</td>
<td>0.0348</td>
<td>0.0212</td>
<td>0.0213</td>
<td></td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>Bias</td>
<td>-0.0299</td>
<td>-0.0102</td>
<td>-0.0316</td>
<td>-0.0108</td>
<td>-0.0018</td>
<td>-0.0025</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0526</td>
<td>0.0346</td>
<td>0.0373</td>
<td>0.0332</td>
<td>0.0366</td>
<td>0.0256</td>
<td>0.0263</td>
<td></td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>Bias</td>
<td>0.0225</td>
<td>0.0245</td>
<td>0.0245</td>
<td>0.0249</td>
<td>0.0263</td>
<td>0.0171</td>
<td>0.0174</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0463</td>
<td>0.0287</td>
<td>0.0287</td>
<td>0.0303</td>
<td>0.0315</td>
<td>0.0190</td>
<td>0.0192</td>
<td></td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>Bias</td>
<td>-0.0318</td>
<td>-0.0138</td>
<td>-0.0256</td>
<td>-0.0037</td>
<td>-0.0101</td>
<td>0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0404</td>
<td>0.0241</td>
<td>0.0237</td>
<td>0.0265</td>
<td>0.0261</td>
<td>0.0161</td>
<td>0.0157</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>Bias</td>
<td>0.0853</td>
<td>0.0360</td>
<td>0.0881</td>
<td>-0.0122</td>
<td>0.0372</td>
<td>-0.0127</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0565</td>
<td>0.0376</td>
<td>0.0408</td>
<td>0.0381</td>
<td>0.0420</td>
<td>0.0256</td>
<td>0.0267</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>Bias</td>
<td>-0.0292</td>
<td>-0.0309</td>
<td>-0.0309</td>
<td>-0.015</td>
<td>0.0262</td>
<td>-0.0118</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0612</td>
<td>0.0427</td>
<td>0.0438</td>
<td>0.0442</td>
<td>0.0466</td>
<td>0.0332</td>
<td>0.0341</td>
<td></td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>Bias</td>
<td>-0.0036</td>
<td>0.0258</td>
<td>0.0197</td>
<td>0.0763</td>
<td>0.0465</td>
<td>0.0740</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0713</td>
<td>0.0569</td>
<td>0.0584</td>
<td>0.0578</td>
<td>0.0610</td>
<td>0.0406</td>
<td>0.0417</td>
<td></td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>Bias</td>
<td>0.0007</td>
<td>0.0016</td>
<td>0.0012</td>
<td>0.0021</td>
<td>-0.0002</td>
<td>-0.0003</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0056</td>
<td>0.0023</td>
<td>0.0024</td>
<td>0.0044</td>
<td>0.0043</td>
<td>0.0014</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>Bias</td>
<td>0.0053</td>
<td>0.0081</td>
<td>0.0080</td>
<td>0.0030</td>
<td>0.0020</td>
<td>-0.0002</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0111</td>
<td>0.0084</td>
<td>0.0092</td>
<td>0.0097</td>
<td>0.0104</td>
<td>0.0033</td>
<td>0.0033</td>
<td></td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>Bias</td>
<td>-0.0044</td>
<td>-0.0019</td>
<td>-0.0046</td>
<td>0.0025</td>
<td>-0.0019</td>
<td>-0.0005</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0057</td>
<td>0.0023</td>
<td>0.0024</td>
<td>0.0040</td>
<td>0.0042</td>
<td>0.0015</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>Bias</td>
<td>-0.0081</td>
<td>-0.0075</td>
<td>-0.0109</td>
<td>0.0021</td>
<td>-0.0030</td>
<td>-0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S.D.</td>
<td>0.0111</td>
<td>0.0094</td>
<td>0.0097</td>
<td>0.0092</td>
<td>0.0097</td>
<td>0.0032</td>
<td>0.0034</td>
<td></td>
</tr>
</tbody>
</table>

Note: $\theta$ is QMLE and $\theta^c$ is QMLE with bias correction.
which return empirical bias and standard deviation of the QML estimates. Some of the biases are about 10% when the sample size is (20,20) and standard deviations are also larger than those in Table 2.2. This is because, unlike QMLE, RQMLE just use part of the information in the model to estimate. Furthermore, the bias corrected estimator can reduce biases of some estimators while it doesn’t work well in some other estimators, e.g. $\Psi_m$. When the sample size increases, estimators’ standard deviations decrease in the rate of $\sqrt{nT}$.

### 2.7 simultaneous equations Dynamic Panel SAR Model

In empirical studies, we often encounter structural model which could be represented by a simultaneous equations model. The multivariate spatial autoregressive model can be treated as a quasi-reduced form for a simultaneous equations SAR model, which is:

$$Y_{nm,t}G_m = W_nY_{nm,t}\Psi_{m,s} + Y_{nm,t-1}P_{m,s} + W_nY_{nm,t-1}\Phi_{m,s} + X_{nk,t}\Pi_{km,s} + C_{nm,s} + \alpha_{m,t,s} \otimes l_n + U_{nm,t};$$

In the above model, $G_m$ is invertible with its diagonal elements being normalized to ones. The relationship between the structural model and quasi-reduced model can be represented as $\Psi_m = \Psi_{m,s}G_m^{-1}$, $P_m = P_{m,s}G_m^{-1}$, $\Phi_m = \Phi_{m,s}G_m^{-1}$, $\Pi_{km} = \Pi_{km,s}G_m^{-1}$, $(\alpha_{m,t,s} \otimes l_n)G_m = (\alpha_{m,t} \otimes l_n$ and $U_{nm,t} = V_{nm,s}G_m$. Let $Z_{nm,t} = [Y_{nm,t}, W_nY_{nm,t}, Y_{nm,t-1}, W_nY_{nm,t-1}X_{nk}]$ and $\gamma_m = (\Gamma_m', -\Psi_{m,s}', -P_{m,s}', -\Phi_{m,s}', -\Pi_{km,s}')'$, the model becomes $Z_{nm,t}\gamma_m = U_{nm,t} + C_{nm,s} + \alpha_{m,t,s} \otimes l_n$.

We define a parameter in the dynamic panel simultaneous equations model is identified if and only if it can be deducted from knowledge of the reduced form.
### Table 2.3: Monte Carlo for MDP-SAR, Robust Estimation

<table>
<thead>
<tr>
<th>TRUE</th>
<th>n,T</th>
<th>20,20</th>
<th>20,40</th>
<th>40,20</th>
<th>40,40</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\hat{\theta}$</td>
<td>$\hat{\theta}^c$</td>
<td>$\hat{\theta}$</td>
<td>$\hat{\theta}^c$</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.7</td>
<td>Bias</td>
<td>0.0331 0.0437</td>
<td>0.0217 0.0234</td>
<td>0.0238 0.0278</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0493 0.0491</td>
<td>0.0339 0.0327</td>
<td>0.0291 0.0282</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0</td>
<td>Bias</td>
<td>-0.0252 -0.0941</td>
<td>-0.0192 -0.0405</td>
<td>-0.0194 -0.0487</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0619 0.1445</td>
<td>0.0425 0.0419</td>
<td>0.0376 0.0405</td>
</tr>
<tr>
<td>$\psi_{21}$</td>
<td>-0.4</td>
<td>Bias</td>
<td>0.0348 0.0940</td>
<td>0.0194 0.0375</td>
<td>0.0302 0.0605</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0542 0.1503</td>
<td>0.0404 0.0410</td>
<td>0.0349 0.0426</td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>0.7</td>
<td>Bias</td>
<td>0.0207 -0.0092</td>
<td>0.0205 0.0094</td>
<td>0.0106 0.0034</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0438 0.0840</td>
<td>0.0305 0.0299</td>
<td>0.0267 0.0291</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1</td>
<td>Bias</td>
<td>-0.0637 -0.0768</td>
<td>0.0127 0.0073</td>
<td>-0.0208 -0.0210</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.1592 0.1746</td>
<td>0.1137 0.1149</td>
<td>0.1241 0.1265</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1</td>
<td>Bias</td>
<td>-0.0379 0.0280</td>
<td>0.0290 0.0493</td>
<td>0.0075 0.0441</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.1911 0.2287</td>
<td>0.1221 0.1235</td>
<td>0.1192 0.1233</td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.7</td>
<td>Bias</td>
<td>-0.1029 -0.0144</td>
<td>-0.0465 -0.0055</td>
<td>-0.0555 -0.0122</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0469 0.0505</td>
<td>0.0309 0.0311</td>
<td>0.0330 0.0347</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0</td>
<td>Bias</td>
<td>0.0375 -0.0064</td>
<td>0.0261 0.0067</td>
<td>0.0447 0.0074</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0419 0.0441</td>
<td>0.0243 0.0244</td>
<td>0.0278 0.0291</td>
</tr>
<tr>
<td>$p_{21}$</td>
<td>-0.4</td>
<td>Bias</td>
<td>-0.0468 -0.0021</td>
<td>-0.0225 -0.0001</td>
<td>-0.0440 0.0022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0437 0.0453</td>
<td>0.0267 0.0271</td>
<td>0.0321 0.0332</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.7</td>
<td>Bias</td>
<td>-0.0351 -0.0044</td>
<td>-0.0154 -0.0021</td>
<td>-0.0281 0.0012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0432 0.0427</td>
<td>0.0256 0.0249</td>
<td>0.0269 0.0265</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.4</td>
<td>Bias</td>
<td>0.0661 -0.0033</td>
<td>0.0240 -0.0055</td>
<td>0.0710 0.0250</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0628 0.0853</td>
<td>0.0456 0.0466</td>
<td>0.0399 0.0431</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>0</td>
<td>Bias</td>
<td>-0.0147 0.0514</td>
<td>-0.0052 0.0257</td>
<td>-0.0217 0.0323</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0706 0.1099</td>
<td>0.0472 0.0464</td>
<td>0.0461 0.0499</td>
</tr>
<tr>
<td>$\phi_{21}$</td>
<td>0.8</td>
<td>Bias</td>
<td>0.0033 -0.0451</td>
<td>0.0006 -0.0187</td>
<td>0.0070 -0.0096</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0644 0.1266</td>
<td>0.0485 0.0502</td>
<td>0.0464 0.0519</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.4</td>
<td>Bias</td>
<td>-0.0037 0.0243</td>
<td>-0.0150 -0.0055</td>
<td>-0.0035 0.0133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0598 0.0767</td>
<td>0.0376 0.0369</td>
<td>0.0372 0.0411</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>1</td>
<td>Bias</td>
<td>-0.0299 0.0252</td>
<td>-0.0010 0.0252</td>
<td>-0.0366 0.0169</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0896 0.0949</td>
<td>0.0613 0.0629</td>
<td>0.0597 0.0630</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.5</td>
<td>Bias</td>
<td>-0.0145 0.0129</td>
<td>-0.0090 0.0039</td>
<td>-0.0096 0.0176</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0669 0.0716</td>
<td>0.0446 0.0457</td>
<td>0.0471 0.0497</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>1</td>
<td>Bias</td>
<td>0.0160 0.0728</td>
<td>0.0408 0.0682</td>
<td>0.0073 0.0632</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.D.</td>
<td>0.0830 0.0888</td>
<td>0.0603 0.0619</td>
<td>0.0573 0.0604</td>
</tr>
</tbody>
</table>

Note: $\hat{\theta}$ is QMLE and $\hat{\theta}^c$ is QMLE with bias correction.
parameters in the multivariate SAR model. Therefore, in this part, we assume the reduced form parameters in the corresponding multivariate SAR model can be identified (it is equivalent to assume Assumptions 2.3.4 and 2.3.5 are satisfied for the corresponding multivariate SAR model).

We consider identification with exclusive restrictions.

**Proposition 2.5.** Suppose \( R_1 \) is a \( r \times (4m + k) \) matrix which represents all exclusive restrictions for the coefficients of the first equation, i.e. \( R_1 \gamma_{1,m} = 0 \). There are no other restrictions. Then, the sufficient and necessary rank condition for identification of the first equation is \( \text{rank}(R_{\gamma_m}) = m - 1 \) and order condition (necessary) is \( R_1 \geq m - 1 \).

The order condition treats not only the exogenous variable \( X_{km} \), but also the predeterminant variables \( Y_{nm,t-1} \) and \( W_n Y_{nm,t-1} \) as well as the spatial lags \( W_n Y_{nm,t} \), as the “exogenous” variables to help identify the structural parameters. Even there is no exogenous variable, time lags or space-time lags, the structural can still be identified since the spatial lags are relevant for the whole system.

The next corollary extends to all linear restrictions but not just exclusive restrictions.

**Proposition 2.6.** Suppose \( \bar{R}_1 \) is a \( r \times (4m + k) \) matrix which represents all linear restrictions including normalization restrictions for the coefficients of the first equation, such that \( \bar{R}_1 \gamma_{1,m} = c \). There is no other restrictions. Then, the sufficient and necessary rank condition for identification of the first equation is \( \text{rank}(\bar{R}_{1\gamma_m}) = m \).
For the whole system, we can include cross-equation linear restrictions. We summarize all within and cross-equation linear restrictions (including normalization) into a matrix $\bar{R}$ such that $\bar{R}vec(\gamma_m) = c$ where $c$ is a $R \times 1$ constant vector and $\bar{R}$ is a $r \times (4m^2 + km)$ matrix.

**Proposition 2.7.** A necessary and sufficient condition to identify the whole system is $\text{rank}\{\bar{R}(I_m \otimes \gamma_m)\} = m^2$.

### 2.8 Application: Market Integration of Yangtze River Basin

The question why China industrialized much later than European countries has received attention for studies of economic growth and regional development. Interregional and international trade are considered as a key condition for economic development, especially for industrialization since it brings in competition, specialization, spillover of information and technology, and economies of scale. However, as indicated by Shiue (2002), Keller and Shiue (2007, 2008) and Yan and Liu (2011), the southern China, including Yangtze River Basin, has demonstrated a high level of market integration in grain market, which suggests that the market integration may only be a necessary condition for industrialization, but it is not sufficient to result in the production revolution. Since in mid-Qing dynasty (18th century) the most important trading goods is grain, all the above researches focus on rice prices in difference cities (or “prefectures” in Shiue (2002)).

However, there are still questions on using rice prices as the only representative of grain prices in studying the market integration. The first reason is on completeness. The rice was the primary food in Yangtze River Basin (Huang,
2009), for example, the percentage of rice consumption in 9 cities on Yangtze Delta counted for about 90% of grain consumption in 1776. However, wheat still counted for about 10% of consumption in these cities. Wheat could be used to make noodles, dumplings and buns and served as a secondary food. A model including both primary food and secondary food takes a more complete grain market into account. The second reason is that other factors, such as government policy, might be one of the reasons of rice prices cointegration in addition to trading. Rice was the primary food in this area and its price is highly related to residents’ everyday life. Considering stability of economy and society as one of the main issues, the government of Qing dynasty would have policies to keep the rice prices stable, such as “Chang-Ping barn22” and transporting rices across cities (Chen, 2004) in order to stabilize rice prices. Therefore, the trading pattern and price changes might not be only due to businesses’ spontaneous trading but also governments’ policies or pressure. However, in Yangtze River Basin, the wheat was not as important as rice. The trading of wheat and the implied relationship among wheat prices in different cities could be attributed to businesses’ spontaneous trading resulting from incentives in markets. Since the aim of the research is to study market integration as the explanation of industrialization, the market integration due to spontaneous commercial behaviour could be what we are looking for.

Therefore, we add one equation of wheat price to Keller and Shiue (2007). The additional equation describe the spatial and intertemporal relationship between the wheat and rice prices in a city, and rice and wheat prices of his neighbours in

22The local governments stored rices in case of significant increase of rice prices due to disaster or other factors.
current period and in last period, in addition to the rice and wheat prices in the city from last period. This model could capture the spatial effects, diffusion and time lagged effects of prices, which may provide evidence in favor of or reject the grain market integration hypothesis.

2.8.1 Data

The data comes from an historical archive, “Gongzhong Liangjian Dan”, translated as “Grain Price Lists in the Palace”. We collect the data from the electronic version from the Database of Grain Prices in Qing Dynasty, maintained by Academia Sinica in Taiwan.\(^{23}\) The data covers 65 cities and 49 years, from 1742 to 1790, during the reign of Emperor Qianlong, one of most powerful emperors in China. The period is considered as stable and no significant structural changes of economy.

The data are collected from middle and lower Yangtze Rive Basin provinces: Anhui, Jiangsu, Zhejiang, Hubei, Hunan, Jiangxi. We exclude the upper Yangtze River Basin since it is geographically far away from others and the transportation to that area could be very costly due to rugged terrain. The middle and lower basin is often considered as the richest area in China where interregional trade was well-developed. We use semi-annual data, collected from February and August of each year, so that there are 98 periods. There are 12% missing data, however, there is no evidence that these data are missing because of regional characteristics. In addition to data on February and August, we also get access to other months’ data, which allow us to interpolate the missing data by averaging nearby months’ prices, since the nearby months often show very similar prices. We have

\(^{23}\)The data is collected monthly. There are some missing data of wheat prices. We use the wheat prices in the closest month as proxies if the data is missing.
three categories of rice, high-quality, mid-quality and low-quality. We use the mid-quality rice prices following Shiue (2002), Keller and Shiue (2007, 2008) and Yan and Liu (2011). We only have one category of wheat. We can access to maximum and minimum of rice and wheat prices so we use the means of them to represent the prices. The data is summarized in table 2.4.

<table>
<thead>
<tr>
<th>Table 2.4: Data Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations:</td>
</tr>
<tr>
<td>Cities:</td>
</tr>
<tr>
<td>Provinces:</td>
</tr>
<tr>
<td>Periods (1742-1790, Semiannual):</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Distance:</td>
</tr>
<tr>
<td>Rice Price (low):</td>
</tr>
<tr>
<td>Rice Price (high):</td>
</tr>
<tr>
<td>Wheat Price (low):</td>
</tr>
<tr>
<td>Wheat Price (high):</td>
</tr>
<tr>
<td>Rice Price:</td>
</tr>
<tr>
<td>Wheat Price:</td>
</tr>
</tbody>
</table>

Note: the distances are measured by kilometers (1km = 0.6213mi).

2.8.2 Model, Results and Robust Check

In Keller and Shiue (2007) and Yu and Lee (2010), the rice price from a city is specified to depend on its own lagged price, the neighbours’ prices and the neighbours’ prices in last period. This is our first equation in the system.

\[ y_{i,t}^r = p_{11} y_{i,t-1}^r + \psi_{11} \sum_{j=1}^{n} w_{i,j} y_{j,t}^r + \phi_{11} \sum_{j=1}^{n} w_{i,j} y_{j,t-1}^r + c_i^r + d_t^r + v_{i,t}^r, \]

where \( p_{11} \) represents the time lagged effect, \( \psi_{11} \) represents the effect of neighbours’ prices (spatial effects), \( \phi_{11} \) represents the effect of neighbours’ prices in last period.
(diffusion), \( c_i \) is the individual fixed effects, \( d_t \) is the time fixed effect, and \( v_{i,t} \) is an idiosyncratic shock.

We add the equation of wheat price so that it becomes a system. The prices of wheat could be influenced by rice prices since wheat is the secondary and substitute food for rice. If there is a change in rice price, the quantity traded of rice may change, which results in a demand shift of wheat. The reverse is unlikely because people’s first choice is rice and the volume of consumption of wheat is much smaller than that of rice. Even a large change of quantity of wheat in trading would have a minor effects on the quantity traded of rice in percentage. As rice was the first choice of principle food according to (Huang, 2009), the price of wheat is assumed to depend not only on its own lagged price, the neighbours’ prices and the neighbours’ prices in last period, but also the lagged rice price and neighbors’ rice prices in current and last period.

\[
y_{i,t}^w = p_{22} y_{i,t-1}^w + \psi_{22} \sum_{j=1}^{n} w_{i,j} y_{j,t}^w + \phi_{22} \sum_{j=1}^{n} w_{i,j} y_{j,t-1}^w + p_{12} y_{i,t-1}^r + \psi_{12} \sum_{j=1}^{n} w_{i,j} y_{j,t}^r + \phi_{12} \sum_{j=1}^{n} w_{i,j} y_{j,t-1}^r + c_i^w + d_t^w + v_{i,t}^w,
\]

where \( p_{12}, \psi_{12} \) and \( \phi_{12} \) represent the cross-variable time lagged effect, spatial effect and diffusion. The idiosyncratic effects in both equations are allowed to depend on each other.

The weight matrix is generated from geographic distances between cities. In eighteenth century, the trade cost was highly related to geographic distance, so was information transmission. We adopt two specifications of weight matrix in Keller and Shiue (2007) and Yu and Lee (2010) by assuming \( w_{i,j} = \exp(-c_m \times \)
Distance(i, j)/100), where distance is measured in kilometers, \( c_m = 1.4 \) and \( c_m = 1.2 \) respectively.

The estimation results are reported in Table 2.5. There are three panels. The first panel (panel A) adopts the Yu and Lee (2010) weight matrix; the second panel (panel B) uses weight matrix introduced by Keller and Shiue (2007); and the last panel (panel C) is without time fixed effects. According to likelihood functions’ value, the first and third specifications fit data well. First, there are evidences that the prices are cointegrated spatially. Since the model is a triangular system, we can test whether the data are spatially cointegrated using \( p^*_{11} \) and \( p^*_{22} \). In panel A, the hypothesis that \( p^*_{11} = 1 \) cannot be rejected; in panel B, the hypothesis can be rejected at 5% significance level but not at 1% significance level; and in panel C, we cannot reject the hypothesis. The second observation is that the time lagged effects and spatial effects are positive while diffusion is negative. The estimator of the rice price equation is similar with Keller and Shiue (2007) in sign and in magnitude. The estimation of the wheat price equation shows that wheat prices are spatially related to their neighbours’ wheat prices and rice prices in current period and in last period. These results suggest that the grain market in Yangtze River Basin, including rice and wheat as commodities, demonstrated high level of market integration in the eighteenth century.

The model can also be used to illustrate how shocks to one variable transmitting to other variables. We adopt the definition of Koop, Pesaran and Potter (1996) to derive the generalized impulse response function, which should be

\[
IR(\tau, \mu_{nm,t}, I_{t-1}) = E(\text{vec}(Y'_{nm,t+\tau})|\mu_{nm,t}, I_{t-1}) - E(\text{vec}(Y'_{nm,t+\tau})|I_{t-1})
= H_{nm}^{\tau-1} S_{nm}^{-1} \text{vec}(\mu_{nm,t}),
\]

97
Table 2.5: Empirical Application: System Estimation Results

<table>
<thead>
<tr>
<th></th>
<th>Panel A</th>
<th></th>
<th>Panel B</th>
<th></th>
<th>Panel C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\theta$</td>
<td>$\theta^c$</td>
<td>$\theta$</td>
<td>$\theta^c$</td>
<td>$\theta$</td>
<td>$\theta^c$</td>
</tr>
<tr>
<td>$p_{11}^*$</td>
<td>0.9850</td>
<td>1.0032</td>
<td>0.9658</td>
<td>0.9600</td>
<td>0.9850</td>
<td>0.9885</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0160)</td>
<td>(0.0160)</td>
<td>(0.0150)</td>
<td>(0.0150)</td>
<td>(0.0064)</td>
<td>(0.0064)</td>
</tr>
<tr>
<td>$p_{12}^*$</td>
<td>0.1136</td>
<td>0.1242</td>
<td>0.1173</td>
<td>0.1150</td>
<td>0.1136</td>
<td>0.1134</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0170)</td>
<td>(0.0170)</td>
<td>(0.0159)</td>
<td>(0.0159)</td>
<td>(0.0088)</td>
<td>(0.0088)</td>
</tr>
<tr>
<td>$p_{22}^*$</td>
<td>0.8565</td>
<td>0.8665</td>
<td>0.8423</td>
<td>0.8376</td>
<td>0.8565</td>
<td>0.8618</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0176)</td>
<td>(0.0176)</td>
<td>(0.0166)</td>
<td>(0.0166)</td>
<td>(0.0121)</td>
<td>(0.0121)</td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.5498</td>
<td>0.5689</td>
<td>0.5696</td>
<td>0.5493</td>
<td>0.5498</td>
<td>0.5700</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0107)</td>
<td>(0.0107)</td>
<td>(0.0108)</td>
<td>(0.0108)</td>
<td>(0.0106)</td>
<td>(0.0106)</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.0676</td>
<td>0.0672</td>
<td>0.0664</td>
<td>0.0661</td>
<td>0.0676</td>
<td>0.0681</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0095)</td>
<td>(0.0095)</td>
<td>(0.0095)</td>
<td>(0.0095)</td>
<td>(0.0094)</td>
<td>(0.0094)</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.6313</td>
<td>0.6468</td>
<td>0.6488</td>
<td>0.6329</td>
<td>0.6313</td>
<td>0.6471</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0097)</td>
<td>(0.0097)</td>
<td>(0.0097)</td>
<td>(0.0097)</td>
<td>(0.0097)</td>
<td>(0.0097)</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.3994</td>
<td>-0.4092</td>
<td>-0.3920</td>
<td>-0.3756</td>
<td>-0.3994</td>
<td>-0.4175</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0215)</td>
<td>(0.0215)</td>
<td>(0.0203)</td>
<td>(0.0203)</td>
<td>(0.0131)</td>
<td>(0.0131)</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>-0.0672</td>
<td>-0.0658</td>
<td>-0.0617</td>
<td>-0.0620</td>
<td>-0.0672</td>
<td>-0.0699</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0213)</td>
<td>(0.0213)</td>
<td>(0.0199)</td>
<td>(0.0199)</td>
<td>(0.0142)</td>
<td>(0.0142)</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.5811</td>
<td>-0.5912</td>
<td>-0.5648</td>
<td>-0.5522</td>
<td>-0.5811</td>
<td>-0.5930</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0201)</td>
<td>(0.0201)</td>
<td>(0.0190)</td>
<td>(0.0190)</td>
<td>(0.0151)</td>
<td>(0.0151)</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.8346</td>
<td>0.8435</td>
<td>0.7881</td>
<td>0.7863</td>
<td>0.8346</td>
<td>0.8360</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0139)</td>
<td>(0.0139)</td>
<td>(0.0133)</td>
<td>(0.0133)</td>
<td>(0.0057)</td>
<td>(0.0057)</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0.1132</td>
<td>0.1227</td>
<td>0.1126</td>
<td>0.1109</td>
<td>0.1132</td>
<td>0.1152</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0174)</td>
<td>(0.0174)</td>
<td>(0.0163)</td>
<td>(0.0163)</td>
<td>(0.0089)</td>
<td>(0.0089)</td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>0.8064</td>
<td>0.8109</td>
<td>0.7583</td>
<td>0.7569</td>
<td>0.8064</td>
<td>0.8077</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0141)</td>
<td>(0.0141)</td>
<td>(0.0135)</td>
<td>(0.0135)</td>
<td>(0.0069)</td>
<td>(0.0069)</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.0091</td>
<td>0.0093</td>
<td>0.0093</td>
<td>0.0092</td>
<td>0.0091</td>
<td>0.0093</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.0019</td>
<td>0.0020</td>
<td>0.0020</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0020</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.0064</td>
<td>0.0065</td>
<td>0.0065</td>
<td>0.0064</td>
<td>0.0064</td>
<td>0.0065</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
</tbody>
</table>

Likelihood | 3.8586 | 3.8580 | 3.8586

Note: $\theta$ is QMLE and $\theta^c$ is QMLE with bias correction. The likelihood is the value of log likelihood function when we substitute the estimates into it.
where \( \tau > 1 \) represents the number of periods after a shock \( \mu_{nm,t} \) and \( I_{t-1} \) summarizes all the information available up to period \( t - 1 \). For example, in Figure 2.1, there is one unit shock in period \( t \) on the price of rice in Anqing City in Anhui Province. The upper graph shows how the shock transmits to the rice prices in 13 cities of this province across time and space. The lower graph shows the propagation of the shock to the wheat prices in 13 cities of the province across time and space. Here we use Panel C bias-corrected estimates to generate the graphs. The graphs show visually how a shock to one variable (here it is the rice price) in a city (here it is Anqing) could be propagated to other variables in this city and its neighbours.

Table 2.6 provides results for some robust checks. Panel D uses the exponential weight matrix with \( c_m = 1 \); panel E uses a step weigh matrix, which assigns \( w_{i,j} = 1 \) if the distance between \( i \) and \( j \) is less than 250km and \( w_{i,j} = 0.05 \) if the distance is less than 500km. Panels F and G check robust estimation using sub-sample: first half periods and second half periods. The observation that markets are integrated does not change in these experiments.

### 2.9 Conclusion

This paper considers identification and estimation of multivariate and simultaneous equations dynamic panel spatial autoregressive models. Behaviours of spatial units are assumed to depend on own time lags, and to respond to their neighbours’ or peers’ behaviours in current period (spatial lags), and in previous period (spatial-time lags). The disturbances in the model are specified with time fixed effects, individual fixed effects and idiosyncratic disturbances.
Figure 2.1: Impulse Response to One Unit Shock on Rice Price in Anqing City, Anhui Province
Table 2.6: Empirical Application: Robust Check

<table>
<thead>
<tr>
<th></th>
<th>Panel D</th>
<th>Panel E</th>
<th>Panel F</th>
<th>Panel G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_m = 1$</td>
<td>250km, 500km</td>
<td>First 1/2 periods</td>
<td>Second 1/2 periods</td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>1.0201</td>
<td>1.2125</td>
<td>0.9889</td>
<td>1.0193</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0175)</td>
<td>(0.0175)</td>
<td>(0.0175)</td>
<td>(0.0175)</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.1123</td>
<td>0.2512</td>
<td>0.1163</td>
<td>0.1338</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0187)</td>
<td>(0.0187)</td>
<td>(0.0184)</td>
<td>(0.0184)</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.8802</td>
<td>0.9770</td>
<td>0.8705</td>
<td>0.8853</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0190)</td>
<td>(0.0190)</td>
<td>(0.0189)</td>
<td>(0.0189)</td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.5506</td>
<td>0.5336</td>
<td>0.5451</td>
<td>0.5634</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0107)</td>
<td>(0.0107)</td>
<td>(0.0107)</td>
<td>(0.0107)</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.0696</td>
<td>0.0589</td>
<td>0.0678</td>
<td>0.0670</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0094)</td>
<td>(0.0094)</td>
<td>(0.0094)</td>
<td>(0.0094)</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.6291</td>
<td>0.6394</td>
<td>0.6165</td>
<td>0.6319</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0097)</td>
<td>(0.0097)</td>
<td>(0.0098)</td>
<td>(0.0098)</td>
</tr>
<tr>
<td>$\phi_{11}$</td>
<td>-0.4284</td>
<td>-0.3632</td>
<td>-0.3768</td>
<td>-0.3814</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0232)</td>
<td>(0.0232)</td>
<td>(0.0231)</td>
<td>(0.0231)</td>
</tr>
<tr>
<td>$\phi_{12}$</td>
<td>-0.0751</td>
<td>-0.0967</td>
<td>-0.0736</td>
<td>-0.0732</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0233)</td>
<td>(0.0233)</td>
<td>(0.0230)</td>
<td>(0.0230)</td>
</tr>
<tr>
<td>$\phi_{22}$</td>
<td>-0.6181</td>
<td>-0.5965</td>
<td>-0.5514</td>
<td>-0.5592</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0216)</td>
<td>(0.0216)</td>
<td>(0.0215)</td>
<td>(0.0215)</td>
</tr>
<tr>
<td>$\psi_{11}$</td>
<td>0.8979</td>
<td>1.0421</td>
<td>0.8206</td>
<td>0.8373</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0146)</td>
<td>(0.0146)</td>
<td>(0.0157)</td>
<td>(0.0157)</td>
</tr>
<tr>
<td>$\psi_{12}$</td>
<td>0.1178</td>
<td>0.2890</td>
<td>0.1220</td>
<td>0.1400</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0191)</td>
<td>(0.0191)</td>
<td>(0.0195)</td>
<td>(0.0195)</td>
</tr>
<tr>
<td>$\psi_{22}$</td>
<td>0.8691</td>
<td>0.9341</td>
<td>0.8054</td>
<td>0.8126</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0149)</td>
<td>(0.0149)</td>
<td>(0.0154)</td>
<td>(0.0154)</td>
</tr>
<tr>
<td>$\sigma_{11}$</td>
<td>0.0092</td>
<td>0.0094</td>
<td>0.0099</td>
<td>0.0101</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>$\sigma_{12}$</td>
<td>0.0020</td>
<td>0.0021</td>
<td>0.0022</td>
<td>0.0022</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>$\sigma_{22}$</td>
<td>0.0064</td>
<td>0.0066</td>
<td>0.0067</td>
<td>0.0069</td>
</tr>
<tr>
<td>S.D.</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
</tbody>
</table>

Likelihood | 3.8552 | 3.8058 | 3.8181 | 3.9842

Note: $\theta$ is QMLE and $\theta^c$ is QMLE with bias correction.
We consider identification for the dynamic model with simultaneous, spatial and dynamic time effects. We treat a multivariate model as a “quasi-reduced” form of a simultaneous equations model. The identification from quasi-reduced form parameters to structural parameters is similar to that of a traditional simultaneous equations model. For the multivariate model, we provide identification conditions: the existence of valid instruments for spatial lags, and regularities of the weight matrix structure from disturbances.

For estimation, we propose the QMLE. We first eliminate time fixed effects and then concentrate out individual fixed effects. The concentrated log likelihood function is then maximized. We establish consistency and asymptotic distribution of the estimator. An asymptotic bias may exist depending on comparison of $T$ and $n$. When $T$ is asymptotically large to $n$, the QMLE is asymptotically normal and is centered around zero. When $T$ is asymptotically proportional to $n$, the QMLE will not center at zero. When $T$ is asymptotically small to $n$, the QMLE is $T$-consistent but its asymptotic limiting distribution may degenerate. We also analyze the case of spatial cointegration. The convergence rate of the QMLE may not be the usual $\sqrt{nT}$. We reparameterize the model by combining all unstable components into a single variable. The estimated coefficients in front of the stable variables are still $\sqrt{nT}$ consistent but the estimated coefficients of the unstable components are $\sqrt{nT^{3}}$ superconsistent. Furthermore, we study a robust estimation method which can be applied to stable case, spatial cointegration case and some spatial explosion cases.

The model with spatial cointegration is applied to study possible grain market integration using a unique historical dataset of grain prices. Previous researches
consider rice prices solely while we add the multivariate feature since wheat is believed to be a substitute for rice. The empirical result shows that rice prices and wheat prices are spatially cointegrated among each other across cities. These results provide evidence of interregional and intertemporal grain market integration for a trading network in the eighteenth-century Yangtze River basin.

Future studies in these related research issues may include nonlinearity and models with dominating stochastic trend. Nonlinearity is important in empirical studies to handle multiple share data, and nonlinear models derived with general objective functions (such as utility function, technology function, etc.). In the current model the deterministic trend dominates, but a model without exogenous regressors and individual effects resembles a VAR model, the stochastic trend may dominate, which would alter convergence rate of estimated parameters and their asymptotic distributions.
Bibliography


Lee, L.f., Yu, J., 2010c. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory 26, 564–597.


Appendix A: Covariance Matrix for QMLE in Chapter 1

A.1 SESAR Model

For the SESAR model (1.1), its second order derivatives are

\[
\frac{\partial^2 \ln L_{nm}}{\partial \text{vec}(C_m) \partial \text{vec}(C_m')} = -\Sigma_{um}^{-1} \otimes (X_n'X_n),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \text{vec}(C_m) \partial \Sigma_{um,ij}} = -[(\Sigma_{um}^{-1}F_{m,ij}\Sigma_{um}^{-1}) \otimes X_n']\text{vec}(U_{nm}(\theta)),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \text{vec}(C_m) \partial \Lambda_{m,ij}} = -[(\Sigma_{um}E_{m,ij}') \otimes (X_n'W_n)]\text{vec}(Y_{nm}),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \text{vec}(C_m) \partial \Gamma_{m,ij}} = [(\Sigma_{um}^{-1}E_{m,ij}') \otimes X_n']\text{vec}(Y_{nm}),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \Sigma_{um,ij} \partial \Sigma_{um,kl}} = \frac{n}{2} \text{Tr}[\Sigma_{um}^{-1}F_{m,ij}\Sigma_{um}^{-1}F_{m,kl}]
\]

\[
- \text{vec}(U_{nm}(\alpha))'[(\Sigma_{um}^{-1}F_{m,ij}\Sigma_{um}^{-1}F_{m,kl}\Sigma_{um}^{-1}) \otimes I_n]\text{vec}(U_{nm}(\theta)),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \Sigma_{um,ij} \partial \Lambda_{m,kl}} = -\text{vec}(U_{nm}(\theta))'[(\Sigma_{um}^{-1}F_{m,ij}\Sigma_{um}^{-1}E_{m,kl}') \otimes W_n]\text{vec}(Y_{nm}),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \Sigma_{um,ij} \partial \Gamma_{m,kl}} = \text{vec}(U_{nm}(\theta))'[(\Sigma_{um}^{-1}F_{m,ij}\Sigma_{um}^{-1}E_{m,kl}') \otimes I_n]\text{vec}(Y_{nm}),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \Lambda_{m,ij} \partial \Lambda_{m,kl}} = -\text{Tr}[S_{nm,s}^{-1}(E_{m,ij}' \otimes W_n)S_{nm,s}^{-1}(E_{m,kl}' \otimes W_n)]
\]

\[
- \text{vec}(Y_{nm})'[(E_{m,ij}\Sigma_{um}^{-1}E_{m,kl}) \otimes (W_n'W_n)]\text{vec}(Y_{nm}),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \Lambda_{m,ij} \partial \Gamma_{m,kl}} = \text{Tr}[S_{nm,s}^{-1}(E_{m,ij}' \otimes W_n)S_{nm,s}^{-1}(E_{m,kl}' \otimes I_n)]
\]

\[
+ \text{vec}(Y_{nm})'[(E_{m,ij}\Sigma_{um}^{-1}E_{m,kl}') \otimes W_n']\text{vec}(Y_{nm}),
\]

\[
\frac{\partial^2 \ln L_{nm}}{\partial \Gamma_{m,ij} \partial \Gamma_{m,kl}} = -\text{Tr}[S_{nm,s}^{-1}(E_{m,ij}' \otimes I_n)S_{nm,s}^{-1}(E_{m,kl}' \otimes I_n)]
\]
\[ -\text{vec}(Y_{nm})'(E_{m,ij}\Sigma^{-1}_{um}E'_{m,kl}) \otimes I_n]\text{vec}(Y_{nm}), \]

for \(i, j, k, l = 1, \ldots, m\), where \(U_{nm}(\theta) = \Gamma_mY_{nm} - W_nY_{nm}\Lambda_m - X_nC_m\).

Let \(\alpha = (\text{vec}(C_m)', \text{vec}(\Lambda_m)', \text{vec}^*(\Sigma_{um})', \text{vec}^*(\Gamma_m)')'\). With constraints, the deep parameter vector will be \(\theta\) such that \(\alpha = \alpha(\theta)\). We note that at the true parameter vector \(\alpha_0 = \alpha(\theta_0)\) and \(E(\partial^2 \ln L_{nm}(\theta_0)) = \frac{\partial \alpha'(\theta_0)}{\partial \theta} E(\frac{\partial^2 \ln L_{nm}(\alpha_0)}{\partial \alpha \partial \alpha'}) \frac{\partial \alpha(\theta_0)}{\partial \theta} \). For the SESAR model, \(E \left[ \left( \frac{\partial \ln L_{nm}(\theta_0)}{\partial \alpha} \right) \left( \frac{\partial \ln L_{nm}(\theta_0)}{\partial \alpha} \right)' \right] = \Omega + \Xi\) where \(\Omega = \begin{pmatrix} \Omega_A & \Omega_D & \Omega_G & \Omega_P \\ \Omega_D' & \Omega_B & \Omega_H & \Omega_R \\ \Omega_G' & \Omega_H' & \Omega_C & \Omega_Q \\ \Omega_P' & \Omega_R' & \Omega_Q' & \Omega_M \end{pmatrix}\) and also \(\Xi = \begin{pmatrix} \Xi_A \\ \Xi_B \\ \Xi_C \\ \Xi_M \end{pmatrix}\). \(\Omega_A\) and \(\Xi_A\) are \(k_xm \times k_xm\) matrices; \(\Omega_B\) and \(\Xi_B\) are \(m^2 \times m^2\) matrices; \(\Omega_C\) and \(\Xi_C\) are \(m(m+1)/2 \times m(m+1)/2\) matrices; \(\Omega_D\) and \(\Xi_D\) are \(k_xm \times m^2\) matrices; \(\Omega_H\) and \(\Xi_H\) are \(m^2 \times m(m+1)/2\) matrices; \(\Omega_G\) and \(\Xi_G\) are \(k_xm \times m(m+1)/2\) matrices; \(\Omega_M\) and \(\Xi_M\) are \((m^2 - m) \times (m^2 - m)\) matrices; \(\Omega_P\) and \(\Xi_P\) are \(k_xm \times (m^2 - m)\) matrices; \(\Omega_Q\) and \(\Xi_Q\) are \(m^2 \times (m^2 - m)\) matrices; \(\Omega_R\) and \(\Xi_R\) are \((m(m+1)/2) \times (m^2 - m)\) matrices. Each of the block matrices has its specific expressions as follows:

1. **Block A:** \(\Omega_A = \Sigma^{-1}_{um0} \otimes (X'_nX_n), \) \(\Xi_A = 0\).

2. **Block B:** The \((i+(j-1)m) \times (k+(l-1)m)\)th element, i.e., \(E \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \alpha_{m,ij}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \alpha_{m,kl}} \right]\), in \(\Omega_B\) and \(\Xi_B\) are

\[
\Omega_{i+(j-1)m,k+(l-1)m,B} = \text{vec}(C_{m0})'(I_m \otimes X'_n)S^{-1}_{vm0}[(E_{m,ij}\Sigma^{-1}_{um0}E'_{m,kl}) \otimes (W'_nW_n)] \\
\times S^{-1}_{vm0}(I_m \otimes X_n)\text{vec}(C_{m0}) \\
+ \text{Tr} \left[ (E'_{m,kl} \otimes W'_n)S^{-1}_{vm0}(E'_{m,ij} \otimes W'_n)S^{-1}_{vm0} \right] \\
+ \text{Tr} \left[ S^{-1}_{vm0}(E_{m,ij}\Sigma^{-1}_{um0}E'_{m,kl}) \otimes (W'_nW_n) \right] \text{vec}(C_{m0}) \otimes I_n 
\]
\[ \Xi_{i+(j-1)m,k+(l-1)m,B} = \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{stpq} - \sigma_{sp}^2 \sigma_{tq}^2 - \sigma_{sq}^2 \sigma_{tp}^2 - \sigma_{st}^2 \sigma_{pq}^2 \right) \times \left( \sum_{h=1}^{n} B_{nst,hh}^{ij} B_{npq,hh}^{kl} \right) \]

where \( B_{nst}^{ij} \) is the \( n \times n \) submatrix with the index \((s,t)\) block in the block matrix

\[ \left( \Sigma_{0}^{-1} E_{m,ij} \otimes W \right) S_{m0}^{-1} \] and \( b_{ns}^{ij} \) is the \( s\)-th subvector in the vector \( \text{vec}(C_0) \) \((I \otimes X'_{0}) S_{m0}^{-1} \left[ (E_{m,ij} \Sigma_{0}^{-1}) \otimes W' \right] \) with length \( 1 \times n \). Recall \( u_{stpq} = \mathbb{E}(v_{ns},v_{it}v_{npq},v_{q}) \) and \( u_{stp} = \mathbb{E}(v_{ns},v_{it}v_{npq},v_{q}) \) for \( i = 1, \ldots, n; s, t, p, q = 1, \ldots, m \).

(3) Block C: The \((i+(j-1)m) \times (k+(l-1)m)\)th element, i.e., \( \mathbb{E} \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \sigma_{ij}} \right] \), in \( \Omega_C \) and \( \Xi_C \) are \( \Omega_{i+(j-1)m,k+(l-1)m,C} = \frac{1}{2} \text{Tr} \left[ \Sigma_{nm}^{-1} F_{m,ij} \Sigma_{nm}^{-1} F_{m,kl} \right] \) and

\[ \Xi_{i+(j-1)m,k+(l-1)m,C} = \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{stpq} - \sigma_{sp}^2 \sigma_{tq}^2 - \sigma_{sq}^2 \sigma_{tp}^2 - \sigma_{st}^2 \sigma_{pq}^2 \right) \times \left( \sum_{h=1}^{n} C_{nst,hh}^{ij} C_{npq,hh}^{kl} \right) \]

where \( C_{nst}^{ij} \) is the \( n \times n \) submatrix with index \((s,t)\) block in \( \left( \Sigma_{nm}^{-1} F_{m,ij} \Sigma_{nm}^{-1} \right) \otimes I_n \).

(4) Block D: The \( i+m \)th column, i.e., \( \mathbb{E} \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \text{vec}(C_0)} \right] \), in \( \Omega_D \) and \( \Xi_D \) are \( \Omega_{s,i+(j-1)m,D} = \left( \Sigma_{nm}^{-1} E_{m,ij} \otimes (X_n W_n) \right) S_{nm0}^{-1} (I \otimes X_n') \text{vec}(C_0) \) and

\[ \Xi_{s,i+(j-1)m,D} = \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} u_{stpq} \left( \sum_{h=1}^{n} d_{ns,h}^{ij} D_{npq,hh}^{ij} \right) \]

where \( D_{npq}^{ij} \) is the \( n \times n \) submatrix with index \((p,q)\) block in \( \left( \Sigma_{nm}^{-1} E_{m,ij} \otimes W_n \right) S_{nm0}^{-1} \) and \( d_{ns}^{ij} \) is the \( s\)-th vector in the \( r+(t-1)k \)th row of \( \Sigma_{nm}^{-1} \otimes X_n' \) with length \( 1 \times n \).

(5) Block G: The \((i+(j-1)m) \times (k+(l-1)m)\)th element, i.e., \( \mathbb{E} \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \sigma_{kl}} \right] \), in \( \Omega_G \) and \( \Xi_G \) are \( \Omega_{i+(j-1)m,k+(l-1)m,G} = 0 \) and

\[ \Xi_{i+(j-1)m,k+(l-1)m,G} = \frac{1}{2} \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{stpq} \left( \sum_{h=1}^{n} d_{ns,h}^{ij} C_{npq,hh}^{kl} \right) \right) \]

111
where $G_{npq}^{ij}$ is the $n \times n$ submatrix with index $(p, q)$ block in $(\Sigma_{um0}^{−1}F_{m,ij} \Sigma_{um0}^{−1}) \otimes I_n$ and $g_{ns}^{ij}$ is the $s$-th vector in the $i + (j−1)k$th row of $\Sigma_{um0}^{−1} \otimes X_n'$ with length $1 \times n$.

(6) Block H: The $(i + (j−1)m) \times (k + (l−1)m)\text{th element, i.e., } \mathbb{E} \left[ \frac{\partial \ln L_{am}(\theta_0)}{\partial \ln L_{am}(\theta_0)} \right],$

in $\Omega_H$ and $\Xi_H$ are $\Omega_{i+(j−1)m,k+(l−1)m} = \text{Tr} \left[ (F_{m,kl} \Sigma_{um0}^{−1} E_{m,ij}) \otimes (W_n) \right] S_{nm0}^{−1}$ and

\[
\Xi_{i+(j−1)m,k+(l−1)m} = \frac{m}{\sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{stpqn}^2 - \sigma_{p,q}^2/\sigma_{q,p}^2 - \sigma_{q,p}^2/\sigma_{p,q}^2 \right)} \times \left( \sum_{r=1}^{n} H_{nst,rr}^{ij} H_{npq,rr}^{kl} \right) + \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} u_{stpqn} \left( h_{nst,rr}^{ij} H_{npq,rr}^{kl} \right),
\]

where $H_{nst,rr}^{ij}$ is the $n \times n$ submatrix with index $(s, t)$ block in $[\Sigma_{um0}^{−1} \otimes \Sigma_{nm0}^{−1}]$, $H_{npq,rr}^{kl}$ is the submatrix with index $(p, q)$ block in $(\Sigma_{um0}^{−1} \otimes \Sigma_{nm0}^{−1}) \otimes \Sigma_{nm0}^{−1}$, and $h_{nst,rr}^{ij}$ is the $s$-th vector in the vector $\text{vec}(C_{m0})' (I_m \otimes X_n') S_{nm0}^{−1} \left[ (E_{m,ij} \Sigma_{um0}^{−1}) \otimes W_n' \right]$ with length $1 \times n$.

(7) Block M: The $(i + (j−1)m) \times (k + (l−1)m)\text{th element, i.e., } \mathbb{E} \left[ \frac{\partial \ln L_{am}(\theta_0)}{\partial \ln L_{am}(\theta_0)} \right],$

in $\Omega_M$ and $\Xi_M$ are

\[
\Omega_{i+(j−1)m,k+(l−1)m} = \text{vec}(C_{m0})' (I_m \otimes X_n') S_{nm0}^{−1} \left[ (E_{m,ij} \Sigma_{um0}^{−1} E_{m,kl}) \otimes (I_n) \right] \times S_{nm0}^{−1} (I_m \otimes X_n) \text{vec}(C_{m0}) + \text{Tr} \left[ (E_{m,kl} \otimes I_n) S_{nm0}^{−1} (E_{m,ij} \otimes I_n) S_{nm0}^{−1} \right] + \text{Tr} \left[ S_{nm0}^{−1} [(E_{m,ij} \Sigma_{um0}^{−1} E_{m,kl}') \otimes I_n] S_{nm0}^{−1} (\Sigma_{um0} \otimes I_n) \right],
\]

\[
\Xi_{i+(j−1)m,k+(l−1)m} = \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{stpqn}^2 - \sigma_{p,q}^2/\sigma_{q,p}^2 - \sigma_{q,p}^2/\sigma_{p,q}^2 \right) \times \left( \sum_{h=1}^{n} M_{nst,hh}^{ij} M_{npq,hh}^{kl} \right) + \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} u_{stpqn} \left( o_{nst,rr}^{ij} M_{npq,rr}^{kl} + o_{nst,rr}^{kl} M_{npq,rr}^{ij} \right),
\]

112
where $M^{ij}_{nst}$ is the $n \times n$ submatrix with index $(s, t)$ block in $\left[ \left( \Sigma_{um0}^{-1} E_{m,ij} \right) \otimes I_n \right] S_{nm0}^{-1}$ and $\sigma_{nst}^i$ is the $s$-th vector in $\text{vec}(C_{m0})'(I_m \otimes X_n')S_{nm0}^{-1} \left[ \left( E_{m,ij} \Sigma_{um0}^{-1} \right) \otimes I_n \right]$ with length $1 \times n$. Recall $u_{stpq} = \mathbb{E}(v_{n,is}v_{n,it}v_{n,jp}v_{n,jq})$ and $u_{stp} = \mathbb{E}(v_{n,is}v_{n,it}v_{n,ip})$ for $i = 1, \ldots, n$; $s, t, p, q = 1, \ldots, m$.

(8) Block P: The $i + (j - 1)m$th column, i.e., $\mathbb{E} \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \text{vec}(C_{m0})} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \Gamma_{ij}} \right]$, in $\Omega_P$ and $\Xi_P$ are $\Omega_{i+(j-1)m,P} = - \left[ \left( \Sigma_{um0}^{-1} E_{m,ij} \right) \otimes X_n' \right] S_{nm0}^{-1} \left( I_m \otimes X_n' \right) \text{vec}(C_{m0})$ and

$$\Xi_{s,i+(j-1)m,P} = - \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} u_{stpq} \left[ \left( \sum_{h=1}^{n} P_{ns,h}^{11} P_{npq,hh}^{ij} \right)' \cdots \left( \sum_{h=1}^{n} P_{ns,h}^{rt} P_{npq,hh}^{ij} \right)' \right] \\
\quad \times \ldots \left( \sum_{h=1}^{n} P_{ns,h}^{k,m} P_{npq,hh}^{ij} \right)' \right]',$$

where $P_{npq}^{ij}$ is the $n \times n$ submatrix with index $(p, q)$ block in $\left[ \left( \Sigma_{um0}^{-1} E_{m,ij} \right) \otimes I_n \right] S_{nm0}^{-1}$ and $P_{npq}^{rt}$ is the $s$-th vector in the $r + (t - 1)k$th row of $\Sigma_{um0}^{-1} \otimes X_n'$ with length $1 \times n$.

(9) Block Q: The $(i+(j-1)m) \times (k+(l-1)m)$th element, i.e., $\mathbb{E} \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \Gamma_{ij}} \right]$, in $\Omega_Q$ and $\Xi_Q$ are $\Omega_{i+(j-1)m,k+(l-1)m,Q} = \text{Tr} \left[ \left( F_{m,kl} \Sigma_{um0}^{-1} E_{m,ij} \right) \otimes I_n \right] S_{nm0}^{-1}$ and

$$\Xi_{i+(j-1)m,k+(l-1)m,H} = \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ \left( u_{stpq} - \sigma_{sp}^2 \sigma_{tp}^2 - \sigma_{sq}^2 \sigma_{tp}^2 - \sigma_{st}^2 \sigma_{pq} \right) \right] \\
\quad \times \left( \sum_{r=1}^{n} Q_{1,nst,rr}^{ij} Q_{2,ntp,rr}^{kl} \right) \\
\quad + \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} u_{stpq} \left( q_{nst,rr}^{ij} Q_{2,ntp,rr}^{kl} \right),$$

where $Q_{1,nst}^{ij}$ is the $n \times n$ submatrix with index $(s, t)$ block in $\left[ \left( \Sigma_{um0}^{-1} E_{m,ij} \right) \otimes I_n \right] S_{nm0}^{-1}$, $Q_{2,ntp,rr}^{kl}$ is the submatrix with index $(p, q)$ block in $\left( \Sigma_{um0}^{-1} F_{m,ij} \Sigma_{um0}^{-1} \right) \otimes I_n$ and $q_{nst,rr}^{ij}$ is the $s$-th vector in the vector $\text{vec}(C_{m0})'(I_m \otimes X_n')S_{nm0}^{-1} \left[ \left( E_{m,ij} \Sigma_{um0}^{-1} \right) \otimes I_n \right]$ with length $1 \times n$. 

113
(10) Block $R$: The $(i+(j-1)m) \times (k+(l-1)m)$ th element, i.e., $E \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \lambda_{m,ij}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial v_{m,kl}} \right]$, in $\Omega_R$ and $\Xi_R$ are

\[
\Omega_{i+(j-1)m,k+(l-1)m,R} = -\text{vec}(C_{m0})'(I_m \otimes X_n')S_{v_m0}^{-1}[(E_{m,ij}\Sigma_{um0}^{-1}E_{m,kl}') \otimes W_n'] \\
\times S_{v_m0}^{-1}(I_m \otimes X_n)\text{vec}(C_{m0}) \quad - \quad \text{Tr} \left[ (E_{m,kl}' \otimes W_n')S_{v_m0}^{-1}(E_{m,ij} \otimes I_n)S_{v_m0}^{-1} \right] \\
- \text{Tr} \left[ S_{v_m0}^{-1}(E_{m,ij}\Sigma_{um0}^{-1}E_{m,kl}') \otimes W_n' \right]S_{v_m0}^{-1}(\Sigma_{um0} \otimes I_n),
\]

$\Xi_{i+(j-1)m,k+(l-1)m,R}$ = $-\sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ u_{stpq} - \sigma_{sp}^2 \sigma_{tq}^2 - \sigma_{sq}^2 \sigma_{tp}^2 - \sigma_{st}^2 \sigma_{pq}^2 \right] \\
\times \left( \sum_{h=1}^{n} R^{ij}_{nst,hh,R^{2kl}_{nsp,hh}} \right) \\
- \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} u_{stpq} \sum_{h=1}^{n} \left( r^{ij}_{nst,hh,R^{2kl}_{nsp,hh}} + r^{2kl}_{nst,hh,R^{ij}_{nsp,hh}} \right),
\]

where $R^{ij}_{nst}$ is the $n \times n$ submatrix with index $(s, t)$ block in $[(\Sigma_{um0}^{-1}E_{m,ij}') \otimes W_n']S_{nm0}^{-1}$, $R^{2ij}_{nst}$ is the $n \times n$ submatrix with index $(s, t)$ block in $[(\Sigma_{um0}^{-1}E_{m,ij}') \otimes I_n]S_{nm0}^{-1}$, $r^{ij}_{nst}$ is the $s$-th vector in $\text{vec}(C_{m0})'(I_m \otimes X_n')S_{v_m0}^{-1'}[E_{m,ij}(\Sigma_{um0}^{-1}) \otimes W_n']$ with length $1 \times n$, and $r^{2ij}_{nst}$ is the $s$-th vector in $\text{vec}(C_{m0})'(I_m \otimes X_n')S_{v_m0}^{-1'}[(E_{m,ij}\Sigma_{um0}^{-1}) \otimes I_n]$ with length $1 \times n$. Recall $u_{stpq} = \mathbb{E}(v_{n,ip}v_{n,ip}v_{n,ip}v_{n,ip})$ and $u_{stpq} = \mathbb{E}(v_{n,ip}v_{n,ip}v_{n,ip})$ for $i = 1, \ldots, n$ and $s, t, p, q = 1, \ldots, m$.

### A.2 MSAR Model

Here we also provide the covariance matrix for the MSAR model. For the general MSAR model (2), $\theta = (\text{vec}(\Pi_m)', \text{vec}(\Psi_m)', \text{vec}(\Lambda_m)', \text{vec}(\Sigma_m)')$, $E \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \lambda_{m,ij}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial v_{m,kl}} \right] = E \left[ \left( \frac{\partial \ln L_{nm}(\theta_0)}{\partial \theta} \right)' \frac{\partial \ln L_{nm}(\theta_0)}{\partial \theta} \right]' = \Omega + \Xi$ where

\[
\Omega = \begin{pmatrix}
\Omega_A & \Omega_D & \Omega_G \\
\Omega'_D & \Omega_B & \Omega_H \\
\Omega'_G & \Omega'_H & \Omega_C
\end{pmatrix}
\quad \text{and} \quad
\Xi = \begin{pmatrix}
\Xi_A & \Xi_D & \Xi_G \\
\Xi'_D & \Xi_B & \Xi_H \\
\Xi'_G & \Xi'_H & \Xi_C
\end{pmatrix}.
\]

114
\(\Omega_A\) and \(\Xi_A\) are \(k_x m \times k_x m\) matrices; \(\Omega_B\) and \(\Xi_B\) are \(m^2 \times m^2\) matrices; \(\Omega_C\) and \(\Xi_C\) are \(m(m+1)/2 \times m(m+1)/2\) matrices; \(\Omega_D\) and \(\Xi_D\) are \(k_x m \times m^2\) matrices; \(\Omega_H\) and \(\Xi_H\) are \(m^2 \times m(m+1)/2\) matrices; \(\Omega_G\) and \(\Xi_G\) are \(k_x m \times m(m+1)/2\) matrices. Each of the block matrices has its specific expression as follows:

(1) Block A: \(\Omega_A = \Sigma_{m0}^{-1} \otimes (X'_n X_n), \quad \Xi_A = 0.\)

(2) Block B

The \((i + (j - 1)m) \times (k + (l - 1)m)\)th element, i.e., \(E_{ij} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \Psi_{n,ij}}\), in \(\Omega_B\) and \(\Xi_B\) are

\[
\begin{align*}
\Omega_{i+(j-1)m,k+(l-1)m,B} &= \text{vec}(\Pi_{m0})' (I_m \otimes X'_n) S_{vm0}^{-1} \left( (E_{m,ij} \Sigma_{nm0}^{-1} E'_{m,kl}) \otimes (W'_n W_n) \right) \\
&\quad \times S_{vm0}^{-1} (I_m \otimes X_n) \text{vec}(\Pi_{m0}) \\
+ & \text{Tr} \left[ (E'_{m,kl} \otimes W'_n) S_{vm0}^{-1} (E'_{m,ij} \otimes W'_n) S_{vm0}^{-1} \right] \\
+ & \text{Tr} \left[ S_{vm0}^{-1} \left( (E_{m,ij} \Sigma_{nm0}^{-1} E'_{m,kl}) \otimes (W'_n W_n) \right) S_{vm0}^{-1} (\Sigma_{vm0} \otimes I_n) \right], \\
\Xi_{i+(j-1)m,k+(l-1)m,B} &= \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{stpq} - \sigma^2_{sp}\sigma^2_{tq} - \sigma^2_{sq}\sigma^2_{tp} - \sigma^2_{st}\sigma^2_{pq} \right) \\
&\times \left( \sum_{h=1}^{n} B_{nst,hh}^{ij} B_{npt,hh}^{kl} \right) + \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} u_{stp} \sum_{h=1}^{n} \left( b_{nsh}^{ij} B_{npt,hh}^{kl} + b_{nsh}^{kl} B_{npt,hh}^{ij} \right),
\end{align*}
\]

where \(B_{nst,hh}^{ij}\) is the \(n \times n\) submatrix with the index \((s, t)\) block in the block matrix

\[
\left[ (\Sigma_{vm0}^{-1} E'_{m,ij}) \otimes W_n \right] S_{nm0}^{-1}
\]

and \(b_{nsh}^{ij}\) is the \(s\)-th subvector in the vector \(\text{vec}(\Pi_{m0})'(I_m \otimes X'_n) S_{vm0}^{-1} \left( (E_{m,ij} \Sigma_{nm0}^{-1}) \otimes W'_n \right)\) with length \(1 \times n\). Recall \(u_{stpq} = \text{E}(v_{n,Is} v_{n,It} v_{n,Ip} v_{n,Iq})\)

and \(u_{stp} = \text{E}(v_{n,Is} v_{n,It} v_{n,Ip})\) for \(i = 1, \ldots, n; s, t, p, q = 1, \ldots, m.\)

(3) Block C
The \((i + (j - 1)m) \times (k + (l - 1)m)\)th element, i.e., \[\mathbb{E} \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \sigma_{ij}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \sigma_{kl}} \right],\] in \(\Omega_C\) and \(\Xi_C\) are

\[
\Omega_{i+(j-1)m,k+(l-1)m,C} = \frac{n}{2} \text{Tr} \left[ \Sigma_{v0}^{-1} F_{m,ij} \Sigma_{v0}^{-1} F_{m,kl} \right],
\]
\[
\Xi_{i+(j-1)m,k+(l-1)m,C} = \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{spq} - \sigma_{sp}^{2} \sigma_{tq}^{2} - \sigma_{sq}^{2} \sigma_{tp}^{2} - \sigma_{st}^{2} \sigma_{pq}^{2} \right)
\times \left( \sum_{h=1}^{n} C_{nst,hh}^{ij} \right),
\]

where \(C_{nst}^{ij}\) is the \(n \times n\) submatrix with index \((s, t)\) block in \((\Sigma_{v0}^{-1} F_{m,ij} \Sigma_{v0}^{-1}) \otimes I_n\).

(4) Block D

The \(i + (j - 1)m\)th column, i.e., \[\mathbb{E} \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial \text{vec}(\Pi_{m0})} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \psi_{kl}} \right],\] in \(\Omega_D\) and \(\Xi_D\) are

\[
\Omega_{*,i+(j-1)m,D} = \left[ (\Sigma_{v0}^{-1} E_{m,ij}') \otimes (\Xi_n' W_n) \right] S_{nm0}^{-1} (I_m \otimes X_n') \text{vec}(\Pi_{m0}),
\]
\[
\Xi_{*,i+(j-1)m,D} = \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{h=1}^{n} d_{spq,hh}^{ij} \left( \sum_{h=1}^{n} d_{ns,h}^{ij} D_{npq,hh}^{ij} \right)',
\]

where \(D_{npq}^{ij}\) is the \(n \times n\) submatrix with index \((p, q)\) block in \([ (\Sigma_{v0}^{-1} E_{m,ij}') \otimes W_n ) S_{nm0}^{-1}\) and \(d_{spq}^{ij}\) is the \(r + (t - 1)k\)th row of \(\Sigma_{v0}^{-1} \otimes X_n'\) with length \(1 \times n\).

(5) Block G

The \((i + (j - 1)m) \times (k + (l - 1)m)\)th element, i.e., \[\mathbb{E} \left[ \frac{\partial \ln L_{nm}(\theta_0)}{\partial g_{ij}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial g_{kl}} \right],\] in \(\Omega_G\) and \(\Xi_G\) are

\[
\Omega_{i+(j-1)m,k+(l-1)m,G} = 0, \quad \Xi_{i+(j-1)m,k+(l-1)m,G} = \frac{1}{2} \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{spq} \left( \sum_{h=1}^{n} g_{ns,h}^{ij} C_{npq,hh}^{kl} \right) \right),
\]

where \(G_{npq}^{ij}\) is the \(n \times n\) submatrix with index \((p, q)\) block in \((\Sigma_{v0}^{-1} F_{m,ij} \Sigma_{v0}^{-1}) \otimes I_n\) and \(g_{ns}^{ij}\) is the \(s\)-th vector in the \(i + (j - 1)k\)th row of \(\Sigma_{v0}^{-1} \otimes X_n'\) with length \(1 \times n\).

(6) Block H
The \((i + (j - 1)m) \times (k + (l - 1)m)\)th element, i.e., \(E^{\text{\(\frac{\partial \ln L_{nm}(\delta_0)}{\partial \psi_{ij}}\)} \frac{\partial \ln L_{nm}(\delta_0)}{\partial \sigma_{kl}}}\), in \(\Omega_H\) and \(\Xi_H\) are

\[
\Omega_{i+(j-1)m,k+(l-1)m,H} = \text{Tr} \left[ (F_{m,kl}^{-1} \Sigma_m^{-1} E_{m,ij}) \otimes (W_n) \right] S_{nm0}^{-1},
\]
\[
\Xi_{i+(j-1)m,k+(l-1)m,H} = \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ (u_{stpq} - \sigma_{sp}^2 \sigma_{pq}^2 - \sigma_{sq}^2 \sigma_{tp}^2 - \sigma_{st}^2 \sigma_{pq}^2) \right] \times \left( \sum_{r=1}^{n} H_{nst,rr}^{ij} H_{npq,rr}^{kl} \right) + \sum_{s=1}^{m} \sum_{t=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} u_{stp} \sum_{r=1}^{n} (h_{ns,r}^{ij} H_{ntp,rr}^{kl}),
\]

where \(H_{nst,rr}^{ij}\) is the \(n \times n\) submatrix with index \((s, t)\) block in \(\left[ (\Sigma_m^{-1} E_{m,ij}) \otimes (W_n) \right] S_{nm0}^{-1}\), \(H_{npq,rr}^{kl}\) is the submatrix with index \((p, q)\) block in \(\left( \Sigma_m^{-1} F_{m,k} \Sigma_m^{-1} \right) \otimes I_n\). and \(h_{ns,r}^{ij}\) is the \(s\)-th vector in \(\text{vec}(\Pi_m)'(I_m \otimes X_n') S_{nm0}^{-1} \left[ (E_{m,ij}^{-1} \Sigma_m^{-1}) \otimes W_n' \right]\) with length \(1 \times n\).
Appendix B: 2SLS and 3SLS estimations with optimal IVs for SESAR

The investigation of the log likelihood functions sheds light on the optimal IVs of the SESAR and MSAR models. In the literature, Baltagi and Deng (2012) derive the optimal IVs in a bivariate spatial panel model for 2SLS and 3SLS estimation, following the method proposed by Lee (2003), without discussing the asymptotic distributions of those estimators. In this section, we consider 2SLS and 3SLS estimations with optimal IVs for the SESAR model, which nests the MSAR model as a special case. We also show the asymptotic normality of estimators and their efficiency. By vectorization of the SESAR model in (1), the reduced form equations are

$$vec(Y_{nm}) = S_{nm}^{-1}(I_m \otimes X_n)vec(\Pi_m) + S_{nm}^{-1}vec(U_{nm}),$$

where

$$S_{nm} = I_{mn} - \Psi_m' \otimes W_n, \quad \Pi_m = C_m \Gamma_m^{-1} \quad \text{and} \quad \Psi_m = \Lambda_m \Gamma_m^{-1}. \quad \text{The optimal IV for } Y_{nm,i},$$

which is the $i$th column of $Y_{nm}$, is

$$(J_i \otimes I_n)S_{nm0}^{-1}(I_m \otimes X_n)vec(\Pi_m);$$

the optimal IV for $W_nY_{nm,i}$ is

$$(J_i \otimes I_n)S_{nm0}^{-1}(I_m \otimes (W_nX_n))vec(\Pi_m)$$

as shown later.

To simplify the notation, with normalizations on the diagonal elements of $\Gamma_m$ being one, let $\Gamma_m = I_m - B_m$ and denote $Y_{nm}^W = W_nY_{nm}$, therefore, $Y_{nm} = Y_{nm}B_m + Y_{nm}^W\Lambda_m + X_nC_m + U_{nm}$. Let $Z_{nm,l} = [Y_{l,nm}, Y_{l,nm}^W, X_{l,nm}]$ represent corresponding matrices of explanatory variables appeared on the right hand side of
the \( l \)th equation, where exclusion restrictions are imposed. Define

\[
Y_{nm}^{IV} = [(J_1 \otimes I_n)S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m)], \quad \ldots, \quad (J_m \otimes I_n)S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m)],
\]

and

\[
Y_{nm}^{WIV} = [(J_1 \otimes I_n)S_{nm0}^{-1}(I_m \otimes (W_n X_n))\text{vec}(\Pi_m)], \quad \ldots, \quad (J_m \otimes I_n)S_{nm0}^{-1}(I_m \otimes (W_n X_n))\text{vec}(\Pi_m)].
\]

\( Y_{l,nm}^{IV} \) and \( Y_{l,nm}^{WIV} \) are corresponding optimal IVs for endogenous explanatory variables appeared in the \( l \)th equation. Let

\[
Z_{OpIV}^{nm,l} = [Y_{nm}^{IV}, Y_{nm}^{WIV}, X_{l,nk}]
\]

and further let

\[
\lim_{n \to \infty} (Z_{OpIV}^{nm,l}Z_{OpIV}^{nm,l})/n = Q_{zz,l}. \quad \beta_l = (B_{m,t}^l, \Lambda_{m,t}^l, C_{km,t}^l)' \quad \text{excluding parameters restricted to zero, is the vector of unknown coefficients of the} \ l \text{th structural equation.}
\]

### B.1 2SLS with Optimal IVs

To derive the asymptotic variance of the 2SLS with optimal IVs, let \( H_{nm}^{OpIV} \) contains all linearly independent columns of \( Z_{nm}^{OpIV} = [Y_{nm}^{WIV}, X_n] \). Assume that

\[
\lim_{n \to \infty} (H_{nm}^{OpIV}H_{nm}^{OpIV})/n = Q_{HH}. \quad \text{Let} \quad P_{H}^{OpIV} = H_{nm}^{OpIV}(H_{nm}^{OpIV}H_{nm}^{OpIV})^{-1}H_{nm}^{OpIV}, \quad \text{a projector matrix to the column space spanned by the optimal IVs. Then, the feasible 2SLS estimator with optimal IVs is} \quad \hat{\beta}_l^{F2SLS} = (Z_{nm,l}^{OpIV}P_{H}^{OpIV}Z_{nm,l})^{-1}Z_{nm,l}^{OpIV}P_{H}^{OpIV}Y_{nm,l}, \]

where \( P_{H}^{OpIV} \) is \( P_{H}^{OpIV} \) evaluated at consistent estimator \( \hat{\beta} \). The following proposition shows that the feasible 2SLS estimator with optimal IVs has the same asymptotic variance as that of optimal IV estimator.

**Proposition B.1.** Under Assumptions 1.1 and 1.7,

\[
\lim_{n \to \infty} \sqrt{n}(\hat{\beta}_l^{F2SLS} - \beta_{l0}) = N(0, \sigma_{l0}^2 Q_{zz,l}).
\]
Proof. $\sqrt{n}(\hat{\beta}^{F2SLS}_T - \beta_T) = \left( \frac{1}{n} Z_{n,m,l}' \tilde{P}_H^{\text{OpIV}} Z_{n,m,l} \right)^{-1} \frac{1}{\sqrt{n}} Z_{n,m,l}' \tilde{P}_H^{\text{OpIV}} U_{n,m,l}$. Consider the first part,

$$\frac{1}{n} Z_{n,m,l}' \tilde{P}_H^{\text{OpIV}} Z_{n,m,l} = \left( \frac{1}{n} Z_{n,m,l}' \tilde{H}_n^{\text{OpIV}} \right) \left( \frac{1}{n} \tilde{H}_n^{\text{OpIV}} \tilde{H}_n^{\text{OpIV}} \right)^{-1} \left( \frac{1}{n} \tilde{H}_n^{\text{OpIV}} Z_{n,m,l} \right),$$

and $Z_{n,m,l} = Z_{n,m,l}^{\text{OpIV}} + [A_{1,l} \text{vec}(U_{nm}), ..., A_{m_l,l} \text{vec}(U_{nm}), B_{1,l} \text{vec}(U_{nm})$,

$$..., B_{m_l^W,l} \text{vec}(U_{nm}), 0, ..., 0],$$

where $A_{1,l}, ..., A_{m_l,l}$ are $n \times mn$ submatrices in $S_{nm0}^{-1}$ and $B_{1,l}, ..., B_{m_l^W,l}$ are $n \times mn$ submatrices in $(I_m \otimes W_n) S_{nm0}^{-1}$. $m_l$ would be the number of columns in $Y_{l,nm}$ and $m_l^W$ would be the number of columns in $Y_{l,nm}^W$ and $\tilde{H}_n^{\text{OpIV}} = H_n^{\text{OpIV}} + \frac{\partial \tilde{H}_n^{\text{OpIV}}}{\partial \beta}(\hat{\beta} - \beta_0)$.

Therefore, due to A’s and B’s being UB,

$$\frac{1}{n} \tilde{H}_n^{\text{OpIV}} Z_{n,m,l} = \frac{1}{n} H_n^{\text{OpIV}} Z_{n,m,l} + \frac{1}{n} (\hat{\beta} - \beta_0) \frac{\partial \tilde{H}_n^{\text{OpIV}}}{\partial \beta} Z_{n,m,l}$$

$$+ \frac{1}{n} H_n^{\text{OpIV}} [A_{1,l} \text{vec}(U_{nm}), ..., A_{m_l,l} \text{vec}(U_{nm}), B_{1,l} \text{vec}(U_{nm}),$$

$$..., B_{m_l^W,l} \text{vec}(U_{nm}), 0, ..., 0] + \frac{1}{n} (\hat{\beta} - \beta_0) \frac{\partial \tilde{H}_n^{\text{OpIV}}}{\partial \beta}$$

$$\times [A_{1,l} \text{vec}(U_{nm}), ..., A_{m_l,l} \text{vec}(U_{nm}), B_{1,l} \text{vec}(U_{nm}), ..., B_{m_l^W,l} \text{vec}(U_{nm}), 0, ..., 0]$$

$$= \frac{1}{n} H_n^{\text{OpIV}} Z_{n,m,l} + o_R(1).$$

Similarly, $\frac{1}{n} \tilde{H}_n^{\text{OpIV}} \tilde{H}_n^{\text{OpIV}} = \frac{1}{n} H_n^{\text{OpIV}} H_n^{\text{OpIV}} + o_R(1)$. So,

$$\frac{1}{n} Z_{n,m,l}' \tilde{P}_H^{\text{OpIV}} Z_{n,m,l} = \left( \frac{1}{n} Z_{n,m,l}' \tilde{H}_n^{\text{OpIV}} \right) \left( \frac{1}{n} H_n^{\text{OpIV}} H_n^{\text{OpIV}} \right)^{-1} \left( \frac{1}{n} H_n^{\text{OpIV}} Z_{n,m,l} \right) + o_R(1)$$

$$= \frac{1}{n} Z_{n,m,l}' Z_{n,m,l} + o_R(1).$$

Furthermore,

$$\frac{1}{n} Z_{n,m,l}' \tilde{P}_H^{\text{OpIV}} U_{n,m,l} = \left( \frac{1}{n} Z_{n,m,l}' \tilde{H}_n^{\text{OpIV}} \right) \left( \frac{1}{n} H_n^{\text{OpIV}} H_n^{\text{OpIV}} \right)^{-1} \left( \frac{1}{n} H_n^{\text{OpIV}} U_{n,m,l} \right) + o_R(1)$$

$$= \frac{1}{n} Z_{n,m,l}' U_{n,m,l} + o_R(1).$$

In sum, $\lim_{n \to \infty} \sqrt{n}(\hat{\beta}^{F2SLS}_T - \beta_T) = N(0, \sigma^2_{\hat{\beta}^{F2SLS}_T} \Omega_{2z,l}^{-1})$. ⪫
The feasible 2SLS estimator with optimal IVs should have the smallest asymptotic variance comparing with other IV estimators. If researchers use IVs, say $H_{nm}$, the asymptotic variance for $\hat{\beta}^{2SLS}_l$ is

$$\sigma_{ll}^2 \lim_{n \to \infty} \left( \frac{1}{n} H'_{nm} Z_{nm,l}^{OpIV} \right)^{-1} \left( \frac{1}{n} H'_{nm} H_{nm} \right) \left( \frac{1}{n} Z_{nm,l}^{OpIV'} H_{nm} \right)^{-1}.$$

The Schwartz inequality guarantees that $\hat{\beta}_l^{F2SLS}$ have the smallest asymptotic variance among all IV estimators. This has indeed justified the claimed optimality of the chosen ‘best’ IVs. Furthermore, we can also consider the feasible IV estimator. The feasible IV estimator with optimal IVs is

$$\hat{\beta}_{l}^{OpIV} = (\tilde{Z}_{nm,l}^{OpIV'} Z_{nm,l})^{-1} \tilde{Z}_{nm,l}^{OpIV'} Y_{nm,l},$$

where $\tilde{Z}_{nm,l}^{OpIV}$ is $Z_{nm,l}^{OpIV'}$ evaluated by $\beta = \hat{\beta}$. The corollary below shows the feasible IV estimator with optimal IVs has the same asymptotic distribution with 2SLS estimator.

**Corollary B.1.** Under Assumptions 1.1 and 1.7, $\lim_{n \to \infty} \sqrt{n}(\hat{\beta}_l^{OpIV} - \beta_0) = N(0, \sigma_{ll}^2 \tilde{Q}_{zz,l}^{-1})$.

**B.2 3SLS with Optimal IVs**

Let $Z_{nm}^D = diag(Z_{nm,l}) = \begin{pmatrix} Z_{nm,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & Z_{nm,m} \end{pmatrix}$ and $\tilde{Z}_{nm}^D = diag(\tilde{Z}_{nm,l})$ where $\tilde{Z}_{nm,l} = \tilde{Z}_{nm,l}^{OpIV} Z_{nm,l}$. The feasible 3SLS estimator is

$$\hat{\beta}^{F3SLS} = (\tilde{Z}_{nm}^D (\tilde{\Sigma}_{um}^{-1} \otimes I_n) Z_{nm}^D)^{-1} \tilde{Z}_{nm}^D (\tilde{\Sigma}_{um}^{-1} \otimes I_n) vec(Y_{nm}).$$

The following proposition provides the asymptotic distribution of the feasible 3SLS estimator:
**Proposition B.2.** Under Assumptions 1.1 and 1.7,

\[
\lim_{n \to \infty} \sqrt{n}(\hat{\beta}^{F3SLS} - \beta_0) = N \left(0, \lim_{n \to \infty} \left[ \text{diag}(Z_{nm,l}^{opIV})'(\Sigma_{um}^{-1} \otimes I_n) \text{diag}(Z_{nm,l}^{opIV}) \right]^{-1} \right).
\]

**Proof.**

\[
\hat{\beta}^{F3SLS} - \beta_0 = \left[ Z_{nm}(\tilde{\Sigma}_{um}^{-1} \otimes I_n)Z_{nm}' \right]^{-1} \tilde{Z}_{nm}(\tilde{\Sigma}_{um}^{-1} \otimes I_n) \text{vec}(U_{nm}).
\]

So,

\[
\sqrt{n}(\hat{\beta}^{F3SLS} - \beta_0) = \left[ \frac{1}{n} Z_{nm}'(\tilde{\Sigma}_{um}^{-1} \otimes \tilde{P}_{H}^{opIV})Z_{nm} \right]^{-1} \frac{1}{\sqrt{n}} Z_{nm}'(\tilde{\Sigma}_{um}^{-1} \otimes \tilde{P}_{H}^{opIV}) \text{vec}(U_{nm}).
\]

The remained proof strategy shares the same arguments as the proof in Proposition B.1 and can be regarded as a technical extension of it.

The feasible 3SLS estimator with optimal IVs should have the smallest asymptotic variance comparing with other 3SLS estimators. By using IVs, say \(H_{nm}\), the asymptotic variance for \(\hat{\beta}_i^{3SLS}\) is

\[
\lim_{n \to \infty} \left[ \text{diag}(H_{nm}Z_{nm,l}^{opIV})'(\Sigma_{um}^{-1} \otimes (H_{nm}'H_{nm})^{-1}) \text{diag}(H_{nm}Z_{nm,l}^{opIV}) \right]^{-1}.
\]

The Schwartz inequality guarantees that \(\hat{\beta}_i^{F3SLS}\) has the smallest asymptotic variance among all 3SLS estimators.
Appendix C: Proofs of Lemmas, Theorems and Corollaries in
Chapter 1

Proof of Lemma 1.1:
Let \( \sup_{\Psi_m} \sup_n \| \Psi_m \otimes W_n \|_\infty = c < 1 \). Apparently, \( \| S_{nm} \|_\infty \leq \| I_{nm} \|_\infty + \| \Psi_m \otimes W_n \|_\infty \leq 1 + c < \infty \). Furthermore, \( S_{nm}^{-1} = (I_{nm} - \Psi_m \otimes W_n)^{-1} = \sum_{i=0}^\infty (\Psi_m \otimes W_n)^i \).

Then, \( \| S_{nm}^{-1} \|_\infty = \| I_{nm} - (\Psi_m \otimes W_n)^{-1} \|_\infty \leq \sum_{i=0}^\infty (\| \Psi_m \otimes W_n \|_\infty)^i = \frac{1}{1-c} < \infty \). So \( S_{nm}^{-1} \) is bounded in row sum uniformly. A similar argument applies for the proof of boundedness in column sum of \( S_{nm} \) and \( S_{nm}^{-1} \).

\[ \square \]

Proof of Proposition 1.1: By the arithmetic and geometric means inequality of eigenvalues,

\[
\frac{1}{mn} \text{Tr} \left[ \left( \sum_{vm}^{*} \otimes I_n \right) S_{nm} S_{nm0}^{-1} (\sum_{vm0}^{*} \otimes I_n) S_{nm0}^{-1} S_{nm}^{'} \right] \\
\geq \left| \left( \sum_{vm}^{*} \otimes I_n \right) S_{nm} S_{nm0}^{-1} (\sum_{vm0}^{*} \otimes I_n) S_{nm0}^{-1} S_{nm}^{'} \right| \frac{1}{mn} .
\]

The equality holds if and only if \( (\sum_{vm}^{*} \otimes I_n) S_{nm} S_{nm0}^{-1} (\sum_{vm0}^{*} \otimes I_n) S_{nm0}^{-1} S_{nm}^{'} = cI_{nm} \),

where \( c \) is a constant, which, by rearrangement, is equivalent to

\[
(I_{nm} - \Psi_m \otimes W_n)(\sum_{vm0}^{*} \otimes I_n)(I_{nm} - \Psi_m \otimes W_n) = c(I_{nm} - \Psi_m \otimes W_n)(\sum_{vm0}^{*} \otimes I_n)(I_{nm} - \Psi_m \otimes W_n) .
\]

Then, \( (\sum_{vm0}^{*} - c\sum_{vm0}^{*}) \otimes I_n + (c\Psi_m \sum_{vm0}^{*} - \Psi_m \sum_{vm0}^{*} - \sum_{vm0}^{*} \Psi_m \otimes W_n + (c\sum_{vm}^{*} \Psi_m - \sum_{vm0}^{*} \Psi_m \otimes W_n + (c\sum_{vm0}^{*} \Psi_m - \sum_{vm0}^{*} \Psi_m \otimes W_n) + (\Psi_m \sum_{vm0}^{*} \Psi_m - c\Psi_m \sum_{vm0}^{*} \Psi_m \otimes W_n = 0 .
\]

123
Now we prove (A). As any linear combination of \( W_n, W'_n \) and \( W'_n W_n \) cannot be proportional to \( I_n \), we must have \( \Sigma^{*}_{vm} = c \Sigma^{*}_{vm0} \) for the above equality. In addition, since \( \text{Tr}(\Sigma^{*}_{vm}) = \text{Tr}(\Sigma^{*}_{vm0}) = 1 \), this condition means \( \Sigma^{*}_{vm0} = \Sigma^{*}_{vm} \). When the condition in (B) is satisfied, the equality holds only if \( \Sigma^{*}_{vm} = c \Sigma^{*}_{vm0} \) and \( \Psi = \Psi_{m0} \), which is equivalent to \( \Sigma^{*}_{vm0} = \Sigma^{*}_{vm} \) and \( \Psi = \Psi_{m0} \) since the trace of a normalized covariance matrix is 1.

\[ \Box \]

**Proof of Proposition 1.2:**

To simplify analysis of the above likelihood function, as \( \Sigma_{vm} \) can be reparameterized by letting \( \sigma^2_m = \text{Tr}(\Sigma_{vm}) \) and \( \Sigma^{*}_{vm} = \Sigma_{vm}/\sigma^2_m \), \( \sigma^2_m \) can be easily concentrated out and the concentrated likelihood function is:

\[
\frac{1}{mn} \ln L_{nm} (\Psi_m, \Pi_m, \Sigma^{*}_{vm}) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2m} \ln |\Sigma^{*}_{vm}| + \frac{1}{mn} \ln |S_{nm}| - \frac{1}{2} \ln \hat{\sigma}^2_m (\Psi_m, \Pi_m, \Sigma^{*}_{vm}),
\]  

(C.1)

where

\[
\hat{\sigma}^2_m (\Psi_m, \Pi_m, \Sigma^{*}_{vm}) = \frac{1}{mn} \left[ S_{nm} \text{vec}(Y_{nm}) - (I_m \otimes X_n) \text{vec}(\Pi_m) \right]' \times \left( \Sigma^{*}_{vm}^{-1} \otimes I_n \right) \left[ S_{nm} \text{vec}(Y_{nm}) - (I_m \otimes X_n) \text{vec}(\Pi_m) \right].
\]

The concentrated expected likelihood function with \( \sigma^2_m \) out follows:

\[
\frac{1}{mn} \mathbb{E} \ln L_{nm} (\Psi_m, \Pi_m, \Sigma^{*}_{vm}) =
-\frac{1}{2} \ln(2\pi) - \frac{1}{2m} \ln |\Sigma^{*}_{vm}| + \frac{1}{mn} \ln |S_{nm}| - \frac{1}{2} \ln \hat{\sigma}^2_mE (\Psi_m, \Pi_m, \Sigma^{*}_{vm}),
\]  

(C.2)

where

\[
\hat{\sigma}^2_mE (\Psi_m, \Pi_m, \Sigma^{*}_{vm}) = \frac{1}{mn} \left[ S_{nm} \Sigma^{*}_{nm0}^{-1}(I_m \otimes X_n) \text{vec}(\Pi_{m0}) - (I_m \otimes X_n) \text{vec}(\Pi_m) \right]' \times \left( \Sigma^{*}_{vm}^{-1} \otimes I_n \right) \left[ S_{nm} \Sigma^{*}_{nm0}^{-1}(I_m \otimes X_n) \text{vec}(\Pi_{m0}) - (I_m \otimes X_n) \text{vec}(\Pi_m) \right] + \frac{1}{mn} \text{Tr} \left[ \left( \Sigma^{*}_{vm}^{-1} \otimes I_n \right) S_{nm} \Sigma^{*}_{nm0}^{-1} \left( \Sigma^{*}_{vm0} \otimes I_n \right) S^{*'}_{nm0} S_{nm} \right].
\]
We can compare the difference between the expected log likelihood function
evaluated at the true parameter vector and that with any feasible parameter vec-
tor:

\[
\frac{1}{mn} \mathbb{E} \ln L_{nm}(\Psi_m, \Pi_m, \Sigma_{v_m}^*) - \frac{1}{mn} \mathbb{E} \ln L_{nm}(\Psi_{m0}, \Pi_{m0}, \Sigma_{v_{m0}}^*)
\]
\[
= - \frac{1}{2m} (\ln |\Sigma_{v_m}^*| - \ln |\Sigma_{v_{m0}}^*|) + \frac{1}{mn} (\ln |S_{nm}| - \ln |S_{n,m0}|)
\]
\[
- \frac{1}{2} (\ln \sigma_{mE}^2(\Psi_m, \Pi_m, \Sigma_{v_m}^*) - \ln \sigma_{m0}^2)
\]
\[
= \frac{1}{2} \left[ \ln \left( (\Sigma_{v_m}^{-1} \otimes I_n) S_{nm} S_{n,m0}^{-1} (\Sigma_{v_{m0}}^* \otimes I_n) S_{n,m0}^{-1} S_{n,m0}^* \right) \right] \left[ \frac{1}{mn} \right]
\]
\[
- \frac{1}{2} \ln \left\{ \frac{1}{mn \sigma_{m}^2} \left[ S_{nm} S_{n,m0}^{-1} (I_m \otimes X_n) \text{vec}(\Pi_{m}) - (I_m \otimes X_n) \text{vec}(\Pi_{m0}) \right] \right\}^t
\]
\[
\times (\Sigma_{v_m}^{-1} \otimes I_n) \left[ S_{nm} S_{n,m0}^{-1} (I_m \otimes X_n) \text{vec}(\Pi_{m}) - (I_m \otimes X_n) \text{vec}(\Pi_{m0}) \right]
\]
\[
+ \frac{1}{mn} \text{Tr} \left( (\Sigma_{v_m}^{-1} \otimes I_n) S_{nm} S_{n,m0}^{-1} (\Sigma_{v_{m0}}^* \otimes I_n) S_{n,m0}^{-1} S_{n,m0}^* \right) \right\}.
\]

We investigate different parts in the above equation as \( n \to \infty \). We firstly look at
the quadratic part of the above equation.

\[
\lim_{n \to \infty} \left[ \frac{1}{mn} \left[ S_{nm} S_{n,m0}^{-1} (I_m \otimes X_n) \text{vec}(\Pi_{m}) - (I_m \otimes X_n) \text{vec}(\Pi_{m0}) \right] \right] \]
\[
\times (\Sigma_{v_m}^{-1} \otimes I_n) \left[ S_{nm} S_{n,m0}^{-1} (I_m \otimes X_n) \text{vec}(\Pi_{m}) - (I_m \otimes X_n) \text{vec}(\Pi_{m0}) \right],
\]

with

\[
S_{nm} S_{n,m0}^{-1} (I_m \otimes X_n) \text{vec}(\Pi_{m}) - (I_m \otimes X_n) \text{vec}(\Pi_{m0})
\]
\[
= [(\Psi_{m0}' - \Psi_m') \otimes W_n] S_{n,m0}^{-1} (I_m \otimes X_n) \text{vec}(\Pi_{m0}) + (I_m \otimes X_n) [\text{vec}(\Pi_{m0}) - \text{vec}(\Pi_{m})]
\]
\[
= (I_m \otimes X_n) [\text{vec}(\Pi_{m0}) - \text{vec}(\Pi_{m})]
\]
\[
+ \left\{ I_m \otimes [(J_1 \otimes I_n) S_{n,m0}^{-1} (I_m \otimes (W_n X_n)) \text{vec}(\Pi_{m0})]ight\} [\text{vec}(\Psi_{m0}) - \text{vec}(\Psi_{m})].
\]
By letting $\beta_m = \text{vec}([\Pi'_m, \Psi'_m])$, 
\[
S_{nm}S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m) - (I_m \otimes X_n)\text{vec}(\Pi_{m0})
= \left\{ I_m \otimes [X_n, (J_1 \otimes I_n)S_{nm0}^{-1}(I_m \otimes (W_nX_n))\text{vec}(\Pi_{m0})],
\ldots, (J_m \otimes I_n)S_{nm0}^{-1}(I_m \otimes (W_nX_n))\text{vec}(\Pi_{m0}) \right\} \times (\beta_{m0} - \beta_m)
= [I_m \otimes A_{n \times (k_x+m)}](\beta_{m0} - \beta_m).
\]
where $A_{n \times (k_x+m)}$ is accordingly defined as in Assumption 1.5. It follows that 
\[
\lim_{n \to \infty} \left[ \frac{1}{mn} \left[ S_{nm}S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m) - (I_m \otimes X_n)\text{vec}(\Pi_{m0}) \right]' \cdot (\Sigma_{vm}^{-1} \otimes I_n) \left[ S_{nm}S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m) - (I_m \otimes X_n)\text{vec}(\Pi_{m0}) \right] \right]
= \lim_{n \to \infty} (\beta_{m0} - \beta_m)' \left[ \Sigma_{vm}^{-1} \otimes \left( \frac{1}{mn} A_{n \times (k_x+m)}'A_{n \times (k_x+m)} \right) \right] (\beta_{m0} - \beta_m) > 0.
\]
Therefore, when Assumption 1.5 holds, the above equation is equal to zero only when all parameters are equal to their true values.

By the arithmetic and geometric means inequality of eigenvalues, 
\[
\frac{1}{mn} \text{Tr} \left[ (\Sigma_{vm}^{-1} \otimes I_n) S_{nm}S_{nm0}^{-1} (\Sigma_{vm0}^{-1} \otimes I_n) S_{nm0}^{-1} S_{nm}' \right] 
\geq \left| \left( \Sigma_{vm}^{-1} \otimes I_n \right) S_{nm}S_{nm0}^{-1} (\Sigma_{vm0}^{-1} \otimes I_n) S_{nm0}^{-1} S_{nm}' \right| \frac{1}{mn}.
\]
As $n \to \infty$, according to Assumption 1.6, the equality holds when $\Sigma_{vm} = \Sigma_{vm0}$ and $\Psi_m = \Psi_{m0}$. With the two parts together, in the limit as $n \to \infty$, Assumption 1.5 guarantees that $\Psi_{m0}$ and $\Pi_{m0}$ are identifiable, while Assumption 1.6 guarantees the identification of $\Sigma_{vm0}$. As those parameters can be identified, we substitute them into $\sigma^2(\Psi_{m0}, \Pi_{m0}, \Sigma_{vm0})$, and $\sigma^2_{vm0}$ will be identified. Q.E.D.

Proof of Corollary 1.1:

Consider the identification in a finite sample. On (I), when $A_{n \times (k_x+m)}$ has full column rank, $\Psi_{m0}$ and $\Pi_{m0}$ can be identified by following the above proof strategies. Proposition 1.1(A) gives the identification of the covariance matrix. On (II),
Proposition 1.1(B) guarantees identification of $\Sigma_{vm0}$ and $\Psi_{m0}$. As $X_n$ has full rank $k_x$, $\Pi_m$ can be identified by substitution via:

$$
\text{vec}(\Pi_{m0}) = [(I_m \otimes X'_n)(\Sigma_{vm0}^{-1} \otimes I_n)(I_m \otimes X_n)]^{-1}(I_m \otimes X'_n)(\Sigma_{vm0}^{-1} \otimes I_n)
\times (I_{nm} - \Psi_{m0}' \otimes W_n)\text{vec}(Y_{nm}).
$$

This completes the proof of (II). \qed

**Proof of Corollary 1.2:**

In the proof of Proposition 1.2, that $\Psi_{m0}$ and $\Pi_{m0}$ can be identified means

$$
\frac{1}{n}(\beta_{m0} - \beta_m)'[\Sigma_{vm0}^{-1} \otimes (A'_{n \times (k_x+m)} A_{n \times (k_x+m)})](\beta_{m0} - \beta_m) = 0
$$

is possible only if $\beta_m = \beta_{m0}$. This requires that $\lim_{n \to \infty} \frac{1}{n} [\Sigma_{vm0}^{-1} \otimes (A'_{n \times (k_x+m)} A_{n \times (k_x+m)})]$ exists and is nonsingular.

By substituting the constraint $\beta_m = R_m \beta_l$ into the above condition, $\beta_l = \beta_{l0}$ if $\lim_{n \to \infty} \frac{1}{n} R_m' [\Sigma_{vm0}^{-1} \otimes (A'_{n \times (k_x+m)} A_{n \times (k_x+m)})]R_m$ exists and is nonsingular. \qed

**Proof of Lemma 1.2:**

$A_m$ is an admissible transformation matrix when

$$
\Phi_{\alpha m} A_1 = 0, \quad \Phi_{\alpha_{1,m}} = \left[ A_{m}' \Gamma_{m}' - A_{m}' A_{m}' - A_{m}' C_{m}' \right]'
$$

satisfies all a priori restrictions on $\alpha_m$. The definition follows Schmidt(1976).

(Sufficiency) Suppose $\text{rank}(\Phi_{\alpha m}) = m-1$. Since $\Phi_{\alpha m} = \Phi[\alpha_1, \alpha_{(-1)}] A_1 = [0, \Phi_{\alpha_{(-1)}}] A_1 = 0$, where $\alpha_{(-1)}$ is a $(2m+k_x) \times (m-1)$ submatrix of $\alpha_m$ excluding the first column.

Therefore, if $[0, \Phi_{\alpha_{(-1)}}] A_1 = 0$, $A_1$ should be a vector whose first element is an arbitrary scalar and whose remaining elements are equal to zero. This means the first equation is identified.
(Necessity) Now assume the first equation is identified. Since the solution \((A_1)\) to the equation \([0, \Phi_{\alpha_{(-1)}}]A_1 = 0\) is a vector whose first element is an arbitrary scalar and whose remaining elements are equal to zero. The column vectors in \(\Phi_{\alpha_{(-1)}}\) should be linearly independent. Thus, \(\text{rank}(\Phi_{\alpha_m}) = m - 1\). □

Proof of Lemma 1.3:

Since \(\bar{\Phi}_{\alpha_{1,m}} = c\) and \(\left( \begin{array}{c} \Psi_m \\ \Pi_m \\ \bar{\Phi} \end{array} \right) \alpha_{1,m} = 0\), \(\left( \begin{array}{c} \Psi_m \\ \Pi_m \\ \bar{\Phi} \end{array} \right) \alpha_{1,m} = \left( \begin{array}{c} 0 \\ c \end{array} \right)\). The true parameter \(\alpha_{1,m0}\) is the solution to above system. It is the only solution if and only if \(\text{rank} \left( \begin{array}{c} \Psi_m \\ \Pi_m \\ \bar{\Phi} \end{array} \right) = 2m + k_x\). Since \(\bar{\Phi} = c\), the columns of the two matrices, \(\alpha_m\) and \(\left( \begin{array}{c} \Psi_m' \\ \Pi_m' \\ \bar{\Phi} \end{array} \right) \alpha_m = \left( \begin{array}{c} \Psi_m' \\ \Pi_m' \\ \bar{\Phi} \end{array} \right) C_1 + \alpha_mC_2\), for some conformably matrices \(C_1\) and \(C_2\), \(\text{rank} \left( \begin{array}{c} \Psi_m \\ \Pi_m \\ \bar{\Phi} \end{array} \right) = \text{rank} \left( \begin{array}{c} \Psi_m' \\ \Pi_m' \\ \bar{\Phi} \end{array} \right) C_1 + \alpha_mC_2\).

Because \(\text{rank}(\alpha_mC_2) = \text{rank} \left( \begin{array}{c} \Psi_m' \\ \Pi_m' \\ \bar{\Phi} \end{array} \right) \begin{array}{c} \alpha_m \\ 0 \end{array} \right) = \text{rank}(C_2)\), It follows that rank of \(C_2\) needs to be \(m\). In addition, \(\bar{\Phi}_{\alpha_m} = C_1' \left( \begin{array}{c} \Psi_m \\ \Pi_m \end{array} \right) \alpha_m + C_2' \alpha_m\). Because \(\alpha_m\) has full rank due to normalization, it follows that \(\text{rank}(\bar{\Phi}_{\alpha_m}) = \text{rank}(C_2) = m\). □

Proof of Lemma 1.4:

Consider any transformation matrix \(A_m\). \(A_m\) being admissible means that \(\bar{\Psi}_{\text{vec}}(\alpha_mA_m) = c\) since \(\bar{\Psi}_{\text{vec}}(\alpha_m) = c\), where \(c\) is a \(R \times 1\) constant vector containing a \(m \times 1\) subvector whose all elements are ones due to normalizations. Therefore, \(\bar{\Psi}(I_m \otimes \alpha_m)\text{vec}(A_m) = c\). As \(\bar{\Psi}(I_m \otimes \alpha_m) = c\), so \(\bar{\Psi}(I_m \otimes \alpha_m)\text{vec}(A_m - I_m) = 0\).
(Sufficiency) Suppose \( \text{rank}\{\bar{\Psi}(I_m \otimes \alpha_m)\} = m^2 \). That means all columns of the matrix are linearly independent. Therefore, \( A_m = I_m \), i.e., the system is identified.

(Necessity) Suppose \( A_m = I_m \) is the unique solution to above equation, then \( \bar{\Psi}(I_m \otimes \alpha_m) \) has full column rank \( m^2 \).

**Proof of Lemma 1.5:**

In the proof we follow the notation of MSAR model for simplicity but the same proof strategies apply to the SESAR model in terms of MSAR with restrictions on parameters. Note that

\[
\frac{1}{mn} \ln L_{mn}(\Psi_m, \Pi_m, \Sigma_{vm}) - \frac{1}{mn} \mathbb{E} \ln L_{mn}(\Psi_m, \Pi_m, \Sigma_{vm}) \\
= -\frac{1}{2} \left[ \ln \hat{\sigma}_m^2(\Psi_m, \Pi_m, \Sigma_{vm}) - \ln \hat{\sigma}_{mE}^2(\Psi_m, \Pi_m, \Sigma_{vm}) \right] \\
= -\frac{1}{2\hat{\sigma}_m^2(\Psi_m, \Pi_m, \Sigma_{vm})} \left[ \hat{\sigma}_m^2(\Psi_m, \Pi_m, \Sigma_{vm}) - \hat{\sigma}_{mE}^2(\Psi_m, \Pi_m, \Sigma_{vm}) \right] .
\]

As

\[
\hat{\sigma}_m^2(\Psi_m, \Pi_m, \Sigma_{vm}) - \hat{\sigma}_{mE}^2(\Psi_m, \Pi_m, \Sigma_{vm}) \\
= \frac{1}{mn} \left[ S_{nm}S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m) + S_{nm}S_{nm0}^{-1}\text{vec}(V_{nm}) - (I_m \otimes X_n)\text{vec}(\Pi_m) \right]' \\
\cdot (\Sigma_{vm}^{-1} \otimes I_n) \left[ S_{nm}S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m) + S_{nm}S_{nm0}^{-1}\text{vec}(V_{nm}) - (I_m \otimes X_n)\text{vec}(\Pi_m) \right] \\
- \frac{1}{mn} \left[ S_{nm}S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m) - (I_m \otimes X_n)\text{vec}(\Pi_m) \right]' (\Sigma_{vm}^{-1} \otimes I_n) \\
\cdot [S_{nm}S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m) - (I_m \otimes X_n)\text{vec}(\Pi_m)] \\
- \frac{1}{mn} \text{Tr} \left[ (\Sigma_{vm}^{-1} \otimes I_n) S_{nm}S_{nm0}^{-1}(\Sigma_{vm0} \otimes I_n) S_{nm0}'S_{nm}' \right] \\
= \frac{1}{mn} \left\{ 2 \left[ S_{nm}S_{nm0}^{-1}(I_m \otimes X_n)\text{vec}(\Pi_m) - (I_m \otimes X_n)\text{vec}(\Pi_m) \right]' (\Sigma_{vm}^{-1} \otimes I_n) \\
\times S_{nm}S_{nm0}^{-1}\text{vec}(V_{nm}) \right\} + \frac{1}{mn} \left\{ \text{vec}(V_{nm})'S_{nm0}'S_{nm}'(\Sigma_{vm}^{-1} \otimes I_n) S_{nm}S_{nm0}^{-1}\text{vec}(V_{nm}) \\
- \text{Tr} \left[ (\Sigma_{vm}^{-1} \otimes I_n) S_{nm}S_{nm0}^{-1}(\Sigma_{vm0} \otimes I_n) S_{nm0}'S_{nm}' \right] \right\} ,
\]

we discuss the two parts separately. By the assumptions that \( S_{nm0}' \) and \( S_{nm}' \) are UB, \( (\Sigma_{vm}^{-1} \otimes I_n) S_{nm}S_{nm0}^{-1} \) and \( S_{nm0}'S_{nm}'(\Sigma_{vm}^{-1} \otimes I_n) S_{nm}S_{nm0}^{-1} \) are UB. So entries in \( \text{vec}(\Pi_m)'(I_m \otimes X_n)'S_{nm0}'S_{nm}(\Sigma_{vm}^{-1} \otimes I_n) S_{nm}S_{nm0}^{-1} \) are uniformly bounded. As the
vectors of disturbances are independent under Assumption 1, for any vector \( a_{nm} \) with uniformly bounded elements, \( \frac{1}{mn} a_{nm}' \text{vec}(V_{nm}) \overset{p}{\rightarrow} 0 \). By Chebyshev’s law of large numbers. Hence,

\[
\frac{1}{mn} \left[ S_{nm}^{-1} (I_m \otimes X_n) \text{vec}(\Pi_{m0}) - (I_m \otimes X_n) \text{vec}(\Pi_{m}) \right]' (\Sigma_{vm}^{-1} \otimes I_n) S_{nm}^{-1} \text{vec}(V_{nm}) \overset{p}{\rightarrow} 0,
\]

uniformly on its parameter space. The uniform convergence follows because the parameters of \( \Psi_m, \Pi_m \) and \( \Sigma_{vm}^{-1} \) appear linearly in the above statistic.

The second part is a quadratic form minus its expectation, where

\[
A_{nm} = S_{nm0}^{-1} \text{vec}(V_{nm})' S_{nm0}^{-1}
\]
as the quadratic matrix. By Lemma 1.7,

\[
\text{Var} \left( \frac{1}{mn} \text{vec}(V_{nm})' A_{nm} \text{vec}(V_{nm}) \right) = \frac{1}{m^2 n^2} \{ \text{Tr} [A_{nm}(\Sigma_{vm0} \otimes I_n)(A_{nm} + A_{nm}'(\Sigma_{vm0} \otimes I_n))]
\]

\[
+ \sum_{k=1}^m \sum_{l=1}^m \sum_{p=1}^m \sum_{q=1}^m (\mu_{klpq} - \sigma_{kp}\sigma_{lq} - \sigma_{kq}\sigma_{lp} - \sigma_{kl}\sigma_{pq}) \left( \sum_{i=1}^n A_{nil,ii} A_{npq,ii} \right) \},
\]

where \( \sum_{i=1}^n (A_{nil,ii} A_{npq,ii}) \leq \sum_{i=1}^n |A_{nil,ii}| |A_{npq,ii}| \leq \text{Tr} [||A_{kl}|| A_{pq}] \), with \( |A_{kl}| \) being a \( n \times n \) submatrix with absolute values of \( A_{nil,ij} \) as its entries. By Lemma A.7 in Lee (2004),

\[
\text{Tr} [A_{nm}(\Sigma_{vm0} \otimes I_n) A_{nm}(\Sigma_{vm0} \otimes I_n)] = O(n). \]

It follows that \( \text{Var} \left( \frac{1}{mn} \text{vec}(V_{nm})' A_{nm} \text{vec}(V_{nm}) \right) = O \left( \frac{1}{n} \right) \). By Chebychev’s inequality, \( \frac{1}{mn} \text{vec}(V_{nm})' A_{nm} \text{vec}(V_{nm}) - \mathbb{E} \left[ \frac{1}{mn} \text{vec}(V_{nm})' A_{nm} \text{vec}(V_{nm}) \right] \)

\( \overset{p}{\rightarrow} 0 \) uniformly on the parameter space. Thus, \( \hat{\sigma}_m^2 (\Psi_m, \Pi_m, \Sigma_{vm}) - \bar{\sigma}_m^2 (\Psi_m, \Pi_m, \Sigma_{vm}) \)

is bounded away from zero. The expected log likelihood function of a pure SAR model (i.e. no exogenous variables in data generating process) is,

\[
-\frac{1}{2} \ln (2\pi) + \frac{1}{2m} \ln |\Sigma_{vm}| + \frac{1}{mn} \ln |S_{nm}| - \frac{1}{2} \ln \hat{\sigma}_{mE,p}^2 (\Psi_m, \Sigma_{vm}), \text{ where } \hat{\sigma}_{mE,p}^2 (\Psi_m, \Pi_m, \Sigma_{vm}) = \frac{1}{mn} \text{Tr} \left( (\Sigma_{vm}^{-1} \otimes I_n) S_{nm} S_{mm0}^{-1} (\Sigma_{vm0} \otimes I_n) S_{nm0}^{-1} \Sigma_{vm0} S_{mm0}^{-1} \right) .
\]

Since

\[
\mathbb{E} \ln L_{nm}^p (\Psi_m, \Sigma_{vm}) \leq \mathbb{E} \ln L_{nm}^p (\Psi_{m0}, \Sigma_{vm0}) ,
\]

we have

\[
-\frac{1}{2} \ln \hat{\sigma}_{mE,p}^2 (\Psi_m, \Sigma_{vm}) \leq -\frac{1}{2} \ln \sigma_{m0}^2 + \frac{1}{2m} \left[ \ln |\Sigma_{vm}| - \ln |\Sigma_{vm0}| \right] + \frac{1}{mn} \left[ \ln |S_{nm0}| - \ln |S_{nm}| \right] .
\]

130
By mean value theorem,

\[
\frac{1}{mn} \left[ \ln |S_{nm0}| - \ln |S_{nm}| \right] = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{m} \left\{ \frac{\partial \ln |\tilde{S}_{nm}|}{\partial \Psi_{ij,m}} \left( \Psi_{ij,m} - \Psi_{ij,m0} \right) \right\}
\]

\[= \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{m} \left\{ - \text{Tr}\left[ (\tilde{S}_{nm}^{-1})'(E'_{m,ij} \otimes W_n) \right] \left( \Psi_{ij,m} - \Psi_{ij,m0} \right) \right\}, \]

where \( \tilde{S}_{nm} = I_{nm} - (c \Psi_{m} + (1-c) \Psi_{m0}) \otimes W_n, c \in (0, 1) \) and \( E_{m,ij} \) is a \( m \times m \) matrix with all zero element except the \( ij \)th element, which is one. Therefore, \( (\tilde{S}_{nm}^{-1})'(E'_{m,ij} \otimes W_n) \) is UB. Since \( \text{Tr}\left[ (\tilde{S}_{nm}^{-1})'(E'_{m,ij} \otimes W_n) \right] = O(n) \) (Lemma A.7, Lee, 2004), and the parameter space is compact, \( \frac{1}{mn} \left[ \ln |S_{nm0}| - \ln |S_{nm}| \right] = O(1) \), uniformly in \( \Psi_m \).

The same method can be applied to show \( \frac{1}{2m} \left[ \ln |\Sigma_{vm}| - \ln |\Sigma_{vm0}| \right] = O(1) \), uniformly in \( \Sigma_{vm} \).

So, \( \hat{\sigma}_{mE,p}^2 (\Psi_m, \Sigma_{vm}) \) is bounded away from zero uniformly on the parameter space. Since \( \sigma_{mE}^2 (\Psi_m, \Pi_m, \Sigma_{vm}) \geq \hat{\sigma}_{mE,p}^2 (\Psi_m, \Sigma_{vm}), \sigma_{mE}^2 (\Psi_m, \Pi_m, \Sigma_{vm}) \) must be bounded away from zero. In turn, these imply that, \( \sigma_{m}^2 (\Psi_m, \Pi_m, \Sigma_{vm}) \) and \( \tilde{\sigma}_{m}^2 (\Psi_m, \Pi_m, \Sigma_{vm}) \) are bounded away from zero in probability, uniformly in its parameter.

In conclusion, \( \frac{1}{mn} \left[ \ln L_{nm} (\Psi_m, \Pi_m, \Sigma_{vm}) - \mathbb{E} \ln L_{nm} (\Psi_m, \Pi_m, \Sigma_{vm}) \right] \xrightarrow{P} 0, \) uniformly in its parameter space. \( \square \)

**Proof of Theorem 1.1:**

The consistency of QMLE follows Lemma 1.5, Proposition 1.2 and the identification uniqueness condition (White (1994)). So, the only remaining task is to establish uniformly equicontinuity of the expected log likelihood function so that the identification uniqueness condition will hold. For generality, we consider the MSAR likelihood (with or without constraints).

First, \( \frac{1}{mn} \frac{\partial \ln |S_{nm}|}{\partial \Psi_{m,ij}} = -\frac{1}{mn} \text{Tr}\left[ S_{nm}^{-1}(E'_{m,ij} \otimes W_n) \right] = O(1), \) uniformly in \( \Psi_m \), since \( S_{nm}^{-1} \) is UB, uniformly in \( \Psi_m \). Suppose that \( S_{nm}^{-1}(E'_{m,ij} \otimes W_n) \)'s row and column sum norms are bounded by the constant \( c \) uniformly in the parameter space of \( \Psi_m \),

\( \frac{1}{mn} \ln |I_{mn} - \Psi_m' \otimes W_n| \) is a Lipschitz function in \( \Psi_m \) with Lipschitz bound \( c/m \) and,
hence, uniformly equicontinuous. Second, so is $\frac{1}{m} \ln |\Sigma_{vm}|$ because $\frac{1}{m} \partial \ln |\Sigma_{vm}| = -\frac{1}{m} \Sigma_{vm}^{-1} F_{m,ij} = O(1)$, where $F_{m,ij}$ is a $m \times m$ matrix with all zero element except the $ij$th and $ji$th elements, which are ones. Since

$$
\ln \hat{\sigma}_{mE}^2 (\Psi_{m1}, \Pi_{km1}, \Sigma_{vm1}) - \ln \hat{\sigma}_{mE}^2 (\Psi_{m2}, \Pi_{km2}, \Sigma_{vm2}) = \frac{1}{\hat{\sigma}_{mE}^2 (\Psi_{m3}, \Pi_{km3}, \Sigma_{vm3})} [\hat{\sigma}_{mE}^2 (\Psi_{m1}, \Pi_{km1}, \Sigma_{vm1}) - \hat{\sigma}_{mE}^2 (\Psi_{m2}, \Pi_{km2}, \Sigma_{vm2})],
$$

and $\hat{\sigma}_{mE}^2 (\Psi_{m3}, \Pi_{km3}, \Sigma_{vm3})$ is bounded away from zero uniformly in its parameter space (proved in Lemma 1.5), its equicontinuity will follow if $\hat{\sigma}_{mE}$ is a Lipschitz function.

Let $V_{nm}(\Psi_{m}, \Pi_{m}) = S_{nm} S_{n m0}^{-1} (I_{m} \otimes X_{n}) \text{vec}(\Pi_{m0}) - (I_{m} \otimes X_{n}) \text{vec}(\Pi_{m})$. Since $V_{nm}(\Psi_{m}, \Pi_{m})$ and $\Sigma_{vm}^{-1} \otimes X_{n}'$ are bounded uniformly in elements, $\frac{\partial \hat{\sigma}_{mE}^2 (\Psi_{m}, \Pi_{m}, \Sigma_{vm})}{\partial \text{vec}(\Pi_{m})} = -\frac{1}{mn} (\Sigma_{vm}^{-1} \otimes X_{n}') V_{nm}(\Psi_{m}, \Pi_{m}) = O(1)$ uniformly in its parameters. Furthermore,

$$
\frac{\partial \hat{\sigma}_{mE}^2 (\Psi_{m}, \Pi_{m}, \Sigma_{vm})}{\partial \sigma_{m,ij}^*} = - \frac{1}{mn} V_{nm}'(\Psi_{m}, \Pi_{m}) [(\Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1}) \otimes I_{n}] V_{nm}(\Psi_{m}, \Pi_{m})
$$

$$
- \frac{1}{mn} \text{Tr} \left[ (\Sigma_{vm0} \otimes I_{n}) S_{nm0}^{-1} S_{nm}' \left( (\Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1}) \otimes I_{n} \right) S_{nm} S_{nm0}^{-1} \right],
$$

and

$$
\frac{\partial \hat{\sigma}_{mE}^2 (\Psi_{m}, \Pi_{m}, \Sigma_{vm})}{\partial \psi_{m,ij}^*} = - \frac{1}{mn} V_{nm}'(\Psi_{m}, \Pi_{m}) [(\Sigma_{vm}^{-1} E_{m,ij}') \otimes W_{n}] S_{mm0}^{-1} (I_{m} \otimes X_{n}) \text{vec}(\Pi_{m0}) - \frac{2}{mn} \text{Tr} \left[ S_{mm0}^{-1} (\Sigma_{vm0} \otimes I_{n}) S_{mm0}^{-1} S_{mm}' [(\Sigma_{vm}^{-1} E_{m,ij}') \otimes W_{n}]) \right],
$$

are both $O(1)$ uniformly in the parameter space. As each component of the expected log likelihood function is uniformly equicontinuous, the expected log likelihood function is uniformly equicontinuous.

\textbf{Proof of Lemma 1.6:}

Here we extend the arguments in Kelejian and Prucha (2001) for the univariate quadratic form to the multivariate one. The derivation relies on the use of a martingale central limit theorem.
Define a $\sigma$-field $\mathcal{F}_{i,n} = \langle v_{n,11}, \ldots, v_{n,i1}, v_{n,12}, \ldots, v_{n,i2}, \ldots, v_{n,1m}, \ldots, v_{n,im} \rangle$ generated by $v_{n,11}, \ldots, v_{n,i1}, v_{n,12}, \ldots, v_{n,i2}, \ldots, v_{n,1m}, \ldots, v_{n,im}$. Denote

$$Q_{i,n} = \sum_{k=1}^{m} b_{nk,i} v_{n,ik} + \sum_{k=1}^{m} \sum_{l=1}^{m} a_{nk,l,i} (v_{n,ik} v_{n,il} - \sigma_{kl}) + v_{n,ik} \sum_{j=1}^{i-1} a_{nk,l,j} v_{n,jl}$$

$$+ v_{n,il} \sum_{j=1}^{i-1} a_{nk,lji} v_{n,jk}.$$ 

So, $Q_i = \sum_{k=1}^{m} b_{nk} V_{nk} + \sum_{k=1}^{m} \sum_{l=1}^{m} [V_{nk} A_{nkl} V_{nl} - \mathbb{E}(V_{nk} A_{nkl} V_{nl})] = \sum_{i=1}^{n} Q_{i,n}$. Because of independence of the random vector $v_{n,i}$ (for different $i = 1, 2, \ldots, n$), we have $\mathbb{E}[Q_{i,n} \mathcal{F}_{i-1,n}] = 0$. So $\{(Q_{i,n}, \mathcal{F}_{i,n}) | 1 \leq i \leq n\}$ is a martingale difference double array. And it is easy to know $\sigma_{Q_i}^2 = \sum_{i=1}^{n} \mathbb{E}Q_i^2$. With $Q_{i,n} = \frac{Q_{i,n}}{\sigma_{Q_i}}$, sufficient conditions for a martingale central limit theorem to establish our result are

$$\sum_{i=1}^{n} \mathbb{E}|Q_{i,n}|^{2+\delta} \overset{p}{\to} 0,$$  

(C.3)

for some $\delta > 0$, and

$$\sum_{i=1}^{n} \mathbb{E}(Q_{i,n}^2 | \mathcal{F}_{i-1,n}) \overset{p}{\to} 1.$$  

(C.4)

We first prove (C.3). For $p, q \in [0, \infty)$ and $1/p + 1/q = 1$,

$$|Q_{i,n}| \leq \sum_{k=1}^{m} |b_{nk,i}| |v_{n,ik}| + \sum_{k=1}^{m} \sum_{l=1}^{m} \left[ |a_{nk,l,i}| (v_{n,ik} v_{n,il} - \sigma_{kl}) + |v_{n,ik}| \sum_{j=1}^{i-1} |a_{nk,l,j}| |v_{n,jl}| + |v_{n,il}| \sum_{j=1}^{i-1} |a_{nk,lji}| |v_{n,jk}| \right]$$

$$= \sum_{k=1}^{m} |b_{nk,i}|^\frac{1}{p} |b_{nk,i}|^\frac{1}{q} |v_{n,ik}| + \sum_{k=1}^{m} \sum_{l=1}^{m} \left[ |a_{nk,l,i}|^\frac{1}{p} |a_{nk,l,i}|^\frac{1}{q} (v_{n,ik} v_{n,il} - \sigma_{kl}) + |v_{n,ik}| \sum_{j=1}^{i-1} |a_{nk,l,j}|^\frac{1}{p} |a_{nk,l,j}|^\frac{1}{q} |v_{n,jl}| + |v_{n,il}| \sum_{j=1}^{i-1} |a_{nk,lji}|^\frac{1}{p} |a_{nk,lji}|^\frac{1}{q} |v_{n,jk}| \right].$$

By Holder's Inequality,

$$|Q_{i,n}|^q \leq \left\{ \sum_{k=1}^{m} (|b_{nk,i}|^\frac{1}{p})^p + \sum_{k=1}^{m} \sum_{l=1}^{m} (|a_{nk,l,i}|^\frac{1}{p})^p \sum_{j=1}^{i-1} (|a_{nk,l,j}|^\frac{1}{q})^q + \sum_{j=1}^{i-1} (|a_{nk,lji}|^\frac{1}{q})^q \right\}^\frac{2}{p}$$

$$\cdot \left\{ \sum_{k=1}^{m} (|b_{nk,i}|^\frac{1}{q})^q + \sum_{k=1}^{m} \sum_{l=1}^{m} (|a_{nk,l,i}|^\frac{1}{q})^q (v_{n,ik} v_{n,il} - \sigma_{kl}) \right\}^q.$$
Just take the first term in the above equation will be bounded by $C$. Because of row and column sums of $A_{nkl}$ are uniformly bounded and $b_{nl}$ is uniformly bounded in elements, we can pick a finite constant $C$ such that, for any $k, l = 1, ... m$, $|b_{nk,i}| \leq \frac{C}{2m}$, $\sum_{j=1}^{i} |a_{nkli,j}| \leq \frac{C}{2m^2}$, and $\sum_{j=1}^{i} |a_{nkli,j}| \leq \frac{C}{2m^2}$. In this case the first term in the above equation will be bounded by $C$. So, 

$$|Q_{i,n}|^q \leq C^q \left\{ \sum_{k=1}^{m} |b_{nk,i}| |v_{n,ik}|^q + \sum_{k=1}^{m} \sum_{l=1}^{m} |a_{nkli,ii}| (v_{n,ik} v_{n,il} - \sigma_{kl})|^q 
+ \sum_{j=1}^{i-1} |a_{nkli,ij}| |v_{n,il}|^q |v_{n,ik}|^q + \sum_{j=1}^{i-1} |a_{nkli,ji}| |v_{n,il}|^q |v_{n,ik}|^q \right\}.$$ 

We have assumed that $\mathbb{E}|v_{n,ik}^2 v_{n,il}^2|^{1+\frac{q}{2}} < D_4$ for any $i = 1, 2, ..., n$ and $k, l = 1, 2, ..., m$. Just take $q = 2 + \delta$, then 

$$\mathbb{E}|Q_{i,n}|^q \leq C^q D_4 \left\{ \sum_{k=1}^{m} |b_{nk,i}| + \sum_{k=1}^{m} \sum_{l=1}^{m} |a_{nkli,ii}| + \sum_{j=1}^{i-1} |a_{nkli,ij}| + \sum_{j=1}^{i-1} |a_{nkli,ji}| \right\}.$$ 

So, by adding them up, 

$$\sum_{i=1}^{n} \mathbb{E}|Q_{i,n}|^q \leq C^q D_4 \left\{ \sum_{k=1}^{m} \sum_{i=1}^{n} |b_{nk,i}| + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{i=1}^{n} |a_{nkli,ii}| 
+ \sum_{j=1}^{i-1} |a_{nkli,ij}| + \sum_{j=1}^{i-1} |a_{nkli,ji}| \right\}.$$ 

Because $\sum_{i=1}^{n} |b_{nk,i}| = O(n)$, $\sum_{i=1}^{n} |a_{nkli,ii}| = O(n)$ (Lemma A.7 Lee, 2004), and $\sum_{j=1}^{i-1} |a_{nkli,ji}| \leq C$, so it follows that $\sum_{i=1}^{n} \mathbb{E}|Q_{i,n}|^{2+\delta} = O(n)$. 

134
Since $\sigma^2_{Q_n}$ is bounded away from zero with $n$-rate, there exists a constant $C_2$ such that, $\frac{\sigma^2_{Q_n} + \delta}{\sigma_{Q_n}} = \left(\frac{1}{n} \sigma^2_{Q_n}\right)^{1+\frac{\delta}{2}} n^{1+\frac{\delta}{2}} \geq C_2 n^{1+\frac{\delta}{2}}$. Therefore,

$$\sum_{i=1}^{n} \mathbb{E}[Q_{i,n}^*]^{2+\delta} = \frac{1}{\sigma_{Q_n}^2} \sum_{i=1}^{n} \mathbb{E}[Q_{i,n}]^{2+\delta} = O(n^{-\delta}) \to 0.$$  

This proves condition (C.3).

Now we prove condition (C.4).

$$Q_{i,n}^2 = \sum_{k=1}^{m} b_{nk,i}^2 v_{n,ik}^2 + 2 \sum_{k=1}^{m} \sum_{l=1}^{k-1} b_{nk,i} b_{nl,i} v_{n,ik} v_{n,il} + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ a_{nk,i} a_{npq,i} \right. \times \left. \left( v_{n,ik} v_{n,il} - \sigma_{kl} \right) \left( v_{n,ip} v_{n,iq} - \sigma_{pq}^2 \right) + v_{n,ik} v_{n,ip} \left( \sum_{j=1}^{i-1} a_{nk,i} v_{n,jl} \right) \left( \sum_{j=1}^{i-1} a_{npq,i} v_{n,jq} \right) \right. \right. \right.$$

$$\left. + v_{n,ik} v_{n,il} \left( \sum_{j=1}^{i-1} a_{nk,i} v_{n,jk} \right) \left( \sum_{j=1}^{i-1} a_{npq,j} v_{n,jk} \right) \right. \right. \right.$$

$$\left. + v_{n,ip} a_{nk,i} \left( v_{n,ik} v_{n,il} - \sigma_{kl} \right) \left( \sum_{j=1}^{i-1} a_{npq,j} v_{n,jq} \right) \right. \right. \right.$$

$$\left. + v_{n,ip} a_{nk,i} \left( v_{n,ik} v_{n,il} - \sigma_{kl} \right) \left( \sum_{j=1}^{i-1} a_{npq,j} v_{n,jp} \right) \right. \right. \right.$$

$$\left. + v_{n,ip} a_{nk,i} \left( v_{n,ik} v_{n,il} - \sigma_{kl} \right) \left( \sum_{j=1}^{i-1} a_{npq,j} v_{n,jk} \right) \right. \right. \right.$$

$$\left. + v_{n,il} a_{npq,i} \left( v_{n,ip} v_{n,iq} - \sigma_{pq}^2 \right) \left( \sum_{j=1}^{i-1} a_{nk,i} v_{n,jl} \right) \right. \right. \right.$$

$$\left. + v_{n,il} a_{npq,i} \left( v_{n,ip} v_{n,iq} - \sigma_{pq}^2 \right) \left( \sum_{j=1}^{i-1} a_{nk,i} v_{n,jk} \right) \right. \right. \right.$$

$$\left. + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \left[ b_{np,i} a_{nk,i} \left( v_{n,ik} v_{n,il} - \sigma_{kl} \right) v_{n,ip} + v_{n,ip} v_{n,ik} b_{np,i} \left( \sum_{j=1}^{i-1} a_{nk,i} v_{n,jl} \right) \right. \right. \right.$$

$$\left. + v_{n,ip} v_{n,il} b_{np,i} \left( \sum_{j=1}^{i-1} a_{nk,i} v_{n,jk} \right) \right].$$

135
Therefore,

\[
\mathbb{E}[Q_{i,n}^2 | \mathcal{F}_{i-1,n}] = \sum_{k=1}^{m} b_{nk,i}^2 \sigma_{kk}^2 + b_{nk,i} b_{nl,i} \sigma_{kl} + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ a_{nkl,ii} a_{npq,ii} \left( \mathbb{E}(v_{n,ik} v_{n,il} v_{n,ip} v_{n,iq}) - \sigma_{kl} \sigma_{pq}^2 \right) \right]
\]

and

\[
\mathbb{E}[Q_{i,n}^2] = \sum_{k=1}^{m} b_{nk,i}^2 \sigma_{kk}^2 + b_{nk,i} b_{nl,i} \sigma_{kl} + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ a_{nkl,ii} a_{npq,ii} \left( \mathbb{E}(v_{n,ik} v_{n,il} v_{n,ip} v_{n,iq}) - \sigma_{kl} \sigma_{pq}^2 \right) \right]
\]

We note that

\[
\sum_{i=1}^{n} \mathbb{E}(Q_{i,n}^2 | \mathcal{F}_{i-1,n}) - 1 = \frac{1}{\sigma_{Q_n}^2} \sum_{i=1}^{n} \left[ \mathbb{E}(Q_{i,n}^2 | \mathcal{F}_{i-1,n}) - \sigma_{Q_n}^2 \right]
\]
\[
\frac{n}{\sigma^2_{Q_n}} \sum_{i=1}^{n} \frac{1}{n} \left[ \mathbb{E}(Q_{i,n}^2 | \mathcal{F}_{i-1,n}) - \mathbb{E}(Q_{i,n}^2) \right],
\]

where

\[
\mathbb{E}(Q_{i,n}^2 | \mathcal{F}_{i-1,n}) - \mathbb{E}(Q_{i,n})
\]

\[
= \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left\{ \sigma_{kp} \left[ \left( \sum_{j=1}^{i-1} a_{nkl,ij} v_{n,jl} \right) \left( \sum_{j=1}^{i-1} a_{npq,ij} v_{n,jq} \right) - \left( \sum_{j=1}^{i-1} a_{nkl,ij} a_{npq,ij} \sigma_{lq}^2 \right) \right] + \sigma_{kq} \left[ \left( \sum_{j=1}^{i-1} a_{nkl,ij} v_{n,jl} \right) \left( \sum_{j=1}^{i-1} a_{npq,ji} v_{n,jq} \right) - \left( \sum_{j=1}^{i-1} a_{nkl,ij} a_{npq,ij} \sigma_{lp} \right) \right]
\]

\[
+ \sigma_{lp} \left[ \left( \sum_{j=1}^{i-1} a_{nkl,ji} v_{n,jk} \right) \left( \sum_{j=1}^{i-1} a_{npq,ij} v_{n,jq} \right) - \left( \sum_{j=1}^{i-1} a_{nkl,ji} a_{npq,ij} \sigma_{kp} \right) \right]
\]

\[
+ \left( \sum_{j=1}^{i-1} a_{nkl,ii} a_{npq,ij} v_{n,jl} \right) + \left( \sum_{j=1}^{i-1} a_{nkl,ij} a_{npq,ji} v_{n,jk} \right)
\]

\[
+ \left( \sum_{j=1}^{i-1} a_{nkl,ji} a_{npq,ij} v_{n,jl} \right) + \left( \sum_{j=1}^{i-1} a_{nkl,ji} a_{npq,ij} \sigma_{lp} \right)
\]

\[
+ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left\{ \sigma_{pk} b_{np,i} \left( \sum_{j=1}^{i-1} a_{nkl,ij} v_{n,jl} \right) + \sigma_{pl} b_{np,i} \left( \sum_{j=1}^{i-1} a_{nkl,ji} v_{n,jl} \right) \right\}.
\]

These terms can be sorted into three categories in which we can concentrate our analysis on the following terms:

\[
H_{a1} = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \sum_{j=1}^{i-1} a_{nkl,ij} v_{n,jl} \right) \left( \sum_{j=1}^{i-1} a_{npq,ij} v_{n,jq} \right) - \left( \sum_{j=1}^{i-1} a_{nkl,ij} a_{npq,ij} \sigma_{lq}^2 \right) \right],
\]

\[
H_{a2} = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \sum_{j=1}^{i-1} a_{nkl,ij} v_{n,jl} \right) \left( \sum_{j=1}^{i-1} a_{npq,ji} v_{n,jq} \right) - \left( \sum_{j=1}^{i-1} a_{nkl,ij} a_{npq,ij} \sigma_{lp} \right) \right],
\]

\[
H_{a3} = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \sum_{j=1}^{i-1} a_{nkl,ji} v_{n,jk} \right) \left( \sum_{j=1}^{i-1} a_{npq,ij} v_{n,jq} \right) - \left( \sum_{j=1}^{i-1} a_{nkl,ji} a_{npq,ij} \sigma_{kp} \right) \right],
\]

\[
H_{a4} = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \sum_{j=1}^{i-1} a_{nkl,ji} v_{n,jk} \right) \left( \sum_{j=1}^{i-1} a_{npq,ji} v_{n,jp} \right) - \left( \sum_{j=1}^{i-1} a_{nkl,ji} a_{npq,ji} \sigma_{kp} \right) \right],
\]

137
\[
H_{b1} = \frac{1}{n} \sum_{i=1}^{n} a_{nkl,ii} \left( \sum_{j=1}^{i-1} a_{npq,ij} v_{n,jq} \right), \quad H_{b2} = \frac{1}{n} \sum_{i=1}^{n} a_{nkl,ii} \left( \sum_{j=1}^{i-1} a_{npq,ji} v_{n,jp} \right),
\]
\[
H_{b3} = \frac{1}{n} \sum_{i=1}^{n} a_{npq,ii} \left( \sum_{j=1}^{i-1} a_{nkl,ij} v_{n,jl} \right), \quad H_{b4} = \frac{1}{n} \sum_{i=1}^{n} a_{npq,ii} \left( \sum_{j=1}^{i-1} a_{nkl,ji} v_{n,jk} \right),
\]
\[
H_{c1} = \frac{1}{n} \sum_{i=1}^{n} b_{np,i} \left( \sum_{j=1}^{i-1} a_{nkl,ij} v_{n,jl} \right) \quad \text{and} \quad H_{c2} = \frac{1}{n} \sum_{i=1}^{n} b_{np,i} \left( \sum_{j=1}^{i-1} a_{nkl,ji} v_{n,jk} \right).
\]

It is easy to observe that all those terms have zero means. Define
\[
H_{1,a1} = \frac{1}{n} \sum_{i=1}^{n} \left[ \sum_{j=1}^{i-1} a_{nkl,ij} a_{npq,ij} \left( v_{n,jl} v_{n,jq} - \sigma_{qy}^2 \right) \right], \quad \text{and}
\]
\[
H_{2,a1} = \frac{1}{n} \sum_{i=1}^{n} \left[ \sum_{j=1}^{i-1} \sum_{r \neq j} a_{nkl,ij} a_{npq,ir} v_{n,jl} v_{n,rq} \right].
\]

Then, \( H_{a1} = H_{1,a1} + H_{2,a1} \) and \( \mathbb{E}(H_{1,a1}) = \mathbb{E}(H_{2,a1}) = 0 \). By rearrangement,
\[
H_{1,a1} = \frac{1}{n} \sum_{i=1}^{n} \left[ \sum_{j=1}^{i-1} a_{nkl,ij} a_{npq,ij} \left( v_{n,jl} v_{n,jq} - \sigma_{qy}^2 \right) \right]
\]
\[
= \frac{1}{n} \sum_{j=1}^{n-1} \sum_{i=j+1}^{n} a_{nkl,ij} a_{npq,ij} \left( v_{n,jl} v_{n,jq} - \sigma_{qy}^2 \right) = \frac{1}{n} \sum_{j=1}^{n-1} (v_{n,jl} v_{n,jq} - \sigma_{qy}^2) \sum_{i=j+1}^{n} a_{nkl,ij} a_{npq,ij}.\]

Denote \( u_j = (v_{n,jl} v_{n,jq} - \sigma_{qy}^2) \sum_{i=j+1}^{n} a_{nkl,ij} a_{npq,ij} \). We note that \( \{u_j\} \) is sequence of independent random variables with 0 mean. Since \( \mathbb{E}|v_{n,jl} v_{n,jq} - \sigma_{qy}^2| \) is bounded by a finite number, say, \( M_4 \), we have \( \mathbb{E}(H_{1,a1}^2) \leq M_4 \frac{1}{n^2} \sum_{j=1}^{n-1} \left( \sum_{i=j+1}^{n} a_{nkl,ij} a_{npq,ij} \right)^2 \).

Since \( A_{nkl} \) and \( A_{npq} \) have uniformly bounded row and column sums, \( \sum_{i=1}^{n} |a_{nkl,ij} a_{npq,ij}| \leq C_2 \), where \( C_2 \) is a constant. Therefore, \( \mathbb{E}(H_{1,a1}^2) \leq M_4 C_2^2 \frac{1}{n} \to 0 \). So \( H_{1,a1} \overset{p}{\to} 0 \). As
\[
H_{2,a1} = \frac{1}{n} \sum_{i=1}^{n} \left[ \sum_{j=1}^{i-1} \sum_{r \neq j} a_{nkl,ij} a_{npq,ir} v_{n,jl} v_{n,rq} \right]
\]
\[
= \frac{1}{n} \sum_{j=1}^{n-1} \sum_{r \neq j} v_{n,jl} v_{n,rq} \sum_{i=\max(j,r)+1}^{n} a_{nkl,ij} a_{npq,ir},
\]
\[138\]
\[ \mathbb{E}(H_{2,a1}^2) = \frac{1}{n^2} \sum_{j=1}^{n-1} \sum_{r \neq j} \sum_{s=1}^{n-1} \sum_{t \neq s} \mathbb{E}(v_{nj}v_{nr,v_{ns}v_{nt,q}}) \]

\[ \times \sum_{i=\max(j,r)+1}^{n} a_{nk,i,j}a_{npq,ir} \sum_{i=\max(s,t)+1}^{n} a_{nk,i,s}a_{npq,it} \]

\[ = \frac{1}{n^2} \sum_{j=1}^{n-1} \sum_{r \neq j} \mathbb{E}(v_{nj}^2v_{nr,r}^2) \left[ \sum_{i=\max(j,r)+1}^{n} a_{nk,i,j}a_{npq,ir} \right] \]

\[ + \frac{1}{n^2} \sum_{j=1}^{n-1} \sum_{r \neq j} \mathbb{E}(v_{nj}v_{nr,v_{nr},v_{nj,q}}) \left[ \sum_{i=\max(j,r)+1}^{n} a_{nk,i,j}a_{npq,ir} \right] \left[ \sum_{i=\max(j,r)+1}^{n} a_{nk,i,r}a_{npq,ij} \right]. \]

Since \( \mathbb{E}(v_{nj}^2v_{nr,r}^2) \) and \( \mathbb{E}(v_{nj}v_{nr,v_{nr},v_{nj,q}}) \) are bounded by a constant, say, \( D_4 \) for any \( j, r = 1, \ldots, n \) and \( l, q = 1, \ldots, m, \)

\[ \mathbb{E}(H_{2,a1}^2) \leq \frac{1}{n^2}D_4 \sum_{j=1}^{n-1} \sum_{r \neq j} \left[ \sum_{i=\max(j,r)+1}^{n} |a_{nk,i,j}a_{npq,ir}| \right]^2 \]

\[ + \frac{1}{n^2}D_4 \sum_{j=1}^{n-1} \sum_{r \neq j} \left[ \sum_{i=\max(j,r)+1}^{n} |a_{nk,i,j}a_{npq,ir}| \right] \left[ \sum_{i=\max(j,r)+1}^{n} |a_{nk,i,r}a_{npq,ij}| \right] \]

\[ \leq \frac{1}{n^2}D_4 \sum_{j=1}^{n} \sum_{r=1}^{n} \left[ \sum_{i=1}^{n} |a_{nk,i,j}a_{npq,ir}| \right]^2 \]

\[ + \frac{1}{n^2}D_4 \sum_{j=1}^{n} \sum_{r=1}^{n} \left[ \sum_{i=1}^{n} |a_{nk,i,j}a_{npq,ir}| \right] \left[ \sum_{i=1}^{n} |a_{nk,i,r}a_{npq,ij}| \right]. \]

Denote \( A_{nk}^{|r|} \) to be a matrix in which all elements are absolute value of corresponding entries of \( A_{nk} \). Then \( \sum_{i=1}^{n} |a_{nk,i,j}a_{npq,ir}| \) is the \( (j, r) \)th element of \( A_{nk}^{|r|}A_{npq}^{|r|} \), say, \( c_{jr} \). So \( \sum_{r=1}^{n} (\sum_{i=1}^{n} |a_{nk,i,j}a_{npq,ir}|)^2 \) is the \( (j, j) \)th element of \( A_{nk}^{|r|}A_{npq}^{|r|}A_{nk}^{|r|}A_{npq}^{|r|} \), which is a UB matrix. Thus, \( \sum_{j=1}^{n} \sum_{r=1}^{n} (\sum_{i=1}^{n} |a_{nk,i,j}a_{npq,ir}|)^2 = O(n) \). In order to get the rate of the other term, the similar logic applies, hence we have the order that \( \sum_{j=1}^{n} \sum_{r=1}^{n} (\sum_{i=1}^{n} |a_{nk,i,j}a_{npq,ir}|) (\sum_{i=1}^{n} |a_{nk,i,r}a_{npq,ij}|) = O(n) \). So, \( H_{2,a1} \overset{p}{\to} 0 \), which means \( H_{a1} \overset{p}{\to} 0 \) by combining the above results. Similarly, all the terms in \( H_a \) can be proved to converge in probability to 0.
Now consider $H_{b1} = \frac{1}{n} \sum_{i=1}^{n} a_{nkli} \left( \sum_{j=1}^{i-1} a_{npqij} v_{njq} \right)$. First, row and column sums of $A_{nkl}$ are uniformly bounded by $C_1$. As,

$$H_{b1} = \frac{1}{n} \sum_{i=1}^{n} a_{nkli} \left( \sum_{j=1}^{i-1} a_{npqij} v_{njq} \right) = \frac{1}{n} \sum_{j=1}^{n} v_{njq} \sum_{i=j+1}^{n} a_{nkli} a_{npqij}.$$

$$E H_{b1}^2 = \frac{1}{n^2} \sum_{j=1}^{n-1} E \left( v_{njq}^2 \right) \left( \sum_{i=j+1}^{n} a_{nkli} a_{npqij} \right)^2 \leq \frac{1}{n^2} \sigma_{qq}^2 \sum_{j=1}^{n-1} \left( \sum_{i=1}^{n} |a_{nkli}| |a_{npqij}| \right)^2 \leq \frac{1}{n^2} \sigma_{qq}^2 \sum_{j=1}^{n-1} (C_1 C_1)^2 = O(n^{-1}).$$

The same method applies for all $H_b$ and $H_c$.

With the preceding analysis, we conclude that

$$\sum_{i=1}^{n} E(Q_{i,i}^2 | F_{i-1,n}) - 1 = \frac{1}{\sigma_{Q_{n}}^2} \sum_{i=1}^{n} \left[ E(Q_{i,i}^2 | F_{i-1,n}) - \sigma_{Q_{n}}^2 \right]$$

$$= \frac{n}{\sigma_{Q_{n}}^2} \sum_{i=1}^{n} \left[ E(Q_{i,i}^2 | F_{i-1,n}) - E(Q_{i,i}) \right] \rightarrow 0.$$

As conditions are satisfied, the martingale central limit theorem provides the result. 

**Proof of Lemma 1.7:**

We first change the form of this linear quadratic form,

$$Q_{n}^{(a,b)} = \sum_{k=1}^{m} b'_{nk} V_{nk} + \sum_{k=1}^{m} \sum_{l=1}^{m} V'_{nk} A_{nkli} V_{nl} - \mathbb{E}(V'_{nk} A_{nkli} V_{nl}),$$

$$Q_{n}^{(c,d)} = \sum_{k=1}^{m} d'_{nk} V_{nk} + \sum_{k=1}^{m} \sum_{l=1}^{m} V'_{nk} C_{nkli} V_{nl} - \mathbb{E}(V'_{nk} C_{nkli} V_{nl}),$$

and,

$$Q_{n}^{(a,b)} Q_{n}^{(c,d)}$$

$$= \left( \sum_{k=1}^{m} b'_{nk} V_{nk} \right) \left( \sum_{k=1}^{m} d'_{nk} V_{nk} \right) + \left( \sum_{k=1}^{m} b'_{nk} V_{nk} \right) \left( \sum_{k=1}^{m} \sum_{l=1}^{m} V'_{nk} A_{nkli} V_{nl} - \mathbb{E}(V'_{nk} A_{nkli} V_{nl}) \right)$$

$$+ \left( \sum_{k=1}^{m} d'_{nk} V_{nk} \right) \left( \sum_{k=1}^{m} \sum_{l=1}^{m} V'_{nk} A_{nkli} V_{nl} - \mathbb{E}(V'_{nk} A_{nkli} V_{nl}) \right)$$

$$+ \left( \sum_{k=1}^{m} \sum_{l=1}^{m} V'_{nk} A_{nkli} V_{nl} - \mathbb{E}(V'_{nk} A_{nkli} V_{nl}) \right) \left( \sum_{k=1}^{m} \sum_{l=1}^{m} V'_{nk} C_{nkli} V_{nl} - \mathbb{E}(V'_{nk} C_{nkli} V_{nl}) \right).$$
Then,
\[
E \left( \sum_{k=1}^{m} b'_{nk} V_{nk} \right) \left( \sum_{l=1}^{m} d'_{nl} V_{nl} \right) = \sum_{k=1}^{m} \sum_{l=1}^{m} \sigma_{kl} b'_{nk} d_{nl},
\]
\[
E \left( \sum_{k=1}^{m} b'_{nk} V_{nk} \right) \left[ \sum_{p=1}^{m} \sum_{l=1}^{m} \left[ V'_{nl} C_{nlp} V_{np} - E(V'_{nl} C_{nlp} V_{np}) \right] \right] = \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \mu_{klp} \sum_{i=1}^{n} b_{nk,i} C_{nlp,ii},
\]
\[
E \left( \sum_{k=1}^{m} d'_{nk} V_{nk} \right) \left[ \sum_{p=1}^{m} \sum_{l=1}^{m} \left[ V'_{nl} A_{nlp} V_{np} - E(V'_{nl} A_{nlp} V_{np}) \right] \right] = \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \mu_{klp} \sum_{i=1}^{n} d_{nk,j} A_{nlp,ii},
\]
\[
E \left( \sum_{k=1}^{m} \sum_{l=1}^{m} \left[ V'_{nk} A_{nklt} V_{nl} - E(V'_{nk} A_{nklt} V_{nl}) \right] \right) \left( \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ V'_{np} C_{npq} V_{nq} - E(V'_{np} C_{npq} V_{nq}) \right] \right) = \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \mu_{klpq} \left( \sum_{i=1}^{n} A_{nklt,i} C_{npq,ii} \right) + \sigma_{kp} \sigma_{lq} \left( \sum_{i=1}^{n} \sum_{j \neq i}^{n} A_{nklt,ij} C_{npq,ij} \right) + \sigma_{kq} \sigma_{lp} \left( \sum_{i=1}^{n} \sum_{j \neq i}^{n} A_{nklt,ij} C_{npq,ji} \right),
\]

The above result is the expansion of matrices stated in the lemma. \(\square\)

**Proof of Lemma 1.8:**

The proof is a technical extension to the counterpart proposition from Lee (2004). First, we prove that \(\frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\hat{\theta})}{\partial \hat{\theta}^2} \right] \overset{p}{\to} 0\), since \(X_n X_n / n \overset{p}{\to} 0\).

(i) \(\frac{1}{n} \frac{\partial^2 \ln L_{nm}(\hat{\theta})}{\partial (C_m)^2} = \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial (C_m)^2} = \frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \lambda_{m,k} \partial \lambda_{m,l}} \right] = 0\), since \(X_n X_n / n \overset{p}{\to} 0\).

(ii) \(\frac{1}{n} \frac{\partial^2 \ln L_{nm}(\hat{\theta})}{\partial \lambda_{m,k} \partial \lambda_{m,l}} = \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \lambda_{m,k} \partial \lambda_{m,l}} = \frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \lambda_{m,k} \partial \lambda_{m,l}} \right] = 0\), since \(X_n X_n / n \overset{p}{\to} 0\).

(iii) Note that \( \text{vec}(U_{nm}(\hat{\theta})) = (\hat{\Gamma}_m' \otimes I_n - \hat{\Lambda}_m' \otimes W_n) \text{vec}(Y_{nm}) - (I_m \otimes X_n) \text{vec}(\hat{C}_m) = \text{vec}(U_{nm}) + (I_m \otimes X_n) (\text{vec}(C_m) - \text{vec}(\hat{C}_m)) + [I_m \otimes (W_n Y_{nm})] (\text{vec}(\Lambda_m) - \text{vec}(\hat{\Lambda}_m)) - (I_m \otimes Y_{nm}) (\text{vec}(\hat{\Gamma}_m) - \text{vec}(\hat{\Gamma}_m)) \), therefore

\[
\frac{1}{n} \frac{\partial^2 \ln L_{nm}(\hat{\theta})}{\partial \lambda_{m,k} \partial \lambda_{m,l}} = \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \lambda_{m,k} \partial \lambda_{m,l}} \overset{p}{\to} 0.
\]
(iv) Since
\[
\frac{1}{n} \text{vec}(U_{nm}(\hat{\theta}))'[(\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um}) \otimes I_n] \text{vec}(U_{nm}(\hat{\theta}))
\]
\[
= \frac{1}{n} \text{vec}(U_{nm}) + (I_m \otimes X_n)(\text{vec}(C_m) - \text{vec}(\tilde{C}_m))
\]
\[
+ \frac{1}{n}((I_m \otimes X_n)(\text{vec}(C_m) - \text{vec}(\tilde{C}_m)))'[(\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um}) \otimes I_n]
\]
\[
\times ((I_m \otimes X_n)(\text{vec}(C_m) - \text{vec}(\tilde{C}_m)))
\]
\[
+ \frac{1}{n}[(I_m \otimes (W_n Y_{nm}))(\text{vec}(\Lambda_m) - \text{vec}(\tilde{\Lambda}_m))]'
\]
\[
[(\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um} F_{m,kl} \tilde{\Sigma}^{-1}_{um}) \otimes I_n]
\]
\[
\times [(I_m \otimes (W_n Y_{nm}))(\text{vec}(\Lambda_m) - \text{vec}(\tilde{\Lambda}_m))]
\]
\[
+ \frac{1}{n}[(I_m \otimes Y_{nm})(\text{vec}(\Gamma_m) - \text{vec}(\tilde{\Gamma}_m))]'
\]
\[
[(\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um} F_{m,kl} \tilde{\Sigma}^{-1}_{um}) \otimes I_n]
\]
\[
\times [(I_m \otimes Y_{nm})(\text{vec}(\Gamma_m) - \text{vec}(\tilde{\Gamma}_m))]
\]
\[
+ \frac{2}{n} \text{vec}(U_{nm})'[(\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um} F_{m,kl} \tilde{\Sigma}^{-1}_{um}) \otimes (W_n Y_{nm})](\text{vec}(\Lambda_m) - \text{vec}(\tilde{\Lambda}_m))
\]
\[
+ \frac{2}{n} \text{vec}(U_{nm})'[(\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um} F_{m,kl} \tilde{\Sigma}^{-1}_{um}) \otimes Y_{nm}](\text{vec}(\Gamma_m) - \text{vec}(\tilde{\Gamma}_m))
\]
\[
+ \frac{2}{n} \text{vec}(U_{nm})'[(\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um} F_{m,kl} \tilde{\Sigma}^{-1}_{um}) \otimes X_n](\text{vec}(C_m) - \text{vec}(\tilde{C}_m))
\]
\[
+ \frac{2}{n}[(I_m \otimes X_n)(\text{vec}(C_m) - \text{vec}(\tilde{C}_m))]'[\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um} F_{m,kl} \tilde{\Sigma}^{-1}_{um}] \otimes I_n
\]
\[
\times [(I_m \otimes (W_n Y_{nm}))(\text{vec}(\Lambda_m) - \text{vec}(\tilde{\Lambda}_m))]
\]
\[
= \frac{1}{n} \text{vec}(U_{nm})'[(\tilde{\Sigma}^{-1}_{um} F_{m,ij} \tilde{\Sigma}^{-1}_{um} F_{m,kl} \tilde{\Sigma}^{-1}_{um}) \otimes I_n] \text{vec}(U_{nm}) + o_p(1),
\]
\[ \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\hat{\theta})}{\partial \Sigma_{nm,kl} \partial \Sigma_{nm,ij}} - \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \Sigma_{nm,kl} \partial \Sigma_{nm,ij}} = \frac{1}{2} \text{Tr} \left[ (\tilde{\Sigma}_{nm}^{-1} F_{m,ij} \tilde{\Sigma}_{nm}^{-1} F_{m,kl} - \Sigma_{nm0}^{-1} F_{m,ij} \Sigma_{nm0}^{-1} F_{m,kl} ) \otimes I_n \right] + \frac{1}{n} \text{vec} (U_{nm})' [(\tilde{\Sigma}_{nm}^{-1} F_{m,ij} \tilde{\Sigma}_{nm}^{-1} F_{m,kl} - (\Sigma_{nm0}^{-1} F_{m,ij} \Sigma_{nm0}^{-1} F_{m,kl}) \otimes I_n ) \text{vec} (U_{nm}) + o_p(1) \xrightarrow{p} 0. \]

(v)

\[ \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\hat{\theta})}{\partial \Sigma_{nm,ijkl} \partial \Lambda_{m,kl}} - \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \Sigma_{nm,ijkl} \partial \Lambda_{m,kl}} = -\frac{1}{n} \text{vec} (U_{nm})' [(\tilde{\Sigma}_{nm}^{-1} F_{m,ij} \tilde{\Sigma}_{nm}^{-1} E_{m,kl} ) \otimes I_n ] \text{vec} (Y_{nm}) - \frac{\text{vec} (C_{m0}) - \text{vec} (\tilde{C}_m)'}{n} (\tilde{\Sigma}_{nm}^{-1} F_{m,ij} \tilde{\Sigma}_{nm}^{-1} E_{m,kl} ) \otimes Y_n' \text{vec} (Y_{nm}) - \text{vec} (\Lambda_{m0}) - \text{vec} (\Lambda_{m0})' (\tilde{\Sigma}_{nm}^{-1} F_{m,ij} \tilde{\Sigma}_{nm}^{-1} E_{m,kl} ) \otimes (Y_n' W_n') \text{vec} (Y_{nm}) \xrightarrow{p} 0. \]

(vi) Since

\[ \text{Tr} [S_{nm}^{-1}(E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,kl} \otimes W_n)] - \text{Tr} [S_{nm0}^{-1}(E'_{m,ij} \otimes W_n) S_{nm0}^{-1}(E'_{m,kl} \otimes W_n)] = \sum_{p=1}^{m} \sum_{q=1}^{m} \text{Tr} [(E'_{m,kl} \otimes W_n) S_{nm}^{-1}(E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,pq} \otimes W_n) S_{nm}^{-1}] (\tilde{\Lambda}_{pq,m} - \Lambda_{pq,m}) \]

\[ + (E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,kl} \otimes W_n) S_{nm}^{-1}(E'_{m,pq} \otimes W_n) S_{nm}^{-1}] (\tilde{\Lambda}_{pq,m} - \Lambda_{pq,m0}) \]

\[ - (E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,kl} \otimes W_n) S_{nm}^{-1}(E'_{m,pq} \otimes I_n) S_{nm}^{-1} \]

\[ + (E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,kl} \otimes W_n) S_{nm}^{-1}(E'_{m,pq} \otimes I_n) S_{nm}^{-1}] (\tilde{\Lambda}_{pq,m} - \Lambda_{pq,m0}) \]

\[ \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\hat{\theta})}{\partial \Lambda_{m,ij} \partial \Lambda_{m,kl}} - \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \Lambda_{m,ij} \partial \Lambda_{m,kl}} = \frac{1}{n} \text{vec} (Y_{nm})' [(E_{m,ij} (\tilde{\Sigma}_{nm}^{-1} - \Sigma_{nm0}^{-1} E_{m,kl} ) \otimes (W_n' W_n)] \text{vec} (Y_{nm}) \]

\[ - (\sum_{p=1}^{m} \sum_{q=1}^{m} \text{Tr} [(E'_{m,kl} \otimes W_n) S_{nm}^{-1}(E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,pq} \otimes W_n) S_{nm}^{-1}] n \times (\tilde{\Lambda}_{pq,m} - \Lambda_{pq,m0}) \]

\[ + (E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,kl} \otimes W_n) S_{nm}^{-1}(E'_{m,pq} \otimes W_n) S_{nm}^{-1}] n \times (\tilde{\Lambda}_{pq,m} - \Lambda_{pq,m0}) \]

\[ + (E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,kl} \otimes W_n) S_{nm}^{-1}(E'_{m,pq} \otimes I_n) S_{nm}^{-1}] n \times (\tilde{\Lambda}_{pq,m} - \Lambda_{pq,m0}) \]

\[ + (E'_{m,ij} \otimes W_n) S_{nm}^{-1}(E'_{m,kl} \otimes W_n) S_{nm}^{-1}(E'_{m,pq} \otimes I_n) S_{nm}^{-1}] n \times (\tilde{\Lambda}_{pq,m} - \Lambda_{pq,m0}) \xrightarrow{p} 0. \]
(vii) Similarly to (ii) (v) and (vi),
\[
\frac{1}{n} \frac{\partial^2 \ln L_{nm}(\tilde{\theta})}{\partial \text{vec}(C_m) \partial \Lambda_{m,ij}} - \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \text{vec}(C_m) \partial \Lambda_{m,ij}} = \\
- \left[ I_{nm} \otimes \left( \frac{X_n W_n Y_{nm}}{n} \right) \right] [\text{vec}((\Sigma_{um}^{-1} - \Sigma_{um0}^{-1}) E_{m,ij}')] \xrightarrow{p} 0,
\]
\[
\frac{1}{n} \frac{\partial^2 \ln L_{nm}(\tilde{\theta})}{\partial \Sigma_{um,ij} \partial \Lambda_{m,kl}} - \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \Sigma_{um,ij} \partial \Gamma_{m,kl}} \xrightarrow{p} 0,
\]
and
\[
\frac{1}{n} \frac{\partial^2 \ln L_{nm}(\tilde{\theta})}{\partial \Lambda_{m,ij} \partial \Lambda_{m,kl}} - \frac{1}{n} \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \Gamma_{m,ij} \partial \Gamma_{m,kl}} \xrightarrow{p} 0.
\]

Therefore, \( \frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\tilde{\theta})}{\partial \theta \partial \theta'} \right] \xrightarrow{p} \frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \theta \partial \theta'} \right]. \)

Next, we prove \( \frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\tilde{\theta})}{\partial \theta \partial \theta'} \right] - \mathbb{E} \left( \frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \theta \partial \theta'} \right] \right) \xrightarrow{p} 0. \) For our model, \( \text{vec}(Y_{nm}) = S_{nm0}^{-1}[(I_m \otimes X_n) \text{vec}(C_{m0}) + \text{vec}(U_{nm})]. \) We evaluate the second order derivatives at the true value of the parameter. The second order derivatives are in Appendix A of the main text. According to Lemma 7, if \( B'_nm, \) a \( mn \times 1 \) vector, is bounded in elements and \( A_{nm}, \) a \( mn \times mn \) matrix, is UB, \( \text{var} \left( \frac{1}{n} B'_nm \text{vec}(U_{nm}) \right) = \frac{1}{n^2} B'_nm (\Sigma_{um} \otimes I_n) B_{nm} = O(\frac{1}{n^2}), \) and
\[
\text{var} \left( \frac{1}{n} \text{vec}(U_{nm})' A_{nm} \text{vec}(U_{nm}) \right) = \frac{1}{n^2} \text{Tr}[A_{nm}(\Sigma_{um} \otimes I_n)(A_{nm} + A_{nm}') (\Sigma_{um} \otimes I_n)]
\]
\[
+ \frac{1}{n^2} \sum_{k=1}^m \sum_{l=1}^m \sum_{p=1}^m \sum_{q=1}^m \left[ (\mu_{kqp} - \sigma_{kp}\sigma_{lq} - \sigma_{kq}\sigma_{lp} - \sigma_{kl}\sigma_{pq}^2) \left( \sum_{i=1}^n A_{nkli} A_{npiq} \right) \right] = O(\frac{1}{n}).
\]
Then, by applying these results into the second order derivatives, Chebyshev’s inequality implies, \( \frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\tilde{\theta})}{\partial \theta \partial \theta'} \right] - \mathbb{E} \left( \frac{1}{n} \left[ \frac{\partial^2 \ln L_{nm}(\theta_0)}{\partial \theta \partial \theta'} \right] \right) \xrightarrow{p} 0. \)

**Proof of Theorem 1.2:**

Utilizing Lemmas 1.6 and 1.7, we can get the asymptotical distribution of \( \frac{1}{\sqrt{n}} \frac{\partial \ln L_{nm}(\theta_0)}{\partial \theta}. \) The convergence of second order derivatives has been derived in Lemma 1.8. The proposition follows.
Appendix D: Some Detailed Derivation and Useful Lemmas in Chapter 2

D.1 Detailed Derivations

Since $B_{nm} = (\Gamma_n \otimes I_m) \text{diag}\{0, ..., 0, (I_m - \omega_{n,n+1}\Psi'_m)^{-1}(P'_m + \omega_{n,n}\Psi'_m)\}(\Gamma_n^{-1} \otimes I_m)$, we have

$$H^h_{nm} = \Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi'_m)^{-1}(P_m + \Phi_m)]^h + B^h_{nm}$$

$$H^h_{nm} S^{-1}_{nm} = \Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi'_m)^{-1}(P_m + \Phi_m)]^h (I_m - \Psi'_m)^{-1} + B^h_{nm} S^{-1}_{nm}$$

$$B^t_{nm} A^h_{nm} S^{-1}_{nm} = B^t_{nm} \{\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi'_m)^{-1}(P_m + \Phi_m)]^h (I_m - \Psi'_m)^{-1}\} + B^{t+h} S^{-1}_{nm}$$

because $B^t_{nm} \{\Gamma_n \mathbb{I}_{n,n} \Gamma_n^{-1} \otimes [(I_m - \Psi'_m)^{-1}(P_m + \Phi_m)]^h (I_m - \Psi'_m)^{-1}\} (\Gamma_n^{-1} \otimes I_m) = 0$, which holds for all $h = 0, 1, \ldots$.

The following shows how to handle the time fixed effects. Since $\text{vec}(D'_{mt}) = l_n \otimes \alpha_{m,t}$. Under the assumptions on $W_n$ and $\Psi_m$, $S^{-1}_{nm}$ can be expanded into an infinite series such that $S^{-1}_{nm} = I_n \otimes I_m + W_n \otimes \Psi'_m + W^2_n \otimes \Psi'^2 + \cdots$. Thus,

$$S^{-1}_{nm} \text{vec}(D'_{mt}) = (I_n \otimes I_m + W_n \otimes \Psi'_m + W^2_n \otimes \Psi'^2 + \cdots)(l_n \otimes \alpha_{m,t}) = l_n \otimes (I_m - \Psi'_m)^{-1} \alpha_{m,t}$$

due to $W_n l_n = l_n$. We note that for any vector or matrix $a_m$ with $m$-rows, $B_{nm}(l_n \otimes a_m) = (\Gamma_n \otimes I_m) \tilde{B}_{nm}(\Gamma_n^{-1} \otimes I_m)(l_n \otimes a_m) = \sqrt{n}(\Gamma_n \otimes I_m) \tilde{B}_{nm}(c_n \otimes a_m) = 0$ because the first diagonal block of $\tilde{B}_{nm}$ is zero. Thus we have $B^h_{nm} \text{vec}(D'_{mt}) = 0$
and $B_{nm}^hS_{nm}^{-1}vec(D'_{m,t}) = 0$ for any $h = 1, 2, \ldots$. It is apparent that $\Gamma_n\mathbb{1}_{n,n}, \Gamma_n^{-1}l_n = \sqrt{n}\Gamma_n\mathbb{1}_{n,n}e_n = \sqrt{n}\Gamma_n e_n = l_n$. Therefore, they imply that $H_{nm}^h vec(D'_{m,t}) = l_n \otimes [(I_m - \Psi^t_m)^{-1}(P_m + \Phi_m)\Gamma_{nm,t}]^h \alpha_{m,t}$, and

$$H_{nm}^h S_{nm}^{-1} vec(D'_{mt}) = l_n \otimes [(I_m - \Psi^t_m)^{-1}(P_m + \Phi_m)\Gamma_{nm,t}]^h (I_m - \Psi^t_m)^{-1} \alpha_{m,t}, \text{ for all } h = 0, 1, 2, \ldots.$$ The preceding feature of $B_{nm}$ gives also

$$B_{nm}^t vec(Y'_{nm,-1}) = B_{nm}^t \sum_{h=0}^{\infty} H_{nm}^h S_{nm}^{-1}[Q_{nm,-h-1} + vec(D'_{m,-h-1})]$$

$$= \sum_{h=0}^{\infty} B_{nm}^{t+h} S_{nm}^{-1}[Q_{nm,-h-1} + vec(D'_{m,-h-1})] = \sum_{h=0}^{\infty} B_{nm}^{t+h} S_{nm}^{-1}Q_{nm,-h-1},$$

and, hence

$$H_{nm}^t vec(Y'_{nm,-1}) = (\Gamma_n\mathbb{1}_{n,n}\Gamma_n^{-1} \otimes [(I_m - \Psi^t_m)^{-1}(P_m + \Phi_m)\Gamma_{nm,t}^t] + B_{nm}^t) vec(Y'_{nm,-1})$$

$$= (\Gamma_n\mathbb{1}_{n,n}\Gamma_n^{-1} \otimes [(I_m - \Psi^t_m)^{-1}(P_m + \Phi_m)\Gamma_{nm,t}^t]) vec(Y'_{nm,-1}) + \sum_{h=0}^{\infty} B_{nm}^{t+h} S_{nm}^{-1}Q_{nm,-h-1}.$$

### D.2 Useful Lemmas

The reduced form of MDP-SAR is

$$vec(Y'_{nm,t}) = \sum_{h=0}^{+\infty} H_{nm}^h S_{nm}^{-1}[(X_{nk,t-h} \otimes I_m) vec(\Pi'_{km}) + vec(C'_{nm})] + l_n \otimes \alpha_{m,t} + vec(V'_{nm,t-h})],$$

Denote $\mathbb{E}(vec(V'_{nm,t}) vec(V'_{nm,t}')) = \Sigma_{v,nn} = I_n \otimes \Sigma_{vm}$, and recall $H_{nm} = S_{nm}^{-1}(P_m \otimes I_n + \Phi^t_m \otimes W_n)$.

The difference between the above expression and that in Appendix of Yu, de Jong and Lee (2008) is that the disturbance terms involve multivariate vectors which are independent identically distributed but elements of the random vector can be correlated. The statistics $\sum_{h=0}^{+\infty} H_{nm}^h S_{nm}^{-1} vec(V'_{nm,t-h})$ is crucial in our analysis for asymptotic properties.

Now we will provide some basic lemmas for relevant statistics. Define $U_{nm,t} = \sum_{h=1}^{+\infty} G_{nm,h} vec(V_{nm,t+1-h})$ and $W_{nm,t} = \sum_{h=1}^{+\infty} H_{nm,h} vec(V_{nm,t+1-h})$. 

146
The first four lemmas state the formula of the first and second moments of this statistics and disturbances forms.

**Lemma D.1.** Under Assumption 2.3.1, when \( t \geq s \),

\[
E(U_{nm,t}'W_{nm,s}') = \sum_{h=1}^{+\infty} G_{nm,t-s+h} \Sigma_{v,nm} H_{nm,h}',
\]

\[
E(U_{nm,t}'W_{nm,s}') = \sum_{h=1}^{+\infty} \text{Tr}[G_{nm,t-s+h} H_{nm,h} \Sigma_{v,nm}].
\]

**Lemma D.2.** Under Assumption 2.3.1,

\[
E[(\text{vec}(V_{nm,t}'))' B_{1nm} \text{vec}(V_{nm,s}'))(\text{vec}(V_{nm,g}'))' B_{2nm} \text{vec}(V_{nm,h}'))] = \]

When \( t = s = g = h \),

\[
\text{Tr}[B_{1nm} \Sigma_{v,nm}(B_{2nm} + B_{2nm}') \Sigma_{v,nm}] + \text{Tr}(B_{1nm} \Sigma_{v,nm}) \text{Tr}(B_{2nm} \Sigma_{v,nm})
\]

\[
+ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2 \sigma_{lq}^2 - \sigma_{kq}^2 \sigma_{lp}^2 - \sigma_{kl}^2 \sigma_{pq}^2) \left( \sum_{i=1}^{n} b_{i,k,1nm} b_{i,pq,2nm} \right);
\]

When \( t = s \neq g = h \), \( \text{Tr}(B_{1nm} \Sigma_{v,nm}) \text{Tr}(B_{2nm} \Sigma_{v,nm}) \);

When \( t = g \neq s = h \), \( \text{Tr}(B_{1nm} \Sigma_{v,nm} B_{2nm}' \Sigma_{v,nm}) \);

When \( t = h \neq s = g \), \( \text{Tr}(B_{1nm} \Sigma_{v,nm} B_{2nm} \Sigma_{v,nm}) \);

Otherwise, 0.

**Lemma D.3.** Under the Assumption 2.3.1,

\[
\text{cov}(U_{nm,t}'W_{nm,t}, U_{nm,s}'W_{nm,s}) = \text{Tr}\left[ +\infty \sum_{g=1}^{H_{nm,t-s+g} \Sigma_{v,nm} H_{nm,g}' \sum_{h=1}^{G_{nm,h} \Sigma_{v,nm} G_{nm,t-s+h}} \right]
\]

\[
+ \text{Tr}\left[ +\infty \sum_{g=1}^{H_{nm,t-s+g} \Sigma_{v,nm} G_{nm,g}' \sum_{h=1}^{H_{nm,h} \Sigma_{v,nm} G_{nm,t-s+h}} \right]
\]

\[
+ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2 \sigma_{lq}^2 - \sigma_{kq}^2 \sigma_{lp}^2 - \sigma_{kl}^2 \sigma_{pq}^2) \left( \sum_{i=1}^{n} b_{i,k,1nm} b_{i,pq,2nm} \right)
\]

\[
\times +\infty \sum_{g=1}^{n} \left( \sum_{i=1}^{m} g_{nm,t-s+g} H_{nm,t-s+g} H_{nm,g} \right)_{1i} \left( G_{nm,g} H_{nm,g} \right)_{1i}^2 .
\]
The order of the variance of a statistic can be applied in establishing uniform convergence and consistency by employing Chebyshev’s inequality.

**Assumption D.2.1.** \( U_{nm,t} = \sum_{h=1}^{\infty} G_{nm,h} \text{vec}(V'_{nm,t+1-h}) = \sum_{h=1}^{\infty} G_{nm,c} G_{nm,d}^h \text{vec}(V'_{nm,t+1-h}) \) and \( W_{nm,t} = \sum_{h=1}^{\infty} H_{nm,h} \text{vec}(V'_{nm,t+1-h}) = \sum_{h=1}^{\infty} H_{nm,c} H_{nm,d}^h \text{vec}(V'_{nm,t+1-h}) \), where \( G_{nm,c}, H_{nm,c}, \sum_{h=1}^{\infty} \text{abs}(G_{nm,d}^h) \) and \( \sum_{h=1}^{\infty} \text{abs}(H_{nm,d}^h) \) are bounded in row and column sum norms.

**Lemma D.4.** Under Assumption 2.3.1 and D.2.1, \( \text{var}(\sum_{t=1}^{T} U_{nm,t} W_{nm,t}) = O(nT) \).

Since fixed effects of individuals are concentrated out by deviations from sample means, the statistic \( \bar{U}_{nm,T} = \frac{1}{T} \sum_{t=1}^{T} U_{nm,T} \) is involved in the concentrated likelihood function. Lemma D.5 formulates its expression.

**Lemma D.5.** Under assumption 2.3.1,

\[
\bar{U}_{nm,T} = \sum_{h=1}^{\infty} \tilde{G}_{nm,h} \text{vec}(V'_{nm,t+1-h}), \quad \bar{W}_{nm,T} = \sum_{h=1}^{\infty} \tilde{H}_{nm,h} \text{vec}(V'_{nm,t+1-h}),
\]

where

\[
\tilde{G}_{nm,h} = \begin{cases} 
\frac{1}{T}(G_{nm,1} + \ldots + G_{nm,h}) = \frac{1}{T} \sum_{g=1}^{h} G_{nm,g} & h \leq T \\
\frac{1}{T} \sum_{g=1}^{T} G_{nm,h-T+g} & h > T 
\end{cases}
\]

The following lemmas, Lemma D.6 and D.7 establish orders of the second moments of relevant statistics and stochastic orders of differences between statistics and their expectations, which are crucial in analyzing the asymptotic properties.

**Lemma D.6.** Under assumptions 2.3.1, 2.3.3 and D.2.1,

\[
(1) \quad E \left[ \frac{1}{nT} \sum_{t=1}^{T} U'_{nm,t} W_{nm,t} \right] = \frac{1}{nT} \sum_{h=1}^{\infty} \text{Tr}[G_{nm,h} H_{nm,h} \Sigma_{v,nm}] = O(1);
\]

\[
\frac{1}{nT} \sum_{t=1}^{T} U'_{nm,t} W_{nm,t} - E \left[ \frac{1}{nT} \sum_{t=1}^{T} U'_{nm,t} W_{nm,t} \right] = O_p \left( \frac{1}{\sqrt{nT}} \right).
\]
\[
E \left[ \frac{1}{n} \bar{U}'_{nm,T} \tilde{W}_{nm,T} \right] = \frac{1}{n} \sum_{h=1}^{\infty} \text{Tr}[\tilde{G}'_{nm,h} \tilde{H}_{nm,h} \Sigma_{v,nn}] = O\left(\frac{1}{T}\right);
\]
\[
\frac{1}{n} \bar{U}'_{nm,T} \tilde{W}_{nm,T} = E \left[ \frac{1}{n} \bar{U}'_{nm,T} \tilde{W}_{nm,T} \right] = O_p\left(\frac{1}{\sqrt{nT}}\right) .
\]

(3)
\[
\frac{1}{nT} \sum_{t=1}^{T} \bar{U}'_{nm,t} \tilde{W}_{nm,t} = E \left[ \frac{1}{nT} \sum_{t=1}^{T} \bar{U}'_{nm,t} \tilde{W}_{nm,t} \right] = O_p\left(\frac{1}{\sqrt{nT}}\right) .
\]

Lemma D.7. Suppose \( D_{nm,t} \) is \( nm \times 1 \) nonstochastic matrix with uniformly bounded elements, under assumptions 2.3.1 and D.2.1,
\[
\frac{1}{nT} \sum_{t=1}^{T} \bar{D}'_{nm,t} \bar{U}_{nm,t} = 0_p(\frac{1}{\sqrt{nT}}),
\]
and
\[
\frac{1}{nT} \sum_{t=1}^{T} \bar{D}'_{nm,t} \bar{U}_{nm,t} = O_p\left(\frac{1}{\sqrt{nT}}\right) .
\]

The nature of dynamic panel model results in time lags. Therefore, define \( \bar{U}_{nm,T} = \frac{1}{T} \sum_{t=0}^{T-1} U_{nm,t} \) and \( \bar{\bar{U}}_{nm,T} = \bar{U}_{nm,t} - \bar{U}_{nm,T} \).

Lemma D.8. Under Assumptions 2.3.1 and D.2.1,
\[
\frac{T}{n} E(\bar{U}'_{nm,T} vec(\tilde{V}'_{nm,T})) = \frac{1}{n} \text{Tr}(\Sigma_{v,nn} G_{nm,c} \sum_{g=1}^{\infty} G_{nm,d}) + O\left(\frac{1}{T}\right);
\]
so \( E(\bar{U}'_{nm,T} vec(\tilde{V}'_{nm,T})) = O\left(\frac{1}{T}\right) \). And
\[
\bar{U}'_{nm,T} vec(\tilde{V}'_{nm,T}) - E(\bar{U}'_{nm,T} vec(\tilde{V}'_{nm,T})) = 0_p(\sqrt{\frac{n}{T^2}}) .
\]

In the analysis of cointegrated model, we decompose the dependent variable into three components \( vec^{(u)}(Y'_{nm,t}), vec^{(s)}(Y'_{nm,t}) \) and \( vec^{(a)}(Y'_{nm,t}) \). Key statistics include \( \sum_{h=0}^{t} (\Gamma_n \Pi_{n,m}^{-1}) \otimes \left[ (I_m - \Psi'_m)^{-1}(P_m + \Phi'_m)^{-1}(I_m - \Psi'_m) \right] vec(C'_{nm}) + vec(V'_{nm,t-h}) + (X_{nk,t-1} \otimes I_m) vec(\Pi_{km}) \).

Assumption D.2.2. Let \( L_{nm,h} = L_{nm,c} L_{nm,d}^h \), where \( L_{nm,c} \) and \( L_{nm,d}^h \) are uniformly bounded in absolute value of row and column sums, uniformly in \( n, m \) and \( h \). Some eigenvalue of \( L_{nm,d}^h \) are one while the others are less than one in absolute value.
We define $X_{nm,t} = \sum_{h=0}^{t-1} L_{nm,h}(X_{nk,t-h} \otimes I_m) vec(\Pi_{km0})$, $C_{nm,t} = \sum_{h=0}^{t-1} L_{nm,h} vec(C_{nm})$ and $V_{nm,t} = \sum_{h=0}^{t-1} L_{nm,h} vec(V_{nm,t-h})$. In addition, define $X_{nm,T} = \frac{1}{T} \sum_{t=1}^{T} X_{nm,t}$ and so are $C_{nm,T}$ and $V_{nm,T}$.

**Lemma D.9.** Under Assumption 2.4.1, 2.4.2 and D.2.2, and $P_{nm}$ is a $n \times m$ matrix with row and column sum bounded uniformly in $nm$ in absolute value,

1. $\frac{1}{nT} \sum_{t=1}^{T} (\bar{X}_{nm,t} + \bar{C}_{nm,t})' P_{nm} (\bar{X}_{nm,t} + \bar{C}_{nm,t}) = O(T^2)$;
2. $\frac{1}{nT} \sum_{t=1}^{T} (\bar{X}_{nm,t} + \bar{C}_{nm,t})' P_{nm} \bar{V}_{nm,t} = O(T)$ and $\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} \bar{V}_{nm,t} - E(\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} \bar{V}_{nm,t}) = O(T);$
3. $E(\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} \bar{V}_{nm,t}) = O(T)$ and $\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} vec(\bar{V}_{nm,t}) - E(\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} vec(\bar{V}_{nm,t})) = O(n)$
4. $\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} vec(\bar{V}_{nm,t}) - E(\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} vec(\bar{V}_{nm,t})) = O(1)$,
5. $\frac{1}{nT} \sum_{t=1}^{T} (\bar{X}_{nm,t} + \bar{C}_{nm,t})' P_{nm} D_{nm,t} = O(T)$;
6. $\frac{1}{nT} \sum_{t=1}^{T} (\bar{X}_{nm,t} + \bar{C}_{nm,t})' \bar{U}_{nm,t} = O(\sqrt{T/n})$;
7. $\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} D_{nm,t} = O(\sqrt{T/n})$;
8. $\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} \bar{U}_{nm,t} - E[\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} \bar{U}_{nm,t}] = O(\sqrt{T/n})$;
9. $\frac{1}{nT} \sum_{t=1}^{T} (\bar{X}_{nm,t} + \bar{C}_{nm,t})' P_{nm} vec(\bar{V}_{nm,t}) = O(\sqrt{T/n})$;
10. $\frac{1}{nT} \sum_{t=1}^{T} \bar{V}_{nm,t} P_{nm} vec(\bar{V}_{nm,t}) = O(\sqrt{T/n})$.

**Proofs**

**Proof of Lemma D.1**

Since we assume $t \geq s$,

\[
\bar{U}_{nm,t} = \sum_{h=1}^{+\infty} G_{nm,h} vec(V'_{nm,t+1-h}) = \sum_{h=1}^{t-s} G_{nm,h} vec(V'_{nm,t+1-h}) + \sum_{h=1}^{+\infty} G_{nm,t-s+h} vec(V'_{nm,s+1-h}),
\]

\[
E(\bar{U}_{nm,t} \bar{W}'_{nm,s})
\]
\[= E \left[ \sum_{h=1}^{t-s} G_{nm,h} \text{vec}(V_{nm,t+1-h}') + \sum_{h=1}^{+\infty} G_{nm,t-s+h} \text{vec}(V_{nm,s+1-h}') \right] \]
\[\times \left[ \sum_{h=1}^{+\infty} H_{nm,h} \text{vec}(V_{nm,s+1-h}') \right]'
\[= E \left[ \sum_{h=1}^{+\infty} G_{nm,t-s+h} \text{vec}(V_{nm,s+1-h}') + \sum_{h=1}^{+\infty} \text{vec}(V_{nm,s+1-h}')' H_{nm,h}' \right]
\[= \sum_{h=1}^{+\infty} G_{nm,t-s+h} \Sigma_{v,nm} H_{nm,h}'. \]

The same method applies to \( E(U_{nm,t} W_{nm,s}) \).

\[\square\]

**Proof of Lemma D.2**

**Proof.**
\[E(\text{vec}(V_{nm,t}') B_{1nm} \text{vec}(V_{nm,s}'))(\text{vec}(V_{nm,g}') B_{2nm} \text{vec}(V_{nm,h}'))\]
\[= E \left( \sum_{i=1}^{n} \sum_{j=1}^{n} v_{i,nm,t} B_{ij,1nm} v_{j,nm,s}' (\sum_{p=1}^{n} \sum_{q=1}^{n} v_{p,nm,g} B_{pq,2nm} v_{q,nm,h}') \right)\]
\[= \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{p=1}^{n} \sum_{q=1}^{n} E(\text{vec}(V_{nm,t} B_{ij,1nm} v_{j,nm,s} v_{p,nm,g} B_{pq,2nm} v_{q,nm,h}'))\]
where \( B_{pq,2nm} \) is the \( pq \)th \( m \times m \) submatrix of \( B_{2nm} \).

When \( t = s = g = h \),
\[\sum_{i=1}^{n} \sum_{p=1}^{n} E(\text{vec}(V_{nm,t} B_{ii,1nm} v_{i,nm,s} v_{p,nm,g} B_{ip,2nm} v_{p,nm,h}'))\]
\[= \sum_{i=1}^{n} \sum_{p=1}^{n} \text{Tr}(B_{ii,1nm} \Sigma_{v,nm}) \text{Tr}(B_{pp,2nm} \Sigma_{v,nm}) = \text{Tr}(B_{1nm} \Sigma_{v,nm}) \text{Tr}(B_{2nm} \Sigma_{v,nm}),\]
where \( i \neq p; \)
\[\sum_{i=1}^{n} \sum_{j=1}^{n} E(\text{vec}(V_{nm,t} B_{ij,1nm} v_{j,nm,s} v_{i,nm,g} B_{ij,2nm} v_{j,nm,h}'))\]
\[= \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Tr}(B_{ij,1nm} \Sigma_{v,nm}) B_{ij,2nm}' \Sigma_{v,nm} = \text{Tr}(B_{1nm} \Sigma_{v,nm} B_{2nm}' \Sigma_{v,nm}),\]
where \( i \neq j; \)
\[\sum_{i=1}^{n} \sum_{j=1}^{n} E(\text{vec}(V_{nm,t} B_{ij,1nm} v_{j,nm,s} v_{j,nm,g} B_{ji,2nm} v_{i,nm,h}'))\]

151
\[
= \sum_{i=1}^{n} \sum_{p=1}^{n} \text{Tr}(B_{ij,1nm} \Sigma_{vm}) \text{Tr}(B_{ji,2nm} \Sigma_{vm}) = \text{Tr}(B_{1nm} \Sigma_{v, nm} B_{2nm} \Sigma_{v, nm}),
\]
where \(i \neq j\); and
\[
\sum_{i=1}^{n} E(v_{i, nm, t} B_{ii,1nm} v'_{i, nm, s} v_{i, nm, g} B_{ii,2nm} v'_{i, nm, h})
\]
\[
= \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2 \sigma_{lq}^2 - \sigma_{kq}^2 \sigma_{lp}^2 - \sigma_{kl}^2 \sigma_{pq}^2) \left( \sum_{i=1}^{n} b_{ii, kl,1nm} b_{ii, pq,2nm} \right).
\]
Therefore
\[
E[(\text{vec}(V'_{nm,t})' B_{1nm} \text{vec}(V'_{nm,s}))(\text{vec}(V'_{nm,g})' B_{2nm} \text{vec}(V'_{nm,h}))]
\]
\[
= \text{Tr}[B_{1nm} \Sigma_{v, nm}(B_{2nm} + B'_{2nm}) \Sigma_{v, nm}] + \text{Tr}(B_{1nm} \Sigma_{v, nm}) \text{Tr}(B_{2nm} \Sigma_{v, nm})
\]
\[
+ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2 \sigma_{lq}^2 - \sigma_{kq}^2 \sigma_{lp}^2 - \sigma_{kl}^2 \sigma_{pq}^2) \left( \sum_{i=1}^{n} b_{ii, kl,1nm} b_{ii, pq,2nm} \right).
\]
When \(t = s \neq g = h\),
\[
E[(\text{vec}(V'_{nm,t})' B_{1nm} \text{vec}(V'_{nm,s}))(\text{vec}(V'_{nm,g})' B_{2nm} \text{vec}(V'_{nm,h}))]
\]
\[
= \text{Tr}(B_{1nm} \Sigma_{v, nm}) \text{Tr}(B_{2nm} \Sigma_{v, nm}).
\]
When \(t = g \neq s = h\)
\[
E[(\text{vec}(V'_{nm,t})' B_{1nm} \text{vec}(V'_{nm,s}))(\text{vec}(V'_{nm,g})' B_{2nm} \text{vec}(V'_{nm,h}))]
\]
\[
= \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} E \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{e=1}^{n} \sum_{f=1}^{n} v_{ik, nm, t} b_{ij, kl,1nm} v_{ji, nm, s} v_{ep, nm, t} b_{ef, pq,2nm} v_{fq, nm, s} \right)
\]
\[
= \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{e=1}^{n} \sum_{f=1}^{n} \sigma_{kp}^2 b_{ij, kl,1nm} \sigma_{lq}^2 b_{ij, pq,2nm} \right)
\]
\[
= \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sigma_{kp}^2 \sigma_{lq}^2 \text{Tr}(B_{kl,1nm} B_{pq,2nm})
\]
\[
= \text{Tr} \left( \begin{pmatrix} B_{11,1nm} & \cdots & B_{1m,1nm} \\ \vdots & \ddots & \vdots \\ B_{m1,1nm} & \cdots & B_{mm,1nm} \end{pmatrix} \begin{pmatrix} \sigma_{11}^2 I_n & \cdots & \sigma_{1m}^2 I_n \\ \vdots & \ddots & \vdots \\ \sigma_{m1}^2 I_n & \cdots & \sigma_{mm}^2 I_n \end{pmatrix} \right)
\]
152
\[
\begin{pmatrix}
B_{11,2nm} & \cdots & B_{1m,2nm} \\
\vdots & \ddots & \vdots \\
B_{m1,2nm} & \cdots & B_{mm,2nm}
\end{pmatrix}
\begin{pmatrix}
\sigma_{11}^2 I_n & \cdots & \sigma_{1m}^2 I_n \\
\vdots & \ddots & \vdots \\
\sigma_{m1}^2 I_n & \cdots & \sigma_{mm}^2 I_n
\end{pmatrix}
\]

\[
= \text{Tr}[B_{1nm} \Sigma_{v,nm} B_{2nm}' \Sigma_{v,nm}].
\]

When \( t = h \neq s = g \), The same method gives

\[
E[(\text{vec}(V'_{nm,t})'B_{1nm}\text{vec}(V'_{nm,s}))(\text{vec}(V'_{nm,g})'B_{2nm}\text{vec}(V'_{nm,h}))]
\]

\[
= \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sigma_{kp}^2 \sigma_{lq}^2 \text{Tr}(B_{kl,1nm} B_{pq,2nm}') = \text{Tr}[B_{1nm} \Sigma_{v,nm} B_{2nm} \Sigma_{v,nm}].
\]

**Proof of Lemma D.3**

**Proof.** \( E(U^t_{nm,t} W_{nm,t}, U^t_{nm,s} W_{nm,s}) = E_1 + E_2 \), where

\[
E_1 = \sum_{h=1}^{t-s} G_{nm,h} \text{vec}(V'_{nm,t+1-h})' \left( \sum_{h=1}^{t-s} H_{nm,h} \text{vec}(V'_{nm,t+1-h}) \right)
\]

\[
\times \left( \sum_{g=1}^{+\infty} G_{nm,g} \text{vec}(V'_{nm,s+1-g})' \left( \sum_{g=1}^{+\infty} H_{nm,g} \text{vec}(V'_{nm,s+1-g}) \right) \right)
\]

\[
= \text{Tr} \left( \sum_{h=1}^{+\infty} G_{nm,h} H_{nm,h} \Sigma_{v,nm} \right) \text{Tr} \left( \sum_{g=1}^{+\infty} G_{nm,g} H_{nm,g} \Sigma_{v,nm} \right)
\]

\[
E_2 = \sum_{g=1}^{+\infty} G_{nm,t-s+g} \text{vec}(V'_{nm,s+1-g})' \left( \sum_{g=1}^{+\infty} H_{nm,t-s+g} \text{vec}(V'_{nm,s+1-g}) \right)
\]

\[
\times \left( \sum_{g=1}^{+\infty} G_{nm,g} \text{vec}(V'_{nm,s+1-g})' \left( \sum_{g=1}^{+\infty} H_{nm,g} \text{vec}(V'_{nm,s+1-g}) \right) \right)
\]

\[
= \sum_{g=1}^{+\infty} G_{nm,s+1-g} \text{vec}(V'_{nm,s+1-g})' G_{nm,t-s+g} H_{nm,t-s+g} \text{vec}(V'_{nm,s+1-g})
\]

\[
\times \sum_{g=1}^{+\infty} G_{nm,s+1-g} \text{vec}(V'_{nm,s+1-g})' G_{nm,g} H_{nm,g} \text{vec}(V'_{nm,s+1-g})
\]

\[
+ E \left[ \sum_{g=1}^{+\infty} \sum_{h \neq g} \text{vec}(V'_{nm,s+1-h})' G_{nm,t-s+h} H_{nm,t-s+g} \text{vec}(V'_{nm,s+1-g}) \right]
\]

153
\[
\times \sum_{g=1}^{+\infty} \sum_{h \neq g} \text{vec}(V'_{nm,s+1-h})' G'_{nm,h} H_{nm,g} \text{vec}(V'_{nm,s+1-g})
\]

\[= E_{21} + E_{22} .\]

\[E_{21} = E \left[ \sum_{g=1}^{+\infty} \text{vec}(V'_{nm,s+1-g})' G'_{nm,t-s+g} H_{nm,t-s+g} \text{vec}(V'_{nm,s+1-g}) \right. \]

\[\times \text{vec}(V'_{nm,s+1-g})' G'_{nm,g} H_{nm,g} \text{vec}(V'_{nm,s+1-g}) \left. \right] \]

\[+ E \left[ \sum_{g=1}^{+\infty} \sum_{h \neq g} \text{vec}(V'_{nm,s+1-h})' G'_{nm,t-s+h} H_{nm,t-s+h} \text{vec}(V'_{nm,s+1-h}) \right. \]

\[\times \text{vec}(V'_{nm,s+1-h})' G'_{nm,g} H_{nm,g} \text{vec}(V'_{nm,s+1-g}) \left. \right] \]

\[= E_{21,a} + E_{21,b} , \]

where

\[E_{21,a} = \sum_{g=1}^{+\infty} \text{Tr}[G'_{nm,t-s+g} H_{nm,t-s+g} \Sigma_{v,nm} (G'_{nm,g} H_{nm,g} + H'_{nm,g} G_{nm,g}) \Sigma_{v,nm}] \]

\[+ \sum_{g=1}^{+\infty} \text{Tr}[G'_{nm,t-s+g} H_{nm,t-s+g} \Sigma_{v,nm}] \text{Tr}[G'_{nm,g} H_{nm,g} \Sigma_{v,nm}] \]

\[+ \sum_{g=1}^{+\infty} \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2 \sigma_{lq}^2 - \sigma_{kp}^2 \sigma_{lp}^2 - \sigma_{kl}^2 \sigma_{pq}^2) \]

\[\times \left( \sum_{i=1}^{n} [G'_{nm,t-s+g} H_{nm,t-s+g}]_{ii} [G'_{nm,g} H_{nm,g}]_{ii} \right) ; \]

\[E_{21,b} = \sum_{g=1}^{+\infty} \sum_{h \neq g} \text{Tr}[G'_{nm,t-s+h} H_{nm,t-s+h} \Sigma_{v,nm}] \text{Tr}[G'_{nm,g} H_{nm,g} \Sigma_{v,nm}] . \]

\[E_{22} = E \left[ \sum_{g=1}^{+\infty} \sum_{h \neq g} \text{vec}(V'_{nm,s+1-h})' G'_{nm,t-s+h} H_{nm,t-s+h} \text{vec}(V'_{nm,s+1-g}) \right. \]

\[\times \text{vec}(V'_{nm,s+1-h})' G'_{nm,g} H_{nm,g} \text{vec}(V'_{nm,s+1-g}) \left. \right] \]

\[+ E \left[ \sum_{g=1}^{+\infty} \sum_{h \neq g} \text{vec}(V'_{nm,s+1-h})' G'_{nm,t-s+h} H_{nm,t-s+h} \text{vec}(V'_{nm,s+1-g}) \right. \]

\[\times \text{vec}(V'_{nm,s+1-h})' G'_{nm,g} H_{nm,g} \text{vec}(V'_{nm,s+1-g}) \left. \right] \]
\[ \times \text{vec}(V_{nm,s+1-g}^t G_{nm,ng}^t H_{nm,h}^t \text{vec}(V_{nm,s+1-h}^t)) \]

\[ = E_{22,a} + E_{22,b} \]

where

\[ E_{22,a} = \sum_{g=1}^{+\infty} \sum_{h \neq g} \text{Tr}(G_{nm,t-s+h}^t H_{nm,t-s+g}^t \Sigma_v \Sigma_v G_{nm,h}^t \Sigma_v) \]

\[ \sum_{g=1}^{+\infty} \sum_{h \neq g} \text{Tr}(G_{nm,t-s+h}^t H_{nm,t-s+g}^t G_{nm,h}^t G_{nm,h}^t \Sigma_v) \]

\[ \text{Proof of Lemma D.4} \]

\[ \text{Proof.} \] (outline)

\[ \text{var} \left( \sum_{t=1}^{T} U_{nm,t}^t \bar{W}_{nm,t} \right) \]

\[ = \sum_{t=1}^{T} \sum_{s=1}^{T} \text{cov}(U_{nm,t}^t, W_{nm,t}^t, U_{nm,s}^t, W_{nm,s}^t) \]

\[ = \sum_{t=1}^{T} \sum_{s=1}^{T} \text{Tr} \left[ \sum_{g=1}^{+\infty} \sum_{h=1}^{+\infty} H_{nm,t-s+g}^t G_{nm,h}^t \Sigma_v \Sigma_v G_{nm,h}^t \Sigma_v \right] \]

\[ + \text{Tr} \left[ \sum_{g=1}^{+\infty} \sum_{h=1}^{+\infty} H_{nm,t-s+g}^t G_{nm,h}^t \Sigma_v \Sigma_v G_{nm,h}^t \Sigma_v \right] \]

\[ + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2 \sigma_{pq}^2 - \sigma_{kl}^2 \sigma_{lp}^2 - \sigma_{kp}^2 \sigma_{lp}^2) \]

\[ \times \left( \sum_{i=1}^{n} [G_{nm,t-s+g}^t H_{nm,t-s+g}^t \Sigma_v^t \Sigma_v [G_{nm,g}^t H_{nm,g}^t]]_{ii} \right) \]

\[ = E_1 + E_2 + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2 \sigma_{pq}^2 - \sigma_{kl}^2 \sigma_{lp}^2 ) E_3. \]

\[ E_1 = \sum_{t=1}^{T} \sum_{s=1}^{T} \text{Tr} \left[ \sum_{g=1}^{+\infty} \sum_{h=1}^{+\infty} H_{nm,t-s+g}^t G_{nm,h}^t \Sigma_v \Sigma_v G_{nm,h}^t \Sigma_v \right] \]

\[ = \sum_{t=1}^{T} \sum_{s=1}^{T} \text{Tr} \left[ H_{nm,c}^t H_{nm,d}^t \left( \sum_{g=1}^{+\infty} H_{nm,d}^t G_{nm,d} G_{nm,d}^t \right) H_{nm,c}^t \right] \]
\times \left\{ G_{nm,c} \left( \sum_{g=1}^{\infty} G_{nm,d}^g \Sigma_{v,nm} G_{nm,d}^g \right) G_{nm,d}^{|t-s|} G_{nm,c}^t \right\}.

Defining $H_{nm,d} = \sum_{g=1}^{\infty} H_{nm,d}^g \Sigma_{v,nm} H_{nm,d}^g$, by Lemma 34 in Yu and Lee (2008),

$$E_1 = \text{Tr} \left( \sum_{t=1}^{T} \sum_{s=1}^{T} G_{nm,d}^{|t-s|} G_{nm,c}^t \Sigma_{v,nm} G_{nm,d}^{|t-s|} H_{nm,d} \Sigma_{v,nm} G_{nm,c}^t \right) = O(nT).$$

$$|E_3| = \left| \sum_{s=1}^{T} \sum_{t=1}^{T + \infty} n \sum_{g=1}^{\infty} \text{Tr} \left[ (H_{nm,t-s+g}^t)^{kl} H_{nm,t-s+g}^t \right] \right| \leq c \sum_{s=1}^{T} \sum_{t=1}^{T + \infty} n \sum_{g=1}^{\infty} \text{Tr} \left[ (H_{nm,t-s+g}^t)^{kl} \right] \leq c \sum_{s=1}^{T} \sum_{t=1}^{T + \infty} n \sum_{g=1}^{\infty} \text{Tr} \left[ (H_{nm,t-s+g}^t)^{kl} \right] = O(nT).$$

\begin{proof}
See Yu, de Jong and Lee (2008), Lemma 29.
\end{proof}

\begin{proof}
(1) has been shown in above lemmas.

(2) The expectation

$$E \left[ \frac{1}{n} \tilde{U}'_{nm,T} \tilde{W}_{nm,T} \right] = \frac{1}{n} \sum_{h=1}^{+\infty} \text{Tr} \left[ \tilde{G}_{nm,h}^t \tilde{H}_{nm,h} \Sigma_{v,nm} \right].$$

$$= \frac{1}{n} \sum_{h=1}^{T} \text{Tr} \left[ \tilde{G}_{nm,h}^t \tilde{H}_{nm,h} \Sigma_{v,nm} \right] + \frac{1}{n} \sum_{h=T+1}^{+\infty} \text{Tr} \left[ \tilde{G}_{nm,h}^t \tilde{H}_{nm,h} \Sigma_{v,nm} \right].$$

$$= \frac{1}{n} \sum_{h=1}^{T} \text{Tr} \left[ \frac{1}{T} \sum_{g=1}^{h} G_{nm,g}^t \frac{1}{T} \sum_{g=1}^{h} H_{nm,g} \Sigma_{v,nm} \right] + \frac{1}{n} \sum_{h=T+1}^{+\infty} \text{Tr} \left[ \frac{1}{T} \sum_{g=1}^{h} G_{nm,h-T+g}^t \frac{1}{T} \sum_{g=1}^{h} H_{nm,h-T+g} \Sigma_{v,nm} \right].$$

$$= \frac{1}{n} (\tilde{E}_1 + \tilde{E}_2).$$

\end{proof}
In above expression, 
\[
\tilde{E}_1 = \sum_{h=1}^{T} \frac{1}{T^2} \text{Tr} \left[ H_{nm,c} \sum_{g=1}^{h} H_{nm,d}^g \sum_{v=1}^{h} G_{nm,d}^g G_{nm,c}^g \right],
\]

When we take the one or infinity norms,
\[
\left\| H_{nm,c} \sum_{g=1}^{h} H_{nm,d}^g \sum_{v=1}^{h} G_{nm,d}^g G_{nm,c}^g \right\| \leq \|H_{nm,c}\| \left\| \sum_{g=1}^{h} \text{abs}(H_{nm,d})^g \| \| \sum_{v=1}^{h} \text{abs}(G_{nm,d})^g \| \| G_{nm,c}^g \| \leq O(1).
\]

Therefore, 
\[
\tilde{E}_1 = O\left(\frac{n}{T}\right).
\]

When we take the one or infinity norms,
\[
\left\| H_{nm,c} \sum_{g=1}^{h} H_{nm,d}^g \sum_{v=1}^{h} G_{nm,d}^g G_{nm,c}^g \right\| \leq \|H_{nm,c}\| \left\| \sum_{g=1}^{h} \text{abs}(H_{nm,d})^g \| \| \sum_{v=1}^{h} \text{abs}(G_{nm,d})^g \| \| G_{nm,c}^g \| \leq O(1).
\]

Therefore, 
\[
\tilde{E}_2 = O\left(\frac{n}{T^2}\right) \text{ and } E\left[ \frac{1}{n} \bar{\Omega}_{nm,T} \bar{W}_{nm,T} \right] = O\left(\frac{1}{T}\right)
\]

\[
\text{var}\left(\bar{\Omega}_{nm,T} \bar{W}_{nm,T}\right) = \text{cov}(\bar{\Omega}_{nm,T} \bar{W}_{nm,T}, \bar{\Omega}_{nm,T} \bar{W}_{nm,T}) = \text{Tr} \left[ \sum_{g=1}^{\infty} \bar{H}_{nm,g} \bar{\Omega}_{nm,T} \bar{H}_{nm,g}^T \sum_{h=1}^{\infty} \bar{G}_{nm,h} \bar{\Omega}_{nm,T} \bar{G}_{nm,h}^T \right]
\]

\[
+ \text{Tr} \left[ \sum_{g=1}^{\infty} \bar{H}_{nm,g} \bar{\Omega}_{nm,T} \bar{G}_{nm,g}^T \sum_{h=1}^{\infty} \bar{H}_{nm,h} \bar{\Omega}_{nm,T} \bar{G}_{nm,h}^T \right]
\]

\[
+ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left( u_{klpq} - \sigma_{kp}^2 \sigma_{lp}^2 - \sigma_{kq}^2 \sigma_{lp}^2 - \sigma_{kl}^2 \sigma_{pq}^2 \right)
\]

\[
\times \left( \sum_{g=1}^{\infty} \left[ \sum_{i=1}^{n} \left( \sum_{j=1}^{n} \bar{G}_{nm,g}^\prime \bar{H}_{nm,g}^\prime \bar{G}_{nm,g}^\prime \bar{H}_{nm,g}^\prime \right) \right] \right).
\]
The above analysis on $\tilde{E}_1$ and $\tilde{E}_2$ can be applied here,

$$\left\| \sum_{g=1}^{T} \tilde{H}_{nm,g} \Sigma_{v,nm} \tilde{G}_{nm,g}' \right\| = O\left( \frac{1}{T} \right); \quad \left\| \sum_{g=T+1}^{+\infty} \tilde{H}_{nm,g} \Sigma_{v,nm} \tilde{G}_{nm,g}' \right\| = O\left( \frac{1}{T^2} \right).$$

$$\sum_{g=1}^{+\infty} \left( \sum_{i=1}^{n} \left[ \tilde{G}_{nm,g} \tilde{H}_{nm,g} \right]_{ii} \right)^{pq} = \sup_{p,q,i} \left( \sum_{g=1}^{+\infty} \left[ \tilde{G}_{nm,g} \tilde{H}_{nm,g} \right]_{ii} \right)^{pq} = O\left( \frac{1}{T} \right) \text{Tr} \left( \sum_{g=1}^{+\infty} \left[ \tilde{G}_{nm,g} \tilde{H}_{nm,g} \right]_{ii} \right)$$

$$= O\left( \frac{1}{T} \right) O\left( \frac{n}{T^2} \right) = O\left( \frac{n}{T^2} \right).$$

Therefore, $\frac{1}{n} \bar{U}'_{nm,T} \bar{W}'_{nm,T} - E \left[ \frac{1}{n} \bar{U}'_{nm,T} \bar{W}'_{nm,T} \right] = O_p \left( \frac{1}{\sqrt{nT^2}} \right)$.

(3) follows (1) and (2).

**Proof of Lemma D.7**

**Proof.** The strategy follows Yu and Lee (2008) Lemma 36.

**Proof of Lemma D.8**

**Proof.**

$$\bar{U}'_{nm,T} = \sum_{h=1}^{+\infty} \tilde{G}_{nm,h} \text{vec}(V'_{nm,t+1-h}),$$

where

$$\tilde{G}_{nm,h} = \begin{cases} 0 & h = 1 \\ \frac{1}{T} \sum_{g=1}^{h-1} G_{nm,g} & 2 \leq h \leq T \\ \frac{1}{T} \sum_{g=0}^{T-1} G_{nm,h-T+g} & h > T \end{cases}$$

and

$$\text{vec}(V'_{nm,T}) = \sum_{h=1}^{+\infty} \tilde{H}_{nm,h} \text{vec}(V'_{nm,t+1-h}),$$

where

$$\tilde{H}_{nm,h} = \begin{cases} \frac{1}{T} I_{nm} & 1 \leq h \leq T \\ 0 & h > T \end{cases}.$$
Therefore,
\[
E(\tilde{U}_{nm,T}^{\prime},vec(\tilde{V}_{nm,T}^{\prime})) = \sum_{h=1}^{T} Tr(\tilde{G}_{nm,h}^{\prime} \frac{1}{T} \Sigma_{v,nm}) = \sum_{h=1}^{T} \frac{1}{T^2} Tr[\Sigma_{v,nm} G_{nm,c} \sum_{g=1}^{h-1} G_{nm,d}^{g}].
\]

Since \(\|\Sigma_{v,nm} G_{nm,c} \sum_{g=1}^{h-1} G_{nm,d}^{g} \| \leq \|\Sigma_{v,nm}\| \|G_{nm,c}\| \|\sum_{g=1}^{h-1} \text{abs}(G_{nm,d})^{g} \| = O(1)\) for infinity or one norm, \(E(\tilde{U}_{nm,T}^{\prime},vec(\tilde{V}_{nm,T}^{\prime})) = O(\frac{n}{T})\). In addition,

\[
E(\tilde{U}_{nm,T}^{\prime},vec(\tilde{V}_{nm,T}^{\prime})) = \frac{1}{T^2} \sum_{h=1}^{T} \frac{1}{T} \sum_{g=1}^{\infty} G_{nm,d}^{g},
\]

\[
\frac{1}{T^2} \sum_{h=1}^{T} \frac{1}{T} \sum_{g=1}^{\infty} \sum_{h=2}^{h-1} G_{nm,d}^{g}.
\]

\[
\frac{1}{T} \sum_{h=0}^{T-1} G_{nm,d}^{h} \sum_{g=1}^{\infty} G_{nm,d}^{g}.
\]

Thus, \(\frac{1}{T^2} \sum_{h=1}^{T} \frac{1}{T} \sum_{g=1}^{\infty} G_{nm,d}^{g} = O(\frac{n}{T^2})\).

\[
\text{var}(\tilde{U}_{nm,T}^{\prime},vec(\tilde{V}_{nm,T}^{\prime})) = \text{Tr} \left[ \sum_{g=1}^{\infty} \tilde{H}_{nm,g} \Sigma_{v,nm} \tilde{H}_{nm,g}^{\prime} \sum_{h=1}^{\infty} \tilde{G}_{nm,h} \Sigma_{v,nm} \tilde{G}_{nm,h}^{\prime} \right]
\]

\[
+ \sum_{g=1}^{\infty} \tilde{H}_{nm,g} \Sigma_{v,nm} \tilde{G}_{nm,g}^{\prime} \sum_{h=1}^{\infty} \tilde{H}_{nm,h} \Sigma_{v,nm} \tilde{G}_{nm,h}^{\prime}
\]

\[
+ \left( u_{klpq} - \sigma_{kl}^{2} \sigma_{pq}^{2} - \sigma_{kl}^{2} \sigma_{pq}^{2} \right) \sum_{g=1}^{\infty} \left( \sum_{i=1}^{n} \tilde{G}_{nm,g}^{i} \tilde{H}_{nm,g}^{i} \tilde{G}_{nm,g}^{i} \tilde{H}_{nm,g}^{i} \right)
\]

where
\[
\sum_{g=1}^{\infty} \tilde{H}_{nm,g} \Sigma_{v,nm} \tilde{H}_{nm,g}^{\prime} = \frac{1}{T} \Sigma_{v,nm};
\]

\[
\sum_{g=1}^{\infty} \tilde{H}_{nm,g} \Sigma_{v,nm} \tilde{G}_{nm,g}^{\prime} = \frac{1}{T} \sum_{g=1}^{T} \Sigma_{v,nm} \tilde{G}_{nm,g}^{\prime};
\]

\[
\sum_{g=1}^{\infty} \left( \sum_{i=1}^{n} \tilde{G}_{nm,g}^{i} \tilde{H}_{nm,g}^{i} \tilde{G}_{nm,g}^{i} \tilde{H}_{nm,g}^{i} \right) = \frac{1}{T} \sum_{g=1}^{\infty} \left( \sum_{i=1}^{n} \tilde{G}_{nm,g}^{i} \tilde{H}_{nm,g}^{i} \tilde{G}_{nm,g}^{i} \tilde{H}_{nm,g}^{i} \right).
\]

Therefore,
\[
\left\| \sum_{g=1}^{\infty} \tilde{H}_{nm,g} \Sigma_{v,nm} \tilde{H}_{nm,g}^{\prime} \sum_{h=1}^{\infty} \tilde{G}_{nm,h} \Sigma_{v,nm} \tilde{G}_{nm,h}^{\prime} \right\|
\]

159
\[
\frac{1}{T} \left\| \sum_{g=1}^{T} \bar{H}_{nm,g} \sum_{v,nm} \bar{H}'_{nm,g} \sum_{h=1}^{+\infty} \sum_{v,nm} \right\| = O\left(\frac{1}{T^2}\right),
\]
\[
\left\| \sum_{g=1}^{+\infty} \bar{H}_{nm,g} \sum_{v,nm} \bar{G}'_{nm,g} \sum_{h=1}^{+\infty} \bar{H}_{nm,h} \sum_{v,nm} \bar{G}'_{nm,h} \right\| \leq \frac{1}{T^2}\left(\left\| \sum_{g=1}^{T} \sum_{v,nm} \bar{G}'_{nm,g} \right\|\right)^2,
\]
where
\[
\left\| \sum_{g=1}^{T} \sum_{v,nm} \bar{G}'_{nm,g} \right\| = \left\| \sum_{h=2}^{h-1} \sum_{v,nm} \frac{1}{T} \sum_{g=1}^{T} G'_{nm,g} \right\| = \frac{1}{T} \left\| \sum_{h=2}^{h-1} \sum_{v,nm} \sum_{g=1}^{h-1} G'_{nm,g} \right\| \leq O(1).
\]
So
\[
\left\| \sum_{g=1}^{+\infty} \bar{H}_{nm,g} \sum_{v,nm} \bar{G}'_{nm,g} \sum_{h=1}^{+\infty} \bar{H}_{nm,h} \sum_{v,nm} \bar{G}'_{nm,h} \right\| \leq O\left(\frac{1}{T^2}\right).\]
For the last term,
\[
\sum_{g=1}^{+\infty} \left( \sum_{i=1}^{n} [\bar{G}'_{nm,g} \bar{H}_{nm,g}]_{ii} [\bar{G}'_{nm,g} \bar{H}_{nm,g}]_{piq}^\top \right) \leq \frac{1}{T^2} \sup_{g,i,p,q} [\bar{G}'_{nm,g}]_{ii}^\top \text{Tr} \left[ \sum_{g=1}^{T} (\bar{G}'_{nm,g})_{ii}^\top \right] = O\left(\frac{n}{T^3}\right),
\]
since
\[
\left\| \sum_{g=2}^{T} \left[ \frac{1}{T} \sum_{h=1}^{g-1} G'_{nm,h} \right]_{kl} \right\| = \left\| \frac{1}{T} \sum_{g=2}^{T} \left[ \sum_{h=1}^{g-1} G'_{nm,h} \right]_{kl} \right\| = O(1).\]
Therefore,
\[
\text{var}(\bar{U}'_{nm,T} \text{vec}(\bar{V}'_{nm,T})) = O\left(\frac{n}{T^2}\right) + O\left(\frac{n}{T^3}\right).
\]
Appendix E: Details on First Order Derivatives and Second Order Derivatives of Quasi-Maximum Likelihood Functions for MDP-SAR

E.1 The Stable Case

Define $E_{m,ij}$ represent zero $m \times m$ matrix but the $(i, j)$th element equals to one and $F_{m,ij}$ represent zero $m \times m$ matrix but the $(i, j)$th and $(j, i)$th elements equal to one.

The first order derivative are

$$
\frac{1}{(n-1)T} \frac{\partial \ln L_{nT,m}}{\partial \text{vec}(\Pi_{km})} = \frac{1}{(n-1)T} \sum_{t=1}^{T} (\tilde{X}_{nk,t} \otimes I_m)'(J_n \otimes \Sigma_{vm}^{-1})\text{vec}(\tilde{V}_{nm,t}''(\theta)),
$$

$$
\frac{1}{(n-1)T} \frac{\partial \ln L_{nT,m}}{\partial \text{vec}(P_{m}''')} = \frac{1}{(n-1)T} \sum_{t=1}^{T} (\tilde{Y}_{nm,t-1} \otimes I_m)'(J_n \otimes \Sigma_{vm}^{-1})\text{vec}(\tilde{V}_{nm,t}''(\theta)),
$$

$$
\frac{1}{(n-1)T} \frac{\partial \ln L_{nT,m}}{\partial \Sigma_{vm,ij}} = \frac{1}{2(n-1)T} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}''(\theta))' \left[(W_n \tilde{Y}_{nm,t}''(\theta) \otimes (J_n \otimes \Sigma_{vm}^{-1}))\text{vec}(\tilde{V}_{nm,t}''(\theta))
\right.
\left. - \frac{1}{2} \text{Tr}[\Sigma_{vm}^{-1}F_{m,ij}] \right],
$$

$$
\frac{1}{(n-1)T} \frac{\partial \ln L_{nT,m}}{\partial \Phi_{m}''} = \frac{1}{(n-1)T} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}''(\theta))' \left[(W_n \tilde{V}_{nm,t}''(\theta) \otimes (J_n \otimes \Sigma_{vm}^{-1}))\text{vec}(\tilde{V}_{nm,t}''(\theta))
\right.
\left. - \frac{1}{n-1} \text{Tr}[S_{nm}^{-1}(W_n J_n \otimes E_{m,ij})] \right].
$$
When evaluated by the true value $\theta_0$, denote \( \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} = \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \theta} \). Suppose \( \bar{X}_{nk,t} = \sum_{g=0}^{+\infty} [H_{nm0}^g \Sigma_{nm0}^{-1}] \vec{v}(\Pi_{m,k0}^g) \), \( \bar{Z}_{nk,T} = 1/T \times \sum_{t=0}^{T-1} \bar{X}_{nk,t} \) and \( \bar{Z}_{nk,T} = \bar{X}_{nk,t} - \bar{Z}_{nk,T} \); \( \bar{V}_{nm,t} = \sum_{g=0}^{+\infty} \Sigma_{nm0}^{-1} H_{nm0}^g \vec{v}(V_{nm,t-g}) \), \( \bar{V}_{nm,T} = 1/T \times \sum_{t=0}^{T-1} \bar{V}_{nm,t} \) and \( \bar{V}_{nm,T} = \bar{V}_{nm,t} - \bar{V}_{nm,T} \); then the first component in the decomposition can be

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \vec{P}_{m,ij}} = \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} \frac{\partial \ln L_{nT,m0}^R}{\partial (\bar{X}_{nk,t} \otimes I_m)} (J_n \otimes \Sigma_{vm0}^{-1}) \vec{v}(V_{nm,t}),
\]

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \Sigma_{vm,ij}} = \frac{1}{2\sqrt{(n-1)T}} \sum_{t=1}^{T} \vec{v}(V_{nm,t}) (J_n \otimes (\Sigma_{vm0}^{-1} F_{m,ij} \Sigma_{vm0}^{-1})) \vec{v}(V_{nm,t}),
\]

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \Phi_{m,ij}} = \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} \vec{v}(V_{nm,t}) (J_n \otimes (\Sigma_{vm0}^{-1} F_{m,ij} \Sigma_{vm0}^{-1})) \vec{v}(V_{nm,t}),
\]

\[
- \frac{\sqrt{T}}{n-1} \text{Tr} \left[ S_{nm}^{-1} (W_n J_n \otimes E_{m,ij}^t) \right] + \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} \vec{v}(V_{nm,t}) (S_{nm0}^{-1} (W_n J_n \otimes E_{m,ij}^t)) \vec{v}(V_{nm,t}),
\]

\[
+ (J_n \otimes P_{m0}^t + W_n \otimes \Phi_{m0}) \bar{Z}_{nk,t-1} \Sigma_{nm0}^{-1} (W_n J_n \otimes (E_{m,ij}^t \Sigma_{vm0}^{-1})) \vec{v}(V_{nm,t})
\]

\[
+ \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} \vec{v}(V_{nm,t}) (J_n \otimes P_{m0}^t + W_n \otimes \Phi_{m0}) \Sigma_{nm0}^{-1} (W_n J_n \otimes (E_{m,ij}^t \Sigma_{vm0}^{-1})) \vec{v}(V_{nm,t})
\]

\[
\times (W_n J_n \otimes (E_{m,ij}^t \Sigma_{vm0}^{-1})) \vec{v}(V_{nm,t}).
\]

And the residuals are

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \vec{P}_{m,ij}} = 0,
\]

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \Phi_{m,ij}} = - \frac{\sqrt{T}}{n-1} \bar{V}_{nm,T} (J_n \otimes (E_{m,ij}^t \Sigma_{vm0}^{-1})) \vec{v}(V_{nm,T}),
\]
Specifically, all the residual terms can be represented as

\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \Phi_{m,ij}} = - \sqrt{\frac{T}{n-1}} \tilde{\psi}_{nm,T}' [(W_n' J_n) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})] \text{vec}(\tilde{V}_{nm,T})' \\
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \Sigma_{vm,ij}} = - \frac{1}{2} \sqrt{\frac{T}{n-1}} \text{vec}(\tilde{V}_{nm,T}') [J_n \otimes (\Sigma_{vm0}^{-1} F_{m,ij} \Sigma_{vm0}^{-1})] \text{vec}(\tilde{V}_{nm,T})' \\
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \Psi_{m,ij}} = - \sqrt{\frac{T}{n-1}} \tilde{\psi}_{nm,T}' (I_n \otimes P_{m0} + W_n \otimes \Phi_{m0}') S_{nm0}' \\
\times [(W_n' J_n) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})] \text{vec}(\tilde{V}_{nm,T})' - \sqrt{\frac{T}{n-1}} \text{vec}(\tilde{V}_{nm,T})' S_{nm0}^{-1} [(W_n' J_n) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})] \text{vec}(\tilde{V}_{nm,T})'.
\]

The second order derivatives at any possible \( \theta \) are

\[
\frac{\partial^2 \ln L_{nT,m}}{\partial \text{vec}(\Pi_{km}') \partial \text{vec}(\Pi_{km}')} = - \sum_{t=1}^{T} (\check{X}_{nk,tl} J_n \check{X}_{nk,t}) \otimes \Sigma_{vm}^{-1},
\]

163
\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \text{vec}(\Pi'_{km}) \partial P_{m,ij}} = - \sum_{t=1}^{T} \left[ (\tilde{X}_{nk,t} J_n) \otimes (\Sigma_{vm}^{-1} F_{m,ij}) \right] \text{vec}(\tilde{Y}_{nm,t-1}),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \text{vec}(\Pi'_{km}) \partial \Phi_{m,ij}} = - \sum_{t=1}^{T} \left[ (\tilde{X}_{nk,t} J_n W_n) \otimes (\Sigma_{vm}^{-1} E'_{m,ij}) \right] \text{vec}(\tilde{Y}_{nm,t-1}),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \text{vec}(\Pi'_{km}) \partial \Sigma_{vm,ij}} = - \sum_{t=1}^{T} \left[ (\tilde{X}_{nk,t} J_n) \otimes (\Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1}) \right] \text{vec}(\tilde{V}_{nm,t}(\theta)),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \text{vec}(\Pi'_{km}) \partial \Psi_{m,ij}} = - \sum_{t=1}^{T} \left[ (\tilde{X}_{nk,t} J_n W_n) \otimes (\Sigma_{vm}^{-1} E'_{m,ij}) \right] \text{vec}(\tilde{Y}_{nm,t}),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial P_{m,ij} \partial P_{m,kl}} = - \sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1})' [J_n \otimes (E_{m,ij} \Sigma_{vm}^{-1} E'_{m,kl})] \text{vec}(\tilde{Y}_{nm,t-1}),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial P_{m,ij} \partial \Phi_{m,kl}} = - \sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1})' [J_n W_n \otimes (E_{m,ij} \Sigma_{vm}^{-1} E'_{m,kl})] \text{vec}(\tilde{Y}_{nm,t-1}),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial P_{m,ij} \partial \Sigma_{vm,kl}} = - \sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1})' [J_n \otimes (E_{m,ij} \Sigma_{vm}^{-1} F'_{m,kl} \Sigma_{vm}^{-1})] \text{vec}(\tilde{Y}_{nm,t-1}),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \Phi_{m,ij} \partial \Phi_{m,kl}} = - \sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1})' [J_n W_n \otimes (E_{m,ij} \Sigma_{vm}^{-1} E'_{m,kl})] \text{vec}(\tilde{Y}_{nm,t-1}),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \Phi_{m,ij} \partial \Sigma_{vm,kl}} = - \sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1})' [W_n' J_n W_n \otimes (E_{m,ij} \Sigma_{vm}^{-1} E'_{m,kl})] \text{vec}(\tilde{Y}_{nm,t}),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \Sigma_{vm,ij} \partial \Sigma_{vm,kl}} = - \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}(\theta))' [J_n \otimes (\Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1} F'_{m,kl} \Sigma_{vm}^{-1})] \text{vec}(\tilde{V}_{nm,t}(\theta))
\]

\[
+ \frac{(n-1)T}{2} \text{Tr}[\Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1} F_{m,kl}],
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \Sigma_{vm,ij} \partial \Psi_{m,kl}} = - \sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t})' [W_n' J_n \otimes (E_{m,kl} \Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1})] \text{vec}(\tilde{V}_{nm,t}(\theta)),
\]

\[
\frac{\partial^2 \ln L_{\nu T,m}}{\partial \Sigma_{vm,ij} \partial \Psi_{m,kl}} = - \sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t})' [(W_n' J_n W_n) \otimes (E_{m,ij} \Sigma_{vm}^{-1} E'_{m,kl} \Sigma_{vm}^{-1})] \text{vec}(\tilde{Y}_{nm,t})
\]

\[= \left( T \text{Tr}[S_{nm}^{-1}(E_{m,ij} \otimes W_{n} J_{n})] S_{nm}^{-1}(E_{m,kl} \otimes W_{n} J_{n}) \right). \]
Details of Information Matrix

The information matrix, \( \Omega_{0,nT} \equiv -\frac{1}{(n-1)T} E \left( \frac{\partial^2 \ln L_{nT,00}}{\partial \theta^2} \right) \), can be decomposed as

\[
\begin{pmatrix}
\Omega_{\Pi \Pi} & \Omega_{\Pi P} & \Omega_{\Pi \Phi} & \Omega_{\Pi \Psi} & \Omega_{\Pi \Sigma} \\
* & \Omega_{P P} & \Omega_{P \Phi} & \Omega_{P \Psi} & \Omega_{P \Sigma} \\
* & * & \Omega_{\Phi \Phi} & \Omega_{\Phi \Psi} & \Omega_{\Phi \Sigma} \\
* & * & * & \Omega_{\Psi \Psi} & \Omega_{\Psi \Sigma} \\
* & * & * & * & \Omega_{\Sigma \Sigma}
\end{pmatrix}
\]

which is a \([mk + 3m^2 + m(m+1)/2] \times [mk + 3m^2 + m(m+1)/2]\) matrix. For each element, \( \Omega_\cdot = \Omega_{\cdot soc} + \Omega_{\cdot soc,R} \), where \( \Omega_{\cdot soc,R} \) is \( O(1/T) \).

1. **Block \( \Omega_{\Pi \Pi} \)**
   
   \[\Omega_{\Pi \Pi}^{soc} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E(\tilde{X}'_{nk,t} J_n \tilde{X}_{nk,t}) \otimes \Sigma^{-1}_{vn0},\]
   
   \[\Omega_{\Pi \Pi}^{soc,R} = 0.\]

2. **Block \( \Omega_{\Pi P} \)**

   The \( \cdot \times (i + (j-1)m) \) elements are
   
   \[\Omega_{\Pi P, \cdot \times (i + (j-1)m)}^{soc} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E\{[(\tilde{X}'_{nk,t} J_n) \otimes (\Sigma_{vn0}^{-1} E_{m,ij}')] \text{vec}(\tilde{Y}'_{nm,t-1})\},\]
   
   \[\Omega_{\Pi P, \cdot \times (i + (j-1)m)}^{soc,R} = 0.\]

3. **Block \( \Omega_{\Pi \Phi} \)**

   The \( \cdot \times (i + (j-1)m) \) elements are
   
   \[\Omega_{\Pi \Phi, \cdot \times (i + (j-1)m)}^{soc} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E\{[(\tilde{X}'_{nk,t} J_n W_n) \otimes (\Sigma_{vn0}^{-1} E_{m,ij}')] \text{vec}(\tilde{Y}'_{nm,t-1})\},\]
   
   \[\Omega_{\Pi \Phi, \cdot \times (i + (j-1)m)}^{soc,R} = 0.\]

4. **Block \( \Omega_{\Pi \Psi} \)**
The \( \cdot \times (i + (j - 1)m) \) elements are

\[
\Omega_{\text{soc}, \times (i + (j - 1)m)}^{\text{P}} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E\{[(\hat{X}'_{nk,t} J_n W_n) \otimes (\Sigma_{vm0}^{-1} E'_{m,ij})] S_{nm0}^{-1} \}
\]

\[
\Omega_{\text{soc}, R, \times (i + (j - 1)m)}^{\text{P}} = -\frac{1}{(n-1)} E\{[(\hat{X}'_{nk,t} J_n W_n) \otimes (\Sigma_{vm0}^{-1} E'_{m,ij})] S_{nm0}^{-1} \}
\]

(5) Block \( \Omega_{\Pi} \)

The \( \cdot \times (i + (j - 1)m) \) elements are

\[
\Omega_{\Pi, \times (i + (j - 1)m)}^{\text{soc}, R} = 0,
\]

\[
\Omega_{\Pi, \times (i + (j - 1)m)}^{\text{soc}, R} = -\frac{1}{(n-1)} E\{[(\hat{X}'_{nk,t} J_n W_n) \otimes (\Sigma_{vm0}^{-1} E'_{m,ij})] vec(\tilde{\nu}'_{nm,t-1}) \}
\]

(6) Block \( \Omega_F \)

The \( (i + (j - 1)m) \times (k + (l - 1)m) \) element is

\[
\Omega_{\Phi, (i + (j - 1)m) \times (k + (l - 1)m)}^{\text{soc}} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E\{vec(\tilde{Y}'_{nm,t-1}) [J_n \otimes \Sigma_{vm0}^{-1} E'_{m,ij}] vec(\tilde{Y}'_{nm,t-1}) \}
\]

\[
\Omega_{\Phi, (i + (j - 1)m) \times (k + (l - 1)m)}^{\text{soc}, R} = 0.
\]

(7) Block \( \Omega_{\Phi} \)

The \( (i + (j - 1)m) \times (k + (l - 1)m) \) element is

\[
\Omega_{\Phi, (i + (j - 1)m) \times (k + (l - 1)m)}^{\text{soc}, R} =
\]

\[
\frac{1}{(n-1)T} \sum_{t=1}^{T} E\{vec(\tilde{Y}'_{nm,t-1}) [J_n \otimes \Sigma_{vm0}^{-1} E'_{m,ij}] vec(\tilde{Y}'_{nm,t-1}) \}
\]

\[
\Omega_{\Phi, (i + (j - 1)m) \times (k + (l - 1)m)}^{\text{soc}, R} = 0.
\]

(8) Block \( \Omega_{\Phi} \)

The \( (i + (j - 1)m) \times (k + (l - 1)m) \) element is

\[
\Omega_{\Phi, (i + (j - 1)m) \times (k + (l - 1)m)}^{\text{soc}} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E\{vec(\tilde{Y}'_{nm,t-1}) [J_n \otimes \Sigma_{vm0}^{-1} E'_{m,ij}] S_{nm0}^{-1} \}
\]

\[
[(I_n \otimes P'_{m0} + W_n \otimes \Phi'_{m0}) vec(\tilde{Y}'_{nm,t-1})] + (\tilde{Y}'_{nm,t-1} \otimes I_m) vec(\tilde{\nu}'_{nm0}) \}
\]

\[
\Omega_{\Phi, (i + (j - 1)m) \times (k + (l - 1)m)}^{\text{soc}, R} = \frac{1}{n-1} E\{vec(\tilde{Y}'_{nm,t-1}) [J_n \otimes \Sigma_{vm0}^{-1} E'_{m,ij}] S_{nm0}^{-1} vec(\tilde{\nu}'_{nm,t-1}) \}
\]
(9) Block $\Omega_{P\Sigma}$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Omega_{P\Sigma,(i+(j-1)m)\times(k+(l-1)m)} = 0,$$

$$\Omega_{P\Sigma,(i+(j-1)m)\times(k+(l-1)m)}^{soc,R} = -\frac{1}{(n-1)} E\{vec(\bar{\bar{Y}}_{nm,T})'[(J_n) \otimes (E_{m,ij}\Sigma_{vm0}^{-1}F_{m,kl}\Sigma_{vm0}^{-1})]vec(\bar{V}_{nm,T}').\}.$$

(10) Block $\Omega_{\Phi\Phi}$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Omega_{\Phi\Phi,(i+(j-1)m)\times(k+(l-1)m)}^{soc} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E\{vec(\bar{\bar{Y}}_{nm,t-1})'[W_n'J_nW_n \otimes (E_{m,ij}\Sigma_{vm}^{-1}E_{m,kl}')]vec(\bar{Y}_{nm,t-1})',\}.$$

$$\Omega_{\Phi\Phi,(i+(j-1)m)\times(k+(l-1)m)}^{soc,R} = 0.$$

(11) Block $\Omega_{\Phi\Psi}$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Omega_{\Phi\Psi,(i+(j-1)m)\times(k+(l-1)m)}^{soc} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E\{vec(\bar{\bar{Y}}_{nm,t-1})'[W_n'J_nW_n \otimes (E_{m,ij}\Sigma_{vm}^{-1}E_{m,kl}')]S_{nm0}^{-1}\}.$$

$$\Omega_{\Phi\Psi,(i+(j-1)m)\times(k+(l-1)m)}^{soc,R} = \frac{1}{n-1} E\{vec(\bar{\bar{Y}}_{nm,T})'[W_n'J_nW_n \otimes (E_{m,ij}\Sigma_{vm}^{-1}E_{m,kl}')]S_{nm0}^{-1}vec(\bar{V}_{nm,T}').\}.$$

(12) Block $\Omega_{\Psi\Sigma}$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Omega_{\Psi\Sigma,(i+(j-1)m)\times(k+(l-1)m)}^{soc} = 0,$$

$$\Omega_{\Psi\Sigma,(i+(j-1)m)\times(k+(l-1)m)}^{soc,R} = -\frac{1}{(n-1)} E\{vec(\bar{\bar{Y}}_{nm,T})'[W_n'J_n] \otimes (E_{m,ij}\Sigma_{vm0}^{-1}F_{m,kl}\Sigma_{vm0}^{-1})]vec(\bar{V}_{nm,T}').\}.$$

(13) Block $\Omega_{\Psi\Psi}$

167
The \((i + (j - 1)m) \times (k + (l - 1)m)\) element is

\[
\Omega_{\Psi_{i+(j-1)m} \times(k+(l-1)m)}^{soc} = \frac{1}{(n-1)T} \sum_{t=1}^{T} E \{ [(I_n \otimes P_{m0} + W_n \otimes \Phi'_{m0}) vec(\tilde{Y}_{nm,t-1}) + (\tilde{X}_{nk,t} \otimes I_m) vec(\Pi'_{km0})]' S_{nm0}^{-1} \}
\]

\[
[W''_n J_n W_n \otimes (E_{m,ij} \Sigma_{vm0}^{-1} E_{m,kl})] S_{nm0}^{-1} 1 \}
\]

\[
\Omega_{\Psi_{i+(j-1)m} \times(k+(l-1)m)}^{soc,R} = -\frac{1}{(n-1)T} \sum_{t=1}^{T} E \{ [(I_n \otimes P_{m0} + W_n \otimes \Phi'_{m0}) vec(\tilde{Y}_{nm,t-1}) + (I_m \otimes \tilde{X}_{nk,t}) vec(\Pi'_{km0})]' S_{nm0}^{-1} \}
\]

\[
[W''_n J_n W_n \otimes (E_{m,ij} \Sigma_{vm0}^{-1} E_{m,kl})] S_{nm0}^{-1} 1 \}
\]

\[
\Omega_{\Psi_{i+(j-1)m} \times(k+(l-1)m)}^{soc,R} = -\frac{1}{(n-1)T} \sum_{t=1}^{T} E \{ [(I_n \otimes P_{m0} + W_n \otimes \Phi'_{m0}) vec(\tilde{Y}_{nm,t-1}) + (I_m \otimes \tilde{X}_{nk,t}) vec(\Pi'_{km0})]' S_{nm0}^{-1} \}
\]

\[
[(E_{m,kl} \Sigma_{vm0}^{-1} E_{m,ij}) \otimes W''_n J_n W_n] S_{nm0}^{-1} 1 \}
\]

\[\text{(14) Block } \Omega_{\Psi_{i+(j-1)m} \times(k+(l-1)m)}\]

The \((i + (j - 1)m) \times (k + (l - 1)m)\) element is

\[
\Omega_{\Psi_{i+(j-1)m} \times(k+(l-1)m)}^{soc} = \frac{1}{n-1} \sum_{t=1}^{T} E \{ [(I_n \otimes P_{m0} + W_n \otimes \Phi'_{m0}) vec(\tilde{Y}_{nm,t-1}) + (I_m \otimes \tilde{X}_{nk,t}) vec(\Pi'_{km0})]' S_{nm0}^{-1} \}
\]

\[
[W''_n J_n W_n \otimes (E_{m,ij} \Sigma_{vm0}^{-1} E_{m,kl})] S_{nm0}^{-1} 1 \}
\]

\[
[(E_{m,kl} \Sigma_{vm0}^{-1} E_{m,ij}) \otimes W''_n J_n W_n] S_{nm0}^{-1} 1 \}
\]

\[\text{(15) Block } \Omega_{\Sigma_{i+(j-1)m} \times(k+(l-1)m)}\]

The \((i + (j - 1)m) \times (k + (l - 1)m)\) element is

\[
\Omega_{\Sigma_{i+(j-1)m} \times(k+(l-1)m)}^{soc} = \frac{1}{2} \sum_{t=1}^{T} E \{ [(I_n \otimes P_{m0} + W_n \otimes \Phi'_{m0}) vec(\tilde{Y}_{nm,t-1}) + (I_m \otimes \tilde{X}_{nk,t}) vec(\Pi'_{km0})]' S_{nm0}^{-1} \}
\]

\[
[W''_n J_n W_n \otimes (E_{m,ij} \Sigma_{vm0}^{-1} E_{m,kl})] S_{nm0}^{-1} 1 \}
\]

\[
[(E_{m,kl} \Sigma_{vm0}^{-1} E_{m,ij}) \otimes W''_n J_n W_n] S_{nm0}^{-1} 1 \}
\]

168
According to Lemma D.6, \( \Omega^{\text{soc},R}_\theta = O(1/T) \).

Now look at the variance of first order conditions.

\[
\frac{1}{(n - 1)T} E \left[ \frac{\partial \ln L_{nT,m0}}{\partial \theta} \frac{\partial \ln L_{nT,m0}}{\partial \theta'} \right] = \Omega^{\text{soc}}_\theta + \Xi_{\theta_0,nT} + \Omega^{\text{loc},R}_\theta.
\]

The first term \( \Omega^{\text{soc}}_\theta \) has been investigated in last part. If we decompose the term \( \Xi_{\theta_0,nT} \) as the same way as that of \( \Omega_{\theta_0,nT} \), then \( \Xi_{\text{III}} = \Xi_{\Pi P} = \Xi_{\Pi \Phi} = \Xi_{\Phi P} = \Xi_{P \Psi} = \Xi_{\Psi P} = 0 \). The other terms are listed below.

(1) Block \( \Xi_{\Pi \Psi} \)

The \(( (j - 1)k + i ) \times (k + (l - 1)m) \) element is

\[
\Xi_{\Pi \Psi,((j-1)k+i) \times (k+(l-1)m)} = \frac{1}{(n - 1)T} \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} u_{spq} \sum_{h=1}^{n} [\lambda_{t, \Pi \Psi}]^{(j-1)k+i} [\Lambda \Psi]^{kl}_{h,h,pq},
\]

where \([\lambda_{t, \Pi \Psi}]^{(j-1)k+i} \) represents the \( s \)-th vector in the \((j-1)k+i \)-th row of \( E((X_{nk,t} J_n) \otimes \Sigma_{vm_0}) \) with length \( 1 \times n \) and \([\Lambda \Psi]^{kl}_{pq} \) is the \( n \times n \) submatrix with index \((p, q)\) of \( mn \times mn \) matrix \( S^{-1}_{nm_0} \left( (W_n' J_n) \otimes (E_{m,kl} \Sigma^{-1}_{vm_0}) \right) \).

(2) Block \( \Xi_{P \Psi} \)

The \(( (i + (j - 1)m) \times (k + (l - 1)m) \) element is

\[
\Xi_{P \Psi,(i+(j-1)m) \times (k+(l-1)m)} = \frac{1}{(n - 1)T} \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} u_{spq} \sum_{h=1}^{n} [\lambda_{t, \Pi \Psi}]^{ij} [\Lambda \Psi]^{kl}_{h,h,pq},
\]

where \([\lambda_{t, \Pi \Psi}]^{ij} \) represents the \( s \)-th vector in the \( i \)-th row of \( E\{ [\tilde{X}_{nk,t-1} + \Psi_{nm,t-1}]' (J_n \otimes (E_{m,ij} \Sigma^{-1}_{vm_0})) \} \) with length \( 1 \times n \) and \([\Lambda \Psi]^{kl}_{pq} \) is the \( n \times n \) submatrix with index \((p, q)\) of \( mn \times mn \) matrix \( S^{-1}_{nm_0} \left( (W_n' J_n) \otimes (E_{m,kl} \Sigma^{-1}_{vm_0}) \right) \).

(3) Block \( \Xi_{\Phi \Psi} \)

The \(( i + (j - 1)m) \times (k + (l - 1)m) \) element is

\[
\Xi_{\Phi \Psi,(i+(j-1)m) \times (k+(l-1)m)} = \frac{1}{(n - 1)T} \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} u_{spq} \sum_{h=1}^{n} [\lambda_{t, \Pi \Psi}^{ij}] [\Lambda \Psi]^{kl}_{h,h,pq},
\]

where \([\lambda_{t, \Pi \Psi}]^{ij} \) represents the \( s \)-th vector in the \( i \)-th row of \( E\{ [\tilde{X}_{nk,t-1} + \Psi_{nm,t-1}]' (W_n' J_n) \otimes (E_{m,ij} \Sigma^{-1}_{vm_0})) \} \) with length \( 1 \times n \) and \([\Lambda \Psi]^{kl}_{pq} \) is the \( n \times n \) submatrix with index \((p, q)\) of \( mn \times mn \) matrix \( S^{-1}_{nm_0} \left( (W_n' J_n) \otimes (E_{m,kl} \Sigma^{-1}_{vm_0}) \right) \).
(4) Block $\Xi_{\Pi^\Sigma}$

The $((j - 1)k + i) \times (k + (l - 1)m)$ element is

$$\Xi_{\Pi^\Sigma,(j-1)k+i}\times(k+(l-1)m) = \frac{1}{(n - 1)T} \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} \left[ u_{spq} \sum_{h=1}^{n} [\lambda_{t,\Pi^\Sigma}]_{h,s}^{(j-1)k+i} [\Lambda_{\Pi^\Sigma}]_{h,h,pq}^{kl} \right],$$

where $[\lambda_{t,\Pi^\Sigma}]^{(j-1)k+i}_{s}$ represents the $s$th vector in the $(j-1)k+i$th row of $E((\tilde{X}_{nk,t}J_n) \otimes \Sigma^{-1}_{vm0})$ with length $1 \times n$ and $[\Lambda_{\Pi^\Sigma}]_{h,h,pq}^{kl}$ is the $n \times n$ submatrix with index $(p, q)$ of $mn \times mn$ matrix $1/2 J_n \otimes (\Sigma_{vm0}^{-1} F_{kl}^{1/2} \Sigma_{vm0}^{-1}).$

(5) Block $\Xi_{P^\Sigma}$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Xi_{P^\Sigma,(i+j-1)m}\times(k+(l-1)m) = \frac{1}{(n - 1)T} \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} \left[ u_{spq} \sum_{h=1}^{n} [\lambda_{t,P^\Sigma}]_{h,s}^{ij} [\Lambda_{P^\Sigma}]_{h,k,h,pq}^{kl} \right],$$

where $[\lambda_{t,P^\Sigma}]_{s}^{ij}$ represents the $s$th vector in the row of $E\{[\tilde{X}_{nk,t-1} + \Sigma_{nm,t-1}^{-1}] (J_n \otimes (E_{m,ij}^{-1} \Sigma_{vm0}))\}$ with length $1 \times n$ and $[\Lambda_{P^\Sigma}]_{h,k,h,pq}^{kl}$ is the $n \times n$ submatrix with index $(p, q)$ of $mn \times mn$ matrix $1/2 (J_n \otimes \Sigma_{vm0}^{-1} F_{kl} \Sigma_{vm0}^{-1}).$

(6) Block $\Xi_{\Phi^\Sigma}$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Xi_{\Phi^\Sigma,(i+j-1)m}\times(k+(l-1)m) = \frac{1}{(n - 1)T} \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} \left[ u_{spq} \sum_{h=1}^{n} [\lambda_{t,\Phi^\Sigma}]_{h,s}^{ij} [\Lambda_{\Phi^\Sigma}]_{h,k,h,pq}^{kl} \right],$$

where $[\lambda_{t,\Phi^\Sigma}]_{s}^{ij}$ represents the $s$th vector in the row of $E\{[\tilde{X}_{nk,t-1} + \Sigma_{nm,t-1}^{-1}] ((W_{n}^{T} J_n) \otimes (E_{m,ij}^{-1} \Sigma_{vm0}))\}$ with length $1 \times n$ and $[\Lambda_{\Phi^\Sigma}]_{h,k,h,pq}^{kl}$ is the $n \times n$ submatrix with index $(p, q)$ of $mn \times mn$ matrix $1/2 (J_n \otimes \Sigma_{vm0}^{-1} F_{kl} \Sigma_{vm0}^{-1}).$

(7) Block $\Xi_{\Phi^\Psi}$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Xi_{\Phi^\Psi,(i+j-1)m}\times(k+(l-1)m) = \frac{1}{(n - 1)T} \sum_{s=1}^{m} \sum_{r=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} \left[ u_{spqr} - \sigma_{sp}^{2} \sigma_{rq}^{2} - \sigma_{sq}^{2} \sigma_{rp}^{2} - \sigma_{sr}^{2} \sigma_{pq}^{2} \right]$$

$$+ \frac{2}{(n - 1)T} \sum_{s=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} \left[ u_{spq} \sum_{h=1}^{n} [\lambda_{t,\Phi^\Psi}]_{h,s}^{ij} [\Lambda_{\Phi^\Psi}]_{h,k,h,pq}^{kl} \right],$$

170
where $[\lambda_t, \psi_s]_{ij}^s$ represents the s\textsuperscript{th} vector in the the row of $E\{[(\tilde{X}_{nk,t} \otimes I_m) vec(\Pi'_{km0}) + (I_n \otimes P'_{m0} + W_n \otimes \Phi'_{m0})\tilde{X}_{nk,t}^{-1}]S_{nm0}^{t-1}[(W'_nJ_n) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\}$ with length $1 \times n$ and $[\Lambda \Psi_{ij}]_{pq}^s$ is the $n \times n$ submatrix with index $(p, q)$ of $mn \times mn$ matrix $S_{nm0}^{t-1}[(W'_nJ_n) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]$.

(8) Block $\Xi_{\Psi}\Sigma$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Xi_{\Psi,\Sigma, (i+(j-1)m) \times (k+(l-1)m)} = \frac{1}{(n-1)T} \sum_{s=1}^{m} \sum_{r=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ (u_{srpq} - \sigma_{sp}^2 \sigma_{rq}^2 - \sigma_{sq}^2 \sigma_{rp}^2 - \sigma_{sr}^2 \sigma_{pq}^2) \right]$$

$$\sum_{t=1}^{T} \sum_{h=1}^{n} [\Lambda_{\psi_s}\Sigma_{ij}]_{hh,sr} [\Lambda_{\psi_s}\Sigma_{ij}]_{hh,pq}$$

$$+ \frac{1}{(n-1)T} \sum_{s=1}^{m} \sum_{r=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \sum_{t=1}^{T} \sum_{h=1}^{n} [u_{srpq} \sum_{q=1}^{n} [\Lambda_{\psi_s}\Sigma_{ij}]_{hh,s} [\Lambda_{\psi_s}\Sigma_{ij}]_{hh,pq}]$$

where $[\lambda_t, \psi_s]_{ij}^s$ represents the s\textsuperscript{th} vector in the the row of $E\{[(\tilde{X}_{nk,t} \otimes I_m) vec(\Pi'_{km0}) + (I_n \otimes P'_{m0} + W_n \otimes \Phi'_{m0})\tilde{X}_{nk,t}^{-1}]S_{nm0}^{t-1}[(W'_nJ_n) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\}$ with length $1 \times n$, $[\Lambda_{\psi_s}\Sigma_{ij}]_{pq}^s$ is the $n \times n$ submatrix with index $(p, q)$ of $mn \times mn$ matrix $S_{nm0}^{t-1}[(E_{m,ij} \Sigma_{vm0}^{-1}) \otimes (W'_nJ_n)]$ and $[\Lambda_{\psi_s}\Sigma_{ij}]_{pq}^s$ is the $n \times n$ submatrix with index $(p, q)$ of $mn \times mn$ matrix $1/2J_n \otimes (\Sigma_{vm0}^{-1}F_{kl} \Sigma_{vm0}^{-1})$.

(9) Block $\Xi_{\Sigma}\Sigma$

The $(i + (j - 1)m) \times (k + (l - 1)m)$ element is

$$\Xi_{\Sigma,\Sigma, (i+(j-1)m) \times (k+(l-1)m)} = \frac{1}{(n-1)T} \sum_{s=1}^{m} \sum_{r=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ (u_{srpq} - \sigma_{sp}^2 \sigma_{rq}^2 - \sigma_{sq}^2 \sigma_{rp}^2 - \sigma_{sr}^2 \sigma_{pq}^2) \right]$$

$$\sum_{t=1}^{T} \sum_{h=1}^{n} [\Lambda_{\Sigma_{ij}}]_{hh,sr} [\Lambda_{\Sigma_{ij}}]_{hh,pq}$$

where $[\Lambda_{\Sigma_{ij}}]_{pq}^s$ is the $n \times n$ submatrix with index $(p, q)$ of $mn \times mn$ matrix $J_n \otimes (\Sigma_{vm0}^{-1}F_{kl} \Sigma_{vm0}^{-1})$.

The difference between

$$\frac{1}{(n-1)T} E \left[ \frac{\partial \ln L_{n,T,m0}}{\partial \theta} \frac{\partial \ln L_{n,T,m0}}{\partial \theta} \right]$$

and $\Omega_{\theta}^{soc} + \Xi_{\theta, nT}$ is $\Omega_{\theta}^{foc,R}$, which has the form

$$- \frac{1}{n-1} \tilde{U}_{nm,T} \tilde{W}_{nm,T}$$
According to Lemma D.6, these terms are $O(1/T)$, therefore,

$$
\frac{1}{(n-1)T} E \left[ \frac{\partial \ln L_{nt,m0}}{\partial \theta} \frac{\partial \ln L_{nt,m0}}{\partial \theta'} \right] = \Omega_{n,T}^{\text{soc}} + \Xi_{\theta_0,nT} + O \left( \frac{1}{T} \right) = \Omega_{\theta_0,nT} + \Xi_{\theta_0,nT} + O \left( \frac{1}{T} \right),
$$

where the last equality is due to $\Omega_{\theta_0,nT}^{\text{soc},R} = O(1/T)$.

### E.2 The Spatial Cointegration Case

The first order conditions are

$$
\frac{1}{(n-1)T} \frac{\partial \ln L_{nt,m}}{\partial \text{vec}(\Pi_{km})} = \frac{1}{(n-1)T} \sum_{t=1}^{T} (\tilde{X}_{nk,t} \otimes I_m)'(J_n \otimes \Sigma_{vm}^{-1}) \text{vec} (\tilde{V}'_{nm,t}(\theta)),
$$

$$
\frac{1}{(n-1)T} \frac{\partial \ln L_{nt,m}}{\partial \text{vec}(P_{m}')} = \frac{1}{(n-1)T} \sum_{t=1}^{T} (\tilde{Y}_{nm,t-1} \otimes I_m)'(J_n \otimes \Sigma_{vm}^{-1}) \text{vec} (\tilde{V}'_{nm,t}(\theta)),
$$

$$
\frac{1}{(n-1)T} \frac{\partial \ln L_{nt,m}}{\partial \Sigma_{vm,ij}} = \frac{1}{(n-1)T} \sum_{t=1}^{T} \text{vec}(\tilde{V}'_{nm,t}(\theta))'[J_n \otimes (\Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1})] \text{vec}(\tilde{V}'_{nm,t}(\theta))

- \frac{1}{2} \text{Tr}[\Sigma_{vm}^{-1} F_{m,ij}],
$$

$$
\frac{1}{(n-1)T} \frac{\partial \ln L_{nt,m}}{\partial \Psi_{m,ij}} = \frac{1}{(n-1)T} \sum_{t=1}^{T} \left[ \text{vec}(\tilde{V}'_{nm,t}(\theta))' (W'_n \otimes I_m) - \text{vec}(\tilde{V}'_{nm,t-1})' \right]

\times \left[ J_n \otimes (E_{m,ij} \Sigma_{vm}^{-1}) \right] \text{vec}(\tilde{V}'_{nm,t}(\theta)) - \frac{1}{n-1} \text{Tr}[\Sigma_{vm}^{-1} (E_{m,ij} \otimes W_n J_n)].
$$

Recall that $J_n \otimes I_m \text{vec}(Y'_{nm,t}) = J_n \otimes I_m \text{vec}(u)(Y'_{nm,t}) + J_n \otimes I_m \text{vec}(u)(Y'_{nm,t})$. Define

$$W_{nm}^u = \Gamma_n \Gamma_{n,n}^{-1} \otimes I_m \quad \text{and} \quad U_0 = (\Gamma_n \Gamma_{n,n}^{-1} \otimes [I_m - \Psi_m')^{-1} (P_m + \Phi_m')^{-1}) \text{vec}(Y'_{nm,-1}),$$

we have $\text{vec}(u)(Y'_{nm,t}) = U_0 + W_{nm}^u X_{nm,t} + W_{nm}^u C_{nm,t} + W_{nm}^u V_{nm,t}$. Suppose $X_{nk,t}^{(s)} = \sum_{g=0}^{+\infty} B_{nm0}^g S_{nm0}^{-1} ((X_{nk,t-g} \otimes I_m) \text{vec}(P_{mk0})$, $\tilde{X}_{nk,t}^{(s)} = 1/T \times \sum_{t=0}^{T-1} X_{nk,t}^{(s)}$ and $\tilde{X}_{nk,t}^{(s)} = X_{nk,t}^{(s)} - \tilde{X}_{nk,t}^{(s)}; \quad Y_{nm,t}^{(s)} = \sum_{g=0}^{+\infty} B_{nm0}^g S_{nm0}^{-1} \text{vec}(V_{nm,t-g})$, $\tilde{Y}_{nm,t}^{(s)} = 1/T \times \sum_{t=0}^{T-1} Y_{nm,t}^{(s)}$ and $\tilde{Y}_{nm,t}^{(s)} = Y_{nm,t}^{(s)} - \tilde{Y}_{nm,t}^{(s)}$; When evaluated by the true value $\theta_0$, we can denote...
\[ \tau^{-1}\frac{\partial \ln L_{nT,m0}}{\partial \theta} = \tau^{-1} \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} + \tau^{-1} \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta}, \]

where

\[ 1 \frac{\partial \ln L_{nT,m0}}{\partial \text{vec}(W'_{km})} = \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} (\tilde{X}_{nk,t} \otimes I_{m})'(J_{n} \otimes \Sigma_{vm0}^{-1})\text{vec}(V'_{nm,t}), \]

\[ 1 \frac{\partial \ln L_{nT,m0}}{\partial P_{m,ij}} = \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} (U_{0} + W_{nm}^u \tilde{X}_{nm,t-1} + W_{nm}^u C_{nm,t-1} + W_{nm}^u V_{nm,t-1} + \tilde{X}_{nk,t-1} + \mathcal{V}_{nm,t-1}^{-1}')(J_{n} \otimes (E_{m,ij} \Sigma_{vm0}^{-1}))\text{vec}(V'_{nm,t}), \]

\[ 1 \frac{\partial \ln L_{nT,m0}}{\partial \Phi_{m,ij}} = \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} \text{vec}(V'_{nm,t})[J_{n} \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\text{vec}(V'_{nm,t}) - \frac{1}{\sqrt{(n-1)T}} \text{Tr}[\Sigma_{vm0}^{-1} F_{m,ij}], \]

\[ 1 \frac{\partial \ln L_{nT,m0}}{\partial \Psi_{m,ij}} = \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} \text{vec}(V'_{nm,t})S'_{nm0}^{-1}[(W_{n} I_{m}) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\text{vec}(V'_{nm,t}) - \frac{\sqrt{T}}{\sqrt{n-1}} \text{Tr}[S_{nm}^{-1} (W_{n} I_{m} \otimes E_{m,ij})] + \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} [(W_{n} \otimes I_{m}) S_{nm0}^{-1} (\tilde{X}_{nk,t} \otimes I_{m})\text{vec}(\Pi'_{km0}) + (W_{n} \otimes I_{m}) S_{nm0}^{-1} (I_{n} \otimes \Phi'_{m0}) - I_{nm} \tilde{X}_{nk,t-1}^{-1}'][J_{n} \otimes (E_{m,ij} \Sigma_{vm0})]\text{vec}(V'_{nm,t}) \]

\[ + \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} [W_{nm}^u (I_{n} \otimes A_{m0}^{-1})(\tilde{X}_{nm,t-1} + \tilde{C}_{nm,t-1} + V_{nm,t-1})] \times [J_{n} \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\text{vec}(V'_{nm,t}) \]

\[ + \frac{1}{\sqrt{(n-1)T}} \sum_{t=1}^{T} \mathcal{V}'_{nm,t-1}[(J_{n} \otimes P_{m0}^u + W_{n} \otimes \Phi'_{m0}) S_{nm0}^{-1} (W_{n} \otimes I_{m}) - I_{nm}] \times [J_{n} \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\text{vec}(V'_{nm,t}). \]

where \( A_{m0}^{-1} = [(I_{m} - \Psi_{m0}'^{-1})(P_{m} + \Phi_{m0})'] - I_{m} \) and \( A_{m0}^{-1} \) is when all parameters are evaluated at the true value. Similarly to Section 3, we know \( \tau^{-1} \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} = \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} = \)
\[
\sqrt{\frac{n-1}{T}} \frac{1}{n-1} \Delta_{R}^{\infty} + O \left( \sqrt{\frac{n}{T^3}} \right) + O_{p} \left( \sqrt{\frac{1}{T}} \right). \quad \text{Specifically, all the residual terms are}
\]
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^{R}}{\partial \text{vec}(\Pi_{km})} = 0,
\]
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^{R}}{\partial P_{m,ij}} = -\sqrt{\frac{n-1}{T^3}} \frac{1}{n-1} \text{Tr} \left[ (J_{n} \otimes E_{m,ij}^{e'}) \sum_{g=0}^{+\infty} B_{nm0}^{g} S_{nm0}^{-1} \right] 
\]
\[
- \sqrt{\frac{n-1}{T}} \frac{1}{nT^2} \text{Tr} \left[ W_{nm}^{u'} (J_{n} \otimes E_{m,ij}^{e'}) \sum_{t=1}^{T-1} \sum_{h=0}^{t-1} I_{n} \otimes [(I_{m} - \Psi_{m0}')^{-1}(P_{m0} + \Phi_{m0})]^h 
\]
\[
(I_{m} - \Psi_{m0}')^{-1} \right] + O \left( \sqrt{\frac{n}{T^3}} \right) + O_{p} \left( \sqrt{\frac{1}{T}} \right),
\]
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^{R}}{\partial \Phi_{m,ij}} = -\sqrt{\frac{n-1}{T}} \frac{1}{n-1} \text{Tr} \left[ ((J_{n}(W_{n} - I_{n})) \otimes E_{m,ij}^{e'}) \sum_{g=0}^{+\infty} B_{nm0}^{g} S_{nm0}^{-1} \right] 
\]
\[
+ O \left( \sqrt{\frac{n}{T^3}} \right) + O_{p} \left( \sqrt{\frac{1}{T}} \right),
\]
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^{R}}{\partial \Sigma_{vm,ij}} = -\sqrt{\frac{n-1}{T}} \text{Tr} [F_{m,ij} \Sigma_{vm0}^{-1}] + O \left( \sqrt{\frac{n}{T^3}} \right) + O_{p} \left( \sqrt{\frac{1}{T}} \right),
\]
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^{R}}{\partial \Psi_{m,ij}} = -\sqrt{\frac{n-1}{T}} \frac{1}{n-1} \text{Tr} \left[ (J_{n} \otimes E_{m,ij}^{e'}) [(W_{n} \otimes I_{m}) S_{nm0}^{-1} (I_{n} \otimes P_{m0}^{e'}) + W_{n} \otimes \Phi_{m0}'] - I_{nm} \sum_{g=0}^{+\infty} H_{nm0}^{g} S_{nm0}^{-1} \right] 
\]
\[
- \sqrt{\frac{n-1}{T}} \frac{1}{n-1} \text{Tr} \left[ ((J_{n}W_{n}) \otimes E_{m,ij}^{e'}) S_{nm0}^{-1} \right] 
\]
\[
+ \frac{1}{T(n-1)} \text{Tr} \left[ (I_{n} \otimes A_{m0}^{(e')}) W_{nm}^{u'} (J_{n} \otimes E_{m,ij}^{e'}) \sum_{t=1}^{T-1} \sum_{h=0}^{t-1} I_{n} \otimes [(I_{m} - \Psi_{m0}')^{-1} (P_{m0} + \Phi_{m0})]^h 
\]
\[
(I_{m} - \Psi_{m0}')^{-1} \right] + O \left( \sqrt{\frac{n}{T^3}} \right) + O_{p} \left( \sqrt{\frac{1}{T}} \right). \]

Applying Lemma 2.4 when \( T \to +\infty \) (\( \Omega_{\theta_{0}}^{co} \) and \( \Xi_{\theta_{0}}^{co} \) are very similar to those in stable case.)

\[
\frac{1}{(n-1)T} \frac{\partial \ln L_{nT,m0}}{\partial \theta} - \frac{1}{(n-1)T} \frac{\partial \ln L_{nT,m0}^{R}}{\partial \theta} \xrightarrow{d} N(0, \Omega_{\theta_{0}}^{co} + \Xi_{\theta_{0}}^{co}).
\]

174
define $n^* = n - n_1$, then the first order derivative are

\[
\frac{1}{n^* T} \frac{\partial \ln L_{nT, m, ro}}{\partial \text{vec}(P'_m)} = \frac{1}{n^* T} \sum_{t=1}^{T} (\tilde{X}_{nk,t} \otimes I_m)'(J^*_n \otimes \Sigma^{-1}_{vm}) \text{vec}(\tilde{V}_{nm,t}')(\tilde{V}_{nm,t}(\theta))
\]

\[
\frac{1}{n^* T} \frac{\partial \ln L_{nT, m, ro}}{\partial \Sigma_{vm, ij}} = \frac{1}{n^* T} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}(\theta))'(J^*_n \otimes \Sigma^{-1}_{vm}) \text{vec}(\tilde{V}_{nm,t}(\theta))
\]

\[
\frac{1}{n^* T} \frac{\partial \ln L_{nT, m, ro}}{\partial \psi_{m, ij}} = \frac{1}{n^* T} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}(\theta))'(W'_n \otimes (E_{m, ij})' \Sigma^{-1}_{vm}) \text{vec}(\tilde{V}_{nm,t}(\theta))
\]

When evaluated by the true value $\theta_0$, we can denote $\frac{1}{\sqrt{n^* T}} \frac{\partial \ln L_{nT, m, ro}}{\partial \theta} = \frac{1}{\sqrt{n^* T}} \frac{\partial \ln L_{nT, m, ro}}{\partial \theta_0} + \frac{1}{\sqrt{n^* T}} \frac{\partial \ln L_{nT, m, ro}}{\partial \theta}$.

Suppose $X_{nk,t} = \sum_{g=0}^{+\infty} B_{nm, g} S_{nm, g} (X_{nk,g} \otimes I_m) \text{vec}(\Pi_{mk,0})$, $\tilde{X}_{nk,t} = 1/T \sum_{t=0}^{T-1} X_{nk,t}$ and $\tilde{X}_{nk,T} = X_{nk,T} - \tilde{X}_{nk,T}$; also, $V_{nm,t} = \sum_{g=0}^{+\infty} S_{nm, g} B_{nm,0} \text{vec}(V_{nm,t-g})$, $\tilde{V}_{nm,T} = 1/T \sum_{t=0}^{T-1} V_{nm,t}$ and $\tilde{V}_{nm,T} = V_{nm,T} - \tilde{V}_{nm,T}$; then the first component in the decomposition can be

\[
\frac{1}{\sqrt{n^* T}} \frac{\partial \ln L_{1T, m, 0, ro}}{\partial \text{vec}(\Pi'_{km})} = \frac{1}{\sqrt{n^* T}} \sum_{t=1}^{T} (\tilde{X}_{nk,t} \otimes I_m)'(J^*_n \otimes \Sigma^{-1}_{vm}) \text{vec}(V_{nm,t})
\]

\[
\frac{1}{\sqrt{n^* T}} \frac{\partial \ln L_{1T, m, 0, ro}}{\partial P_{m, ij}} = \frac{1}{\sqrt{n^* T}} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t-1}(\theta) + V_{nm,t-1})(J^*_n \otimes (E_{m, ij})' \Sigma^{-1}_{vm}) \text{vec}(V_{nm,t})
\]

\[
\frac{1}{\sqrt{n^* T}} \frac{\partial \ln L_{1T, m, 0, ro}}{\partial \psi_{m, ij}} = \frac{1}{\sqrt{n^* T}} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}(\theta) + V_{nm,t})(W'_n \otimes (E_{m, ij})' \Sigma^{-1}_{vm}) \text{vec}(V_{nm,t})
\]

\[
\frac{1}{\sqrt{n^* T}} \frac{\partial \ln L_{1T, m, 0, ro}}{\partial \Sigma_{vm, ij}} = \frac{1}{2 \sqrt{n^* T}} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}(\theta) + V_{nm,t})(J^*_n \otimes (\Sigma^{-1}_{vm} F_{m, ij}) \Sigma^{-1}_{vm}) \text{vec}(V_{nm,t})
\]

175
\[-\frac{\sqrt{n^* T}}{2} \operatorname{Tr}[\Sigma_{vm0}^{-1} F_{m,ij}], \]
\[
\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Psi_{m,ij}^T} = \frac{1}{\sqrt{n^* T}} \sum_{t=1}^{T} \text{vec}(V'_{nm,t}) S'_{nm0}^{-1} \left[(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,t})
\]
\[-\sqrt{\frac{n}{n_T}} \operatorname{Tr}[S'_{nm0}^{-1} (W_n^* \otimes E'_{m,ij}) + \frac{1}{\sqrt{n^* T}} \sum_{t=1}^{T} \text{vec}(\Pi'_{km0}) \right]
\[
+ (I_n \otimes P'_{m0} + W_n \otimes \Phi'_{m0}) \tilde{X}_{nk,t-1} \text{vec}(\Pi'_{km0})
\]
\[
+ \frac{1}{\sqrt{n^* T}} \sum_{t=1}^{T} \text{vec}(V'_{nm,t}) - \sqrt{n_T} \sum_{t=1}^{T} \text{vec}(V'_{nm,T}) S'_{nm0}^{-1} [(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial P_{m,ij}^T} = -\sqrt{\frac{T}{n^*}} \Psi_{nm,T}' \left[(J_n^* \otimes (E_{m,ij} \Sigma_{vm0}^{-1})) \text{vec}(V'_{nm,T})
\]
\[-\sqrt{\frac{T}{n^*}} \Psi_{nm,T}' \left[(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,T}), \]
\[-\sqrt{\frac{T}{n^*}} \Psi_{nm,T}' \left[(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,T})\]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Pi_{km0}} = 0, \]
\[-\sqrt{\frac{T}{n^*}} \Psi_{nm,T}' \left[(J_n^* \otimes (E_{m,ij} \Sigma_{vm0}^{-1})) \text{vec}(V'_{nm,T})
\]
\[-\sqrt{n^* T} \Psi_{nm,T}' \left[(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,T})\]
\[-\sqrt{n^* T} \Psi_{nm,T}' \left[(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,T})\]
\[-\sqrt{n_T} \sum_{t=1}^{T} \text{vec}(V'_{nm,T}) S'_{nm0}^{-1} [(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Sigma_{vm0}^T} = \frac{1}{2} \sqrt{\frac{T}{n^*}} \text{vec}(V'_{nm,T}) \left[J_n^* \otimes (\Sigma_{vm0}^{-1} F_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,T})\]
\[-\sqrt{n_T} \sum_{t=1}^{T} \text{vec}(V'_{nm,T}) S'_{nm0}^{-1} [(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Phi_{m,ij}^T} = -\sqrt{\frac{T}{n^*}} \Psi_{nm,T}' \left[(I_n \otimes P'_{m0} + W_n \otimes \Phi'_{m0}) \right] S'_{nm0}^{-1} [(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Pi_{km0}} = 0, \]
\[-\sqrt{\frac{n_T}{T}} \Psi_{nm,T}' \left[(J_n^* \otimes (E_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,T})\]
\[-\sqrt{n_T} \sum_{t=1}^{T} \text{vec}(V'_{nm,T}) S'_{nm0}^{-1} [(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Phi_{m,ij}^T} = -\sqrt{\frac{n_T}{T}} \Psi_{nm,T}' \left[(I_n \otimes P'_{m0} + W_n \otimes \Phi'_{m0}) \right] S'_{nm0}^{-1} [(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\]

And the residuals are
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Psi_{m,ij}^T} = 0, \]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial P_{m,ij}^T} = -\sqrt{\frac{T}{n^*}} \Psi_{nm,T}' \left[J_n^* \otimes (E_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,T}), \]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Phi_{m,ij}^T} = -\sqrt{\frac{T}{n^*}} \Psi_{nm,T}' \left[(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})\right] \text{vec}(V'_{nm,T}), \]
\[-\sqrt{\frac{T}{n^*}} \Psi_{nm,T}' \left[(I_n \otimes P'_{m0} + W_n \otimes \Phi'_{m0}) \right] S'_{nm0}^{-1} [(W_n^* J_n^*) \otimes (E_{m,ij} \Sigma_{vm0}^{-1})]\]

Specifically, all the residual terms can be represented as
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Psi_{m,ij}^T} = 0, \]
\[-\sqrt{\frac{n^* T}{T}} \frac{1}{n^*} \operatorname{Tr} \left[\left(\frac{J_n^* \otimes (E_{m,ij} \Sigma_{vm0}^{-1})}{\sum_{g=0}^{+\infty} B_{nm0}^g S_{nm0}^{-1}}\right)\right] \]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial P_{m,ij}^T} = -\sqrt{\frac{n^* T}{T}} \frac{1}{n^*} \operatorname{Tr} \left[\left((J_n^* W_n) \otimes (E_{m,ij} \Sigma_{vm0}^{-1}) \sum_{g=0}^{+\infty} B_{nm0}^g S_{nm0}^{-1}\right)\right] \]
\[-\frac{1}{\sqrt{n^* T}} \frac{\partial \ln I_{nT,m0,ro}^R}{\partial \Phi_{m,ij}^T} = -\sqrt{\frac{n^* T}{T}} \frac{1}{n^*} \operatorname{Tr} \left[\left((J_n^* W_n) \otimes (E_{m,ij} \Sigma_{vm0}^{-1}) \sum_{g=0}^{+\infty} B_{nm0}^g S_{nm0}^{-1}\right)\right] \]

176
\[
\frac{1}{\sqrt{n^*T}} \frac{\partial \ln L_{nT,m0,ro}^R}{\partial \Sigma_{em,ij}} = -\sqrt{n^*/T} \text{Tr}[F_{m,ij} \Sigma_{em0}^{-1}] + O\left(\sqrt{n^*/T^3}\right) + O_p\left(\sqrt{1/T}\right),
\]
\[
\frac{1}{\sqrt{n^*T}} \frac{\partial \ln L_{nT,m0,ro}^R}{\partial \Psi_{m,ij}} = -\sqrt{n^*/T} \frac{1}{n^*} \text{Tr} \left[ ((J_n^*W_n) \otimes E_{m,ij}^') S_{nm0}^{-1} (I_n \otimes P_m' + W_n \otimes \Phi_m') \right] - \sqrt{n^*/T} \frac{1}{n^*} \text{Tr} \left[ ((J_n^*W_n) \otimes E_{m,ij}^') S_{nm0}^{-1} \right] + O\left(\sqrt{n^*/T^3}\right) + O_p\left(\sqrt{1/T}\right).
\]

The second order derivatives at any possible \(\theta\) are

\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial \text{vec}(\Pi'_{km}) \partial \text{vec}(\Pi'_{km})} = -\sum_{t=1}^{T} (\tilde{X}_{nk,t}J_n^*X_{nk,t}) \otimes \Sigma_{em}^{-1},
\]
\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial \text{vec}(\Pi'_{km}) \partial P_{m,ij}} = -\sum_{t=1}^{T} [(\tilde{X}_{nk,t}J_n^*W_n) \otimes (\Sigma_{em}^{-1}E_{m,ij}')] \text{vec}(\tilde{Y}_{nm,t-1}'),
\]
\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial \text{vec}(\Pi'_{km}) \partial \Phi_{m,ij}} = -\sum_{t=1}^{T} [(\tilde{X}_{nk,t}J_n^*W_n) \otimes (\Sigma_{em}^{-1}F_{m,ij} \Sigma_{em}^{-1})] \text{vec}(\tilde{V}_{nm,t}(\theta)),
\]
\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial \text{vec}(\Pi'_{km}) \partial \Sigma_{em,ij}} = -\sum_{t=1}^{T} [(\tilde{X}_{nk,t}J_n^*W_n) \otimes (\Sigma_{em}^{-1}E_{m,ij}')] \text{vec}(\tilde{Y}_{nm,t}'),
\]
\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial P_{m,ij} \partial P_{m,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1}')(J_n^* \otimes (E_{m,ij} \Sigma_{em}^{-1}E_{m,kl}')) \text{vec}(\tilde{Y}_{nm,t-1}'),
\]
\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial P_{m,ij} \partial \Phi_{m,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1}')(J_n^*W_n \otimes (E_{m,ij} \Sigma_{em}^{-1}E_{m,kl}')) \text{vec}(\tilde{Y}_{nm,t-1}'),
\]
\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial P_{m,ij} \partial \Sigma_{em,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1}')(J_n^*W_n \otimes (E_{m,ij} \Sigma_{em}^{-1}E_{m,kl}')) \text{vec}(\tilde{V}_{nm,t}(\theta)),
\]
\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial P_{m,ij} \partial \Psi_{m,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1}')(J_n^*W_n \otimes (E_{m,ij} \Sigma_{em}^{-1}E_{m,kl}')) \text{vec}(\tilde{V}_{nm,t}),
\]
\[
\frac{\partial^2 \ln L_{nT,m,ro}}{\partial \Phi_{m,ij} \partial \Phi_{m,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1}')(W_n'J_n^*W_n \otimes (E_{m,ij} \Sigma_{em}^{-1}E_{m,kl}')) \text{vec}(\tilde{V}_{nm,t-1}),
\]

177
\[ \frac{\partial^2 \ln L_{nT,m,ro}}{\partial \Phi_{m,ij} \partial \Sigma_{vm,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1}'(W_n J_n^* \otimes (E_{m,ij} \Sigma_{vm}^{-1} F_{m,kl} \Sigma_{vm}^{-1})) \text{vec}(\tilde{V}_{nm,t}(\theta))) \]

\[ \frac{\partial^2 \ln L_{nT,m,ro}}{\partial \Phi_{m,ij} \partial \Psi_{m,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t-1}'(W_n J_n^* W_n \otimes (E_{m,ij} \Sigma_{vm}^{-1} F_{m,kl} \Sigma_{vm}^{-1})) \text{vec}(\tilde{V}_{nm,t})) \]

\[ \frac{\partial^2 \ln L_{nT,m,ro}}{\partial \Sigma_{vm,ij} \partial \Sigma_{vm,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t}(\theta)) [J_n^* \otimes (\Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1} F_{m,kl} \Sigma_{vm}^{-1})] \text{vec}(\tilde{V}_{nm,t}(\theta)) \]

\[ + \frac{(n - n_1)T}{2} \text{Tr}[\Sigma_{vm}^{-1} F_{m,ij} \Sigma_{vm}^{-1} F_{m,kl}] \]

\[ \frac{\partial^2 \ln L_{nT,m,ro}}{\partial \Sigma_{vm,ij} \partial \Psi_{m,kl}} = -\sum_{t=1}^{T} \text{vec}(\tilde{Y}_{nm,t}(\theta))(W_n J_n^* W_n \otimes (E_{m,ij} \Sigma_{vm}^{-1} F_{m,kl} \Sigma_{vm}^{-1})) \text{vec}(\tilde{Y}_{nm,t}(\theta)) \]

\[ -T \text{Tr}[(S_{nm}^\dagger E_{m,ij} \otimes W_n^\dagger) S_{nm}^\dagger (E_{m,kl} \otimes W_n^\dagger)]. \]
Appendix F: Proofs of Lemmas, Theorems and Corollaries in Chapter 2

This section collects proofs of important lemmas, theorems and corollaries in Chapter 2.

**Proof of Proposition 2.1**

Define

\[ Q_{nT,m} = E \left[ \frac{1}{(n-1)Tm} \ln L^*_n T,m \right] \]

\[ = -\frac{1}{2} \ln(2\pi) + \frac{1}{(n-1)m} \ln |S_{nm}| - \frac{1}{(n-1)m} \ln |I_m - \Psi^\prime_m| - \frac{1}{2m} \ln |\Sigma_{vm}| \]

\[-\frac{1}{2(n-1)Tm} \sum_{t=1}^{T} E(\text{vec}(\tilde{V}'_{nm,t}(\theta))' (J_n \otimes \Sigma_{vm}^{-1}) \text{vec}(\tilde{V}'_{nm,t}(\theta))).\]

Since \( \text{vec}(\tilde{V}'_{nm,t}(\theta))' (J_n \otimes \Sigma_{vm}^{-1}) \text{vec}(\tilde{V}'_{nm,t}(\theta)) = \text{vec}(\tilde{V}'_{nm,t}(\theta))' (F_{n,n-1} \otimes I_m)(\Sigma_{vm}^{-1} \otimes I_n-1)(F'_{n,n-1} \otimes I_m) \text{vec}(\tilde{V}'_{nm,t}(\theta)), \)

\[(F'_{n,n-1} \otimes I_m) \text{vec}(\tilde{V}'_{nm,t}(\theta)) = (F'_{n,n-1} \otimes I_m)S_{nm} \text{vec}(\tilde{Y}'_{nm,t}) \]

\[-(F'_{n,n-1} \otimes I_m)(I_n \otimes P'_{m} + W_{n} \otimes \Phi'_{m}) \text{vec}(\tilde{Y}'_{nm,t-1}) \]

\[-(F'_{n,n-1} \otimes I_m)(\tilde{X}_{nk,t} \otimes I_m) \text{vec}(\Pi'_{km}) \]

\[ = S_{nm}^{*}(F'_{n,n-1} \otimes I_m) \text{vec}(\tilde{Y}'_{nm,t}) - (I_{n-1} \otimes P_{m} + W_{n}^{*} \otimes \Phi'_{m})(F'_{n,n-1} \otimes I_m) \text{vec}(\tilde{Y}'_{nm,t-1}) \]

\[-(F'_{n,n-1} \otimes I_m)(\tilde{X}_{nk,t} \otimes I_m) \text{vec}(\Pi'_{km}).\]
Substitute \((F'_{n,n-1} \otimes I_m) vec(\tilde{Y}'_{nm,t})\) into the model, note that
\[
(F'_{n,n-1} \otimes I_m) vec(\tilde{V}'_{nm,t}(\theta)) = S_{nm}^{*} S_{nm}^{-1} ((I_{n-1} \otimes P'_m + W'_n \otimes \Phi'_m)(F'_{n,n-1} \otimes I_m) vec(\tilde{Y}'_{nm,t-1}) + \left(I_{n-1} \otimes P'_m + W'_n \otimes \Phi'_m\right) + (F'_{n,n-1} \otimes I_m) vec(\tilde{V}'_{nm,t})) - (I_{n-1} \otimes P'_m + W'_n \otimes \Phi'_m)(F'_{n,n-1} \otimes I_m) vec(\tilde{Y}'_{nm,t-1}) - (F'_{n,n-1} \tilde{X}_{nk,t} \otimes I_m) vec(\Pi'_{km}).
\]

Since \((F'_{n,n-1} \otimes I_m) S_{nm} = S_{nm}^{*} (F'_{n,n-1} \otimes I_m), (F'_{n,n-1} \otimes I_m)(I_{n} \otimes P'_m + W_n \otimes \Phi'_m) = (I_{n-1} \otimes P'_m + W'_n \otimes \Phi'_m)(F'_{n,n-1} \otimes I_m), (F'_{n,n-1} \otimes I_m) S_{nm} = (F'_{n,n-1} \otimes I_m) S_{nm}(J_n \otimes I_m)\) and \((F'_{n,n-1} \otimes I_m) S_{nm}^{-1} = (F'_{n,n-1} \otimes I_m) S_{nm}^{-1}(J_n \otimes I_m)\), we can decompose the
\[
\frac{1}{2(n-1)Tm} \sum_{t=1}^{T} E(vec(\tilde{V}'_{nm,t}(\theta))' (J_n \otimes \Sigma_{vm}^{-1}) vec(\tilde{V}'_{nm,t}(\theta)))
\]
into three part:
\[
G_1(\theta, \Sigma_{vm}) + G_2(\theta, \Sigma_{vm}) + G_3(\theta, \Sigma_{vm})
\]
where,
\[
G_1(\theta, \Sigma_{vm}) = \frac{1}{2(n-1)Tm} (\xi_0 - \xi)' \left[ E \left[ \sum_{t=1}^{T} \left( \begin{array}{ccc} \tilde{Z}'_{nm,t} J_n \tilde{Z}_{nm,t} & \tilde{Z}'_{nm,t} J_n \tilde{A}_{zt,t} \\ \tilde{A}'_{zt,t} J_n \tilde{Z}_{nm,t} & \tilde{A}'_{zt,t} J_n \tilde{A}_{zt,t} \end{array} \right) \right] \otimes \Sigma_{vm}^{-1} \right] (\xi_0 - \xi);
\]
\[
G_2(\theta, \Sigma_{vm}) = \frac{1}{2(n-1)Tm} \sum_{t=1}^{T} E \left[ vec(\tilde{V}'_{nm,t})' S_{nm}^{-1} S_{nm}^{*} (J_n \otimes \Sigma_{vm}) S_{nm} S_{nm0} vec(\tilde{V}'_{nm,t}) \right];
\]
\[
G_3(\theta, \Sigma_{vm}) = \frac{1}{(n-1)Tm} \sum_{t=1}^{T} E \left\{ S_{nm} S_{nm0} ((I_n \otimes P'_m + W_n \otimes \Phi'_m) vec(\tilde{Y}'_{nm,t-1}) + (\tilde{X}_{nk,t} \otimes I_m) vec(\Pi'_{km}))) - (I_n \otimes P'_m + W_n \otimes \Phi'_m) vec(\tilde{Y}'_{nm,t-1}) - (\tilde{X}_{nk,t} \otimes I_m) vec(\Pi'_{km}) \right\}';
\]

180
and $\xi = (\text{vec}(P_m'), \text{vec}(\Psi_m'), \text{vec}(\Pi_{km}'), \text{vec}(\Psi_m'))'$. $G_1(\theta, \Sigma_{vm})$ is easy to handle. For $G_2(\theta, \Sigma_{vm})$,

\[
\frac{1}{2(n-1)T} \sum_{t=1}^{T} \mathbb{E} \left[ \text{vec}(\tilde{V}_{nm,t}') (\Sigma_{nm}^{-1}) \text{vec}(\tilde{V}_{nm,t}) \right]
\]

\[
= \frac{1}{2(n-1)T} \sum_{t=1}^{T} \text{vec}(V_{nm,t}') (\Sigma_{nm}^{-1}) \text{vec}(V_{nm,t}) - \frac{1}{2(n-1)m} \text{vec}(\tilde{V}_{nm,T}') (\Sigma_{nm}^{-1}) \text{vec}(\tilde{V}_{nm,T})
\]

\[
= \frac{1}{(n-1)T} \sum_{t=1}^{T} \text{vec}(V_{nm,t}') (\Sigma_{nm}^{-1}) \text{vec}(V_{nm,t}) - \frac{1}{2(n-1)m} \text{vec}(\tilde{V}_{nm,T}') (\Sigma_{nm}^{-1}) \text{vec}(\tilde{V}_{nm,T})
\]

$G_3(\theta, \Sigma_{vm})$ includes, $\frac{1}{(n-1)T} \sum_{t=1}^{T} D' \text{vec}(\tilde{V}_{nm,t})$ and $\frac{1}{(n-1)T} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t-1}) B_{nm} \text{vec}(\tilde{V}_{nm,t})$, and the lemma 2.1 states that $E(\frac{1}{(n-1)T} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{nm,t-1}) B_{nm} \text{vec}(\tilde{V}_{nm,t})) = O(1/T)$.

Therefore, this term $G_3(\theta, \Sigma_{vm}) = O(1/T)$.

We can concentrate $\sigma^2_m$ out from the likelihood function. Let $\sigma^2_m = \text{Tr}(\Sigma_{vm})$ and $\Sigma^*_{vm} = \Sigma_{vm} / \sigma^2_m$. The first order condition of the likelihood function with respect to $\sigma^2_m$ gives

\[
\sigma^2_m(\theta, \Sigma^*_{vm}) = \frac{1}{(n-1)T} \sum_{t=1}^{T} \text{vec}(\tilde{V}_{t}(\theta)') (\Sigma_{vm}^{-1}) \text{vec}(\tilde{V}_{t}(\theta))
\]

\[
\sigma^2_{mE}(\theta, \Sigma^*_{vm}) = 2(G_1(\theta, \Sigma^*_{vm}) + G_2(\theta, \Sigma^*_{vm}) + G_3(\theta, \Sigma^*_{vm})).
\]

Therefore, the concentrated expected likelihood function

\[
Q_{nT,m}^c = -\frac{1}{2} \ln(2\pi) + \frac{1}{2(n-1)m} \ln |S_{nm}| - \frac{1}{(n-1)m} \ln |I_m - \Psi'_m|
\]

\[
- \frac{1}{2m} \ln |\Sigma^*_{vm}| - \frac{1}{2} \ln \sigma^2_{mE}(\theta, \Sigma^*_{vm}).
\]

The analysis of identification can be separated into two parts.

\[
Q_{nT,m}^c - Q_{nT,m0}^c = \frac{1}{2(n-1)m} \ln |S_{nm}| - 2 \ln |S_{nm0}| - 2 \ln |I_m - \Psi'_m| + 2 \ln |I_m - \Psi'_m|
\]

\[
- (n-1) \ln |\Sigma^*_{vm}| + (n-1) \ln |\Sigma^*_{vm}| - \frac{1}{2} \ln \sigma^2_{mE}(\theta, \Sigma^*_{vm}) - \ln \left( \frac{T-1}{T} \sigma^2_{m0} \right),
\]

where

\[
\ln \sigma^2_{mE}(\theta, \Sigma^*_{vm}) - \ln \left( \frac{T-1}{T} \sigma^2_{m0} \right)
\]

181
\[
= \ln \left[ \frac{2(G_1(\theta, \Sigma_{\text{vm}}^*) + G_2(\theta, \Sigma_{\text{vm}}^*) + G_3(\theta, \Sigma_{\text{vm}}^*))}{T - 1} \sigma_{m0}^2 \right]
\]

\[
= \ln \left[ \frac{1}{(n - 1)(T - 1)m} \right] \left( \xi_0 - \xi \right)^{\prime} \left[ E \left[ \sum_{t = 1}^{T} \left( \begin{array}{c}
\tilde{Z}_{\text{nm},t} J_n \tilde{Z}_{\text{nm},t} \\
\tilde{A}_{z,t} J_n \tilde{A}_{z,t}
\end{array} \right) \right] \right] \otimes \Sigma_{\text{vm}}^{-1}
\]

\[
\times (\xi_0 - \xi) + \frac{1}{(n - 1)m} \text{Tr}[S_{nm0}^{-1} S_{nm}(J_n \otimes \Sigma_{\text{vm}}^{-1}) S_{nm} S_{nm0}^{-1}(I_n \otimes \Sigma_{\text{vm}})] + O \left( \frac{1}{T} \right)
\]

where

\[
\text{Tr}[S_{nm0}^{-1} S_{nm}(J_n \otimes \Sigma_{\text{vm}}^{-1}) S_{nm} S_{nm0}^{-1}(I_n \otimes \Sigma_{\text{vm}})]
\]

\[
= \text{Tr}[S_{nm0}^{-1} S_{nm}(J_n \otimes \Sigma_{\text{vm}}^{-1}) S_{nm} S_{nm0}^{-1}(I_n \otimes \Sigma_{\text{vm}})],
\]

since \( (F_{n,n-1}^\prime \otimes I_m) S_{nm} = (F_{n,n-1}^\prime \otimes I_m) S_{nm} (J_n \otimes I_m) \) and \( (F_{n,n-1}^\prime \otimes I_m) S_{nm}^\prime = (F_{n,n-1}^\prime \otimes I_m) S_{nm}^\prime (J_n \otimes I_m). \) Also,

\[
= \frac{1}{2(n - 1)m} \left[ 2 \ln |S_{nm}| - 2 \ln |S_{nm0}| - 2 \ln |J_n - \Psi_m^\prime| + 2 \ln |I_m - \Psi_{m0}^\prime| \right] - (n - 1) \ln |\Sigma_{\text{vm}}^*| + (n - 1) \ln |\Sigma_{\text{vm0}}^*|
\]

\[
= \frac{1}{2(n - 1)m} \ln |S_{nm0}^{-1} S_{nm} (I_{n-1} \otimes \Sigma_{\text{vm}}^{-1}) S_{nm}^* S_{nm0}^{-1}(I_{n-1} \otimes \Sigma_{\text{vm0}}^*)|
\]

\[
= \frac{1}{2(n - 1)m} \ln |S_{nm0}^{-1} S_{nm} (J_n \otimes \Sigma_{\text{vm}}^{-1}) S_{nm} S_{nm0}^{-1}(J_n \otimes \Sigma_{\text{vm0}})|
\]

since \( S_{nm}^* = (F_{n,n-1}^\prime \otimes I_m) S_{nm} (F_{n,n-1} \otimes I_m). \)

By the arithmetic and geometric means inequality of eigenvalues,

\[
|S_{nm0}^{-1} S_{nm} (J_n \otimes \Sigma_{\text{vm}}^{-1}) S_{nm} S_{nm0}^{-1}(J_n \otimes \Sigma_{\text{vm}})|^{\frac{1}{m(n - 1)}}
\]

\[
- \frac{1}{m(n - 1)} \text{Tr} \left[ S_{nm0}^{-1} S_{nm} (J_n \otimes \Sigma_{\text{vm}}^{-1}) S_{nm}^* S_{nm0}^{-1}(J_n \otimes \Sigma_{\text{vm}}) \right] < 0.
\]

And when \( T \) goes to infinity,

\[
= \frac{1}{(n - 1)(T - 1)m} (\xi_0 - \xi)^{\prime} \left[ E \left[ \sum_{t = 1}^{T} \left( \begin{array}{c}
\tilde{Z}_{\text{nm},t} J_n \tilde{Z}_{\text{nm},t} \\
(W_n \tilde{A}_{z,t}) J_n (W_n \tilde{A}_{z,t})
\end{array} \right) \right] \right] \otimes \Sigma_{\text{vm}}^{-1} (\xi_0 - \xi) = 0
\]

182
if and only if $\xi = \xi_0$. Therefore we know $Q^c_{nT,m} \leq Q^c_{nT,m0}$ when $T$ tend to infinity by assumptions 2.3.4 and 2.3.5. Furthermore, the equality hold if and only if all parameters are given their true value.

Proof of Lemma 2.1

\[
(J_n \otimes I_m)\text{vec}(\tilde{Y}'_{nm,t-1}) = (J_n \otimes I_m) \sum_{g=0}^{+\infty} H_{nm}^g [S_{nm0}^{-1}(I_m \otimes \tilde{X}_{nk,t-g-1})\text{vec}(\Pi'_{km}) + S_{nm0}^{-1}\text{vec}(\tilde{V}_{nm,t-g-1})]
\]

\[
= D_{x,t} + \tilde{U}_{nm,t-1,y},
\]

where $D_{x,t}$ is a column with uniformly bounded elements and $\tilde{U}_{nm,t-1,y}$ is a random column vector sharing the properties with $\tilde{U}_{nm,t}$ (and $\tilde{U}_{nm,t}$) defined in Appendix.

\[
\frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m)\text{vec}(\tilde{Y}'_{nm,t-1})]'B_{nm}\text{vec}(\tilde{V}'_{nm,t})
\]

\[
= \frac{1}{nT} \sum_{t=1}^{T} (D'_{x,t}\text{vec}(\tilde{V}'_{nm,t}) + \tilde{U}'_{nm,t-1,y}\text{vec}(\tilde{V}'_{nm,t})),
\]

and

\[
\frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m)\text{vec}(\tilde{Y}'_{nm,t-1})]'B_{nm}[(J_n \otimes I_m)\text{vec}(\tilde{Y}'_{nm,t-1})]
\]

\[
= \frac{1}{nT} \sum_{t=1}^{T} (D'_{x,t}D_{x,t} + 2D'_{x,t} \tilde{U}_{nm,t-1,y} + \tilde{U}'_{nm,t-1,y} \tilde{U}_{nm,t-1,y}).
\]

According to Lemma D.6, Lemma D.7 and Lemma D.8,

\[
E \left\{ \frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m)\text{vec}(\tilde{Y}'_{nm,t-1})]'B_{nm}\text{vec}(\tilde{V}'_{nm,t}) \right\} = O\left( \frac{1}{T} \right);
\]

\[
E \left\{ \frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m)\text{vec}(\tilde{Y}'_{nm,t-1})]'B_{nm}[(J_n \otimes I_m)\text{vec}(\tilde{Y}'_{nm,t-1})] \right\} = O(1).
\]
\[
\frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m)\text{vec}(\tilde{Y}_{nm,t-1})]'B_{nm}\text{vec}(\tilde{V}_{nm,t})
\]
\[
- E \left[ \frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m)\text{vec}(\tilde{Y}_{nm,t-1})]'B_{nm}\text{vec}(\tilde{V}_{nm,t}) \right] = O_p(\frac{1}{\sqrt{nT}});
\]
\[
\frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m)\text{vec}(\tilde{Y}_{nm,t-1})]'B_{nm}[(J_n \otimes I_m)\text{vec}(\tilde{Y}_{nm,t-1})] \]
\[
- E \left[ \frac{1}{nT} \sum_{t=1}^{T} [(J_n \otimes I_m)\text{vec}(\tilde{Y}_{nm,t-1})]'B_{nm}[(J_n \otimes I_m)\text{vec}(\tilde{Y}_{nm,t-1})] \right] = O_p(\frac{1}{nT}).
\]

\[\Box\]

**Proof of Lemma 2.2**

Since
\[
\text{var}\left[ \frac{1}{nT} \sum_{t=1}^{T} \text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,t}) \right]
\]
\[
= \frac{1}{n^2T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \text{cov}(\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,t}), \text{vec}(V_{nm,s}')B_{nm}\text{vec}(V_{nm,s}))
\]
\[
= \frac{1}{n^2T^2} \sum_{t=1}^{T} [E(\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,t})\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,t}))
\]
\[
- E(\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,t})E(\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,t})))) = O(\frac{1}{nT}),
\]
the statement (1) can be proved by Chebyshev’s inequality. Furthermore,
\[
\text{var}\left[ \frac{1}{n} \text{vec}(\tilde{V}_{nm,T}')B_{nm}\text{vec}(\tilde{V}_{nm,T}) \right]
\]
\[
= \frac{1}{n^2} \text{var}\left( \frac{1}{T} \sum_{t=1}^{T} \text{vec}(V_{nm,t}')B_{nm} \frac{1}{T} \sum_{t=1}^{T} \text{vec}(V_{nm,t}) \right)
\]
\[
= \frac{1}{n^2T^4} \sum_{t=1}^{T} E(\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,t})\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,t}))
\]
\[
+ \sum_{t=1}^{T} \sum_{s \neq t} E(\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,s})\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,s}))
\]
\[
+ \sum_{t=1}^{T} \sum_{s \neq t} E(\text{vec}(V_{nm,t}')B_{nm}\text{vec}(V_{nm,s})\text{vec}(V_{nm,s}')B_{nm}\text{vec}(V_{nm,t})) \right] = O(\frac{1}{nT^2}).
\]
Again, Chebyshev’s inequality proves (2).

\[\Box\]

**Proof of Proposition 2.2**

184
In the likelihood function,
\[
(F'_{n,n-1} \otimes I_m)\text{vec}(\tilde{V}'_{nm,t}) = (F'_{n,n-1} \otimes I_m)[S_{nm}\text{vec}(\tilde{Y}'_{nm,t})
- (I_n \otimes P'_m + W_{n} \otimes \Phi'_m)\text{vec}(\tilde{Y}'_{nm,t}) - (\tilde{X}_{nk,t} \otimes I_m)\text{vec}(\Pi'_{km})]
= (F'_{n,n-1} \otimes I_m)[(S_{nm}S_{nm}^{-1}(I_n \otimes P'_m + W_{n} \otimes \Phi'_m) - (I_n \otimes P'_m + W_{n} \otimes \Phi'_m))\text{vec}(\tilde{Y}'_{nm,t-1})
+ S_{nm}S_{nm}^{-1}(\tilde{X}_{nk,t} \otimes I_m)\text{vec}(\Pi'_{km}) + S_{nm}S_{nm}^{-1}\text{vec}(\tilde{V}'_{nm,t})].
\]

Therefore,
\[
\text{vec}(\tilde{V}'_{nm,t})' (J_n \otimes I_m)\text{vec}(\tilde{V}'_{nm,t}) = \text{vec}(\tilde{V}'_{nm,t})' D_1 \text{vec}(\tilde{V}'_{nm,t}) + \text{vec}(\tilde{V}'_{nm,t})' D_2 \text{vec}(\tilde{V}'_{nm,t}) + \text{vec}(\tilde{V}'_{nm,t})' D_3 \text{vec}(\tilde{V}'_{nm,t}) + D_4 \text{vec}(\tilde{V}'_{nm,t}) + D_5 \text{vec}(\tilde{V}'_{nm,t}) + D_6,
\]
where $D_1$, $D_2$ and $D_3$ are constant $nm \times nm$ matrices with bounded row and column sum norms uniformly in the parameter space, $D_4$ and $D_5$ are constant $nm \times 1$ vectors with bounded elements uniformly in the parameter space, and entries in $D_6$ is $O(n)$. Therefore, utilizing Lemma 2.1 and 2.2, the difference between that $rac{1}{nTm} \sum_{t=1}^{T} \text{vec}(\tilde{V}'_{nm,t})' (J_n \otimes I_m)\text{vec}(\tilde{V}'_{nm,t})$ and its expectation is $o_p(1)$ uniformly on the parameter space. Therefore, the uniform convergence is shown. \hfill \Box

**Proof of Theorem 2.1**

Proposition 2.1 and Lemma 2.2 show that identification and uniform convergence parts respectively, the only left part is to show that $Q_{nT,m}$ is uniformly equicontinuous. The method is similar to counterpart of cross-sectional case in Chapter 1.

Since
\[
Q_{nT,m} = -\frac{1}{2} \ln(2\pi) + \frac{1}{(n-1)m} \ln |S_{nm}| - \ln |I_m - \Psi'_m| - \frac{1}{2m} \ln |\Sigma_{vm}|
+ G_1(\theta, \Sigma_{vm}) + G_2(\theta, \Sigma_{vm}) + G_3(\theta, \Sigma_{vm}),
\]

185
\[ G_2(\theta, \Sigma_{vm}), -\frac{1}{(n-1)m} \ln |S_{nm}| - \ln |I_m - \Psi_m'| \] and \[ \frac{1}{2m} \ln |\Sigma_{vm}| \] has been shown in Chapter 1. We need to check \[ G_1(\theta, \Sigma_{vm}) \] and \[ G_3(\theta, \Sigma_{vm}) \] .

\begin{equation*}
G_1(\theta, \Sigma_{vm}) = (\xi_0 - \xi)' \left[ \Sigma_{vm}^{-1} \otimes \mathbb{E} \left[ \frac{1}{2(n-1)Tm} \sum_{t=1}^{T} \left( \tilde{z}_{nm,t}' J_n \tilde{z}_{nm,t}, \tilde{z}_{nm,t}' J_n (W_n \tilde{A}_{z,t}) \right) \right] \right] \times (\xi_0 - \xi);
\end{equation*}

Following Lemma 2.1, the expectation is \( O(1) \), in addition, the term is a polynomial of parameters, which means is equicontinuous uniformly in parameter space.

\begin{equation*}
G_3(\theta, \Sigma_{vm}) = \frac{1}{(n-1)Tm} \sum_{t=1}^{T} \mathbb{E} \left[ S_{nm0}^{-1} S_{nm0}' ((I_n \otimes P_m' + W_n \otimes \Phi_m') \text{vec}(\tilde{Y}_{nm,t-1}') + (\tilde{X}_{nk,t} \otimes I_m) \text{vec}(\Pi_{km}') - (I_n \otimes P_m' + W_n \otimes \Phi_m') \text{vec}(\tilde{Y}_{nm,t-1}') - (\tilde{X}_{nk,t} \otimes I_m) \text{vec}(\Pi_{km}'))' \right] \times (J_n \otimes \Sigma_{vm}^{-1}) S_{nm0}^{-1} S_{nm0}' \text{vec}(\tilde{V}_{nm,t}').
\end{equation*}

\[ G_3(\theta, \Sigma_{vm}) \] is \( O(1/T) \), it is a polynomial of parameters, we prove it is equicontinuous uniformly in parameter space.

Therefore, we establish the consistency of QML estimators.

\[ \square \]

**Proof of Lemma 2.3**

Since \( \mathbb{U}_{nm,t-1} = \sum_{g=1}^{+\infty} G_{nm,g} \text{vec}(V_{nm,t-g}') \), the \((k-1)n + i\)th entry is \( u_{nk,t-1,i} = e_{(k-1)n+i,nm}' \mathbb{U}_{nm,t-1} = \sum_{g=1}^{+\infty} e_{(k-1)n+i,nm}' G_{nm,g} \text{vec}(V_{nm,t-g}') \). So,

\[ \mathbb{E}|u_{nk,t-1,i}|^4 \]

186
Define $G_{nm,gh} = G'_{(k-1)n+i,nm}G_{nm,h}$, we have

$$
E|u_{nk,t-1,i}|^4 = \sum_{g=1}^{+\infty} [\mathbb{E}[\sum_{n=1}^{m} \sum_{m=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{f_{pq}} - \sigma_f^2 \sigma_q^2 - \sigma_f^2 \sigma_p^2 - \sigma_f^2 \sigma_{pq}^2) (\sum_{j=1}^{n} G_{nm,gg}^f)] G_{nm,gg}^g] ] + 3 \left[ \sum_{h=1}^{+\infty} e'_{(k-1)n+i,nm} G_{nm,h} \sum_{n=1}^{m} G_{nm,h} e'_{(k-1)n+i,nm} \right]^2.
$$

In above statements,

$$[G_{nm,gg}]_{jj}^{fl} = e'_{(k-1)n+i,nm} G_{nm,gh} e'_{(k-1)n+i,nm} G_{nm,gh} e'_{(k-1)n+i,nm}.$$

Therefore,

$$[G_{nm,gg}]_{jj}^{fl} [G_{nm,gg}]_{jj}^{pq} = [G_{nm,gg}]_{ij}^{kf} [G_{nm,gg}]_{ij}^{kl} [G_{nm,gg}]_{ij}^{lp} [G_{nm,gg}]_{ij}^{lk}. $$

and

$$\sum_{h=1}^{+\infty} e'_{(k-1)n+i,nm} G_{nm,h} \sum_{n=1}^{m} G_{nm,h} e'_{(k-1)n+i,nm} = \left[ \sum_{h=1}^{+\infty} G_{nm,h} \sum_{n=1}^{m} G_{nm,h} e'_{(k-1)n+i,nm} \right]_{ii}^{kk}. $$

So, $E|u_{nk,t-1,i}|^4 = O(1)$.

**Proof of Lemma 2.4**

Since $vec(V_{nm,t}) = (V_{n,1,t}, ..., V_{nm,t})'$, we can decompose $U_{nm,t-1}, D_{nm,t}$ and $B_{nm}$ into $m \times 1$ subvectors and $m \times m$ submatrices:

$$Q_{nT,m} = \sum_{t=1}^{T} \left[ \sum_{i=1}^{n} (U_{n,t-1,i} + d_{n,t,i}) v_{n,t,i} + \sum_{i=1}^{n} \sum_{j=1}^{n} (v_{n,t,i} B_{n,ij} v_{n,t,i} - E(v_{n,t,i} B_{n,ij} v'_{n,t,i})) \right].$$
where \( v_{n,t,i} = (v_{1n,t,i}, v_{2n,t,i}, \ldots, v_{nm,t,i}) \) is the \( i \)th row of \( V_{nm,t} \). Furthermore, let
\[
z_{m,t,i} = \sum_{k=1}^{m} (u_{nk,t-1,i} + d_{nk,t,i})' v_{n,t,ik} + \sum_{k=1}^{m} \sum_{l=1}^{m} [b_{nkl,ii} (v_{n,t,ik} v_{n,t,il} - \sigma_{kl}^2)] + v_{n,t,ik} \sum_{j=1}^{i-1} b_{nkl,ij} v_{n,t,j} + v_{n,t,il} \sum_{j=1}^{i-1} b_{nkl,ji} v_{n,t,j},
\]
then \( Q_{nT,m} = \sum_{t=1}^{T} \sum_{i=1}^{n} z_{m,t,i} \).

Define a \( \sigma \)-algebra \( \mathcal{F}_{n,t,i} = \sigma(v_{n,1,1}, v_{n,1,2}, \ldots, v_{n,1,n}, v_{n,2,1}, \ldots, v_{n,2,n}, \ldots, v_{n,t-1,1}, \ldots, v_{n,t-1,n}) \), then \( E(z_{m,t,i} | \mathcal{F}_{n,t,i-1}) = 0 \) and \( E(z_{m,t,1} | \mathcal{F}_{n,t-1,n}) = 0 \). Therefore, the \( \{z_{m,t,i}, \mathcal{F}_{n,t,i}\} \) forms a martingale difference array. We may apply the martingale central limit theorem to construct the specific central limit theorem for them. This requires us to prove: (1) \( \sum_{t=1}^{T} \sum_{i=1}^{n} E|z_{m,t,1}|^{2+\delta}/\sigma_{QnT,m}^{2+\delta} \to 0 \) and (2) \( \sum_{t=1}^{T} \sum_{i=1}^{n} E[z_{m,t,1}^{2+\delta}/\sigma_{QnT,m}^{2+\delta} - 1] \to 0 \).

We show (1) first.
\[
z_{m,t,i} \leq \sum_{k=1}^{m} (|u_{nk,t-1,i}| + |d_{nk,t,i}|)' v_{n,t,ik} + \sum_{k=1}^{m} \sum_{l=1}^{m} [b_{nkl,ii} |v_{n,t,ik} v_{n,t,il} - \sigma_{kl}^2|]
\]
\[
+ |v_{n,t,ik}| \sum_{j=1}^{i-1} |b_{nkl,ij}| |v_{n,t,j}| + |v_{n,t,il}| \sum_{j=1}^{i-1} |b_{nkl,ji}| |v_{n,t,j}|
\]
\[
\leq \sum_{k=1}^{m} (|u_{nk,t-1,i}| + |d_{nk,t,i}|)' v_{n,t,ik} + \sum_{k=1}^{m} \sum_{l=1}^{m} [b_{nkl,ii} |v_{n,t,ik} v_{n,t,il} - \sigma_{kl}^2|
\]
\[
+ |v_{n,t,ik}| \sum_{j=1}^{i-1} |b_{nkl,ij}| |v_{n,t,j}| + |v_{n,t,il}| \sum_{j=1}^{i-1} |b_{nkl,ji}| |v_{n,t,j}|,
\]
for \( 1/p + 1/q = 1 \). By Holder’s Inequality,
\[
|z_{m,t,i}|^{q} \leq \left[ \sum_{k=1}^{m} (|u_{nk,t-1,i}| + |d_{nk,t,i}|)'^{p} + \sum_{k=1}^{m} \sum_{l=1}^{m} [b_{nkl,ii}] + \sum_{j=1}^{i-1} |b_{nkl,ij}| + \sum_{j=1}^{i-1} |b_{nkl,ji}| \right]^{\frac{2p}{p}}
\]
\[
\times \left[ \sum_{k=1}^{m} |v_{n,t,ik}|^{q} + \sum_{k=1}^{m} \sum_{l=1}^{m} + \sum_{k=1}^{m} \sum_{l=1}^{m} [b_{nkl,ii}] |v_{n,t,ik} v_{n,t,il} - \sigma_{kl}^2|^{q}
\]
\[
+ |v_{n,t,ik}|^{q} \sum_{j=1}^{i-1} |b_{nkl,ij}| |v_{n,t,j}|^{q} + |v_{n,t,il}|^{q} \sum_{j=1}^{i-1} |b_{nkl,ji}| |v_{n,t,j}|^{q} \right]^{\frac{2q}{q}}.
\]
The second part is bounded when \( q = 2 + \delta \) by, say, \( D_{L} \), since the forth moment exists (by Assumption 1).

\( ^{25} \)Note that we can define \( \mathcal{F}_{n,t,0} = \mathcal{F}_{n,t-1,n} \) to simplify the notations.
In the first part, since \( B_{nm} \) is bounded by \( D_b \) in row and column sums,

\[
\sum_{k=1}^{m} \sum_{l=1}^{m} \left[ |b_{nkli}| + \sum_{j=1}^{i-1} |b_{nklij}| + \sum_{j=1}^{i-1} |b_{nklij}| \right] < 2m^2 D_b.
\]

By Jessen's Inequality (Generalized \( C_r \)-inequality \( \left| \frac{\sum_{k=1}^{m+1} X_k}{m+1} \right|^r \leq \sum_{k=1}^{m+1} |X_k|^r \) for \( r > 1 \)),

\[
\sum_{k=1}^{m} \left( |u_{nk,t-1,i}| + |d_{nk,t,i}| \right)^p + \sum_{k=1}^{m} \sum_{l=1}^{m} \left[ |b_{nkli}| + \sum_{j=1}^{i-1} |b_{nklij}| + \sum_{j=1}^{i-1} |b_{nklij}| \right] \leq (m + 1)^{\frac{p}{r} - 1} \sum_{k=1}^{m} (|u_{nk,t-1,i}| + |d_{nk,t,i}|)^q + (m + 1)^{\frac{q}{r} - 1} (2m^2 D_b)^\frac{q}{r}.
\]

In addition, \( (|u_{nk,t-1,i}| + |d_{nk,t,i}|)^q \leq 2^{q-1} \left( |u_{nk,t-1,i}|^q + |d_{nk,t,i}|^q \right) \). So \( |z_{mt,i}|^q \leq D_L (m + 1)^{\frac{q}{r} - 1} \left( \sum_{k=1}^{m} (2^{q-1} |u_{nk,t-1,i}|^q + 2^{q-1} D_u) + (2m^2 D_b)^\frac{q}{r} \right) \), \( E|z_{mt,i}|^{2+\delta} \leq O(E|u_{mt-1,i}|^{2+\delta}) + O(1) \). According to Lemma 2.3, \( E|u_{mt-1,i}|^{2+\delta} = O(1) \), which means

\[
\frac{1}{\sigma_{Q_{n,T,m}}^2} \sum_{t=1}^{T} \sum_{i=1}^{n} E|z_{mt,i}|^{2+\delta} = O((nT)^{-\delta/2}) \rightarrow 0.
\]

Now we show (2). Since \( \sigma_{Q_{n,T,m}}^2 = \sum_{t=1}^{T} \sum_{i=1}^{n} E(z_{mt,i}^2) \), by Chapter 1,

\[
Q_{i,n} = \sum_{k=1}^{m} d_{nk,i} v_{nk,i} + \sum_{k=1}^{m} \sum_{l=1}^{m} b_{nkli} (v_{nk,i} v_{nl,i} - \sigma_{kl}^2) + v_{nk,i} \sum_{j=1}^{i-1} b_{nkij} v_{nl,j} + v_{nl,i} \sum_{j=1}^{i-1} b_{nkji} v_{nk,j}.
\]

Now we define

\[
R_{z^2,t,i} \equiv \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} \left[ b_{nkli} b_{nqij} (v_{n,t,ik} v_{n,t,il} - \sigma_{kl}^2)(v_{n,t,ip} v_{n,t,jq} - \sigma_{pq}^2)
\right.
\]

\[
+ v_{n,t,ik} v_{n,t,ip} \left( \sum_{j=1}^{i-1} b_{nkij} v_{nl,t,j} \right) \left( \sum_{j=1}^{i-1} b_{nqij} v_{nq,t,j} \right)
\]

\[
+ v_{n,t,ik} v_{n,t,ip} \left( \sum_{j=1}^{i-1} b_{nkij} v_{nl,t,j} \right) \left( \sum_{j=1}^{i-1} b_{nqij} v_{nq,t,j} \right)
\]

\[
+ v_{n,t,ik} v_{n,t,ip} \left( \sum_{j=1}^{i-1} b_{nkij} v_{nk,t,j} \right) \left( \sum_{j=1}^{i-1} b_{nqij} v_{nq,t,j} \right)
\]

\[
+ v_{n,t,ik} v_{n,t,ip} \left( \sum_{j=1}^{i-1} b_{nkij} v_{nk,t,j} \right) \left( \sum_{j=1}^{i-1} b_{nqij} v_{nq,t,j} \right).
\]
Therefore,

\[
 z_{m,t,i} = \sum_{k=1}^{m} (u_{nk,t-1,i} + d_{nk,t,i})^2 v_{n,t,ik}^2 + 2 \sum_{k=1}^{m} \sum_{l=1}^{k-1} (u_{nk,t-1,i} + d_{nk,t,i})(u_{nl,t-1,i} + d_{nl,t,i}) v_{n,t,ik} v_{n,t,il} \\
 + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \left[ (u_{np,t-1,i} + d_{np,t,i}) b_{nkli} i (v_{n,t,ik} v_{n,t,il} - \sigma_{kl}^2) v_{n,t,ip} \\
 + v_{n,t,ik} v_{n,t,ip} (u_{np,t-1,i} + d_{np,t,i}) \left( \sum_{j=1}^{i-1} b_{nkli,j} v_{nl,t,j} \right) \\
 + v_{n,t,il} v_{n,t,ip} (u_{np,t-1,i} + d_{np,t,i}) \left( \sum_{j=1}^{i-1} b_{nkli,j} v_{nk,t,j} \right) \right] + R_{z,t,i}.
\]

Take expectation,

\[
 E(z_{m,t,i}^2 | \mathbf{\eta}_{n,t,i-1}) = \sum_{k=1}^{m} (u_{nk,t-1,i} + d_{nk,t,i})^2 \sigma_k^2 + 2 \sum_{k=1}^{m} \sum_{l=1}^{k-1} (u_{nk,t-1,i} + d_{nk,t,i}) \\
 \times (u_{nl,t-1,i} + d_{nl,t,i}) \sigma_k \sigma_l + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \left[ (u_{np,t-1,i} + d_{np,t,i}) b_{nkli} i u_{klp} \\
 + \sigma_{pk}(u_{np,t-1,i} + d_{np,t,i}) \left( \sum_{j=1}^{i-1} b_{nkli,j} v_{nl,t,j} \right) \right] \\
 + \sigma_{pl}(u_{np,t-1,i} + d_{np,t,i}) \left( \sum_{j=1}^{i-1} b_{nkli,j} v_{nk,t,j} \right) \\
 + E(R_{z,t,i} | \mathbf{\eta}_{n,t,i-1}).
\]

\[
 \sigma_{Q,t,m}^2 = \sum_{t=1}^{T} E \left[ \left( \sum_{n,m} D_{n,m} v_{nm,t} - \text{Tr}(B_{nm} \Sigma_v \Sigma_v^t) \right)^2 \right] \\
 = \sum_{t=1}^{T} E \left[ \left( \sum_{n,m} D_{n,m} v_{nm,t} \right)^2 + \left( \text{vec}(V_{nm,t}') - \text{vec}(V_{nm,t}) \right)^2 \right] \\
 + \text{Tr}(B_{nm} \Sigma_v \Sigma_v^t)^2.
\]
\[ + \text{Tr}(B_{nm}\Sigma_{v,nm})^2 + 2(D'_{nm,t}\text{vec}(V''_{nm,t}))(\text{vec}(V''_{nm,t})'B_{nm}\text{vec}(V''_{nm,t})) \\
- 2(\text{vec}(V''_{nm,t})'B_{nm}\text{vec}(V''_{nm,t})) \text{Tr}(B_{nm}\Sigma_{v,nm}) \right]. \]

In the above expression,

\[ E(U'_{nm,t-1}\text{vec}(V''_{nm,t}))^2 = E[E(\text{vec}(V''_{nm,t})'U_{nm,t-1}U'_{nm,t-1}\text{vec}(V''_{nm,t})|F_{n,t,0})] \]

\[ = E(\text{Tr}(U_{nm,t-1}U'_{nm,t-1}\Sigma_{v,nm})) = E(U'_{nm,t-1}\Sigma_{v,nm}U_{nm,t-1}) \]

\[ = \sum_{h=1}^{+\infty} \text{Tr}(G'_{nm,h}\Sigma_{v,nm}G_{nm,h}\Sigma_{v,nm}); \]

\[ E(D'_{nm,t}\text{vec}(V''_{nm,t}))^2 = \text{Tr}(D_{nm,t}D'_{nm,t}\Sigma_{v,nm}); \]

\[ E(\text{vec}(V''_{nm,t})'B_{nm}\text{vec}(V''_{nm,t}))^2 = \text{Tr}(B_{nm}\Sigma_{v,nm}(B_{nm} + B'_{nm})\Sigma_{v,nm}) + \text{Tr}(B_{nm}\Sigma_{v,nm})^2 \]

\[ + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2\sigma_{ql}^2 - \sigma_{kq}^2\sigma_{pl}^2 - \sigma_{kl}^2\sigma_{pq}^2) \left( \sum_{i=1}^{n} b_{nkli,ii}b_{npq,ii} \right) ; \]

\[ E(D'_{nm,t}\text{vec}(V''_{nm,t})'B_{nm}\text{vec}(V''_{nm,t})) = \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{n} d_{np,i}b_{nkli,ii}. \]

Therefore,

\[ \sigma_{q,t,m,r}^2 = \sum_{t=1}^{T} \sum_{h=1}^{+\infty} \text{Tr}(G'_{nm,h}\Sigma_{v,nm}G_{nm,h}\Sigma_{v,nm}) + D'_{nm,t}\Sigma_{v,nm}D_{nm,t} \]

\[ + \text{Tr}(B_{nm}\Sigma_{v,nm}(B_{nm} + B'_{nm})\Sigma_{v,nm}) \]

\[ + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2\sigma_{ql}^2 - \sigma_{kq}^2\sigma_{pl}^2 - \sigma_{kl}^2\sigma_{pq}^2) \left( \sum_{i=1}^{n} b_{nkli,ii}b_{npq,ii} \right) \]

\[ + \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{n} u_{kpl} \sum_{i=1}^{n} d_{np,i}b_{nkli,ii} \right]. \]

In addition,

\[ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{i=1}^{n} (u_{nk,t-1,i} + d_{nk,t,i})(u_{nl,t-1,i} + d_{nl,t,i})\sigma_{kl}^2 \]

\[ = \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{i=1}^{n} (u_{nk,t-1,i}u_{nl,t-1,i} + d_{nk,t,i}u_{nl,t-1,i} + u_{nk,t-1,i}d_{nl,t,i} + d_{nl,t,i}d_{nl,t,i})\sigma_{kl}^2 \]

\[ = \sum_{k=1}^{m} \sum_{l=1}^{m} (U'_{nk,t-1}\sigma_{kl}^2I_nU_{nl,t-1} + U'_{nk,t-1}\sigma_{kl}^2I_nD_{nl,t} + D'_{nk,t}\sigma_{kl}^2I_nU_{nl,t-1} + D'_{nk,t}\sigma_{kl}^2I_nD_{nl,t}) \]

\[ = U'_{nm,t-1}\Sigma_{v,nm}U_{nm,t-1} + 2D'_{nm,t}\Sigma_{v,nm}U_{nm,t-1} + D'_{nm,t}\Sigma_{v,nm}D_{nm,t}, \]

191
\[
\sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{i=1}^{n} (u_{np,t-1,i} b_{nkl,ii} u_{klp}) = \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{i=1}^{n} \sum_{q=1}^{n} (u_{np,t-1,i} b_{nkl,ij} u_{klp}) = B_{nm}^{'} n_{nm,t-1},
\]

where \( B_{nm} \) has bounded entries,

\[
\sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{i=1}^{n} \sigma_{pk}^2 (u_{np,t-1,i} + d_{np,t,i}) \left( \sum_{j=1}^{i-1} b_{nkl,ij} v_{nl,t,j} \right)
\]

\[
= \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sigma_{pk}^2 (U_{np,t-1} + D_{np,t}) B_{nkl} V_{nm,l}
\]

\[
= (U_{nm,t-1} + D_{nm,t}) \sigma_{v,nm} B_{nm}^{'} vec(V_{nm,t}),
\]

and

\[
\sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{i=1}^{n} \sigma_{pk}^2 (u_{np,t-1,i} + d_{np,t,i}) \left( \sum_{j=1}^{i-1} b_{nkl,ij} v_{nk,t,j} \right)
\]

\[
= vec(V_{nm,t})^T B_{nm}^{+} \sigma_{v,nm} (U_{nm,t-1} + D_{nm,t}).
\]

Now, (i)

\[
\sum_{t=1}^{T} \sum_{i=1}^{n} E \left[ R_{z,t,i} \left| F_{n,t,i-1} \right. \right] - \sum_{t=1}^{T} \left( Tr(B_{nm} \sigma_{v,nm} (B_{nm} + B_{nm}^{'} \sigma_{v,nm})
\]

\[
+ \sum_{k=1}^{m} \sum_{l=1}^{m} \sum_{p=1}^{m} \sum_{q=1}^{m} (u_{klpq} - \sigma_{kp}^2 \sigma_{q}^2 - \sigma_{kp}^2 \sigma_{q}^2 - \sigma_{kp}^2 \sigma_{q}^2) \left( \sum_{i=1}^{n} b_{nkl,ii} b_{npq,ii} \right) \right) = o_p(1)
\]

by Chapter 1.

(ii)

\[
\frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j=1}^{n} (U_{nm,t-1} \sigma_{v,nm} U_{nm,t-1} - \frac{1}{nT} \sum_{t=1}^{T} \sum_{h=1}^{+\infty} Tr(G_{nm,h} \sigma_{v,nm} G_{nm,h} \sigma_{v,nm})) \rightarrow o_p(1)
\]

due to Lemma D.6 (1).

(iii) \( \frac{1}{nT} \sum_{t=1}^{T} B_{nm}^{'} nm \rightarrow o_p(1) \) by lemma D.7.

(iv) \( (U_{nm,t-1} + D_{nm,t}) \sigma_{v,nm} B_{nm}^{'} vec(V_{nm,t}) \rightarrow o_p(1) \) due to lemma D.6 and D.7.

Therefore, \( \sum_{t=1}^{T} \sum_{i=1}^{n} E \left[ z_{m,t,1}^2 \left| F_{n,t,i-1} \right. \right] / \sigma_{q,nT}^2 - 1 \rightarrow 0. \]

**Proof of Lemma 2.5**
(1) The difference between \( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m}}{\partial \theta \partial \theta'} \) and \( \frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m0}}{\partial \theta \partial \theta'} \) has the forms
\[ D_1(\theta) - D_1(\theta_0) \equiv \frac{1}{(n-1)T} \sum_{t=1}^{T} [U_{1nm,t}(\theta)'] \times [C \cap (B(\Sigma_{vm}))]U_{2nm,t}(\theta) - U_{1nm,t}(\theta_0)' [C \cap (B(\Sigma_{vm0}))]U_{2nm,t}(\theta_0) \]
\[ D_2(\theta) - D_2(\theta_0) = Tr(\Sigma_{vm}B_1\Sigma_{vm}B_2) - Tr(\Sigma_{vm0}B_0\Sigma_{vm0}B_2) \]
\[ D_3(\theta) - D_3(\theta_0) = \frac{1}{n-1} [Tr(S_{nm}B_3S_{nm}B_4) - Tr(S_{nm0}B_3S_{nm0}B_4)] \]
where \( C \) and \( B_j \) (hereafter for any \( j \in \mathbb{Z} \) in this proof) are matrices with bounded row and column sum norms, \( B(\Sigma_{vm}) \) equals to \( B_5\Sigma_{vm}B_6 \) or \( B_5\Sigma_{vm}B_6\Sigma_{vm}B_7 \) or \( B_5\Sigma_{vm}B_6\Sigma_{vm}B_7\Sigma_{vm}B_8 \) and \( U_{jnm,t}(\theta) = vec(\tilde{V}_{nm,t}(\theta)) \) or \( vec(\tilde{Y}_{nm,t}'(\theta)) \) or \( vec(\tilde{Y}_{nm,t}'(\theta)) \).

Since
\[ D_j(\theta) - D_j(\theta_0) = \frac{1}{(n-1)T} \sum_{\theta_j} \frac{\partial^3 \ln L_{nT,m}(\tilde{\theta})}{\partial \theta \partial \theta' \partial \theta_j} (\theta_j - \theta_{j0}) \]
where \( \theta_j \) is the \( j \)th entry in vector \( \theta \) and \( \tilde{\theta} \) is in between \( \theta \) and \( \theta_0 \) due to the Mean Value Theorem, what we need to show is that \( \frac{1}{(n-1)T} \frac{\partial^3 \ln L_{nT,m}(\tilde{\theta})}{\partial \theta \partial \theta' \partial \theta_j} = O_p(1) \).

The derivative with respect to \( \sigma^2_{kl} \) are
\[ \frac{\partial D_1(\theta)}{\partial \sigma^2_{kl}} = \frac{1}{(n-1)T} \sum_{t=1}^{T} U_{1nm,t}(\theta)'[C \cap (B^*(\Sigma_{vm}))]U_{2nm,t}(\theta) \]
\[ \frac{\partial D_2(\theta)}{\partial \sigma^2_{kl}} = Tr(\Sigma_{vm}B_1\Sigma_{vm}B_2\Sigma_{vm}B_3^*) \]
where the matrix \( B^*(\Sigma_{vm}) \) could equal to \( B_5\Sigma_{vm}B_6\Sigma_{vm}B_7 \) or \( B_5\Sigma_{vm}B_6\Sigma_{vm}B_7\Sigma_{vm}B_8 \) or \( B_5\Sigma_{vm}B_6\Sigma_{vm}B_7\Sigma_{vm}B_8 \) and \( B_3^* \) is another bounded matrix. According to Lemma D.7 and 2.1, both \( \frac{\partial D_1(\theta)}{\partial \sigma^2_{kl}} \) and \( \frac{\partial D_2(\theta)}{\partial \sigma^2_{kl}} \) are \( O_p(1) \).

Let \( \alpha_j \) represent the \( j \)th element in vector \( (vec(P_{km})', vec(P_m)', vec(\Phi_m)) \), then the derivative is
\[ \frac{\partial D_1(\theta)}{\partial \alpha_j} = \frac{1}{(n-1)T} \sum_{t=1}^{T} \left[ \frac{\partial U_{1nm,t}(\theta)'}{\partial \alpha_j} [C \cap (B(\Sigma_{vm}))]U_{2nm,t}(\theta) + U_{1nm,t}(\theta) \right] \times [C \cap (B(\Sigma_{vm}))] \]
\[ \frac{\partial U_{2nm,t}(\theta)'}{\partial \alpha_j} \].

193
So we have Proof of Theorem 2.2 Lemma D.6, D.7 and 2.1. Since by applying central limit theorem on the first order conditions, where 
\[
\begin{align*}
\frac{\partial \text{vec}(\tilde{Y}'_{nm,t}(\theta))}{\partial \pi_{ij}} &= -(I_n \otimes E'_{km,ij}) \text{vec}(\tilde{X}_{nk,t}), \\
\frac{\partial \text{vec}(\tilde{Y}'_{nm,t}(\theta))}{\partial \phi_{ij}} &= -W_n \otimes E'_{m,ij}) \text{vec}(\tilde{Y}'_{nm,t-1}).
\end{align*}
\]
According to Lemma D.7 and 2.1, both \[
\frac{\partial U_{1nm,t}(\theta)'}{\partial \alpha}[(C \otimes (B(\Sigma^{-1}_{1m})))U_{2nm,t}(\theta) + U_{1nm,t}(\theta)'] \text{and } \frac{\partial U_{2nm,t}(\theta)'}{\partial \Psi_{m,ij}} \text{ are } O_p(1).
\]
\[
\begin{align*}
\frac{\partial D_1(\theta)}{\partial \Psi_{m,ij}} &= \frac{\partial D_3(\theta)}{\partial \Psi_{m,ij}} = \frac{1}{n-1} \text{Tr}(S_{nm}^{-1}B_1S_{nm}^{-1}B_2S_{nm}^{-1}B_3^*).
\end{align*}
\]
where \[
\text{Tr}(S_{nm}^{-1}B_1S_{nm}^{-1}B_2S_{nm}^{-1}B_3^*) = O(\eta) \text{ by Lee and Yang (2014).}
\]
In the former term, \[
\text{vec}(\tilde{Y}'_{nm,t}) = S_{nm}^{-1}[(I_n \otimes P_m' + W_n \otimes \Phi'_m)\text{vec}(\tilde{Y}_{nm,t-1}) + (\tilde{X}_{nk,t} \otimes I_m)\text{vec}(\Pi'_{km}) + \text{vec}(\tilde{D}_{nm,t}) + \text{vec}(\tilde{V}'_{nm,t})],
\]
which is sufficient to guarantee that \[
\frac{\partial D_3(\theta)}{\partial \Psi_{m,ij}} = O_p(1) \text{ by Lemma D.6, D.7 and 2.1.}
\]
\(2)\ is the direct result from Lemma 2.1 and 2.2. \hfill \square

**Proof of Theorem 2.2**

The distribution of \(\hat{\theta}\) is derived from
\[
\sqrt{(n-1)T} (\hat{\theta} - \theta) = \left( -\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m}(\bar{\theta})}{\partial \bar{\theta} \partial \bar{\theta}'} \right) \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta}.
\]
We have
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}}{\partial \theta} = \frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \theta} \overset{d}{\to} N(0, \Omega_{\theta_0} + \Xi_{\theta_0})
\]
by applying central limit theorem on the first order conditions, where
\[
\frac{1}{\sqrt{(n-1)T}} \frac{\partial \ln L_{nT,m0}^R}{\partial \theta} = \sqrt{\frac{n-1}{T}} \Delta_R + O_p(\max(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}})).
\]
Therefore, we know
\[
\sqrt{(n-1)T} (\hat{\theta} - \theta) = O_p(1)[O_p(1) + O_p(1) \sqrt{\frac{n-1}{T}} + O_p(\max(\sqrt{\frac{n}{T^3}}, \sqrt{\frac{1}{T}}))].
\]
So \(\hat{\theta} - \theta = O_p(\max(\frac{1}{\sqrt{T}}, \frac{1}{T}))\). The second order derivatives \[\frac{1}{(n-1)T} \frac{\partial^2 \ln L_{nT,m}(\bar{\theta})}{\partial \bar{\theta} \partial \bar{\theta}'} = \Omega_{\theta_0} + O_p(\max(\frac{1}{\sqrt{T^3}}, \frac{1}{T}))\] according to lemma 2.5 and above result.

194
In sum,
\[ \sqrt{(n-1)}T(\hat{\theta} - \theta) = (\Omega_{\theta_0} + O_p(\max(\frac{1}{\sqrt{nT}}, \frac{1}{T})))^{-1} \]
\[ \times \left( \frac{1}{\sqrt{(n-1)}T} \frac{\partial \ln L_{nT,m0}}{\partial \theta} + \sqrt{\frac{n-1}{T}} \Delta_R + O_p(\max(\frac{n}{T^3}, \frac{1}{T})) \right) \]
\[ = \Omega_{\theta_0}^{-1} \frac{1}{\sqrt{(n-1)}T} \frac{\partial \ln L_{nT,m0}}{\partial \theta} + \sqrt{\frac{n-1}{T}} \Delta_R + O_p(\max(\frac{n}{T^3}, \frac{1}{T})), \]
which means
\[ \sqrt{(n-1)}T(\hat{\theta} - \theta) = \sqrt{\frac{n-1}{T}} \Delta_R - O_p(\max(\frac{n}{T^3}, \frac{1}{T})) \]
\[ = o_p(1) + \Omega_{\theta_0}^{-1} \frac{1}{\sqrt{(n-1)}T} \frac{\partial \ln L_{nT,m0}}{\partial \theta} \xrightarrow{d} N(0, \Omega_{\theta_0}^{-1}(\Omega_{\theta_0} + \Xi_{\theta_0})\Omega_{\theta_0}^{-1}). \]

Multiplying \( \sqrt{\frac{T}{n-1}} \) on both side when \( n/T \to \infty \), we could arrive at \( T(\hat{\theta} - \theta_0) - \Delta_\theta \xrightarrow{p} 0. \)

**Proof of Theorem 2.3**

According to Theorem 2.2, we need to use two conditions, (i) \( \frac{n}{T^3} \to 0 \) and (ii) \( \sqrt{\frac{n}{T}} \left( \frac{1}{(n-1)}T \frac{\partial^2 \ln L_{nT,m}(\hat{\theta})}{\partial \theta \partial \theta'} \right) - \Delta_R(\hat{\theta}) - \Omega_{\theta_0}^{-1} \Delta_R \xrightarrow{p} 0 \). The condition (i) has been assumed. For condition (ii), we can derive it from the relationship
\[ \left( E \left( \frac{1}{(n-1)}T \frac{\partial^2 \ln L_{nT,m}(\hat{\theta})}{\partial \theta \partial \theta'} \right) \right)^{-1} = \Omega_{\theta_0}^{-1} + O_p \left( \max(\frac{1}{\sqrt{nT}}, \frac{1}{T}) \right). \]

Therefore, what we need to show is \( \sqrt{\frac{n}{T}} \Omega_{\theta_0}^{-1}(\Delta_R(\hat{\theta}) - \Delta_R) = o_p(1) \), which can be shown by proving \( \frac{\partial \Delta_R(\hat{\theta})}{\partial \theta} < \infty \) since \( \hat{\theta} - \theta = O_p(\max(\frac{1}{\sqrt{nT}}, \frac{1}{T})) \).

The general form of \( \Delta_R(\hat{\theta}) \) has three parts: \( \frac{1}{n-1} \text{Tr}(B_{1,nm}(\hat{\theta}) \sum_{g=0}^{\infty} H_{nm}^g(\hat{\theta}) S_{nm}^{-1}(\hat{\theta})) \), \( \text{Tr}(F_{m,ij} \Sigma_{vm}^{-1}) \) and \( \frac{1}{n-1} \text{Tr}((J_nW_n) \otimes E_{m,ij} S_{nm}^{-1}(\hat{\theta})) \).

It is obviously \( \frac{\partial \text{Tr}(F_{m,ij} \Sigma_{vm}^{-1})}{\partial \theta} < \infty \) and \( \frac{1}{n-1} \frac{\partial \text{Tr}((J_nW_n) \otimes E_{m,ij} S_{nm}^{-1}(\hat{\theta}))}{\partial \theta} < \infty \) due to the uniform boundedness of row and column sum norms of \( W_n \) and \( S_{nm}^{-1} \). For the first
term, $\frac{1}{n-1} \operatorname{Tr}(B_{1, nm}(\bar{\theta}) \sum_{g=0}^{\infty} H_{nm}^g(\bar{\theta}) S_{nm0}^{-1}(\bar{\theta}))$, we have

$$\frac{\partial H_{nm}}{\partial \Psi_{ij}} = \sum_{g=1}^{\infty} \sum_{h=0}^{g-1} H_{nm}^h S^{-1} (W_n \otimes E_{m,ij}) H_{nm}^{g-1-h} H_{nm},$$

$$\frac{\partial H_{nm}}{\partial \Phi_{ij}} = \sum_{g=1}^{\infty} \sum_{h=0}^{g-1} H_{nm}^h S^{-1} (W_n \otimes E_{m,ij}) H_{nm}^{g-1-h},$$

$$\frac{\partial H_{nm}}{\partial P_{ij}} = \sum_{g=1}^{\infty} \sum_{h=0}^{g-1} H_{nm}^h S^{-1} (I_n \otimes E_{m,ij}) H_{nm}^{g-1-h},$$

$$\frac{\partial S_{nm}^{-1}}{\partial \Psi_{ij}} = S_{nm}^{-1} (W_n \otimes E_{m,ij}) S_{nm}^{-1}.$$

Utilizing Assumption 2.3.6, $\frac{1}{n-1} \operatorname{Tr}(B_{1, nm}(\bar{\theta}) \sum_{g=0}^{\infty} H_{nm}^g(\bar{\theta}) S_{nm0}^{-1}(\bar{\theta})) = O_p(1).$ □

**On Proofs of Theorems 2.7, 2.8 and 2.9**

The proofs are similar to Proofs of Theorems 2.1, 2.2 and 2.3. This is because the transformation,

$$\text{vec}(Y'_{nm,t}) = (\Lambda_1^{-\frac{1}{2}} R_1'(I_n - W_n) \otimes I_m) \text{vec}(Y'_{nm,t})$$

$$= (\Lambda_1^{-\frac{1}{2}} R_1'(I_n - W_n) \otimes I_m) \text{vec}^{(s)}(Y'_{nm,t}).$$

Assumption 2.5.5 restricts the boundedness of transformation matrix $J_n^\dagger$, therefore, one can follow the proofs in stable case with symbolic modifications to show the properties of consistency and asymptotic normality of the RQMLE and the bias-corrected RQMLE. □