Three Essays on Panel Data Models in Econometrics

Lina Lu

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2017

ABSTRACT

Three Essays on Panel Data Models in Econometrics

Lina Lu

My dissertation consists of three chapters that focus on panel data models in econometrics and under high dimensionality; that is, both the number of individuals and the number of time periods are large. This high dimensionality is widely applicable in practice, as economists increasingly face large dimensional data sets. This dissertation contributes to the methodology and techniques that deal with large data sets.

All the models studied in the three chapters contain a factor structure, which provides various ways to extract information from large data sets. Chapter 1 and Chapter 2 use the factor structure to capture the comovement of economic variables, where the factors represent the common shocks and the factor loadings represent the heterogeneous responses to these shocks. Common shocks are widely present in the real world, for example, global financial shocks, macroeconomic shocks and energy price shocks. In applications where common shocks exist, failing to capture these common shocks would lead to biased estimation. Factor models provide a way to capture these common shocks. In contrast to Chapter 1 and Chapter 2, Chapter 3 directly focuses on the factor model with the loadings being constrained, in order to reduce the number of parameters to be estimated.

In addition to the common shocks effect, Chapter 1 considers two other effects: spatial effects and simultaneous effects. The spatial effect is present in models where dependent variables are spatially interacted and spatial weights are specified based on location and distance, in a geographic space or in more general economic, social or network spaces. The

simultaneous effect comes from the endogeneity of the dependent variables in a simultaneous equations system, and it is important in many structural economic models. A model including all these three effects would be useful in various fields.

In estimation, all the three chapters propose quasi-maximum likelihood (QML) based estimation methods and further study the asymptotic properties of these estimators by providing a full inferential theory, which includes consistency, convergence rate and limiting distribution. Moreover, I conduct Monte-Carlo simulations to investigate the finite sample performance of these proposed estimators.

Specifically, Chapter 1 considers a simultaneous spatial panel data model with common shocks. Chapter 2 studies a panel data model with heterogenous coefficients and common shocks. Chapter 3 studies a high dimensional constrained factor model.

In Chapter 1, I consider a simultaneous spatial panel data model, jointly modeling three effects: simultaneous effects, spatial effects and common shock effects. This joint modeling and consideration of cross-sectional heteroskedasticity result in a large number of incidental parameters. I propose two estimation approaches, a QML method and an iterative generalized principal components (IGPC) method. I develop full inferential theories for the two estimation approaches and study the trade-off between the model specifications and their respective asymptotic properties. I further investigate the finite sample performance of both methods using Monte-Carlo simulations. I find that both methods perform well and that the simulation results corroborate the inferential theories. Some extensions of the model are considered. Finally, I apply the model to analyze the relationship between trade and GDP using a panel data over time and across countries.

Chapter 2 investigates efficient estimation of heterogeneous coefficients in panel data models with common shocks, which have been a particular focus of recent theoretical and empirical literature. It proposes a new two-step method to estimate the heterogeneous coefficients. In the first step, a QML method is first conducted to estimate the loadings and idiosyncratic variances. The second step estimates the heterogeneous coefficients by using the structural relations implied by the model and replacing the unknown parameters with their QML estimates. Further, Chapter 2 establishes the asymptotic theory of the estimator, including consistency, asymptotic representation, and limiting distribution. The two-step estimator is asymptotically efficient in the sense that it has the same limiting distribution as the infeasible generalized least squares (GLS) estimator. Intensive Monte-Carlo simulations show that the proposed estimator performs robustly in a variety of data setups.

Chapter 3 documents the estimation and inferential theory of high dimensional constrained factor models. Factor models have been widely used in practice. However, an undesirable feature of a high dimensional factor model is that the model has too many parameters. An effective way to address this issue, proposed in Tsai and Tsay (2010), is to decompose the loadings matrix by a high-dimensional known matrix multiplying with a low-dimensional unknown matrix, which Tsai and Tsay (2010) name the constrained factor models. Chapter 3 proposes a QML method to estimate the model and develops the asymptotic properties of its estimators. A new statistic is proposed for testing the null hypothesis of constrained factor models against the alternative of standard factor models. Partially constrained factor models are also investigated. Monte-Carlo simulations confirm the theoretical results and show that the QML estimators and the proposed new statistic perform well in finite samples. Chapter 3 also considers the extension to an approximate constrained factor model where the idiosyncratic errors are allowed to be weakly dependent processes.

Contents

List of Tables

Acknowledgements

I would like to thank my committee members, Jushan Bai, Serena Ng, Bernard Salanié, Christoph Rothe and Yang Feng. First of all, I offer my deepest gratitude to my advisor, Jushan Bai, for his constant guidance and support throughout the last six years. I am also very grateful to Serena Ng for her invaluable directions and comments throughout my research and great encouragement when I went through serious family loss. I would like to express my warm and sincere thanks to Bernard Salanié for his insightful advice and suggestions. I benefited greatly from the patient and helpful comments of Christoph Rothe. Also, I appreciate much from the support and thoughtful feedback from Yang Feng.

I am also deeply indebted to Sokbae (Simon) Lee, José L. Montiel-Olea, Miikka Rokkanen, Jón Steinsson, Alireza Tahbaz-Salehi, and David E. Weinstein, for their invaluable advice. Discussions with and comments by Zheli He, Kerem Tuzcuoglu, Teck Yong Tan and other seminar participants at Columbia University greatly improved this dissertation. I also wish to thank my classmates for their constant help and encouragement throughout my research and life at Columbia University.

Finally, I want to thank my family for their unconditional love and support.

Chapter 1

Simultaneous Spatial Panel Data Models with Common Shocks

1.1 Introduction

In this chapter, I consider a simultaneous spatial panel data model, jointly modeling three effects: simultaneous effects, spatial effects and common shock effects.[1] First, the simultaneous effect comes from the endogeneity of the dependent variables in a simultaneous equation system, and is important in many structural economic modeling. Second, the spatial effect is present in models where dependent variables are spatially interacted and spatial weights matrices are specified based on location and distance, in a geographic space or in more general economic, social or production network spaces. Third, common shocks stem from a common factor structure in panel data models, where the dependent variables' responses to shocks (i.e., factors) are heterogeneous and captured by the factor loadings.

That the model includes all three effects is useful in various fields. For example, this framework can be applied to analyze the relationship between trade volume and gross domestic product (GDP) within and across countries, a prominent research topic in international trade and macroeconomics. Within a country, trade volume is endogenously correlated with GDP, which can be regarded as a simultaneous effect. Across countries, a country's trade volume (or GDP) might be affected by other countries' trade volumes (or GDPs) through trade and financial linkages. This type of impact can be viewed as a spatial effect. Moreover, a global financial shock or a common energy shock might affect all countries' trade volumes and GDPs, which is referred to as a common shock effect. The model can also be applied in social network studies such as peer effects analysis in applied microeconomics, or in regional

[1]Related literature and studies of these three effects are provided in the end of the introduction section.

economic studies.[2]

In this chapter, I consider the following simultaneous spatial panel data model, combining all three effects, with both a large time dimension T and a large cross-sectional dimension N:

$$\begin{aligned} y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + \sum_{p=1}^{k_1} x_{1itp}\beta_{1p} + \lambda_i' f_t + e_{1it} \\ y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + \sum_{q=1}^{k_2} x_{2itq}\beta_{2q} + \psi_i' f_t + e_{2it} \end{aligned} \tag{1.1.1}$$

where $i = 1, 2, \ldots, N$; $t = 1, 2, \ldots, T$; y_{1it} and y_{2it} are the dependent variables for cross-section i at time t; $x_{1itp}, p = 1, 2, \ldots, k_1$ and $x_{2itq}, q = 1, 2, \ldots, k_2$ are explanatory variables, with their coefficients denoted as β_{1p} and β_{2q}, respectively; f_t is an r-dimensional vector of unobservable common shocks, termed the common factor; λ_i and ψ_i are the corresponding r-dimensional vectors of unobservable heterogeneous responses to the common shocks, termed the factor loadings; for $l = 1, 2$, $W_l = (w_{lij})_{N\times N}$ is a pre-specified spatial weights matrix whose diagonal elements w_{lii} are 0;[3] e_{1it} and e_{2it} are the idiosyncratic errors; and α_{1i} and α_{2i} are the intercepts. In model (1.1.1), taking the y_{1it} equation as an example, the term $(\gamma_1 y_{2it})$ captures the simultaneous effect from y_{2it} to y_{1it}, $(\rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt})$ captures the spatial effect, and $(\lambda_i' f_t)$ captures the common shock effect. The $(\lambda_i' f_t)$ part can also be viewed as an interactive fixed effect, which is more general than an additive fixed effect

[2] For the peer effects studies as in Cohen-Cole et al. (2013) and Liu (2014), the common factor structure in my model can be used to capture unobservable individual characteristics which have time-varying impacts on individuals' decisions or choices. For these regional economic studies as in Jeanty et al. (2010), Baltagi and Bresson (2011), Gebremariam et al. (2011) and Hauptmeier et al. (2012), the common shocks can capture macroeconomic shocks which have heterogeneous impacts on local economies.

[3] More details of the weights can be found in Remark 1.2.1 in this chapter.

and provides a flexible way to model cross-sectional and serial correlations.[4] My interest is estimating the key coefficients $(\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1, \beta_2)$ and analyzing the asymptotic properties of their estimates.

In the econometrics literature, to the best of my knowledge, no existing paper jointly models these three effects. However, recently, a few papers consider two types of models combining two of these three effects. The first type is a spatial panel data model with common shocks in a single-equation system. Extending this type of model to a simultaneous equation system would make it applicable when multiple dependent variables are simultaneously interdependent, e.g., the above trade and GDP case. In estimating this type of model, Pesaran and Tosetti (2011) implement the same common correlated effects (CCE) estimation used in Pesaran (2006), while Bai and Li (2014a) propose a quasi-maximum likelihood (QML) method and Kuersteiner and Prucha (2015) use a generalized method of moments (GMM).[5] However, these estimation methods cannot be directly applied to my model due to the additional simultaneous structure of my framework. In addition, in the above trade and GDP example, if we use a single-equation system to study the effect of trade on GDP, their endogeneity would make the existing estimation methods in these papers inconsistent.

The second type is a spatial model in a simultaneous equation system but without the

[4]The interactive fixed effects have been widely considered in the econometric literature, see Pesaran (2006), Bai (2009), Pesaran and Tosetti (2011), Bai and Li (2014b), and to name a few.

[5]The difference between the first two papers is that Pesaran and Tosetti (2011) specify the spatial interaction of the unobservable errors, while Bai and Li (2014a) specify the spatial interaction of the observable dependent variables. Thus, the CCE method cannot be applied to the model studied in Bai and Li (2014a), due to the endogeneity of the dependent variables. Kuersteiner and Prucha (2015) is based on a dynamic case where the dependent variable also depends on its previous value. In the estimation, Kuersteiner and Prucha (2015) first perform a quasi-transformation to eliminate the common shocks and then implement GMM.

common shock effect. Two estimation methods have been studied for this type of model, instrumental variable (IV) methods (see Kelejian and Prucha (2004), Cohen-Cole et al. (2013), Baltagi and Deng (2015) and Liu (2014))[6] and QML methods (see Baltagi and Bresson (2011), Wang et al. (2014) and Yang and Lee (2017)).[7] However, neither approach can be directly applied to my model due to the additional common shock effect. In all these papers, the errors are assumed to be idiosyncratic (i.e., uncorrelated over time and cross section), which is too strong in applications, and potential correlation of the errors would cause their estimation methods to be inconsistent. Augmenting this type of model with common shock effects can make it reasonable to assume that the new errors are idiosyncratic, since the common shocks would capture the correlations in the original errors, making the new errors idiosyncratic.[8]

In this chapter, I focus on model (1.1.1). I present its estimation method and the corresponding asymptotic properties of the estimators. Under the joint presence of these three effects, there exist a large number of incidental parameters. In addition, I allow for cross-sectional heteroskedasticity in the errors, which is useful and important in spatial models[9] but gives rise to further incidental parameters due to the large number of variance param-

[6]All of these papers focus on cross-sectional data, except Baltagi and Deng (2015), which is based on a panel data setting with random effects.

[7]Baltagi and Bresson (2011) propose a QML method to estimate a spatial seemingly unrelated regression panel data model with spatially correlated errors. Both Wang et al. (2014) and Yang and Lee (2017) are based on cross-sectional data with homoskedasticity. By comparison, Wang et al. (2014) implement a limited QML method without cross-equation correlation of the errors, while Yang and Lee (2017) consider a full information QML method allowing the errors to be correlated across equations.

[8]This common shock effect is an important feature to be implemented by various techniques, as noted in Pesaran (2006), Pesaran and Tosetti (2011), Bai (2009), Bai and Li (2014a) and Castagnetti et al. (2015).

[9]On inference, see Anselin (1988), Lin and Lee (2010), Kelejian and Prucha (2010), Bai and Li (2014a) and Baltagi and Deng (2015).

eters. To estimate the model, I propose two different approaches: a QML method and an iterative generalized principal components (IGPC) method. I show that both methods can effectively deal with the incidental parameters in model (1.1.1). For each method, I derive a full inferential theory for its estimators, which includes consistency, convergence rates and limiting distributions. To investigate finite sample performance, I conduct Monte Carlo simulations. I find that both methods perform well and that the simulation results corroborate the inferential theories derived in this paper. Furthermore, some extensions of the model are discussed. Finally, I apply the model to analyze the causal relationship between trade and GDP, taking into account spatial effects and global common shock effects.

Comparing the two approaches, I show that there is a trade-off between the model specification and the asymptotic property of the estimator. In the QML approach, I specify a model for the explanatory variables assuming that they are also affected by the common shocks and follow a common factor structure. The same specification of the explanatory variables has been considered in many papers; see Pesaran (2006), Bai and Li (2014b) and Castagnetti et al. (2015). Based on the fully specified model of the dependent and explanatory variables, I consider an objective function, which is the likelihood function if the factors and errors are assumed to be i.i.d normal distributed. Since the normality assumption is not required in this paper, this approach is referred to as the QML method. In computing its estimator (QMLE), the expectation maximization (EM) algorithm is implemented. Note that I estimate the sample variance of the common factors instead of the factors themselves. The inferential theory shows that the QMLE is consistent, and its limiting distribution is unbiased (i.e., centered at zero) and has a smaller variance than that of the IGPC estimator. The gains of unbiasedness and more efficiency of the QMLE come at the cost of fully specifying the model of both the dependent and the explanatory variables.

In the IGPC approach, I do not specify the model for the explanatory variables but allow them to be arbitrarily correlated with the common factors and loadings, which is a more general approach than that used in QML. Unlike the treatment of the factors in the QML approach, I treat these as parameters and estimate them directly. In the estimation, I consider an objective function which is the likelihood function if errors are assumed to be i.i.d normal distributed, though such normality is not required in this paper. I then propose the IGPC method which is an iterative method based on the first-order conditions derived from the objective function. I call this estimation procedure the IGPC since one of the first-order conditions involves a generalized principal components method, and the word "generalized" stems from the heteroskedasticity assumption. The IGPC estimator (IGPCE) is consistent. Compared to the QMLE, the limiting distribution of the IGPCE is biased (i.e., not centered at zero) and has a larger variance. The cost of the bias and less efficiency of the IGPCE is offset by the gain of a more flexible model specification for the explanatory variables. In addition, based on the limiting distribution of the IGPCE, a bias-corrected IGPCE is obtained.

In Section 1.6, I apply the model to explore the relationship between trade volume and GDP using a panel data over time and across countries. My model is able to address endogeneity between trade and GDP, which is a well-known problem encountered in analyzing their relationship, as noted in Helpman (1988), Bradford and Chakwin (1993), Rodrik (1995), Winters (2004) and Winters and Masters (2013). Thus far, economists have been making efforts to construct valid IVs for trade to tackle the endogeneity problem. For inferences, see Frankel and Romer (1999), Feyrer (2009), Felbermayr and Groschl (2013) and Ortega and Peri (2014). However, the validity of these IVs is still questionable. Unlike those papers that use a single-equation approach, I study the same type of question by modeling trade and GDP as a system of simultaneous equations and taking into account the endogeneity between them naturally. Moreover, despite their importance, global common shocks have not been well captured in the existing literature, whereas they can be captured using my model through a factor structure. Additionally, my model incorporates the spatial effect through international trade, which is implied from gravity theory as noted in Helpman (1987) and Anderson and van Wincoop (2003). In estimating the model, I implement the IGPC method, which does not need IVs. The empirical results show that all three effects emphasized in the model play important roles: 1) trade and GDP mutually and positively affect each other within a country (i.e., the simultaneous effect); 2) there exist spatial effects across countries for both trade and GDP (i.e., the spatial effect); and 3) global common shocks cannot be ignored. The key finding is that the elasticity of GDP with respect to trade is approximately 0.1, while Feyrer (2009) finds an elasticity of approximately 0.5 using an IV approach.

Related literature. In both the empirical and theoretical literature, many papers consider the three effects separately. First, regarding spatial models, two estimation methods have been considered so far. One is the generalized method of moments (GMM) (see Kelejian and Prucha (1998, 1999, 2010), among others), and the other is the QML method (see Anselin (1988), Lee (2004), Yu et al. (2008), Lee and Yu (2010a,b), Yu and Lee (2010), among others). Spatial models can be applied in many fields, such as spatial propogation of macroeconomic shocks in europe (Dewachter et al. (2012)), propogation of monetary policy shocks through production network (Ozdagli and Weber (2016)), international trade (Baltagi et al. (2008), Lawless (2009), and Rauch and Trindade (2002)), interregional trade (Keller and Shiue (2007)), banking and finance (Arezki et al. (2011) and Korte and Steffen (2014)), public economics (Egger et al. (2005)), transportation research (Frazier and Kockelman (2005)), good demand (Baltagi and Li (2006)), and agricultural economics (Druska and Horrace (2004)), among others.

Second, various methods have been studied for panel data models with common shocks. For instance, Pesaran (2006) propose CCE estimation; both Bai (2009) and Moon and Weidner (2017) consider a principle components (PC) method; Ahn et al. (2013) use GMM, and Bai and Li (2014a, 2015) implement QML. Regarding applications, common shocks models can be used in economic forecasting (Stock and Watson (2002a,b)), time trends modeling (Kneip et al. (2012)), analyzing spillovers in private returns to R&D (Eberhardt et al. (2013)), asset pricing (Bai and Ando (2015a)), and so on.

Third, for simultaneous panel data models, IV approaches have been widely implemented; see Baltagi (1981), Balestra and Varadharajan-Krishnakumar (1987), Cornwell et al. (1992), Baltagi and Li (1992), among others. In practice, simultaneous panel data models can be applied to earnings studies (such as the income-schooling-ability simultaneous equations model considered in Chamberlain (1977a,b), Chamberlain and Griliches (1975), and Griliches (1979)), trade economics (Egger and Pfaffermayr (2004) and Serlenga and Shin (2007)), finance (Chen et al. (2006)) and operational management (Jain et al. (2013)).

In the application of spatial models, although many existing examples are based on a single-equation setup, spatial models with simultaneous equations have received more attention lately and have been widely used in various areas. For instance,these models have been in regional science studies of housing economics (Jeanty et al. (2010); Baltagi and Bresson (2011)); environmental and health economics (Ho and Hite (2008)); the determinants of local growth (interactions among migration, employment and income; see Gebremariam et al. (2011)); fiscal policy analysis (Hauptmeier et al. (2012) focus on fiscal competition over taxes and public input provisions, and Allers and Elhorst (2011) focus on the interactions between governments expenditures); and agricultural economics (Wu and Lin (2010)). Moreover, simultaneous spatial models can be applied in social network studies, such as the multi-choice games in Cohen-Cole et al. (2013), Goldsmith-Pinkham and Imbens (2013) and Liu (2014). It would be potentially useful to apply my model to these areas by allowing common shocks to control for cross-sectional or serial correlations.

This chapter proceeds as follows. I present the QML approach in Section 1.2 and the IGPC method in Section 1.3. In each section, I describe the model specification, assumptions, objective function, first-order conditions, inferential theory and computing algorithm. Then, in Section 1.4, I report the Monte Carlo simulation results for both approaches. Some extensions of the model are considered in Section 1.5, and an application is provided in Section 1.6. Finally, Section 1.7 concludes. Important notation is provided in Appendix A and B, and some proofs are presented in Appendix C. Other technical proofs and additional simulation results are provided in the supplementary material. Throughout the paper, $\|A\|$ is defined as the Frobenius norm of A, where $\|A\| = [\mathrm{tr}(A'A)]^{1/2}$ for any $m \times n$ matrix A. In addition, $\dot{a}_t$ represents the de-meaned version of a column vector a_t, defined as $\dot{a}_t =$

$a_t - \frac{1}{T}\sum_{t=1}^{T} a_t$, and M_{ab} is defined as $M_{ab} = \frac{1}{T}\sum_{t=1}^{T} \dot{a}_t \dot{b}_t'$ for any column vectors a_t and b_t.

1.2 First approach: the QML method

In the first approach, in addition to model (1.1.1), I specify a model of the explanatory variables by assuming that they are also affected by the common shocks and follow a factor structure. Such specification of the explanatory variables is applicable and widely considered, see Pesaran (2006), Bai and Li (2014b) and Castagnetti et al. (2015). Then, based on the fully specified model of both the dependent and explanatory variables, I consider the likelihood-based objective function and propose the QML method. In the estimation, I do not estimate the common factor f_t itself but its sample variance. Further, I develop a full inferential theory of its estimator and provide its computation algorithm. Some simulations results of this QML approach are presented in Section 1.4.

1.2.1 Model description and assumptions

In this section, in addition to model (1.1.1), I specify the model for the explanatory variables assuming that they are affected by the common shocks and following a factor structure of f_t, described as follows:

$$\begin{aligned} x_{1itp} &= \nu_{1ip} + \phi_{1ip}' f_t + v_{1itp}, \quad p = 1, 2, \ldots, k_1 \\ x_{2itq} &= \nu_{2iq} + \phi_{2iq}' f_t + v_{2itq}, \quad q = 1, 2, \ldots, k_2 \end{aligned} \tag{1.2.1}$$

where ϕ_{1ip} is an r-dimensional factor loading, representing the heterogeneous response of x_{1itp} to the common factor f_t; ϕ_{2iq} is defined in a similar way. Therefore, in the first approach, I consider a fully specified model of both dependent and explanatory variables, combining (1.1.1) and (1.2.1).

Let $x_{1it} = (x_{1it1}, x_{1it2}, \ldots, x_{1itk_1})'$, $\beta_1 = (\beta_{11}, \beta_{12}, \ldots, \beta_{1k_1})'$, $\nu_1 = (\nu_{11}, \nu_{12}, \ldots, \nu_{1N})'$, $\phi_{1i} = (\phi_{1i1}, \phi_{1i2}, \ldots, \phi_{1ik_1})$, $v_{1it} = (v_{1it1}, v_{1it2}, \ldots, v_{1itk_1})'$, and define $x_{2it}, \beta_2, \nu_2, \phi_{2i}, v_{2it}$ in a similar way. Then, I can rewrite the model (1.1.1) and (1.2.1) as follows:

$$\begin{aligned}
y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x'_{1it}\beta_1 + \lambda'_i f_t + e_{1it} \\
y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x'_{2it}\beta_2 + \psi'_i f_t + e_{2it} \\
x_{1it} &= \nu_{1i} + \phi'_{1i} f_t + v_{1it} \\
x_{2it} &= \nu_{2i} + \phi'_{2i} f_t + v_{2it}
\end{aligned} \tag{1.2.2}$$

Let $\mu_i = (\alpha_{1i}, \alpha_{2i}, \nu'_{1i}, \nu'_{2i})'$, $L_i = (\lambda_i, \psi_i, \phi_{1i}, \phi_{2i})$, and $\epsilon_{it} = (e_{1it}, e_{2it}, v'_{1it}, v'_{2it})'$. I can rewrite the above model as:

$$\begin{bmatrix} y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} - x'_{1it}\beta_1 \\ y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} - x'_{2it}\beta_2 \\ x_{1it} \\ x_{2it} \end{bmatrix} = \mu_i + L'_i f_t + \epsilon_{it}$$

Denote $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta'_1, \beta'_2)$, $k = k_1 + k_2$, and $\bar{k} = k + 2$. Let $D(\delta)$ be an $N\bar{k} \times N\bar{k}$

matrix whose (i, j) subblock, denoted by $D_{ij}(\delta)$, a $\bar{k} \times \bar{k}$ matrix, is equal to:

$$D_{ij}(\delta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 & -\beta_1' & 0 \\ -\gamma_2 & 1 & 0 & -\beta_2' \\ 0 & 0 & I_{k_1} & 0 \\ 0 & 0 & 0 & I_{k_2} \end{bmatrix} & \text{if } \; i = j \\ \begin{bmatrix} -\rho_1 w_{1ij} & 0 & 0 & 0 \\ 0 & -\rho_2 w_{2ij} & 0 & 0 \\ 0 & 0 & 0_{k_1} & 0 \\ 0 & 0 & 0 & 0_{k_2} \end{bmatrix} & \text{if } \; i \neq j \end{cases} \tag{1.2.3}$$

Now model (1.2.2) can be further transformed into the following matrix form (also a factor-structured model):

$$D(\delta) z_t = \mu + L f_t + \epsilon_t \tag{1.2.4}$$

where $z_t = (z_{1t}, z_{2t}, \ldots, z_{Nt})'$, with $z_{it} = (y_{1it}, y_{2it}, x_{1it}', x_{2it}')'$, $L = (L_1, L_2, \ldots, L_N)'$, $\mu = (\mu_1', \mu_2', \ldots, \mu_N')'$, and $\epsilon_t = (\epsilon_{1t}', \epsilon_{2t}', \ldots, \epsilon_{Nt}')'$. This matrix form will be used throughout the first approach.

Throughout the paper, I assume that the number of factors r is fixed and known. In the simulation section, I propose a modified information criterion based on Bai and Ng (2002) to determine r for each of the two approaches.

Assumptions

To analyze model (1.2.2), I assume that there exists a constant $C > 0$ sufficiently large such that the following assumptions hold.

Assumption A: The factor f_t can be either fixed constants or random variables such that

A.1 Let $\dot{f}_t = f_t - \frac{1}{T}\sum_{t=1}^{T} f_t$, and $M_{ff} = T^{-1}\sum_{t=1}^{T} \dot{f}_t\dot{f}_t'$ be the sample variance of f_t. If f_t is fixed, I assume that $\|f_t\| \leq C$ for all t and $M_{ff} \to \Omega_F$. If f_t are random variables, I assume that $E(\|f_t\|^4) \leq C$ for all i and $M_{ff} \xrightarrow{p} \Omega_F$, where Ω_F is some positive definite matrix.

A.2 If f_t are random variables, I assume f_t to be independent of ϵ_{is} for all t and s.

Assumption B: The loading L_i can be either fixed constants or random variables such that

B.1 If L_i is fixed, I assume that $\|L_i\| \leq C$ for all i and $\frac{1}{N}L'\Sigma_{\epsilon\epsilon}^{-1}L \to \Omega_L$. If L_i are random variables, I assume that $E(\|L_i\|^4) \leq C$ for all i and $\frac{1}{N}L'\Sigma_{\epsilon\epsilon}^{-1}L \xrightarrow{p} \Omega_L$, where $\Sigma_{\epsilon\epsilon}$ is defined in Assumption B, and Ω_L is some positive definite matrix.

B.2 If L_i are random variables, I assume L_i to be independent of the idiosyncratic errors ϵ_{jt} for all i and j.

Assumptions A and B allow both the loadings and the common factors to be either fixed or random, which results in a model that is more general and applicable in various empirical studies.

Assumption C: The idiosyncratic errors $\epsilon_{it} = (e_{1it}, e_{2it}, v_{1it}', v_{2it}')'$ are such that

C.1 e_{lit} is independent and identically distributed over t and uncorrelated over i, with $E(e_{lit}) = 0$ and $E(e_{lit}^8) \leq \infty$ for all $l = 1, 2$, $i = 1, \cdots, N$ and $t = 1, \cdots, T$. Let σ_{li}^2 denote the variance of e_{lit}. I assume $C^{-1} \leq \sigma_{li}^2 \leq C$.

C.2 e_{1it} is independent of e_{2js} for all (i, j, t, s). Let Σ_{iie} denote the variance matrix of $e_{it} = (e_{1it}, e_{2it})'$, so I have $\Sigma_{iie} = \text{diag}(\sigma_{1i}^2, \sigma_{2i}^2)$, a diagonal 2×2 matrix. Let Σ_{ee} denote the variance matrix of $e_t = (e_{1t}', e_{2t}', \ldots, e_{Nt}')'$. Then, $\Sigma_{ee} = \text{diag}(\Sigma_{11e}, \Sigma_{22e}, \ldots, \Sigma_{NNe})$ is a diagonal $2N \times 2N$ matrix.

C.3 v_{lit} is independent and identically distributed over t and uncorrelated over i, with $E(v_{lit}) = 0$ and $E(\|v_{lit}\|^4) \leq \infty$ for all (l, i, t). Let Σ_{iivl} denote the variance matrix of v_{lit} and assume that all eigenvalues of Σ_{iivl} are uniformly bounded (UB) for all l and i. In addition, v_{1it} is independent of v_{2js} for all (i, j, t, s). Let $v_{it} = (v_{1it}, v_{2it})'$ and assume v_{it} is independent of e_{js} for all (i, j, t, s).

C.4 Let Σ_{ii} denote the variance matrix of ϵ_{it}, so I have $\Sigma_{ii} = \text{diag}(\sigma_{1i}^2, \sigma_{2i}^2, \Sigma_{iiv1}, \Sigma_{iiv2})$, a block-diagonal $\bar{k} \times \bar{k}$ matrix, where $\bar{k} = k_1 + k_2 + 2$. Let $\Sigma_{\epsilon\epsilon}$ denote the variance matrix of ϵ_t. Then, $\Sigma_{\epsilon\epsilon} = \text{diag}(\Sigma_{11}, \Sigma_{22}, \ldots, \Sigma_{NN})$ is a block-diagonal $N\bar{k} \times N\bar{k}$ matrix.

Assumption C is that the variance of the idiosyncratic errors ϵ_t is a block-diagonal matrix, extending traditional factor analysis wherein a diagonal matrix is assumed instead. Moreover, Assumption C allows cross-sectional heteroskedasticity, which extends exiting studies with simultaneous spatial models where homoskedasticity is assumed, such as Kelejian and Prucha (2004), Baltagi and Bresson (2011), Wang et al. (2014), Baltagi and Deng (2015), Liu (2014) and Yang and Lee (2017). Note that neither Σ_{iiv1} nor Σ_{iiv2} need be diagonal, meaning that the k_1 components within the error v_{1it} can be correlated with each other.

This is also the case for v_{2it}.

Assumption D: The underlying value $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')'$ satisfies $\|\delta\| \leq C$.

Assumption E: Compactness of estimates.

E.1 The variances Σ_{ii} for all i and M_{ff} are estimated in a compact set, i.e., all the eigenvalues of $\hat{\Sigma}_{ii}$ and $\hat{M}_{ff}$ are in an interval $[C^{-1}, C]$.

E.2 The key parameters $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')'$ are estimated in a compact set $\mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \times \mathcal{A}_4 \times \mathcal{A}_5 \times \mathcal{A}_6 \subset \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{k_1} \times \mathbb{R}^{k_2}$, where $\mathbb{R}$ is the set of real numbers.

Assumption E requires that the variance parameters are estimated in a compact set. Compactness is a condition for theoretical analysis, which is usually used when the objective function is highly nonlinear, for instance, in Newey and McFadden (1994), Jennrich (1969) and Wu (1981). I impose Assumption E here since the objective functions considered in both approaches presented in this paper are highly nonlinear. However, I do not require restrictions on the factor loading L_i.

Assumption F: Aassumptions about some important matrices.

F.1 The transformation matrix $D(\delta)$ is invertible.

F.2 W_1 and W_2 are constant $N \times N$ weights matrices with diagonal elements being zero.

F.3 Let $P_1 = (I_N - \rho_1 W_1)$ and $P_2 = (I_N - \rho_2 W_2)$. Then, I assume all matrices $P_1, P_2, (I - \gamma_1\gamma_2 P_1^{-1} P_2^{-1})$ and $(I - \gamma_1\gamma_2 P_2^{-1} P_1^{-1})$ are invertible.

F.4 Let $B_{12} = (I - \gamma_1\gamma_2 P_1^{-1} P_2^{-1})^{-1}$ and $B_{21} = (I - \gamma_1\gamma_2 P_2^{-1} P_1^{-1})^{-1}$. I assume that the row and column sums of matrices $W_1, W_2, P_1^{-1}, P_2^{-1}, B_{12}$ and B_{21} are all UB in absolute value.

Assumptions F.1–F.4 are standard in the spatial econometrics literature, for instance, Kelejian and Prucha (2004), Lee (2004), Yu et al. (2008), Bai and Li (2014a) and Yang and Lee (2017). The invertibility of $D(\delta)$ (Assumption F.1) is standard in spatial models when using the QML method, which guarantees that the first-order conditions of δ exist and the system has an equilibrium. Assumption F.2 is a standard normalization assumption for weights matrices. Assumption F.3 guarantees the invertibility of key matrices that will be used frequently in the theoretical analysis. The UB condition in Assumption F.4 keeps the degree of spatial correlation manageable and will be used in the theoretical analysis, especially in the consistency analysis.

Remark 1.2.1. In empirical applications, weights can be defined in many ways. Let w_{ij} be the entry of an $N \times N$ weights matrix W. The weight w_{ij} measures the presence and strength of an interaction between location i and j in a geographic space, or more generally, w_{ij} can be interpreted as the strength of a link between nodes i and j or between observations i and j in an economics or social network space. In applications, w_{ij} is usually a decreasing function of distance, as higher weights are assigned to closer observations than to distant observations. The most popular weighting scheme in practice is K-nearest neighbor weights, where location i is only affected by its K-nearest neighbors; more details about this scheme will be given in the simulation section. In the simplest case, when $K = 1$, the weights matrix is binary, where $w_{ij} = 1$ if i and j are neighbors (sharing a common boundary), and $w_{ij} = 0$

otherwise. The choice of weights matrix always depends on the empirical application. [10]

Remark 1.2.2. There is an alternative way to write the UB condition defined in Assumption F.4. First, an equivalent way to say that an m-by-n matrix A is UB in absolute row sum and column sum is to assume that $\limsup_{N\to\infty} \|A\|_\infty < \infty$ and $\limsup_{N\to\infty} \|A\|_1 < \infty$, where $\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|$ represents the maximum absolute row-sum of A; $\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|$ represents the maximum absolute column sum of A, where a_{ij} is the (i,j)th element of A. Second, a set of three conditions $\limsup_{N\to\infty} \|W_1\|_\infty \leq 1$, $\limsup_{N\to\infty} \|W_1\|_1 \leq 1$ and $|\rho_1| < 1$, imply that P_1^{-1} is UB. This is because by definition of $P = (I_N - \rho_1 W_1)$,

$$\limsup_{N\to\infty} \|P^{-1}\|_\infty \leq \limsup_{N\to\infty} \sum_{j=0}^{\infty} (\|\rho_1 W_1\|_\infty)^j \leq \frac{1}{1-\rho_1} < \infty$$

and

$$\limsup_{N\to\infty} \|P^{-1}\|_1 \leq \limsup_{N\to\infty} \sum_{j=0}^{\infty} (\|\rho_1 W_1\|_1)^j \leq \frac{1}{1-\rho_1} < \infty$$

Further, a set of sufficient conditions for the assumption that B_{12} is UB can be that $\limsup_{N\to\infty} \|P_1^{-1}P_2^{-1}\|_\infty \leq 1$, $\limsup_{N\to\infty} \|P_1^{-1}P_2^{-1}\|_1 \leq 1$ and $|\gamma_1\gamma_2| < 1$. Similar arguments can be made for P_2^{-1} and B_{21}.

Assumption G: Let $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$, for all $\eta^\dagger = (\rho_1^\dagger, \rho_2^\dagger, \gamma_1^\dagger, \gamma_2^\dagger,) \in \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \times \mathcal{A}_4$, with $\eta^\dagger \neq \eta$. One of the following two conditions holds:

[10] For example, in geographic spatial models, Ho and Hite (2008) uses a binary weights matrix, where the weight w_{ij} is nonzero only if i and j are neighbors. Jeanty et al. (2010) consider two choices of weights. One defines w_{ij} as a binary distance-based weight, as w_{ij} equals one only if the distance is smaller than a certain distance threshold and zero otherwise. The other defines the weight w_{ij} as an inverse distance function d_{ij}^{-a}, where d_{ij} measures the distance, and a is a dampening coefficient indicating how fast the weight decreases with distance. Furthermore, Cohen-Cole et al. (2013) and Liu (2014) consider the weights in the multi-choice game framework of a social network model.

G.1 For $\beta_1 \neq 0$ and $\beta_2 \neq 0$,

$$\liminf_{N\to\infty} \ \mathbb{M}_a > 0$$

where $\mathbb{M}_a = \begin{bmatrix} a_1 & 0 & b_1 & 0 \\ 0 & a_2 & 0 & b_2 \\ b_1 & 0 & c_1 & 0 \\ 0 & b_2 & 0 & c_2 \end{bmatrix}$, with $a_1, a_2, b_1, b_2, c_1, c_2$ being scalars and functions of $(\rho_1, \rho_2, \gamma_1, \gamma_2)$, depending on N, as defined in Table A.12 in Appendix A.1.[11]

G.2

$$\liminf_{N\to\infty} \ \mathbb{M} > 0$$

where $\mathbb{M}$ is a 4×4 matrix, depending on $N, \eta, \eta^\dagger$ and variances $(\sigma_{1j}^2, \sigma_{2j}^2)$ $(j = 1, 2, \ldots, N)$. Its $(i,j)th$ entry is defined as $\mathbb{M}_{ij} = \frac{1}{N}\text{tr}(M_i M_j')$, where each M_l is an $N\times N$ matrix, for $l = 1, 2, 3, 4$, defined in Table A.12 in Appendix A.1, and $\text{tr}(\cdot)$ is the trace operator.[12] The condition G.2 is equivalent to that, matrix $\mathbb{M}$ is positive definite for all N.

Remark 1.2.3. Assumption G imposes identification conditions for the key coefficients $\delta = (\rho_1, \rho_2, \gamma_1, \gamma_2)'$. Specifically, Assumption G.1 depends on $\beta \neq 0$, while Assumption G.2 does not depend on β. Conditions (G.1) and (G.2) are related to Assumption 8 and 9 respectively in Lee (2004), and the two conditions in Assumption 8 in Yu et al. (2008), but differ in that they impose homoskedasticity and use a single-equation setup, while I allow cross-sectional heteroskedasticity and focus on a simultaneous equation system together with

[11]All $a_1, a_2, b_1, b_2, c_1, c_2$ involve matrices $W_1, W_2, P_1, P_2, B_{12}, B_{21}$ and $\mathbb{G}_l$ $(l = 1, \ldots, 4)$. The $N \times N$ matrices $\mathbb{G}_l$ $(l = 1, \ldots, 4)$ are defined in Table A.11 in Appendix A.1

[12]Compared to the definitions of $a_1, a_2, b_1, b_2, c_1, c_2$, the matrices M_i not only depend on $W_1, W_2, P_1, P_2, B_{12}, B_{21}, \mathbb{G}_l$, but also on the variances $(\sigma_{1j}^2, \sigma_{2j}^2)$ $(j = 1, 2, \ldots, N)$.

common shocks in this paper. Assumption G is also related to Lemmas 2, 3 and 4 in Yang and Lee (2017), but the difference is that they consider a cross-sectional simultaneous spatial model without common shocks and assume homoskedasticity. As shown in Appendix C, condition G.2 is related to the unique solution to $\mathcal{T}_{1N}(\rho_1^\dagger, \rho_2^\dagger, \gamma_1^\dagger, \gamma_2^\dagger, \sigma_{11}^{\dagger 2}, \sigma_{21}^{\dagger 2}, \ldots, \sigma_{1N}^{\dagger 2}, \sigma_{2N}^{\dagger 2}) = 0$, with:

$$\mathcal{T}_{1N}(\rho_1^\dagger, \rho_2^\dagger, \gamma_1^\dagger, \gamma_2^\dagger, \sigma_{11}^{\dagger 2}, \sigma_{21}^{\dagger 2}, \ldots, \sigma_{1N}^{\dagger 2}, \sigma_{2N}^{\dagger 2}) = -\frac{1}{2N}\mathrm{tr}[\mathcal{R}^\dagger \Sigma_{ee} \mathcal{R}^{\dagger\prime} \Sigma_{ee}^{\dagger -1}] + \frac{1}{2N}\ln[\mathcal{R}^\dagger \Sigma_{ee} \mathcal{R}^{\dagger\prime} \Sigma_{ee}^{\dagger -1}] + 1$$

where $\mathcal{R}^\dagger = \Upsilon(\eta^\dagger)\Upsilon(\eta)^{-1}$, with $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$ and $\eta^\dagger = (\rho_1^\dagger, \rho_2^\dagger, \gamma_1^\dagger, \gamma_2^\dagger)$.

Remark 1.2.4. The intuition behind the above identification condition is that, if there are explanatory variables x_1 and x_2, the model can be identified based on condition G.1; if not, the model can still be identified if either spatial effect or cross-sectional heteroskedasticity exists, implied from condition G.2.

Remark 1.2.5. The matrix $\mathbb{M}_a$ in Assumption G.1 is positive-definite if and only if that

$$a_1 > 0, (a_1 c_1 - b_1^2) > 0, a_2 > 0, (a_2 c_2 - b_2^2) > 0.$$

Furthermore, a sufficient condition for $a_1 > 0$ is following denoted as **(GS.1.1)**: there exists a positive constant ε such that at least one of the following two conditions holds:

$$\frac{1}{N}\left[\mathrm{tr}(\Sigma_{1ee}^{-1}\mathbb{G}_1\Sigma_{1ee}\mathbb{G}_1') + \mathrm{tr}(\mathbb{G}_1^2) - 2\sum_{i=1}^{N}[\mathbb{G}_{1,ii}]^2\right] > \varepsilon$$

$$\frac{1}{N}\left[\mathrm{tr}(\Sigma_{1ee}^{-1}W_1\mathbb{G}_4\Sigma_{1ee}\mathbb{G}_4'W_1') + \mathrm{tr}[(W_1\mathbb{G}_4)^2] - 2\sum_{i=1}^{N}[(W_1\mathbb{G}_4)_{ii}]^2\right] > \varepsilon$$

where $\Sigma_{1ee} = \text{diag}(\sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{1N}^2)$; the $N \times N$ matrices $\mathbb{G}_1$ and $\mathbb{G}_4$ are defined in Table A.11 in Appendix A; $\mathbb{G}_{1,ii}$ denotes the (i,i)th entry of matrix $\mathbb{G}_1$, which is similar for $(W_1\mathbb{G}_1)_{ii}$. To see this, it can be shown that the above condition $\frac{1}{N}\Big[\text{tr}(\Sigma_{1ee}^{-1}\mathbb{G}_1\Sigma_{1ee}\mathbb{G}_1') + \text{tr}(\mathbb{G}_1^2) - 2\sum_{i=1}^N[\mathbb{G}_{1,ii}]^2\Big] > \varepsilon$ implies $\frac{1}{2N}\sum_{i=1}^N\sum_{j=1,j\neq i}^N[\mathbb{G}_{1,ij}]^2 > \varepsilon_1$ for some positive constant ε_1. Similarly, the above condition $\frac{1}{N}\Big[\text{tr}(\Sigma_{1ee}^{-1}W_1\mathbb{G}_4\Sigma_{1ee}\mathbb{G}_4'W_1') + \text{tr}[(W_1\mathbb{G}_4)^2] - 2\sum_{i=1}^N[(W_1\mathbb{G}_4)_{ii}]^2\Big] > \varepsilon$ implies that $\frac{1}{2N}\sum_{i=1}^N\sum_{j=1,j\neq i}^N[(W_1\mathbb{G}_4)_{ij}]^2 > \varepsilon_2$ for some positive constant ε_2. Then, summarizing the preceding analysis, together with the definition of $a_1 = \frac{1}{2N}\sum_{i=1}^N\sum_{j=1,j\neq i}^N\Big[[\mathbb{G}_{1,ij}]^2 + [(W_1\mathbb{G}_4)_{ij}]^2\Big]$, it follows that condition **(GS.1.1)** implies $a_1 > 0$. Similar arguments can be made for the other conditions involved in the above sufficient condition for (G.1). More details can be found in the supplementary material.

Normalization conditions for factors and factor loadings

In the factor analysis literature, it is well known that the factors and corresponding loadings can only be identified up to a rotation. The model considered in this paper can be regarded as an extension of the factor model and has the same rotational indeterminacy problem. Thus, in this section, I introduce a set of normalization conditions (NC) for both factors and factor loadings in order to facilitate the inference analysis.

Model (1.2.4) can be alternatively written as follows:

$$\begin{aligned} D(\delta)z_t &= \mu + Lf_t + \epsilon_t \qquad (1.2.5)\\ &= (\mu + L\bar{f}) + L(f_t - \bar{f}) + \epsilon_t \\ &= \underbrace{(\mu + L\bar{f})}_{\mu^\star} + \underbrace{\left(LM_{ff}^{1/2}R\right)}_{L^\star}\underbrace{\left(R'M_{ff}^{-1/2}(f_t - \bar{f})\right)}_{f_t^\star} + \epsilon_t \end{aligned}$$

where R is an orthogonal matrix that consists the eigenvectors of $M_{ff}L'\Sigma_{\epsilon\epsilon}^{-1}LM_{ff}$ arranged in descending order. Let $\mu^\star, L^\star$ and $f_t^\star$ be the new intercepts, new loadings and new factors, respectively, as defined in the above equation. Then, model (1.2.4) is equivalent to:

$$D(\delta)z_t = \mu^\star + L^\star f_t^\star + \epsilon_t$$

where $\frac{1}{T}\sum_{t=1}^{T} f_t^\star = 0$, $\frac{1}{T}\sum_{t=1}^{T} f_t^\star f_t^{\star\prime} = I_r$, and $\frac{1}{N}L^{\star\prime}\Sigma_{\epsilon\epsilon}^{-1}L^\star$ is a diagonal matrix. Therefore, without loss of generality, I can impose the following NC for the factors and factor loadings in model (1.2.4):

NC.1 $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t = 0$

NC.2 $M_{ff} = \frac{1}{T}\sum_{t=1}^{T}(f_t - \bar{f})(f_t - \bar{f})' = I_r$

NC.3 $\frac{1}{N}L'\Sigma_{\epsilon\epsilon}^{-1}L = Q_N$, where Q_N is a diagonal matrix with its distinct diagonal elements arranged in descending order.

Remark 1.2.6. As shown later, NC.3 is not needed for the QML estimation of the regression coefficients δ, but it is needed to identify the factors and factor loadings. Under this NC, the orthogonal matrix R in (1.2.5), which is associated with the rotational indeterminacy of factors and factor loadings, now can be uniquely determined up to a column sign change. In addition, NC.3 simplifies the asymptotic analysis of the QMLE of δ.

Remark 1.2.7. In the factor analysis literature, the above NC are commonly used in maximum likelihood estimation; see, for instance, Anderson (2003). There are other NC to deal with rotational indeterminacy; see Bai and Li (2012) and Bai and Ng (2013).[13] For the QML approach in this paper, different NC will induce different estimates of the sample variance of factors M_{ff} and loadings L, which are the nuisance parameters in this paper, but they will not change the estimates of the key parameters δ and $\Sigma_{\epsilon\epsilon}$.

1.2.2 Objective function and first-order conditions

Let $\theta_1 = (\delta, L, \Sigma_{\epsilon\epsilon})$ be the parameters to be estimated. In this approach, I consider the following objective function:

$$\mathcal{L}_1(\theta_1) = -\frac{1}{2N}\ln|\Sigma_{zz}| + \frac{1}{N}\ln|D| - \frac{1}{2N}\mathrm{tr}[DM_{zz}D'\Sigma_{zz}^{-1}] \tag{1.2.6}$$

where $\Sigma_{zz} = LL' + \Sigma_{\epsilon\epsilon}$; $D = D(\delta)$ is given in equation (1.2.3); and $M_{zz} = \frac{1}{T}\sum_{t=1}^{T}\dot{z}_t\dot{z}_t'$ is the data matrix. The above objective function is the likelihood function if f_t and ϵ_t are assumed to be i.i.d. normal. Without such assumption, function (1.2.6) is referred to as the quasi-likelihood function.[14] The QMLE denoted as $\hat{\theta} = (\hat{\delta}, \hat{L}, \hat{\Sigma}_{\epsilon\epsilon})$ is defined as the maximizer of the above objective function:

$$\hat{\theta} = \underset{\theta_1 \in \Theta_1}{\mathrm{argmax}}\ \mathcal{L}_1(\theta_1)$$

where Θ_1 is the parameter space specified by Assumptions E and G, NC.1, NC.2 and NC.3. By the definition of D, as shown in Lemma A.1, $\det(D) = \det(\Upsilon(\eta))$, where

---

[13] Bai and Li (2012) consider five different sets of identification conditions and derive the inferential theories of the the corresponding QMLEs. Bai and Ng (2013) discuss three different sets of identification conditions for static factors in the PC analysis.

[14] In this paper, such normality assumption of ϵ_t is not required, as shown in both the theoretical analysis and the simulation section, the QML method is robust for different underlying distributions of errors.

$\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)'$, and $\Upsilon(\eta)$ is a $2N \times 2N$ matrix, with its (i, j)th block, a 2×2 matrix, equal to:

$$\Upsilon_{ij}(\eta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{bmatrix} & \text{if } \; i = j \\ \begin{bmatrix} -\rho_1 w_{1ij} & 0 \\ 0 & -\rho_2 w_{2ij} \end{bmatrix} & \text{if } \; i \neq j \end{cases} \tag{1.2.7}$$

Compared to $D = D(\delta)$, $\Upsilon(\eta)$ only depends on η, without involving β_1 and β_2. Replacing $\det(D)$ with $\det(\Upsilon(\eta))$ in (1.2.6) implies the following alternative objective function, which will simplify the derivation of the first-order conditions thereafter:

$$\mathcal{L}_1(\theta_1) = -\frac{1}{2N} \ln |\Sigma_{zz}| + \frac{1}{N} \ln |\Upsilon(\eta)| - \frac{1}{2N} \text{tr}[D M_{zz} D' \Sigma_{zz}^{-1}] \tag{1.2.8}$$

where only the last part, $-\frac{1}{2N}\text{tr}[D M_{zz} D' \Sigma_{zz}^{-1}]$, involves β_1, β_2. Based on the above expression, we can derive the following first-order conditions for θ_1.

The first-order condition for L is:

$$\hat{L}' \hat{\Sigma}_{\epsilon\epsilon}^{-1} (\hat{D} M_{zz} \hat{D}' - \hat{\Sigma}_{zz}) = 0 \tag{1.2.9}$$

where $\hat{D} = D(\hat{\delta})$. The first-order condition for $\Sigma_{\epsilon\epsilon}$ is:

$$\hat{D} M_{zz} \hat{D}' - \hat{\Sigma}_{zz} = \mathbb{W} \tag{1.2.10}$$

where $\mathbb{W}$ is an $N\bar{k} \times N\bar{k}$ matrix ($\bar{k} = k_1 + k_2 + 2$) whose ith $\bar{k} \times \bar{k}$ diagonal subblock denoted as $\mathbb{W}_{ii}$ is such that the diagonal entries of the upper-left 2×2 are zeros. Regarding the lower-right $(k_1+k_2)\times(k_1+k_2)$ submatrix of $\mathbb{W}_{ii}$, all entries of the upper-left $k_1 \times k_1$ and the lower-right $k_2 \times k_2$ are zeros. The rest of the elements of $\mathbb{W}$ are unspecified. The unspecified elements of $\mathbb{W}$ correspond to the zero elements of $\Sigma_{\epsilon\epsilon}$.

The first-order condition for ρ_1 is:

$$\begin{aligned}\frac{1}{N}\mathrm{tr}\Big(\Upsilon(\hat{\eta})^{-1}\cdot\Upsilon_{\rho_1}\Big) &+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\ddot{y}_{1it}}{\hat{\sigma}_{1i}^2}(\dot{y}_{1it}-\hat{\rho}_1\ddot{y}_{1it}-\hat{\gamma}_1\dot{y}_{2it}-\dot{x}_{1it}'\hat{\beta}_1)\\ &- \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\ddot{y}_{1it}}{\hat{\sigma}_{1i}^2}\hat{\lambda}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t = 0\end{aligned} \tag{1.2.11}$$

where $\ddot{y}_{pit} = \sum_{j=1}^{N} w_{pij}\dot{y}_{pjt}$ for $p = 1, 2$, $\hat{G} = (I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L})^{-1}$, and Υ_{ρ_1} is the partial derivative of $\Upsilon(\eta)$ with respect to ρ_1, which is a constant $2N \times 2N$ matrix dependent only on weights W_1. Specifically, the (i, j)th subblock of Υ_{ρ_1} is a 2×2 matrix denoted by $(\Upsilon_{\rho_1})_{ij}$, which equals $0_{2\times2}$ if $i = j$ and $(-w_{1ij}, 0; 0, 0)$ otherwise.

The first-order condition for ρ_2 is:

$$\begin{aligned}\frac{1}{N}\mathrm{tr}\Big(\Upsilon(\hat{\eta})^{-1}\cdot\Upsilon_{\rho_2}\Big) &+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\ddot{y}_{2it}}{\hat{\sigma}_{2i}^2}(\dot{y}_{2it}-\hat{\rho}_2\ddot{y}_{2it}-\hat{\gamma}_2\dot{y}_{1it}-\dot{x}_{2it}'\hat{\beta}_2)\\ &- \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\ddot{y}_{2it}}{\hat{\sigma}_{2i}^2}\hat{\psi}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t = 0\end{aligned} \tag{1.2.12}$$

where Υ_{ρ_2} is the partial derivative of $\Upsilon(\eta)$ with respect to ρ_2, which is a constant $2N \times 2N$ matrix dependent only on weights W_2. Specifically, the (i, j)th subblock of Υ_{ρ_2} is a 2×2 matrix denoted by $(\Upsilon_{\rho_2})_{ij}$ and equal to $0_{2\times2}$ if $i = j$ and $(0, 0; 0, -w_{2ij})$ otherwise.

The first-order condition for γ_1 is:

$$\begin{aligned}\frac{1}{N}\mathrm{tr}\Big(\Upsilon(\hat{\eta})^{-1}\cdot\Upsilon_{\gamma_1}\Big)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\dot{y}_{2it}}{\hat{\sigma}_{1i}^2}(\dot{y}_{1it}-\hat{\rho}_1\ddot{y}_{1it}-\hat{\gamma}_1\dot{y}_{2it}-\dot{x}_{1it}'\hat{\beta}_1)\\ -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\dot{y}_{2it}}{\hat{\sigma}_{1i}^2}\hat{\lambda}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t=0\end{aligned} \tag{1.2.13}$$

where Υ_{γ_1} is the partial derivative of $\Upsilon(\eta)$ with respect to γ_1, which is a constant $2N\times 2N$ matrix. Specifically, the (i,j)th subblock of Υ_{γ_1} is a 2×2 matrix denoted by $(\Upsilon_{\gamma_1})_{ij}$ and equal to $0_{2\times 2}$ if $i\neq j$ and $(0,-1;0,0)$ otherwise.

The first-order condition for γ_2 is:

$$\begin{aligned}\frac{1}{N}\mathrm{tr}\Big(\Upsilon(\hat{\eta})^{-1}\cdot\Upsilon_{\gamma_2}\Big)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\dot{y}_{1it}}{\hat{\sigma}_{2i}^2}(\dot{y}_{2it}-\hat{\rho}_2\ddot{y}_{2it}-\hat{\gamma}_2\dot{y}_{1it}-\dot{x}_{2it}'\hat{\beta}_2)\\ -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\dot{y}_{1it}}{\hat{\sigma}_{2i}^2}\hat{\psi}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t=0\end{aligned} \tag{1.2.14}$$

where Υ_{γ_2} is the partial derivative of $\Upsilon(\eta)$ with respect to γ_2, which is a constant $2N\times 2N$ matrix. Specifically, the (i,j)th subblock of Υ_{γ_2} is a 2×2 matrix denoted by $(\Upsilon_{\gamma_2})_{ij}$ and equal to $0_{2\times 2}$ if $i\neq j$ and $(0,0;-1,0)$ otherwise.

The first-order condition for β_1 is:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{1i}^2}\dot{x}_{1it}(\dot{y}_{1it}-\hat{\rho}_1\ddot{y}_{1it}-\hat{\gamma}_1\dot{y}_{2it}-\dot{x}_{1it}'\hat{\beta}_1)-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{1i}^2}\dot{x}_{1it}\hat{\lambda}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t=0 \tag{1.2.15}$$

The first-order condition for β_2 is:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{2i}^2}\dot{x}_{2it}(\dot{y}_{2it}-\hat{\rho}_2\ddot{y}_{2it}-\hat{\gamma}_2\dot{y}_{1it}-\dot{x}_{2it}'\hat{\beta}_2)-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_{2i}^2}\dot{x}_{2it}\hat{\psi}_i'\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{D}\dot{z}_t=0 \tag{1.2.16}$$

The above first-order conditions are useful in the derivation of the asymptotic properties, including the convergence rate and limiting distributions of the QMLE $\hat{\delta}$. They are involved neither in the proof of consistency nor in the computation of the QMLE. The QMLEs are computed via the expectation maximization (EM) algorithm, which does not need to solve these first-order conditions, but the EM solutions satisfy these conditions (proof is provided in the supplementary material).

1.2.3 Asymptotic properties of the QMLE

In this section, I first show that the QMLE is consistent and then present its convergence rates. Further, I provide the asymptotic representation and limiting distributions of the QMLE.

Proposition 1.2.1. ***(Consistency)*** *Under Assumptions A–G, when $N, T \to \infty$, for $\delta = (\eta', \beta_1', \beta_2')'$, I have:*

$$\hat{\delta}-\delta=o_p(1)$$

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2=o_p(1)$$

In addition, if NC.1–NC.3 hold, I have:

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{L}_i - L_i\|^2 = o_p(1)$$

Remark 1.2.8. In order to derive asymptotic properties, I need to specify $\det(D)$, D^{-1} and DD^{-1}, where D is the high dimensional transformation matrix and makes the theoretical analysis complicated. The number of incidental parameters goes to infinity when $N, T \to \infty$ brings additional complex.

Based on the consistency result, I further derive the rates of convergence of the QMLE.

Theorem 1.2.1. ***(Convergence rates)*** *Under Assumptions A–G, when $N, T \to \infty$, I have:*

$$\hat{\delta} - \delta = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-3/2})$$

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = O_p(T^{-1})$$

In addition, if NC.1–NC.3 hold, I have:

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1}\|\cdot\|\hat{L}_i - L_i\|^2 = O_p(T^{-1})$$

Remark 1.2.9. From Theorem 1.2.1, it can be seen that the QMLE of δ is $\sqrt{T}$-consistent even when N is finite, implying that the QML method still works when N is finite. Under fixed N, however, the asymptotic representation and limiting distribution of the QMLE will change. Theorem 1.2.1 also implies that based on the result that $\hat{\delta}-\delta$ has a faster convergence rate, the limiting distributions of $\mathrm{vec}(\hat{L}_i - L_i)$ and $\mathrm{vech}(\hat{\Sigma}_{ii} - \Sigma_{ii})$ are not affected by the estimation of δ and are the same as those in the pure factor model without regressors. Thus, in the following, I provide only the asymptotic representation of $\hat{\delta}$, excluding the estimated loadings and variances.[15]

In order to state the asymptotic representation of $\hat{\delta}$, I introduce the following notation:

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \dots & \Omega_{16} \\ \Omega_{21} & \Omega_{22} & \dots & \Omega_{26} \\ \dots & \dots & \dots & \dots \\ \Omega_{61} & \Omega_{62} & \dots & \Omega_{66} \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{bmatrix}$$

where the details of each Ω and ε entry are given in Tables A4 and A5, respectively, of Appendix A.1. Then, I have the following theorem.

Theorem 1.2.2. ***(Asymptotic representation)*** *Under Assumptions A–G, when $N, T \to \infty$ and $\sqrt{N}/T \to 0$, I have:*

$$\sqrt{NT}(\hat{\delta} - \delta) = \Omega^{-1}\sqrt{NT}\varepsilon + o_p(1)$$

[15] Bai and Li (2012) provide asymptotic representations and limiting distributions of the QMLE of the loadings and variances.

Remark 1.2.10. The above expression is equivalent to:

$$\sqrt{NT}\begin{bmatrix}\hat{\rho}_1-\rho_1\\ \hat{\rho}_2-\rho_2\\ \hat{\gamma}_1-\gamma_1\\ \hat{\gamma}_2-\gamma_2\\ \hat{\beta}_1-\beta_1\\ \hat{\beta}_2-\beta_2\end{bmatrix}=\begin{bmatrix}\Omega_{11} & \Omega_{12} & \dots & \Omega_{16}\\ \Omega_{21} & \Omega_{22} & \dots & \Omega_{26}\\ \dots & \dots & \dots & \dots\\ \Omega_{61} & \Omega_{62} & \dots & \Omega_{66}\end{bmatrix}^{-1}\cdot\sqrt{NT}\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \varepsilon_4\\ \varepsilon_5\\ \varepsilon_6\end{bmatrix}+o_p(1)$$

From the calculation of Ω and $\text{var}(\sqrt{NT}\varepsilon)$ in the supplementary material, I show that Ω is symmetric and that $\Omega=\text{var}(\sqrt{NT}\varepsilon)$, implying the following corollary.

Corollary 1.2.1. ***(Limiting distribution)*** *Under the assumptions of Theorem 1.2.2, I have:*

$$\sqrt{NT}(\hat{\delta}-\delta)\xrightarrow{d}N(0,\Omega_{QML})$$

where $\Omega_{QML}=\lim_{N\to\infty}\Omega^{-1}$.

Remark 1.2.11. To gain an intuitive understanding of the asymptotic expression in Theorem (1.2.2), consider the following simultaneous spatial panel data model without common shocks:

$$\begin{aligned}y_{1it}&=\alpha_{1i}+\rho_1\sum_{j=1}^{N}w_{1ij}y_{1jt}+\gamma_1 y_{2it}+v_{1it}'\beta_1+e_{1it}\\ y_{2it}&=\alpha_{2i}+\rho_2\sum_{j=1}^{N}w_{2ij}y_{2jt}+\gamma_2 y_{1it}+v_{2it}'\beta_2+e_{2it}\end{aligned}\tag{1.2.17}$$

where e_{1it}, e_{2it}, v_{1it} and v_{2it} satisfy the same conditions as in Assumption C, but v_{1it} and

v_{2it} are assumed to be observable (kind of the regressors). Conditional on v_{1it} and v_{2it}, the quasi-likelihood function of the above model (1.2.17), assuming normality of the errors (after concentrating out α_{1i} and α_{2i}), is:

$$
\begin{aligned}
\mathcal{L}'(\theta) = &- \frac{1}{2N}\sum_{i=1}^{N} \ln \sigma_{1i}^2 - \frac{1}{2N}\sum_{i=1}^{N} \ln \sigma_{2i}^2 + \frac{1}{N} \ln |\Delta(\eta)| \\
&- \frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_{1i}^2}(\dot{y}_{1it} - \rho_1 \ddot{y}_{1it} - \gamma_1 \dot{y}_{2it} - \dot{v}_{1it}'\beta_1)^2 \\
&- \frac{1}{2N}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_{2i}^2}(\dot{y}_{2it} - \rho_2 \ddot{y}_{2it} - \gamma_1 \dot{y}_{1it} - \dot{v}_{2it}'\beta_2)^2
\end{aligned}
\tag{1.2.18}
$$

where $\Delta(\eta) = I_{2N} - P$, with $P = \begin{bmatrix} \rho_1 W_1 & \gamma_1 I_N \\ \gamma_2 I_N & \rho_2 W_2 \end{bmatrix}$, and $\ddot{y}_{pit}$ is defined as in (1.2.11). Let $\tilde{\theta} = (\tilde{\rho}_1, \tilde{\rho}_2, \tilde{\gamma}_1, \tilde{\gamma}_2, \tilde{\sigma}_{11}^2, \ldots, \tilde{\sigma}_{21}^2, \ldots, \tilde{\sigma}_{1N}^2, \tilde{\sigma}_{2N}^2)$ be the QMLE of the above likelihood function. It can be shown that $(\tilde{\delta} - \delta)$ has the same asymptotic representation as in Theorem (1.2.2), which implies that the QML method can help address the endogenous parts of x_{1it} and x_{2it}, as they are affected by the common factors.

Remark 1.2.12. From Corollary 1.2.1, it can be seen that the limiting variance of the QMLE is not of a sandwich form, indicating that the QMLE is asymptotically efficient for simultaneous spatial panel models under cross-sectional heteroskedasticity. However, the situation becomes different when homoskedasticity is imposed instead, where the limiting variance of the QMLE would have a sandwich form. More details follow.

Consider model (1.2.2) but assume homoskedasticity. Then, the asymptotic expression for the QMLE (estimating homoskedastic variances) becomes:

$$\sqrt{NT}(\tilde{\delta} - \delta) = \tilde{\Omega}^{-1}\sqrt{NT}\tilde{\varepsilon} + o_p(1)$$

where $\tilde{\delta}$ is the QMLE of δ under homoskedasticity; $\tilde{\Omega}^{-1}$ and $\tilde{\varepsilon}$ are defined in Table 6 and Table 7, respectively, of Appendix A.1. Note that $\tilde{\Omega}$ and $\tilde{\varepsilon}$ are different from Ω and ε in the heteroskedastic case. More importantly, $\tilde{\varepsilon}$ now involves e_{1it}^2 and e_{2it}^2, while ε does not, implying that the limiting variance of $\tilde{\delta} - \delta$ will depend on the kurtosis of e_{1it} and e_{2it}. However, $\tilde{\Omega}$ does not depend on such kurtosis, so the limiting variance of $\tilde{\delta} - \delta$ has a sandwich form, unless normality of the errors is assumed. As shown in Corollary 1.2.1, the limiting variance of the QMLE under the heteroskedasticity assumption is not of a sandwich form, regardless of normality. This is a meaningful finding, demonstrating two important advantages of imposing the heteroskedasticity assumption. First, it makes the limiting variance of the QMLE robust to the underlying distributions of the errors; second, it eliminates potential inconsistency when homoskedasticity is incorrectly imposed.

1.2.4 Computation of the QMLE

To compute the QMLE, I propose a computing algorithm for the QMLE of model (1.2.2) combining the usual maximization procedures with the EM algorithm. Let $\theta^{(s)} = (\eta^{(s)}, \beta_1^{(s)}, \beta_2^{(s)}, L^{(s)}, \Sigma_{\epsilon\epsilon}^{(s)})$ with $\eta^{(s)} = (\rho_1^{(s)}, \rho_2^{(s)}, \gamma_1^{(s)}, \gamma_2^{(s)})$ denote the estimated value at the sth iteration. My updating procedures consist of two steps. In the first step, I update $L, \Sigma_{\epsilon\epsilon}$, β_1 and β_2 according to the

EM algorithm:

$$L^{(s+1)} = \Big[\frac{1}{T}\sum_{t=1}^{T} E(D\dot{z}_t f_t'|\theta^{(s)})\Big]\Big[\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t'|\theta^{(s)})\Big]^{-1} \tag{1.2.19}$$

$$\begin{aligned}\Sigma_{\epsilon\epsilon}^{(s+1)} &= \mathrm{Dg}\Big[D^{(s)} M_{zz} D^{(s)\prime} - L^{(s+1)} L^{(s)\prime} (\Sigma_{zz}^{(s)})^{-1} D^{(s)} M_{zz} D^{(s)\prime}\Big] \\ &= \mathrm{Dg}\Big\{\Big[I_{N(k_1+k_2+2)} - L^{(s+1)} L^{(s)\prime} (\Sigma_{zz}^{(s)})^{-1}\Big] D^{(s)} M_{zz} D^{(s)\prime}\Big\}\end{aligned} \tag{1.2.20}$$

and

$$\begin{aligned}\beta_1^{(s+1)} = & \Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{(\sigma_{1i}^{(s+1)})^2}\dot{x}_{1it}\dot{x}_{1it}'\Big]^{-1} \\ & \times \Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{(\sigma_{1i}^{(s+1)})^2}\dot{x}_{1it}\Big(\dot{y}_{1it} - \rho_1^{(s)}\sum_{j=1}^{N} w_{1ij}\dot{y}_{1jt} - \gamma_1^{(s)}\dot{y}_{2it} - \lambda_i^{(s+1)\prime} f_t^{(s)}\Big)\Big]\end{aligned} \tag{1.2.21}$$

$$\begin{aligned}\beta_2^{(s+1)} = & \Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{(\sigma_{2i}^{(s+1)})^2}\dot{x}_{2it}\dot{x}_{2it}'\Big]^{-1} \\ & \times \Big[\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{(\sigma_{2i}^{(s+1)})^2}\dot{x}_{2it}\Big(\dot{y}_{2it} - \rho_2^{(s)}\sum_{j=1}^{N} w_{2ij}\dot{y}_{2jt} - \gamma_2^{(s)}\dot{y}_{1it} - \psi_i^{(s+1)\prime} f_t^{(s)}\Big)\Big]\end{aligned} \tag{1.2.22}$$

where Dg is the operator that sets the entries of its argument to zero if their counterparts in $E(\epsilon_t\epsilon_t')$ are zeros; $(\sigma_{1i}^{(s+1)})^2$ is the $[(i-1)(k_1+k_2+2)+1]$th diagonal element of $\Sigma_{\epsilon\epsilon}^{(s+1)}$, and $(\sigma_{2i}^{(s+1)})^2$ is the $[(i-1)(k_1+k_2+2)+2]$th diagonal element of $\Sigma_{\epsilon\epsilon}^{(s+1)}$; $\lambda_i^{(s+1)}$ is the transpose of the $[(i-1)(k_1+k_2+2)+1]$th row of $L^{(s+1)}$, and $\psi_i^{(s+1)}$ is the transpose of the

$[(i-1)(k_1+k_2+2)+2]$th row of $L^{(s+1)}$. In addition:

$$\frac{1}{T}\sum_{t=1}^{T} E(D\dot{z}_t f_t'|\theta^{(s)}) = D^{(s)} M_{zz} D^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} L^{(s)} \tag{1.2.23}$$

$$\frac{1}{T}\sum_{t=1}^{T} E(f_t f_t'|\theta^{(s)}) = I_r - L^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} L^{(s)} + L^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} D^{(s)} M_{zz} D^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} L^{(s)} \tag{1.2.24}$$

and

$$f_t^{(s)} = L^{(s)\prime}(\Sigma_{zz}^{(s)})^{-1} D^{(s)} \dot{z}_t \tag{1.2.25}$$

In the second step, η is updated by maximizing (1.2.6) with respect to η at $\beta_1 = \beta_1^{(s+1)}$,$\beta_2 = \beta_1^{(s+1)}$, $L = L^{(s+1)}$ and $\Sigma_{\epsilon\epsilon} = \Sigma_{\epsilon\epsilon}^{(s+1)}$ with an initial value of η at $\eta^{(s)}$. The two-step procedure suggested above is a version of the Expectation/Conditional Maximization Either (ECME) procedure in Liu and Rubin (1994). Combining these two steps, I obtain $\theta^{(s+1)} = (\eta^{(s+1)}, \beta_1^{(s+1)}, \beta_2^{(s+1)}, L^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)})$. The iteration continues until $\|\theta^{(s+1)} - \theta^{(s)}\|$ is smaller than a preset tolerance.

This two-step iterative procedure guarantees that the value of the objective function (1.2.6) in each iteration does not decrease. This is because in the first step, letting $\eta = \eta^{(s)}$ be fixed and drawing on the standard theory of the EM algorithm, (for the inference, see Dempster et al. (1977) and McLachlan and Krishnan (1997)), I have the following inequality:

$$\mathcal{L}(\eta^{(s)}, \beta_1^{(s+1)}, \beta_2^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}) \geq \mathcal{L}(\eta^{(s)}, \beta_1^{(s)}, \beta_2^{(s)}, \Phi^{(s)}, \Sigma_{\epsilon\epsilon}^{(s)}) \tag{1.2.26}$$

In the second step, by the definition of $\eta^{(s+1)}$, I have the following inequality:

$$\mathcal{L}(\eta^{(s+1)}, \beta_1^{(s+1)}, \beta_2^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}) \geq \mathcal{L}(\eta^{(s)}, \beta_1^{(s+1)}, \beta_2^{(s+1)}, \Phi^{(s+1)}, \Sigma_{\epsilon\epsilon}^{(s+1)}) \tag{1.2.27}$$

In the supplementary material, I show that the limit of the iterated solution satisfies the first-order conditions (1.2.9)–(1.2.16) and hence possesses the local optimality property.

In the simulation results reported in the next section, I use the within-group estimator as the starting value for $\eta^{(1)}, \beta_1^{(1)}, \beta_2^{(1)}$, ignoring the endogeneity problem and the common shock effect. Then, let the initial values of $L^{(1)}$ and $\Sigma_{\epsilon\epsilon}^{(1)}$ be the maximizer of (1.2.6) given $\eta = \eta^{(1)}$, $\beta_1 = \beta_1^{(1)}$ and $\beta_2 = \beta_2^{(1)}$.

1.3 Second approach: the IGPC method

In the second approach, I do not specify the model for the explanatory variables but allow them to be arbitrarily correlated with the common factors, the factor loadings or both, which is more general than the model specification considered in the first approach. Regarding the common factor f_t, I treat it as parameter and estimate it directly instead of estimating its sample variance as in the first approach. For this estimation, I propose an iterative approach based on a generalized principal components (GPC) method. Furthermore, I derive a full inferential theory of its estimator, the IGPCE, as in the first approach. Finally, I describe the computation of the IGPCE. The simulation results are provided in Section 4.

1.3.1 Model description and assumptions

In the second approach, I study model (1.1.1) without specifying the model for the explanatory variables. Using the same definitions of $x_{1it}, x_{2it}, \beta_1$ and β_2 given in (1.2.2), I can rewrite (1.1.1) as follows:

$$\begin{aligned} y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x_{1it}' \beta_1 + \lambda_i' f_t + e_{1it} \\ y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x_{2it}' \beta_2 + \psi_i' f_t + e_{2it} \end{aligned} \tag{1.3.1}$$

Let $\alpha_i = (\alpha_{1i}, \alpha_{2i})'$, $x_{it} = \begin{bmatrix} x_{1it} & 0 \\ 0 & x_{2it} \end{bmatrix}$, $\beta = (\beta_1', \beta_2')'$, $\Gamma_i = (\lambda_i, \psi_i)$, and $e_{it} = (e_{1it}, e_{2it})'$. I can then rewrite model (1.3.1) as:

$$\begin{bmatrix} y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} \\ y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} \end{bmatrix} = \alpha_i + x_{it}' \beta + \Gamma_i' f_t + e_{it}$$

Using the same notation, $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$ and $\Upsilon(\eta)$, as in (1.2.7) in the first approach and letting $y_{it} = (y_{1it}, y_{2it})'$, model (1.3.1) can be transformed to:

$$\sum_{j=1}^{N} \Upsilon_{ij}(\eta) y_{jt} = \alpha_i + x_{it}' \beta + \Gamma_i' f_t + e_{it} \tag{1.3.2}$$

Finally, let $Y_t = (y_{1t}', y_{2t}', \ldots, y_{Nt}')'$, $X_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$, $\alpha = (\alpha_1', \alpha_2', \ldots, \alpha_N')'$, $\Gamma = (\Gamma_1, \Gamma_2, \ldots, \Gamma_N)'$, and $e_t = (e_{1t}', e_{2t}', \ldots, e_{Nt}')'$. I can then rewrite model (1.3.1) in the following matrix form:

$$\Upsilon(\eta) Y_t = \alpha + X_t \beta + \Gamma f_t + e_t \tag{1.3.3}$$

Assumptions

In addition to the assumptions made in Section 2.1.1 in the first approach, I impose the following additional assumptions to facilitate the analysis in this second approach. Assume that there is a sufficiently large constant $C > 0$ such that the following assumptions hold.

Assumption H: The explanatory variables x_{lit} can be either fixed constants or random variables for $l = 1, 2$. If x_{lit} is fixed, I assume $\|x_{lit}\| \leq C$ for all i and t. If x_{lit} are random variables, I assume $E(\|x_{lit}\|^4) \leq C$ for all i and t; in addition, x_{lit} is independent of the idiosyncratic error e_{mjs} for all (l, m, i, j, t, s).

Assumption H is newly imposed on the explanatory variables, since in this approach, I do not specify a model for them. To analyze model (1.3.1), I need to make the above assumption.

Assumption A$'$.2: If f_t are random variables, I assume f_t is independent of e_{is} for all t and s.

Assumption B$'$: The loading Γ_i can be either fixed constants or random variables such that

B$'$.1 If Γ_i is fixed, I assume that $\|\Gamma_i\| \leq C$ for all i and $\frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma \to \Omega_\Gamma$. If Γ_i are random variables, I assume that $E(\|\Gamma_i\|^4) \leq C$ for all i and $\frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma \xrightarrow{p} \Omega_\Gamma$, where Σ_{ee} is defined in Assumption C, and Ω_Γ is some positive definite matrix.

B$'$.2 If Γ_i are random variables, I assume that Γ_i is independent of the idiosyncratic errors e_{jt} for all i and j.

Assumption B$'$ is similar to Assumption B, but it is based on the new loading Γ, which is part of the loading L in the first approach. Since L contains Γ, Assumption B.2 implies Assumption B$'$.2 but not vice versa. However, Assumption B.1 cannot imply Assumption B$'$.1, and vice versa.

Assumption E$'$: Compactness of the estimates.

E$'$.1 The variances σ_{1i} and σ_{2i} for $i = 1, 2, \dots, N$ are all estimated in compact sets, i.e., all variances σ_{1i} and σ_{2i} are estimated in an interval $[C^{-1}, C]$.

Assumption E.1 implies Assumption E$'$.1, since σ_{1i} and σ_{2i} are parts of Σ_{ii}. However, I do not need the compactness assumption of the estimate of M_{ff} here because in the second approach, I estimate the factor f_t itself instead of M_{ff}. Moreover, the compactness of the estimate of f_t is not required due to the nature of this estimation approach.

To state the following Assumption G$'$, let $\Im$ be the parameter space for Γ and Σ_{ee}, satisfying the assumptions and NC (which will be included in the following Section 3.1.2):

$$\Im = \left\{ \theta = (\Gamma, \Sigma_{ee}) \middle| C^{-1} \leq \sigma_{1i}^2 \leq C, C^{-1} \leq \sigma_{2i}^2 \leq C, \forall i; \frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma = I_r \right\}$$

Assumption G$'$: One of the following two conditions holds:

G$'$.1 For $\beta_1 \neq 0$ and $\beta_2 \neq 0$, the matrix $\mathbb{D}_a = \begin{bmatrix} \mathbb{D}_b & \zeta \\ \zeta' & \mathbb{D}_c \end{bmatrix}$ is positive definite on $\Im$ for all N, where the $k \times k$ matrix $\mathbb{D}_b$, 4×4 matrix $\mathbb{D}_c$ and $k \times 4$ matrix ζ are all defined in Appendix A.2.

G$'$.2 For all $\eta^\dagger = (\rho_1^\dagger, \rho_2^\dagger, \gamma_1^\dagger, \gamma_2^\dagger,) \in \mathcal{A}_1 \times \mathcal{A}_2 \times \mathcal{A}_3 \times \mathcal{A}_4$, with $\eta^\dagger \neq \eta$, both $\mathbb{M}$ and $\mathbb{D}_b$ are positive definite on $\Im$ for all N, where the 4×4 matrix $\mathbb{M}$ is defined the same as in Assumption G.2.

Remark 1.3.1. The intuition behind Assumption G$'$ is similar to that behind Assumption G. In addition, the first part of condition G$'$.2 involving matrix $\mathbb{M}$ is same as Assumption G.2. However, Assumption G$'$.2 includes additional positive definite condition on $\mathbb{D}_b$ because now I estimate the factor f_t itself instead of its sample variance, which introduces more incidental parameters.

Normalization conditions for factors and factor loadings

As in Section 2.1.2, I introduce a set of NC to facilitate the inference analysis in the second approach. Note that model (1.3.3) can always be written as:

$$\Upsilon(\eta)Y_t = \underbrace{(\alpha + \Gamma\bar{f})}_{\alpha^\dagger} + X_t\beta + \underbrace{\Gamma Q^{-1/2}}_{\Gamma^\dagger}\underbrace{Q^{1/2}(f_t - \bar{f})}_{f_t^\dagger} + e_t$$

where $Q = \frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma$, and $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t$. Using the definitions of $\alpha^\dagger$, $\Gamma^\dagger$ and $f_t^\dagger$ given in the above expression, I can treat them as the new intercept, new loading and new factor, respectively. Then, it can be seen that $\sum_{t=1}^{T} f_t^\dagger = 0$, and $\frac{1}{N}\Gamma^{\dagger\prime}\Sigma_{ee}^{-1}\Gamma^\dagger = I_r$. Thus, without loss of generality, in addition to NC.1 stated in Section 2.1.2, I can impose the following NC:

NC.4: $\frac{1}{N}\Gamma'\Sigma_{ee}^{-1}\Gamma = I_r$, where Σ_{ee} is defined in Assumption C.2.

1.3.2 Objective function and first-order conditions

In this approach, I allow the explanatory variables x_{1it} and x_{2it} to be arbitrarily correlated with the loading Γ_i and factor f_t. I treat both Γ_i and f_t as parameters and estimate them together.

Using the same definitions of $\delta, \eta, \Sigma_{ee}$ and $\Upsilon(\eta)$ as in the first approach, let $F = (f_1, f_2, \ldots, f_T)'$ and $\theta_2 = (\delta, \Gamma, \Sigma_{ee})$. I thus consider the following objective function in this approach:

$$\begin{aligned}\mathcal{L}_2^*(\theta_2, \alpha, F) = &- \frac{1}{2NT}\sum_{t=1}^{T}\Big(\Upsilon(\eta)Y_t - \alpha - X_t\beta - \Gamma f_t\Big)'\Sigma_{ee}^{-1}\Big(\Upsilon(\eta)Y_t - \alpha - X_t\beta - \Gamma f_t\Big)\\ &- \frac{1}{2N}\ln|\Sigma_{ee}| + \frac{1}{N}\ln|\Upsilon(\eta)|\end{aligned} \tag{1.3.4}$$

The above expression can be viewed as the quasi-likelihood function by assuming the normality of e_{it}. Given δ, Γ and Σ_{ee}, it is easy to see that α and f_t maximize the above function $\mathcal{L}_2^*(\theta_2, \alpha, F)$ at:

$$\alpha = \Upsilon(\eta)\bar{Y} - \bar{X}\beta - \Gamma\bar{f} \tag{1.3.5}$$

and

$$f_t = (\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}\Gamma'\Sigma_{ee}^{-1}(\Upsilon(\eta)\dot{Y}_t - \dot{X}_t\beta) \tag{1.3.6}$$

where $\bar{Y} = \frac{1}{T}\sum_{t=1}^{T} Y_t$, $\bar{X} = \frac{1}{T}\sum_{t=1}^{T} X_t$, $\bar{f} = \frac{1}{T}\sum_{t=1}^{T} f_t$, $\dot{Y}_t = Y_t - \bar{Y}$, and $\dot{X}_t = X_t - \bar{X}$. Substituting the above two formulas into $\mathcal{L}_2^*(\theta_2, \alpha, F)$ to concentrate out α and f_t, the objective function becomes:

$$\mathcal{L}_2(\theta_2) = -\frac{1}{2NT}\sum_{t=1}^{T}\Big(\Upsilon(\eta)\dot{Y}_t - \dot{X}_t\beta\Big)'\ddot{M}\Big(\Upsilon(\eta)\dot{Y}_t - \dot{X}_t\beta\Big) - \frac{1}{2N}\ln|\Sigma_{ee}| + \frac{1}{N}\ln|\Upsilon(\eta)| \tag{1.3.7}$$

where $\ddot{M} = \Sigma_{ee}^{-1} - \Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}\Gamma'\Sigma_{ee}^{-1} = \Sigma_{ee}^{-1} - \frac{1}{N}\Sigma_{ee}^{-1}\Gamma\Gamma'\Sigma_{ee}^{-1}$ with the second equality due to NC.4.

Let $\tilde{\theta}_2 = (\tilde{\delta}, \tilde{\Gamma}, \tilde{\Sigma}_{ee})$ be the maximizer of the above objective function, defined as:

$$\tilde{\theta}_2 = \underset{\theta_2 \in \Theta_2}{\text{argmax}}\ \mathcal{L}_2(\theta_2)$$

where Θ_2 is the parameter space specified by Assumptions E.2, E$'$.1, G$'$, NC.1 and NC.4.

Based on the above $\mathcal{L}_2(\theta_2)$, the first-order conditions for θ_2 can be derived as following (1.3.8)-(1.3.15). To compute $\tilde{\theta}_2$, I propose an iterative estimation procedure based on these first-order conditions. Since the first-order condition for loading Γ_i involves the generalized principal components (GPC) methodology, this estimation approach is referred to as an iterative generalized principal components method (IGPC) and its estimator (IGPCE) is denoted as $\check{\theta} = (\check{\delta}, \check{\Gamma}, \check{\Sigma}_{ee})$. The following are the first-order conditions for θ_2.

The first-order condition for Γ is:

$$\left[\frac{1}{NT}\sum_{t=1}^{T}\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\right)\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\right)'\right]\check{\Sigma}_{ee}^{-1}\check{\Gamma} = \check{\Gamma}\check{V} \tag{1.3.8}$$

where $\check{V}$ is a diagonal $r \times r$ matrix consisting of the first r largest eigenvalues of the $2N \times 2N$ matrix $\check{D}_{\Gamma} = \frac{1}{NT}\sum_{t=1}^{T}\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\right)\left(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\right)'\check{\Sigma}_{ee}^{-1}$. Here, $\check{\Gamma}$ contains the r eigenvectors associated with these r eigenvalues in $\check{V}$. Thus, the computation algorithm using the above equation is referred to as the GPC method, where the word "generalized" stems from the assumption of heteroskedasticity of the errors.

The first-order condition for σ_{1i}^2 is:

$$\check{\sigma}_{1i}^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\dot{y}_{1it} - \check{\rho}_1\ddot{y}_{1it} - \check{\gamma}_1\dot{y}_{2it} - \dot{x}_{1it}'\check{\beta}_1 - \check{\lambda}_i'\check{f}_t\right)^2 \tag{1.3.9}$$

and the first-order condition for σ_{2i}^2 is:

$$\check{\sigma}_{2i}^2 = \frac{1}{T}\sum_{t=1}^{T}\Big(\dot{y}_{2it} - \check{\rho}_2\ddot{y}_{2it} - \check{\gamma}_2\dot{y}_{1it} - \dot{x}_{2it}'\check{\beta}_2 - \check{\psi}_i'\check{f}_t\Big)^2 \tag{1.3.10}$$

where $\dot{y}_{1it} = y_{1it} - \frac{1}{T}\sum_{s=1}^{T} y_{1is}$, $\dot{y}_{2it}$, $\dot{x}_{1it}$ and $\dot{x}_{2it}$ are defined in a similar way; $\ddot{y}_{1it} = \sum_{j=1}^{N} w_{1ij}\dot{y}_{1jt}$, and $\ddot{y}_{2it} = \sum_{j=1}^{N} w_{2ij}\dot{y}_{2jt}$; and

$$\check{f}_t = (\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\check{\Gamma})^{-1}\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\Big(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\Big) = \frac{1}{N}\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\Big(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\Big)$$

The first-order condition for ρ_1 is:

$$-\frac{1}{NT}\sum_{t=1}^{T}\Big(\Upsilon_{\rho_1}\cdot\dot{Y}_t\Big)'\widehat{\ddot{M}}\Big(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\Big) + \frac{1}{N}\mathrm{tr}\Big\{\Upsilon(\check{\eta})^{-1}\cdot\Upsilon_{\rho_1}\Big\} = 0 \tag{1.3.11}$$

where Υ_{ρ_1} is defined in (1.2.11) and $\widehat{\ddot{M}} = \check{\Sigma}_{ee}^{-1} - \check{\Sigma}_{ee}^{-1}\check{L}(\check{L}'\check{\Sigma}_{ee}^{-1}\check{L})^{-1}\check{L}'\check{\Sigma}_{ee}^{-1} = \check{\Sigma}_{ee}^{-1} - \frac{1}{N}\check{\Sigma}_{ee}^{-1}\check{L}\check{L}'\check{\Sigma}_{ee}^{-1}$. The first-order condition for ρ_2 is:

$$-\frac{1}{NT}\sum_{t=1}^{T}\Big(\Upsilon_{\rho_2}\cdot\dot{Y}_t\Big)'\widehat{\ddot{M}}\Big(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\Big) + \frac{1}{N}\mathrm{tr}\Big\{\Upsilon(\check{\eta})^{-1}\cdot\Upsilon_{\rho_2}\Big\} = 0 \tag{1.3.12}$$

where Υ_{ρ_2} is defined in (1.2.12). The first-order condition for γ_1 is:

$$-\frac{1}{NT}\sum_{t=1}^{T}\Big(\Upsilon_{\gamma_1}\cdot\dot{Y}_t\Big)'\widehat{\ddot{M}}\Big(\Upsilon(\check{\eta})\dot{Y}_t - \dot{X}_t\check{\beta}\Big) + \frac{1}{N}\mathrm{tr}\Big\{\Upsilon(\check{\eta})^{-1}\cdot\Upsilon_{\gamma_1}\Big\} = 0 \tag{1.3.13}$$

where Υ_{γ_1} is defined in (1.2.13). The first-order condition for γ_2 is:

$$-\frac{1}{NT}\sum_{t=1}^{T}\Big(\Upsilon_{\gamma_2}\cdot\dot{Y}_t\Big)'\widehat{\ddot{M}}\Big(\Upsilon(\check{\eta})\dot{Y}_t-\dot{X}_t\check{\beta}\Big)+\frac{1}{N}\mathrm{tr}\Big\{\Upsilon(\check{\eta})^{-1}\cdot\Upsilon_{\gamma_2}\Big\}=0 \tag{1.3.14}$$

where Υ_{γ_2} is defined in (1.2.14). The first-order condition for $\beta=(\beta_1',\beta_2')'$ is:

$$\frac{1}{NT}\sum_{t=1}^{T}\dot{X}_t'\widehat{\ddot{M}}\Big(\Upsilon(\check{\eta})\dot{Y}_t-\dot{X}_t\check{\beta}\Big)=0 \tag{1.3.15}$$

More details about the computation of the IGPCE are given in Section 3.4. These first-order conditions will be used in the derivation of the asymptotic properties of the IGPCE in Section 3.3.

1.3.3 Asymptotic properties of the IGPCE

In this section, I first show that the IGPCE is consistent and then derive its convergence rates, asymptotic representation and limiting distributions.

Proposition 1.3.1. ***(Consistency)*** *Under Assumptions A.1, A$'$.2, B$'$, C.1, C.2, D, E$'$.1, E.2, F, G$'$ and H, when $N,T\to\infty$, I have:*

$$\check{\delta}-\delta=o_p(1)$$

$$\frac{1}{N}\sum_{i=1}^{N}||\check{\Sigma}_{iie}-\Sigma_{iie}||^2=o_p(1)$$

$$\frac{1}{N}\Gamma'\widehat{\ddot{M}}\Gamma=o_p(1)$$

where $\delta = (\eta', \beta')' = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')'$, $\Sigma_{iie} = \text{diag}(\sigma_{1i}^2, \sigma_{2i}^2)$, *and*
$\widehat{\ddot{M}} = \check{\Sigma}_{ee}^{-1} - \check{\Sigma}_{ee}^{-1}\check{\Gamma}(\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\check{\Gamma})^{-1}\check{\Gamma}'\check{\Sigma}_{ee}^{-1} = \check{\Sigma}_{ee}^{-1} - \frac{1}{N}\check{\Sigma}_{ee}^{-1}\check{\Gamma}\check{\Gamma}'\check{\Sigma}_{ee}^{-1}$.

Based on the consistency result, I derive the rates of convergence.

Theorem 1.3.1. ***(Convergence rates)*** *Let* $H = \frac{1}{NT}\check{V}^{-1}(\check{\Gamma}'\check{\Sigma}_{ee}^{-1}\Gamma)(F'F)$. *Under Assumptions A.1, A′.2, B′, C.1, C.2, D, E′.1, E.2, F, G′ and H, when* $N, T \to \infty$, *I have:*

$$\begin{aligned} \check{\delta} - \delta =& O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-3/2}) \\ \frac{1}{N}\sum_{i=1}^{N}||\check{\Sigma}_{iie} - \Sigma_{iie}||^2 =& O_p(N^{-2}) + O_p(T^{-1}) \\ \frac{1}{N}\sum_{i=1}^{N}||\check{\Gamma}_i - H\Gamma_i||^2 =& O_p(N^{-2}) + O_p(T^{-1}) \end{aligned}$$

Remark 1.3.2. In the convergence rate for $\hat{\check{\delta}}$, note that there is a bias term of order $O_p(N^{-1})$, but there is no such bias term in the QMLE in the first approach. This bias term comes from the additional incidental parameters involved in the treatment of the common shocks. In this approach, I treat f_t as parameter and estimate it directly, whereas I estimate its sample variance instead in the first approach. Similarly, the extra term $O_p(N^{-2})$ included in the average convergence rates for $\check{\Sigma}_{iie}$ and $\check{\Gamma}_i$ occurs for the same reason. Because these extra terms depend only on N, the IGPCE is no longer consistent under fixed N, which is different from the QMLE. However, it is still true that since $\check{\delta} - \delta$ has a faster convergence rate, the limiting distributions of $\text{vec}(\check{\Gamma}_i - \Gamma_i)$ and $\text{vech}(\check{\Sigma}_{iie} - \Sigma_{iie})$ are the same as in the case of no regressors.

Theorem 1.3.2. ***(Asymptotic representation)*** *Under Assumptions A.1, A′.2, B′, C.1,* *C.2, D, E′.1, E.2, F, G′ and H, when $N, T \to \infty$ and $\sqrt{N}/T \to 0$, $\sqrt{T}/N \to 0$, I have:*

$$\sqrt{NT}(\check{\delta} - \delta + b) = \mathbb{D}^{-1}\sqrt{NT}\xi + o_p(1)$$

where $\delta = (\eta', \beta')' = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')'$; $b, \mathbb{D}$ and ξ are defined as follows.

Let $k = k_1 + k_2$ and $\tilde{k} = k + 4$, and the $\tilde{k} \times 1$ vector ξ is defined as:

$$\xi = \frac{1}{NT} \begin{bmatrix} \sum_{t=1}^T (\dot{X}_t\beta + \Gamma f_t)' Q_1' \ddot{M} e_t - \sum_{t=1}^T \sum_{s=1}^T (\dot{X}_t\beta + \Gamma f_t)' Q_1' \ddot{M} e_s \pi_{st} + \varphi_1 \\ \sum_{t=1}^T (\dot{X}_t\beta + \Gamma f_t)' Q_2' \ddot{M} e_t - \sum_{t=1}^T \sum_{s=1}^T (\dot{X}_t\beta + \Gamma f_t)' Q_2' \ddot{M} e_s \pi_{st} + \varphi_2 \\ \sum_{t=1}^T (\dot{X}_t\beta + \Gamma f_t)' Q_3' \ddot{M} e_t - \sum_{t=1}^T \sum_{s=1}^T (\dot{X}_t\beta + \Gamma f_t)' Q_3' \ddot{M} e_s \pi_{st} + \varphi_3 \\ \sum_{t=1}^T (\dot{X}_t\beta + \Gamma f_t)' Q_4' \ddot{M} e_t - \sum_{t=1}^T \sum_{s=1}^T (\dot{X}_t\beta + \Gamma f_t)' Q_4' \ddot{M} e_s \pi_{st} + \varphi_4 \\ \sum_{t=1}^T \dot{X}_t' \ddot{M} e_t - \sum_{t=1}^T \sum_{s=1}^T \dot{X}_t' \ddot{M} e_s \pi_{st} \end{bmatrix}$$

where $\ddot{M} = \Sigma_{ee}^{-1} - \frac{1}{N}\Sigma_{ee}^{-1}\Gamma\Gamma'\Sigma_{ee}^{-1}$; $\pi_{st} = f_s'(F'F)^{-1}f_t$; $Q_1 = -\Upsilon_{\rho_1}\Upsilon(\eta)^{-1}$; $Q_2 = -\Upsilon_{\rho_2}\Upsilon(\eta)^{-1}$; $Q_3 = -\Upsilon_{\gamma_1}\Upsilon(\eta)^{-1}$; and $Q_4 = -\Upsilon_{\gamma_2}\Upsilon(\eta)^{-1}$, with all $\Upsilon_{\rho_1}, \Upsilon_{\rho_2}, \Upsilon_{\gamma_1}, \Upsilon_{\gamma_2}$ being constant matrices defined as in the first-order conditions (1.2.11)–(1.2.14) in Section 2.2. For $p = 1, 2, 3, 4$, the scalar φ_p is defined as:

$$\varphi_p = \frac{1}{NT}\sum_{t=1}^T e_t' Q_p^{o\prime} \Sigma_{ee}^{-1} e_t$$

where Q_p^o is an $2n \times 2N$ matrix that is obtained by setting all the diagonal elements of Q_p

to zero. The $k \times 1$ vector b is defined as:

$$b = \mathbb{D}^{-1} \begin{bmatrix} \frac{1}{N}\mathrm{tr}[\Gamma' Q_1^o \Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}] \\ \frac{1}{N}\mathrm{tr}[\Gamma' Q_2^o \Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}] \\ \frac{1}{N}\mathrm{tr}[\Gamma' Q_3^o \Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}] \\ \frac{1}{N}\mathrm{tr}[\Gamma' Q_4^o \Sigma_{ee}^{-1}\Gamma(\Gamma'\Sigma_{ee}^{-1}\Gamma)^{-1}] \\ 0_{k\times 1} \end{bmatrix} = \mathbb{D}^{-1} \begin{bmatrix} \frac{1}{N^2}\mathrm{tr}[\Gamma' Q_1^o \Sigma_{ee}^{-1}\Gamma] \\ \frac{1}{N^2}\mathrm{tr}[\Gamma' Q_2^o \Sigma_{ee}^{-1}\Gamma] \\ \frac{1}{N^2}\mathrm{tr}[\Gamma' Q_3^o \Sigma_{ee}^{-1}\Gamma] \\ \frac{1}{N^2}\mathrm{tr}[\Gamma' Q_4^o \Sigma_{ee}^{-1}\Gamma] \\ 0_{k\times 1} \end{bmatrix}$$

The $\tilde{k} \times \tilde{k}$ matrix $\mathbb{D}$ is defined as:

$$\mathbb{D} = \begin{bmatrix} \mathbb{D}_\eta + \Phi & \vartheta \\ \vartheta' & \mathbb{D}_\beta \end{bmatrix}$$

where the 4×4 matrices $\mathbb{D}_\eta$ and Φ, the $4 \times k$ matrix ϑ and the $k \times k$ matrix $\mathbb{D}_\beta$ are defined in Appendix A.2.

Remark 1.3.3. There is a bias term b of order $O_p(\frac{1}{N})$ in the IGPCE $\check{\delta}$ due to the treatment of the common shocks. The IGPC approach estimates f_t itself instead of its sample variance, which introduces more incidental parameters to the time dimension. As a comparison, the first approach estimates the sample variance of f_t instead, and its corresponding QMLE is unbiased in terms of limiting distribution.

Remark 1.3.4. There is some connection between matrix $\mathbb{D}$ in the above theorem and $\mathbb{D}_a$ in Assumption G′. Note that all matrices $\tilde{X}_l$ $(l = 1, 2, \ldots, \tilde{k})$ involved in $\mathbb{D}_a$ and $(\dot{X}_{1p}, \dot{X}_{2q})$ $(p = 1, 2, \ldots, k_1; q = 1, 2, \ldots, k_2)$ involved in $\mathbb{D}$ are defined on the explanatory variables, different only in the ordering and presentation. Thus, I can simplify $\mathbb{D}_b$ (part of $\mathbb{D}_a$, which is associated with the identification of β) to $\mathbb{D}_\beta$ (part of $\mathbb{D}$, which is associated with the limiting variance of $\check{\beta}$). However, the matrix $\mathbb{D}_c$ (part of $\mathbb{D}_a$), which depends on the explanatory variables, does not equal $\mathbb{D}_\eta$ (part of $\mathbb{D}$), which depends on the dependent variables.

In the supplementary material, I show that $\mathbb{D}^{-1/2}\sqrt{NT}\xi \xrightarrow{d} N(0, I_{\tilde{k}})$ under the same conditions as in Theorem 1.3.2, which implies the following corollary.

Corollary 1.3.1. ***(Limiting distribution)*** *Under the assumptions of Theorem 1.3.2, when $N, T \to \infty$ and $T/N \to \kappa > 0$, I have:*

$$\sqrt{NT}(\hat{\delta} - \delta) \xrightarrow{d} N(-b^\diamond, \Omega_{IGPC})$$

where $\Omega_{IGPC} = \text{plim}_{N,T\to\infty}\mathbb{D}^{-1}$ and:

$$b^\diamond = \text{plim}_{N,T\to\infty}\left\{\mathbb{D}^{-1}\begin{bmatrix}\kappa\frac{1}{N}\text{tr}[\Gamma' Q_1^o \Sigma_{ee}^{-1}\Gamma] \\ \kappa\frac{1}{N}\text{tr}[\Gamma' Q_2^o \Sigma_{ee}^{-1}\Gamma] \\ \kappa\frac{1}{N}\text{tr}[\Gamma' Q_3^o \Sigma_{ee}^{-1}\Gamma] \\ \kappa\frac{1}{N}\text{tr}[\Gamma' Q_4^o \Sigma_{ee}^{-1}\Gamma] \\ 0_{k\times 1}\end{bmatrix}\right\}$$

Remark 1.3.5. Regarding the limiting disribution, IGPCE has a bias term $b^\diamond$ while QMLE does not. In terms of efficiency, the limiting variance of the IGPCE is larger than that of the QMLE when the explanatory variables indeed follow the specification in QML, as shown in the supplementary material and such finding is also confirmed by the simulation results in Section 4. The better asymptotic performance of the QMLE compared to the IGPCE comes at a cost: the QML approach restricts the model for the explanatory variables, whereas the IGPC approach does not.

Bias-corrected estimator: Using the asymptotic representation of $\check{\delta}$ stated in Theorem 1.3.2, I can construct a bias-corrected estimator for δ by substitution as follows:

$$\check{\delta}^{\star} = \check{\delta} + \check{b}$$

where $\check{b}$ is the estimator of b by replacing all the true parameters $\delta, \Gamma, F, \Sigma_{ee}$ with their IGPCEs $\check{\delta}, \check{\Gamma}, \check{F}, \check{\Sigma}_{ee}$. Then, I have the following limiting distribution for the bias-corrected estimator $\check{\delta}^{\star}$.

Theorem 1.3.3. ***(Limiting distributions for bias-corrected estimators)*** *Under the assumptions of Theorem 1.3.2, when $N, T \to \infty$ and $T/N \to \kappa > 0$, I have:*

$$\sqrt{NT}(\check{\delta}^{\star} - \delta) \xrightarrow{d} N(0, \Omega_{IGPC})$$

where Ω_{IGPC} is defined in Corollary 1.3.1.

1.3.4 Computation of the IGPCE

Computation of the IGPCE involves an iterative procedure based on the first-order conditions (1.3.8)–(1.3.15) in Section 3.2. In sth iteration, I update $\check{\theta}^{(s+1)} = (\check{\Gamma}^{(s+1)}, \check{\Sigma}_{ee}^{(s+1)}, \check{\beta}^{(s+1)}, \check{\eta}^{(s+1)})$, where $\check{\eta}^{(s+1)} = (\check{\rho}_1^{(s+1)}, \check{\rho}_2^{(s+1)}, \check{\gamma}_1^{(s+1)}, \check{\gamma}_2^{(s+1)})$ as follows.

In the first step, $\check{\Gamma}^{(s+1)}$ is computed as the first r eigenvectors associated with the first r largest eigenvalues of the $2N \times 2N$ matrix $\check{D}_{\Gamma}^{(s)} = \frac{1}{NT}\sum_{t=1}^{T}\left(\Upsilon(\check{\eta}^{(s)})\dot{Y}_t - \dot{X}_t\check{\beta}^{(s)}\right)\left(\Upsilon(\check{\eta}^{(s)})\dot{Y}_t - \right.$

$\dot{X}_t\check{\beta}^{(s)}\Big)'\big(\check{\Sigma}_{ee}^{(s)}\big)^{-1}$.

In the second step, update $\check{\Sigma}_{ee}^{(s+1)}$ according to:

$$(\check{\sigma}_{1i}^{(s+1)})^2 = \frac{1}{T}\sum_{t=1}^{T}\Big(\dot{y}_{1it} - \check{\rho}_1^{(s)}\ddot{y}_{1it} - \check{\gamma}_1^{(s)}\dot{y}_{2it} - \dot{x}_{1it}'\check{\beta}_1^{(s)} - \check{\lambda}_i^{(s+1)\prime}\check{f}_t^{(s)}\Big)^2$$

and

$$(\check{\sigma}_{2i}^{(s+1)})^2 = \frac{1}{T}\sum_{t=1}^{T}\Big(\dot{y}_{2it} - \check{\rho}_2^{(s)}\ddot{y}_{2it} - \check{\gamma}_2^{(s)}\dot{y}_{1it} - \dot{x}_{2it}'\check{\beta}_2^{(s)} - \check{\psi}_i^{(s+1)\prime}\check{f}_t^{(s)}\Big)^2$$

where

$$\check{f}_t^{(s)} = \frac{1}{N}\check{\Gamma}^{(s+1)\prime}(\check{\Sigma}_{ee}^{(s)})^{-1}\Big(\Upsilon(\check{\eta}^{(s)})\dot{Y}_t - \dot{X}_t\check{\beta}^{(s)}\Big)$$

In the third step, update $\check{\beta}^{(s+1)}$ according to:

$$\check{\beta}^{(s+1)} = \Big(\dot{X}_t'\widehat{\ddot{M}}^{(s+1)}\dot{X}_t\Big)^{-1}\Big(\dot{X}_t'\widehat{\ddot{M}}^{(s+1)}\Upsilon(\check{\eta}^{(s)})\dot{Y}_t\Big)^{-1}$$

where $\widehat{\ddot{M}}^{(s+1)} = (\check{\Sigma}_{ee}^{(s+1)})^{-1} - \frac{1}{N}(\check{\Sigma}_{ee}^{(s+1)})^{-1}\check{\Gamma}^{(s+1)}\check{\Gamma}^{(s+1)\prime}(\check{\Sigma}_{ee}^{(s+1)})^{-1}$.

In the final step, update $\check{\eta}^{(s+1)} = (\check{\rho}_1^{(s+1)}, \check{\rho}_2^{(s+1)}, \check{\gamma}_1^{(s+1)}, \check{\gamma}_2^{(s+1)})$ by directly maximizing the likelihood $\mathcal{L}_2(\theta_2)$ (1.3.7) with respect to η at $\Gamma = \check{\Gamma}^{(s+1)}, \Sigma_{ee} = \check{\Sigma}_{ee}^{(s+1)}$ and $\beta = \check{\beta}^{(s+1)}$. Combining these steps, I obtain $\check{\theta}^{(s+1)} = (\check{\Gamma}^{(s+1)}, \check{\Sigma}_{ee}^{(s+1)}, \check{\beta}^{(s+1)}, \check{\eta}^{(s+1)})$. The iteration continues until the distance $\|\check{\theta}^{(s+1)} - \check{\theta}^{(s)}\|$ is smaller than a preset tolerance.

In the simulation results reported in Section 4, similar to the QML approach, I use the within-group estimator as the starting value for $\check{\beta}^{(1)}$ and $\check{\eta}^{(1)}$. Then let $\check{\Gamma}^{(1)}, \check{\Sigma}_{ee}^{(1)}$ be the solution according to the above first and second steps, given $\check{\beta}^{(1)}, \check{\eta}^{(1)}$. The simulation results show that the IGPCE performs well in finite sample and corroborate its asymptotic properties, as derived in this paper.

1.4 Finite sample properties via simulations

In this section, I investigate the finite sample performance of both approaches by Monte Carlo simulation. The simulation results reported in the following sections show that both approaches work well and corroborate the inferential theories derived in this paper.

1.4.1 Data generating processes

I consider two different data generating processes (DGPs). Both DGPs follow the model of the dependent variables in (1.1.1), but they use different specifications of the explanatory variables. DGP1 generates the explanatory variables according to (1.2.1), while DGP2 does not follow (1.2.1).

Specifically, both DGP1 and DGP2 generate the dependent variables according to:

$$
\begin{aligned}
y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x_{1it}\beta_1 + \lambda_i' f_t + e_{1it} \\
y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x_{2it}\beta_2 + \psi_i' f_t + e_{2it}
\end{aligned}
$$

$$
((\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1, \beta_2) = (0.2, 0.2, 0.2, 0.2, 1, 2))
$$

where $f_t = (f_{t1}, f_{t2})'$, $\lambda_i = (\lambda_{i1}, \lambda_{i2})'$, and $\psi_i = (\psi_{i1}, \psi_{i2})'$. The variables $\alpha_{1i}, \alpha_{2i}, \lambda_{il}, \psi_{il}$ and f_{tl} are all i.i.d. $N(0, 1)$. I generate the errors e_{1it} and e_{2it} with cross-sectional heteroskedasticity. I set $e_t = \sqrt{\text{diag}(\Xi^\dagger)}\varepsilon_t^\dagger$, where $\varepsilon_t^\dagger$ is a $2N$ dimensional column vector with all the elements being $(\chi_2^2 - 2)/2$ independently, where χ_2^2 denotes the chi-squared distribution with two degrees of freedom, and normalized to zero mean and unit variance. In addition, $\Xi^{\dagger}$ is a $2N$ dimensional column vector, whose mth element is set to:

$$\Xi^{\dagger}_m = 0.1 + \frac{1-\eta_m}{\eta_m}\iota'_m\iota_m, \qquad m = 1, 2, \ldots, 2N$$

where η_m is drawn from U$[0.1, 0.9]$, and ι_m is the transpose of the mth row of Γ; the constant 0.1 keeps the variance away from zero.

The spatial weights matrices W_1 and W_2 are generated based on the "q ahead and q behind" framework, similar to that in Kelejian and Prucha (1999), Baltagi and Deng (2015), among others. In the "q ahead and q behind" framework, all the individuals are arranged in a circle, and each individual is affected only by the q individuals immediately in front and immediately behind it with equal weight. Then, the weight matrix is row normalized to ensure that the sum of each row is equal to 1. Thus, the non-zero weight equals $\frac{1}{2q}$.[16] I consider two setups of the spatial weights matrix, "1 ahead and 1 behind" and "5 ahead and 5 behind", for both the QML and the IGPC approaches.

DGP1 and DGP2 differ in the generation of the explanatory variables, as explained in the following.

[16]In the simulation, I set the cross-sectional dimension N to be larger than $2q$, so the non-zero weight under the "q ahead and q behind" framework is always $\frac{1}{2q}$. In practice, when $N < 2q$, then the weights matrix under the "q ahead and q behind" is defined as follows: all the diagonal elements are 0, and all the off-diagonal elements are $\frac{1}{(N-1)}$.

DGP1

DGP1 generates them according to model (1.2.1) in the following specification:

$$x_{1it} = \nu_{1i} + a_1 \phi_{1i}' f_t + v_{1it}$$
$$x_{2it} = \nu_{2i} + a_2 \phi_{2i}' f_t + v_{2it}, \qquad (a_1 = a_2 = 1)$$

where f_t is the same as in the above model of y; $\phi_{1i} = (\phi_{1i1}, \phi_{1i2})'$, and $\phi_{2i} = (\phi_{2i1}, \phi_{2i2})'$, where $\phi_{1il} = \lambda_{il} + u_{1il}$ and $\phi_{2il} = \psi_{il} + u_{2il}$. All the variables $\nu_{1i}, \nu_{2i}, u_{1il}$ and u_{2il} are i.i.d. with $N(0,1)$. I also generate the errors v_{1it} and v_{2it} with cross-sectional heteroskedasticity by setting $v_t = \sqrt{\text{diag}(\Xi^\star)}\varepsilon_t^\star$, where $\varepsilon_t^\star$ is a $2N$ dimensional vector generated as in $\varepsilon_t^\dagger$; $\Xi^\star$ is a $2N$ dimensional column vector generated similar to $\Xi^\dagger$ but Γ is replaced with $\Gamma^\star$, where $\Gamma^\star = (\Gamma_1^\star, \Gamma_2^\star, \dots, \Gamma_N^\star)'$, with $\Gamma_i^\star = (\phi_{1i}, \phi_{2i})$.

DGP2

DGP2 generates the explanatory variables based on DGP1 but with truncation (similarly to Moon et al. (2014)), specified as follows:

$$x_{1it} = \left[\nu_{1i} + a_1 \phi_{1i}' f_t + v_{1it}\right]\left[\nu_{1i} + a_1 \phi_{1i}' f_t + v_{1it} > -3.5\right]$$
$$x_{2it} = \left[\nu_{2i} + a_2 \phi_{2i}' f_t + v_{2it}\right]\left[\nu_{2i} + a_2 \phi_{2i}' f_t + v_{2it} > -3.5\right], \qquad (a_1 = a_2 = 1)$$

where the variables ν_{li}, ϕ_{li}, f_t and v_{1it} are the same as in DGP1. With truncation, the explanatory variables are no longer a factor structure of common shocks f_t as in model (1.2.1).

Remark 1.4.1. The errors in the above two DGPs are non-normal and skewed. I also consider the cases when the errors have normal or student's t distributions. The corresponding simulation results for both approaches are provided in the supplementary material.

1.4.2 Finite sample performance of the QMLE

In this section, I provide the simulation results of the QML approach based on the above two DGPs and both setups of the spatial weights matrix.

In addition, since the number of factors r is usually unknown in practice, I propose a likelihood-based information criterion following Bai and Ng (2002) to determine it in the QML approach. Specifically, r is determined by:

$$\hat{r} = \underset{0 \leq m \leq r_{max}}{\text{argmin}} \, IC(m) \tag{1.4.1}$$

with

$$IC(m) = \frac{1}{2N\bar{k}} \ln \left| \hat{L}^m \hat{L}^{m\prime} + \hat{\Sigma}_{\epsilon\epsilon}^m \right| - \frac{1}{N\bar{k}} \ln \left| \Upsilon(\hat{\eta}^m) \right| + m \frac{N\bar{k} + T}{2N\bar{k}T} \ln[\min(N\bar{k}, T)]$$

where $\bar{k} = k_1 + k_2 + 2$, and $(\hat{\eta}^m, \hat{L}^m, \hat{\Sigma}_{\epsilon\epsilon}^m)$ are the QMLE of $(\eta, L, \Sigma_{\epsilon\epsilon})$ when the number of factors is set to m. In the simulation, I take DGP1, for example, and report the percentage of r values that are correctly estimated by (1.4.1) (set $r_{\max} = 4$) based on 1000 repetitions in the third row of Tables 1.1 and 1.2. The results show that the percentage of correctly estimated r values is very high and equal to or close to 100% for different combinations of (N, T) and different setups of the weights matrix. I then conduct simulations for the QMLE by assuming that the true number of factors is known in both DGP1 and DGP2.

Tables 1.1–1.3 present the simulation results of the QMLE based on 1000 repetitions. Both biases and root mean square errors (RMSE) are reported. From the results for both DGP1 and DGP2, I find that the biases are small and the RMSE decrease as the sample increases, indicating that the QMLE performs well and is consistent. Moreover, the simulation results corroborate the asymptotic properties of the QMLE derived in this paper. Additional simulation results based on different distributions of errors (normal and student's t distributions) are reported in the supplementary material, which confirm that QMLE has good finite sample properties and is robust to different distributions of errors.

Table 1.1: The performance of QMLE under DGP1 & "1 ahead and 1 behind" weights matrix

N		25	50	100	25	50	100
T		50	50	50	100	100	100
$\% \hat{r} = r$		99.8	100.0	100.0	99.8	100.0	100.0
ρ_1	Bias	0.0005	0.0002	0.0002	0.0003	0.0001	0.0002
	RMSE	0.0070	0.0043	0.0027	0.0048	0.0030	0.0020
ρ_2	Bias	0.0001	0.0002	0.0001	0.0002	0.0000	0.0002
	RMSE	0.0043	0.0026	0.0017	0.0030	0.0019	0.0012
γ_1	Bias	0.0000	0.0000	0.0001	0.0000	0.0001	0.0001
	RMSE	0.0035	0.0020	0.0013	0.0023	0.0014	0.0009
γ_2	Bias	-0.0001	0.0005	0.0004	0.0002	0.0003	0.0003
	RMSE	0.0057	0.0033	0.0023	0.0038	0.0025	0.0016
β_1	Bias	0.0002	0.0000	0.0001	0.0000	0.0001	-0.0001
	RMSE	0.0090	0.0059	0.0035	0.0065	0.0039	0.0026
β_2	Bias	-0.0001	-0.0001	-0.0001	0.0001	-0.0003	-0.0002
	RMSE	0.0101	0.0058	0.0037	0.0071	0.0041	0.0027

Table 1.2: The performance of QMLE under DGP1 & “5 ahead and 5 behind” weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
$\% \hat{r} = r$		99.8	100.0	100.0	99.8	100.0	100.0
ρ_1	Bias	-0.0001	-0.0001	0.0003	0.0002	0.0000	0.0000
	RMSE	0.0150	0.0076	0.0050	0.0088	0.0056	0.0036
ρ_2	Bias	0.0002	0.0003	0.0001	0.0001	-0.0002	0.0002
	RMSE	0.0083	0.0050	0.0033	0.0059	0.0036	0.0022
γ_1	Bias	-0.0001	0.0000	0.0001	-0.0001	0.0001	0.0001
	RMSE	0.0036	0.0020	0.0013	0.0023	0.0014	0.0009
γ_2	Bias	-0.0002	0.0003	0.0003	0.0001	0.0002	0.0001
	RMSE	0.0057	0.0033	0.0023	0.0039	0.0026	0.0016
β_1	Bias	0.0003	0.0000	0.0001	0.0000	0.0001	-0.0001
	RMSE	0.0090	0.0058	0.0034	0.0065	0.0039	0.0026
β_2	Bias	0.0000	-0.0001	0.0000	0.0002	-0.0002	-0.0001
	RMSE	0.0102	0.0058	0.0037	0.0071	0.0041	0.0026

Table 1.3: The performance of QMLE under DGP2 & “1 ahead and 1 behind” weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
ρ_1	Bias	0.0015	-0.0002	-0.0001	-0.0007	0.0002	0.0003
	RMSE	0.0098	0.0058	0.0038	0.0060	0.0041	0.0026
ρ_2	Bias	-0.0011	0.0003	-0.0001	0.0003	0.0003	0.0001
	RMSE	0.0070	0.0043	0.0030	0.0047	0.0034	0.0020
γ_1	Bias	0.0006	0.0001	0.0003	0.0000	-0.0005	0.0001
	RMSE	0.0058	0.0035	0.0023	0.0044	0.0024	0.0016
γ_2	Bias	-0.0001	0.0008	0.0000	0.0005	0.0001	0.0002
	RMSE	0.0060	0.0053	0.0031	0.0056	0.0037	0.0024
β_1	Bias	-0.0010	-0.0024	-0.0005	-0.0027	-0.0013	-0.0009
	RMSE	0.0154	0.0118	0.0062	0.0103	0.0062	0.0046
β_2	Bias	-0.0019	-0.0017	-0.0009	0.0004	-0.0018	-0.0014
	RMSE	0.0167	0.0098	0.0062	0.0110	0.0076	0.0052

1.4.3 Finite sample performance of the IGPCE

In this section, I present the simulation results of the IGPC approach based on the above two DGPs and both setups of the spatial weights matrix.

Similarly to the QML approach, I propose an information criterion adapting the ideas in Bai and Ng (2002) to determine the number of factors r for the IGPC approach, specified as:

$$\check{r} = \underset{0 \leq m \leq r_{max}}{\operatorname{argmin}} IC_2(m) \tag{1.4.2}$$

with

$$IC_2(m) = \frac{1}{2N\bar{k}} \sum_{i=1}^{N} \left(\ln\left[(\check{\sigma}_{1i}^m)^2\right] + \ln\left[(\check{\sigma}_{2i}^m)^2\right] \right) - \frac{1}{N\bar{k}} \ln|\Upsilon(\check{\eta}^m)| + m\frac{N\bar{k}+T}{2N\bar{k}T} \ln[\min(N\bar{k}, T)]$$

and

$$(\check{\sigma}_{1i}^m)^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\dot{y}_{1it} - \check{\rho}_1^m \ddot{y}_{1it} - \check{\gamma}_1^m \dot{y}_{2it} - \dot{x}_{1it}'\check{\beta}_1^m - \check{\lambda}_i^{m\prime} \check{f}_t^m\right)^2$$

$$(\check{\sigma}_{2i}^m)^2 = \frac{1}{T}\sum_{t=1}^{T}\left(\dot{y}_{2it} - \check{\rho}_2^m \ddot{y}_{2it} - \check{\gamma}_2^m \dot{y}_{1it} - \dot{x}_{2it}'\check{\beta}_2^m - \check{\psi}_i^{m\prime} \check{f}_t^m\right)^2$$

where $\bar{k} = 2$; $(\check{\eta}^m, \check{\beta}_1^m, \check{\beta}_2^m, \check{\lambda}_i^m, \check{f}_t^m)$ are the IGPCEs of $(\eta, \beta_1, \beta_2, \lambda_i, f_t)$ when the number of factors is set to m. Again, I set $r_{\max} = 4$ and report the percentage of r values correctly estimated by (1.4.2) based on 1000 repetitions for DGP1. From the results shown in the third row of Tables 1.4 and 1.5, it can be seen that the percentage is high for most choices of (N, T) (except small (N, T)) and setups of the weights matrix. Although the percentage is slightly lower than that in the QML approach, it is a reasonably good choice in practice when researchers prefer to allow the explanatory variables to be arbitrarily correlated with

Table 1.4: The performance of IGPCE under DGP1 & "1 ahead and 1 behind" weights matrix

N		25	50	100	25	50	100
T		50	50	50	100	100	100
% $\hat{r} = r$		78.60%	88.40%	88.00%	84.40%	86.80%	90.00%
ρ_1	Bias	-0.0025	-0.0017	-0.0011	-0.0024	-0.0019	-0.0014
	RMSE	0.0118	0.0073	0.0049	0.0091	0.0071	0.0058
ρ_2	Bias	-0.0011	-0.0006	-0.0005	-0.0009	-0.0006	-0.0006
	RMSE	0.0071	0.0046	0.0032	0.0057	0.0035	0.0028
γ_1	Bias	-0.0012	-0.0005	-0.0005	-0.0012	-0.0009	-0.0007
	RMSE	0.0066	0.0035	0.0026	0.0070	0.0037	0.0030
γ_2	Bias	-0.0064	-0.0030	-0.0016	-0.0052	-0.0033	-0.0027
	RMSE	0.0251	0.0155	0.0067	0.0215	0.0176	0.0108
β_1	Bias	0.0002	0.0006	0.0003	0.0006	0.0003	0.0003
	RMSE	0.0121	0.0073	0.0043	0.0080	0.0054	0.0033
β_2	Bias	0.0035	0.0014	0.0006	0.0024	0.0017	0.0011
	RMSE	0.0158	0.0097	0.0051	0.0130	0.0096	0.0055

the factors and loadings. Therefore, I assume that r is known in the simulation of the IGPCE under both DGPs.

Tables 1.4–1.6 state the simulation results of the IGPCE based on 1000 repetitions. The results show that both the bias and the RMSE of the IGPCE are small in terms of the sample size and the magnitude of the true underlying parameters across different combinations of (N, T) and different choices of the weights matrix. In addition, the RMSE of the IGPCE declines as sample becomes larger, indicating that the IGPCE is consistent. Thus, the simulation results indicate that the IGPCE works well in a finite sample.

Table 1.5: The performance of IGPCE under DGP1 & "5 ahead and 5 behind" weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
$\% \hat{r} = r$		77.40%	87.80%	86.00%	85.00%	88.40%	88.00%
ρ_1	Bias	-0.0021	-0.0012	-0.0010	-0.0023	-0.0014	-0.0007
	RMSE	0.0186	0.0110	0.0073	0.0142	0.0084	0.0053
ρ_2	Bias	-0.0010	-0.0005	-0.0002	-0.0006	-0.0003	-0.0004
	RMSE	0.0123	0.0073	0.0047	0.0092	0.0051	0.0033
γ_1	Bias	-0.0016	-0.0007	-0.0006	-0.0012	-0.0010	-0.0007
	RMSE	0.0099	0.0038	0.0030	0.0058	0.0039	0.0026
γ_2	Bias	-0.0061	-0.0027	-0.0017	-0.0059	-0.0030	-0.0024
	RMSE	0.0236	0.0133	0.0074	0.0238	0.0137	0.0091
β_1	Bias	0.0001	0.0004	0.0001	0.0004	0.0002	0.0002
	RMSE	0.0122	0.0074	0.0042	0.0080	0.0052	0.0031
β_2	Bias	0.0032	0.0012	0.0005	0.0024	0.0013	0.0010
	RMSE	0.0153	0.0090	0.0049	0.0146	0.0077	0.0048

Table 1.6: The performance of IGPCE under DGP2 & "1 ahead and 1 behind" weights matrix

	N	25	50	100	25	50	100
	T	50	50	50	100	100	100
ρ_1	Bias	-0.0025	-0.0009	-0.0018	-0.0011	-0.0027	-0.0017
	RMSE	0.0162	0.0126	0.0096	0.0106	0.0095	0.0068
ρ_2	Bias	-0.0012	-0.0014	-0.0001	-0.0006	-0.0008	-0.0003
	RMSE	0.0109	0.0086	0.0061	0.0099	0.0061	0.0043
γ_1	Bias	-0.0020	-0.0015	-0.0021	-0.0021	-0.0007	-0.0009
	RMSE	0.0096	0.0078	0.0055	0.0075	0.0046	0.0041
γ_2	Bias	-0.0065	-0.0046	-0.0038	-0.0036	-0.0014	-0.0026
	RMSE	0.0280	0.0150	0.0119	0.0231	0.0085	0.0101
β_1	Bias	0.0011	0.0028	0.0047	0.0022	0.0012	0.0012
	RMSE	0.0254	0.0173	0.0123	0.0151	0.0102	0.0076
β_2	Bias	0.0021	0.0022	0.0028	0.0014	0.0018	0.0002
	RMSE	0.0251	0.0166	0.0138	0.0185	0.0125	0.0102

1.4.4 Comparison of the performance of both approaches

A comparison of both approaches reveals that: (1), when the explanatory variables x are correctly specified, QML performs better than IGPC, otherwise the superiority of QML is weakened; (2), IGPC is robust for different underlying models of x.

Based on DGP1 when the model of explanatory variables is correctly specified in the QML approach, I find that the bias of IGPCE is relatively obvious compared to QMLE whose bias is close to zero. This finding is consistent with the inferential theory that IGPCE has a bias term in its limiting distribution while QMLE does not. At the same time, the RMSE of IGPCE is slightly larger than that of QMLE, implying that QMLE is more efficient than IGPCE.

Based on DGP2 when the explanatory variables do not satisfy the specification as in the QML approach but are still affected by the common shocks, QML performs slightly worse than in DGP1, while the performance of IGPC is similar to that in DGP1. This implies that IGPCE is robust to different underlying specifications of the explanatory variables, while QMLE is more sensitive. In addition, the superiority of QMLE is weakened when the explanatory variables are not correctly specified.

1.5 Some extensions

In this section, I discuss four important and useful extensions of model (1.1.1) with a brief summary as follows: (1) models with additional explanatory variables, denoted x_{3it}, which affect both dependent variables y_1 and y_2; (2) models with time-invariant and common regressors; (3) models with spatial autoregressive (SAR) errors; (4) models with additional spatial lags: the dependent variable y_1 is affected not only by its own spatial lag but also by the spatial lag of y_2, and vice versa.[17]

1.5.1 Models with additional common explanatory variables

In model (1.1.1), I consider two different sets of explanatory variables x_{1it} and x_{2it} in the y_{1it} and y_{2it} equations, respectively. In this section, I augment model (1.1.1) with additional common explanatory variables, denoted x_{3it}, which affect both y_{1it} and y_{2it}, as follows:

$$\begin{aligned} y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + \sum_{p=1}^{k_1} x_{1itp}\beta_{1p} + \sum_{l=1}^{k_3} x_{3itl}\beta_{3l} + \lambda_i' f_t + e_{1it} \\ y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + \sum_{q=1}^{k_2} x_{2itq}\beta_{2q} + \sum_{l=1}^{k_3} x_{3itl}\beta_{4l} + \psi_i' f_t + e_{2it} \end{aligned} \tag{1.5.1}$$

where $x_{3itl}, (l = 1, 2, \ldots, k_3)$ are additional explanatory variables. I then propose both QML and IGPC approaches for the above extension.

Extension 1 using QML approach

For model (1.5.1), I assume the additional explanatory variable x_{3it} is also affected by the common factor f_t and follows the same factor structure model as (1.2.1):

$$x_{3itl} = \nu_{3il} + \phi_{3il}' f_t + v_{3itl}, \quad l = 1, 2, \ldots, k_3 \tag{1.5.2}$$

[17] For each extension, I discuss both QML and IGPC estimation methods with modification. The large sample theory can be derived, but is much more involved.

Then, the extended model combining (1.5.1), (1.2.1) and (1.5.2) can be rewritten as:

$$\begin{bmatrix} y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} - x'_{1it}\beta_1 - x'_{3it}\beta_3 \\ y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} - x'_{2it}\beta_2 - x'_{3it}\beta_4 \\ x_{1it} \\ x_{2it} \\ x_{3it} \end{bmatrix} = \mu_i^{\dagger} + L_i^{\dagger\prime} f_t + \epsilon_{it}^{\dagger}$$

where $x_{3it} = (x_{3it1}, x_{3it2}, \ldots, x_{3it,k_3})'$; $\beta_3 = (\beta_{31}, \beta_{32}, \ldots, \beta_{3,k_3})'$, which is similar for β_4 and v_{3it}; $\mu_i^{\dagger} = (\alpha_{1i}, \alpha_{2i}, \nu'_{1i}, \nu'_{2i}, \nu'_{3i})'$; $L_i^{\dagger} = (\lambda_i, \psi_i, \phi_{1i}, \phi_{2i}, \phi_{3i})$; and $\epsilon_{it}^{\dagger} = (e_{1it}, e_{2it}, v'_{1it}, v'_{2it}, v'_{3it})'$. Let $\delta^{\dagger} = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta'_1, \beta'_2, \beta'_3, \beta'_4)$, $k^{\dagger} = k_1 + k_2 + k_3$ and $\bar{k}^{\dagger} = k^{\dagger} + 2$. I can then rewrite the above model into the same framework used in (1.2.4):

$$D^{\dagger}(\delta^{\dagger}) z_t^{\dagger} = \mu^{\dagger} + L^{\dagger} f_t + \epsilon_t^{\dagger} \tag{1.5.3}$$

where $z_t^{\dagger} = (z_{1t}^{\dagger}, z_{2t}^{\dagger}, \ldots, z_{Nt}^{\dagger})'$ with $z_{it}^{\dagger} = (y_{1it}, y_{2it}, x'_{1it}, x'_{2it}, x'_{3it})'$, $L^{\dagger} = (L_1^{\dagger}, L_2^{\dagger}, \ldots, L_N^{\dagger})'$, $\mu^{\dagger} = (\mu_1^{\dagger\prime}, \mu_2^{\dagger\prime}, \ldots, \mu_N^{\dagger\prime})'$ and $\epsilon_t^{\dagger} = (\epsilon_{1t}^{\dagger\prime}, \epsilon_{2t}^{\dagger\prime}, \ldots, \epsilon_{Nt}^{\dagger\prime})'$. The new transformation matrix $D^{\dagger}(\delta^{\dagger})$ is an $N\bar{k}^\dagger \times N\bar{k}^\dagger$ matrix whose (i,j) subblock, denoted by $D^\dagger_{ij}(\delta^\dagger)$, a $\bar{k}^\dagger \times \bar{k}^\dagger$ matrix, equals:

$$D^\dagger_{ij}(\delta^\dagger) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 & -\beta_1' & 0 & -\beta_3' \\ -\gamma_2 & 1 & 0 & -\beta_2' & -\beta_4' \\ 0 & 0 & I_{k_1} & 0 & 0 \\ 0 & 0 & 0 & I_{k_2} & 0 \\ 0 & 0 & 0 & 0 & I_{k_3} \end{bmatrix} & \text{if } \ i = j \\ \begin{bmatrix} -\rho_1 w_{1ij} & 0 & 0 & 0 & 0 \\ 0 & -\rho_2 w_{2ij} & 0 & 0 & 0 \\ 0 & 0 & 0_{k_1} & 0 & 0 \\ 0 & 0 & 0 & 0_{k_2} & 0 \\ 0 & 0 & 0 & 0 & 0_{k_3} \end{bmatrix} & \text{if } \ i \neq j \end{cases} \tag{1.5.4}$$

Similarly, I propose the QML method to estimate this extended model. In order to derive the inferential theory of its QMLE, the key is to study the determinant and the inverse matrix of the transformation matrix $D^\dagger(\delta^\dagger)$. Then, a similar analytical approach can be used in the analysis of the QMLE. Let η and $\Upsilon(\eta)$ be defined as in Section 2.2. Then, it can be verified that $\det(D^\dagger(\delta^\dagger)) = \det(\Upsilon(\eta))$. Let $V^\dagger(\delta^\dagger)$ denote the inverse matrix of $D^\dagger(\delta^\dagger)$, which is an $N\bar{k}^\dagger \times N\bar{k}^\dagger$ matrix. Its (i,j)th subblock is denoted by $V^\dagger_{ij}$, a $\bar{k}^\dagger \times \bar{k}^\dagger$ matrix, and has the same expression as in Lemma A.2 of the supplementary material, with the same definition of F_{ij} but a different $\beta^\dagger$. Here, $\beta^{\dagger\prime} = \begin{bmatrix} \beta_1' & 0 & \beta_3' \\ 0 & \beta_2' & \beta_4' \end{bmatrix}$. Lemma A.3 of the supplementary material still holds but with the preceding definition of $\beta^\dagger$. Based on the preceding analysis, the inferential analysis for this extended model can be studied in a similar way as that for model

(1.2.2).

Extension 1 using IGPC approach

In this approach, I consider model (1.5.1) without specifying the model for all explanatory variables and propose the IGPC method.

Let $x_{it}^{\dagger} = \begin{bmatrix} x_{1it}' & 0 & x_{3it}' & 0 \\ 0 & x_{2it}' & 0 & x_{3it}' \end{bmatrix}'$, and $\beta^{\dagger} = (\beta_1', \beta_2', \beta_3', \beta_4')'$. Then, model (1.5.1) can be rewritten as:

$$\begin{bmatrix} y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} \\ y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} \end{bmatrix} = \alpha_i + x_{it}^{\dagger\prime} \beta^{\dagger} + \Gamma_i' f_t + e_{it} \tag{1.5.5}$$

where all α_i, Γ_i and e_{it} are defined as in Section 3.1. With the same η and $\Upsilon(\eta)$, I can rewrite the above model into the form of (1.3.2) with the new $x_{it}^{\dagger}$ and $\beta^{\dagger}$. Then, I propose the same IGPC method for this extended model.

1.5.2 Models with time-invariant and common regressors

In applications, it is common to observe some time-invariant regressors (i.e., not varying with t, such as gender, race and education in microeconomic earnings studies) and some common regressors (i.e., not varying with individual i, such as unemployment rates, aggregate price index representing trends and other macroeconomic policy variables). Therefore, in this section, I extend model (1.1.1) to include some time-invariant and common regressors using both QML and IGPC approaches as follows.

Extension 2 using QML approach

In this approach, I allow the regression coefficients of the time-invariant regressors to be time varying and the coefficients of the common regressors to be individual dependent (varying with i). In addition, I allow both x_{1it} and x_{2it} to be affected by the time-invariant regressors r_i and the common regressors p_t using a factor structure specification. Specifically, I consider the following extended model:

$$
\begin{aligned}
y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + \sum_{p=1}^{k_1} x_{1itp}\beta_{1p} + r_i' h_{1t} + \tau_{1i}' p_t + \lambda_i' f_t + e_{1it} \\
y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + \sum_{q=1}^{k_2} x_{2itq}\beta_{2q} + r_i' h_{2t} + \tau_{2i}' p_t + \psi_i' f_t + e_{2it} \\
x_{1itp} &= \nu_{1ip} + r_i' s_{1tp} + \eta_{1ip}' p_t + \phi_{1ip}' f_t + v_{1itp}, \quad p = 1, 2, \ldots, k_1 \\
x_{2itq} &= \nu_{2iq} + r_i' s_{2tq} + \eta_{2iq}' p_t + \phi_{2iq}' f_t + v_{2itq}, \quad q = 1, 2, \ldots, k_2
\end{aligned}
\tag{1.5.6}
$$

where r_i represents a vector of observable time-invariant variables, and p_t represents a vector of observable common variables.

The above model can be rewritten as follows:

$$
\begin{bmatrix}
y_{1it} - \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} - \gamma_1 y_{2it} - x_{1it}'\beta_1 \\
y_{2it} - \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} - \gamma_2 y_{1it} - x_{2it}'\beta_2 \\
x_{1it} \\
x_{2it}
\end{bmatrix}
= \mu_i + L_i^{\dagger\prime} f_t^{\dagger} + \epsilon_{it}
$$

where $f_t^\dagger = (h_{1t}', h_{2t}', p_t', s_{1t1}', \ldots, s_{1t,k_1}', s_{2t1}', \ldots, s_{2t,k_2}', f_t')'$, and

$$L_i^{\dagger\prime} = \begin{bmatrix} r_i' & 0 & \tau_{1i}' & 0 & 0 & \lambda_i' \\ 0 & r_i' & \tau_{2i}' & 0 & 0 & \psi_i' \\ 0 & 0 & \eta_{1i}' & I_{k_1} \otimes r_i' & 0 & \phi_{1i}' \\ 0 & 0 & \eta_{2i}' & 0 & I_{k_2} \otimes r_i' & \phi_{2i}' \end{bmatrix}$$

with $\eta_{1i} = (\eta_{1i1}, \ldots, \eta_{1i,k_1})$ and $\eta_{2i} = (\eta_{2i1}, \ldots, \eta_{2i,k_2})$.

The above model specification is similar to that in Section 1.2, with the difference that some components of the common factors $f_t^\dagger$ and some components of the factor loadings $L_i^\dagger$ are now observable. The QML method can still be implemented for this extension but with modifications (for the observable components of $f_t^\dagger$ and $L_i^\dagger$, the QMLE does not estimate them but fixes them at the observed value). The asymptotic properties of the QMLE can be analyzed in a similar way as for the basic model (1.2.2), with attention to the fact that some components of $f_t^\dagger$ and $L_i^\dagger$ are observable.

Extension 2 using IGPC approach

In this section, I do not specify a model for time-invariant or common regressors but allow them to be arbitrarily correlated with the common shocks. In addition, I treat them as explanatory variables with constant coefficients and specify the extended model as follows:

$$\begin{aligned} y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x_{1it}'\beta_1 + r_i'\beta_3 + p_t'\beta_5 + \lambda_i' f_t + e_{1it} \\ y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x_{2it}'\beta_2 + r_i'\beta_4 + p_t'\beta_6 + \psi_i' f_t + e_{2it} \end{aligned} \quad (1.5.7)$$

where r_i is a vector of observable time-invariant variables, and p_t is a vector of observable common variables. The above model can be rewritten in the same form as (1.5.5) with different definitions of $x_{it}^\dagger$ and $\beta^\dagger$: $x_{it}^\dagger = \begin{bmatrix} x'_{1it} & 0 & r'_i & 0 & p'_t & 0 \\ 0 & x'_{2it} & 0 & r'_i & 0 & p'_t \end{bmatrix}$ and $\beta^\dagger = (\beta'_1, \beta'_2, \ldots, \beta'_6)'$. Again, I propose the IGPC method for this extended model, and the corresponding inferential theory can be studied in a similar way as in Section 1.3.

1.5.3 Models with SAR disturbances

In the spatial econometric literature, SAR disturbances have received much attention and are considered an important part of spatial models. Based on model (1.1.1), which only considers spatial correlations in the dependent variables, I now develop a more general model by including additional spatial correlations on the errors (i.e., SAR errors) in the following model specification:

$$\begin{aligned} y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^N w_{1ij} y_{1jt} + \gamma_1 y_{2it} + x'_{1it}\beta_1 + \lambda'_i f_t + u_{1it}, \quad u_{1it} = \pi_1 \sum_{j=1}^N m_{1ij} u_{1jt} + e_{1it} \\ y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^N w_{2ij} y_{2jt} + \gamma_2 y_{1it} + x'_{2it}\beta_2 + \psi'_i f_t + u_{2it}, \quad u_{2it} = \pi_2 \sum_{j=1}^N m_{2ij} u_{2jt} + e_{2it} \end{aligned} \tag{1.5.8}$$

where m_{1ij} and m_{2ij} are spatial weights involved in the SAR disturbances. The above model can be rewritten as:

$$\begin{aligned} Y_{1t} &= \alpha_1 + \rho_1 W_1 Y_{1t} + \gamma_1 Y_{2t} + X_{1t}\beta_1 + \Lambda' f_t + U_{1t}, \quad U_{1t} = \pi_1 M_1 U_{1t} + e_{1t} \\ Y_{2t} &= \alpha_2 + \rho_2 W_2 Y_{2t} + \gamma_2 Y_{1t} + X_{2t}\beta_2 + \Psi' f_t + U_{2t}, \quad U_{2t} = \pi_2 M_2 U_{2t} + e_{2t} \end{aligned} \tag{1.5.9}$$

where Y_{1t} is an $N \times 1$ vectors, defined as $Y_{1t} = (y_{11t}, y_{12t}, \ldots, y_{1Nt})'$, which is similar for $Y_{2t}, \alpha_1, \alpha_2, U_{1t}, U_{2t}, e_{1t}$ and e_{2t}; $X_{1t} = (x_{11t}, x_{12t}, \ldots, x_{1Nt})'$ is $N \times k_1$; $X_{2t} = (x_{21t}, x_{22t}, \ldots, x_{2Nt})'$ is $N \times k_2$; $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$; and $\Psi = (\psi_1, \psi_2, \ldots, \psi_N)'$. Both W_1 and W_2 are $N \times N$ weights matrices associated with the spatial effects of the dependent variables, while M_1 and M_2 are $N \times N$ weights matrices representing the additional spatial correlations in the errors. Baltagi and Deng (2015) consider the above model (1.5.9) without the common shocks parts Λf_t and Ψf_t. Furthermore, they impose an error component specification instead of the SAR structure and assume cross-sectional homoskedasticity of the errors, while I allow cross-sectional heteroskedasticity here.[18]

To transform (1.5.9) into the framework of (1.1.1), premultiply $I_N - \pi_1 M_1$ on both sides of the Y_{1t} equation and premultiply $I_N - \pi_2 M_2$ on both sides of the Y_{2t} equation. Then, I obtain:

$$\begin{aligned} Y_{1t} =& (\alpha_1 - \pi_1 M_1 \alpha_1) + \rho_1 W_1 Y_{1t} + \pi_1 M_1 Y_{1t} - \rho_1 \pi_1 M_1 W_1 Y_{1t} + \gamma_1 Y_{2t} - \gamma_1 \pi_1 M_1 Y_{2t} \\ &+ X_{1t}\beta_1 - \pi_1 M_1 X_{1t} \beta_1 + (\Lambda - \pi_1 M_1 \Lambda) f_t + e_{1t} \\ Y_{2t} =& (\alpha_2 - \pi_2 M_2 \alpha_2) + \rho_2 W_2 Y_{2t} + \pi_2 M_2 Y_{2t} - \rho_2 \pi_2 M_2 W_2 Y_{2t} + \gamma_2 Y_{1t} - \gamma_2 \pi_2 M_2 Y_{1t} \\ &+ X_{2t}\beta_2 - \pi_2 M_2 X_{2t} \beta_2 + (\Psi - \pi_2 M_2 \Psi) f_t + e_{2t} \end{aligned} \tag{1.5.10}$$

Note that $(\alpha_1 - \pi_1 M_1 \alpha_1)$ is a free parameter, so I can treat it as a new α_1, Similarly, treat $(\Lambda - \pi_1 M_1 \Lambda)$ as a new Λ, $(\alpha_2 - \pi_2 M_2 \alpha_2)$ as a new α_2 and $(\Psi - \pi_2 M_2 \Psi)$ as a new Ψ. Then,

[18] The error component specification in Baltagi and Deng (2015) is described as $U_{lt} = \varphi_l + \varepsilon_{lt}$, where $\varphi_l \sim i.i.d.(0, \sigma^2_{\varphi ll} I_N)$ and $\varepsilon_{lt} \sim i.i.d.(0, \sigma^2_{\varepsilon ll} I_N)$, for $l = 1, 2$.

(1.5.10) can be rewritten as:

$$
\begin{aligned}
Y_{1t} =& \alpha_1 + \rho_1 W_1 Y_{1t} + \pi_1 M_1 Y_{1t} - \rho_1 \pi_1 M_1 W_1 Y_{1t} + \gamma_1 Y_{2t} - \gamma_1 \pi_1 M_1 Y_{2t} \\
& + X_{1t}\beta_1 - \pi_1 M_1 X_{1t}\beta_1 + \Lambda f_t + e_{1t} \\
Y_{2t} =& \alpha_2 + \rho_2 W_2 Y_{2t} + \pi_2 M_2 Y_{2t} - \rho_2 \pi_2 M_2 W_2 Y_{2t} + \gamma_2 Y_{1t} - \gamma_2 \pi_2 M_2 Y_{1t} \\
& + X_{2t}\beta_2 - \pi_2 M_2 X_{2t}\beta_2 + \Psi f_t + e_{2t}
\end{aligned} \tag{1.5.11}
$$

This can be further rewritten as:

$$
\begin{aligned}
y_{1it} =& \alpha_{1i} + \rho_1 \Big(\sum_{j=1}^{N} w_{1ij} y_{1jt} \Big) + \pi_1 \Big(\sum_{j=1}^{N} m_{1ij} y_{1jt} \Big) - \rho_1 \pi_1 \Big(\sum_{j=1}^{N} \sum_{l=1}^{N} m_{1ij} w_{1jl} y_{1lt} \Big) \\
& + \gamma_1 y_{2it} - \gamma_1 \pi_1 \Big(\sum_{j=1}^{N} m_{1ij} y_{2jt} \Big) + x'_{1it}\beta_1 - \pi_1 \Big(\sum_{j=1}^{N} m_{1ij} x'_{1jt} \Big) \beta_1 + \lambda'_i f_t + e_{1it} \\
y_{2it} =& \alpha_{2i} + \rho_2 \Big(\sum_{j=1}^{N} w_{2ij} y_{2jt} \Big) + \pi_2 \Big(\sum_{j=1}^{N} m_{2ij} y_{2jt} \Big) - \rho_2 \pi_2 \Big(\sum_{j=1}^{N} \sum_{l=1}^{N} m_{2ij} w_{2jl} y_{2lt} \Big) \\
& + \gamma_2 y_{1it} - \gamma_2 \pi_2 \Big(\sum_{j=1}^{N} m_{2ij} y_{1jt} \Big) + x'_{2it}\beta_2 - \pi_2 \Big(\sum_{j=1}^{N} m_{2ij} x'_{2jt} \Big) \beta_2 + \Psi'_i f_t + e_{2it}
\end{aligned} \tag{1.5.12}
$$

Then, I analyze the above model using the following two approaches.

Extension 3 using QML approach

I assume that the explanatory variables follow the same model as (1.2.1). Then, combining (1.5.12) and (1.2.1), I can rewrite this extended model in the same framework as (1.2.4):

$$
D^{\dagger}(\delta^{\dagger}) z_t = \mu + L f_t + \epsilon_t \tag{1.5.13}
$$

where $\delta^\dagger = (\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2', \pi_1, \pi_2)'$; z_t, μ, L and ϵ_t are defined in the same way as in Section 1.2; and $D^\dagger(\delta^\dagger)$ is an $N\bar{k} \times N\bar{k}$ transformation matrix whose (i,j)th subblock, denoted by $D_{ij}^\dagger(\delta^\dagger)$, is defined as:

$$D_{ij}^\dagger(\delta^\dagger) = \begin{cases} \begin{bmatrix} 1+\rho_1\pi_1 m_{1i*}w_{1*i} & -\gamma_1 & -\beta_1' & 0 \\ -\gamma_2 & 1+\rho_2\pi_2 m_{2i*}w_{2*i} & 0 & -\beta_2' \\ 0 & 0 & I_{k_1} & 0 \\ 0 & 0 & 0 & I_{k_2} \end{bmatrix} & \text{if } i = j \\ \begin{bmatrix} \boldsymbol{d}_{1,ij} & \gamma_1\pi_1 m_{1ij} & \pi_1 m_{1ij}\beta_1' & 0 \\ \gamma_2\pi_2 m_{2ij} & \boldsymbol{d}_{2,ij} & 0 & \pi_2 m_{2ij}\beta_2' \\ 0 & 0 & 0_{k_1} & 0 \\ 0 & 0 & 0 & 0_{k_2} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{1.5.14}$$

where $\boldsymbol{d}_{1,ij} = -\rho_1 w_{1ij} - \pi_1 m_{1ij} + \rho_1\pi_1 m_{1i*}w_{1*j}$ and $\boldsymbol{d}_{2,ij} = -\rho_2 w_{2ij} - \pi_2 m_{2ij} + \rho_2\pi_2 m_{2i*}w_{2*j}$, and m_{li*} is the ith row of matrix M_l and w_{l*j} is the jth column of matrix W_l, for $l = 1, 2$.

Note that model (1.5.13) is similar to model (1.2.4) but with a new $\delta^\dagger$ (including additional parameters π_1, π_2 due to the SAR errors) and a more complicated transformation matrix $D^\dagger(\delta^\dagger)$. The QML method can be easily implemented in this extended model. To develop the inferential theory, similar to the derivation for model (1.2.2), the key is to specify the determinant and the inverse matrix of $D^\dagger(\delta^\dagger)$. Let $\eta^\dagger = (\rho_1, \rho_2, \gamma_1, \gamma_2, \pi_1, \pi_2)'$ and $\Upsilon^\dagger(\eta^\dagger)$ be a $2N \times 2N$ matrix whose (i,j)th subblock, denoted by $\Upsilon^{\dagger}_{ij}(\eta^{\dagger})$, is a 2×2 matrix that equals:

$$\Upsilon^{\dagger}_{ij}(\eta^{\dagger}) = \begin{cases} \begin{bmatrix} 1 + \rho_1 \pi_1 m_{1i*} w_{1*i} & -\gamma_1 \\ -\gamma_2 & 1 + \rho_2 \pi_2 m_{2i*} w_{2*i} \end{bmatrix} & \text{if } \; i = j \\ \begin{bmatrix} \boldsymbol{d}_{1,ij} & \gamma_1 \pi_1 m_{1ij} \\ \gamma_2 \pi_2 m_{2ij} & \boldsymbol{d}_{2,ij} \end{bmatrix} & \text{if } \; i \neq j \end{cases} \tag{1.5.15}$$

where $\boldsymbol{d}_{1,ij}$ and $\boldsymbol{d}_{2,ij}$ are the same as in the definition of $D^{\dagger}_{ij}(\delta^{\dagger})$. Then, it can be verified that $\det(D^{\dagger}(\delta^{\dagger})) = \det(\Upsilon^{\dagger}(\eta^{\dagger}))$. Furthermore, let $V^{\dagger}(\delta^{\dagger})$ be the inverse matrix of $D^{\dagger}(\delta^{\dagger})$. Then, its (i,j)th block, a $(k+2) \times (k+2)$ matrix $(k = k_1 + k_2)$ denoted by $V^{\dagger}_{ij}(\delta)$ has a closed form, which is equal to:

$$V^{\dagger}_{ij}(\delta^{\dagger}) = \begin{cases} \begin{bmatrix} F^{\dagger}_{ii} & F^{\dagger}_{ii}\beta' \\ 0 & I_k \end{bmatrix} & \text{if } \; i = j \\ \begin{bmatrix} F^{\dagger}_{ij} & F^{\dagger}_{ij}\beta' \\ 0 & 0_{k \times k} \end{bmatrix} & \text{if } \; i \neq j \end{cases} \tag{1.5.16}$$

where $\beta = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix}$, and $F^{\dagger}_{ij}$ is the (i,j)th 2×2 block of the inverse matrix of $\Upsilon^{\dagger}(\eta)$. The QML method can be used to estimate model (1.5.13), and the inferential theory can be studied similarly to basic model (1.2.2), together with the preceding results for the determinant and inverse of $D^{\dagger}(\eta^{\dagger})$.

Extension 3 using IGPC approach

In this approach, I do not specify the model for the explanatory variables.

Using the same notation as in Section 1.3.1, let $\Upsilon^X(\delta^\dagger)$ be a $2N \times 2N$ matrix, with its (i, j)th subblock denoted as $\Upsilon^X_{ij}(\delta^\dagger)$, a 2×2 matrix that equals:

$$\Upsilon^X_{ij}(\eta^\dagger) = \begin{cases} \begin{bmatrix} \beta_1' & 0 \\ 0 & \beta_2' \end{bmatrix} & \text{if } \ i = j \\ \begin{bmatrix} -\pi_1 m_{1ij}\beta_1' & 0 \\ 0 & -\pi_2 m_{2ij}\beta_2' \end{bmatrix} & \text{if } \ i \neq j \end{cases} \tag{1.5.17}$$

Then, the extended model (1.5.12) can be rewritten similarly to (1.3.2):

$$\sum_{j=1}^{N} \Upsilon^\dagger_{ij}(\eta^\dagger) y_{jt} = \alpha_i + \sum_{j=1}^{N} \Upsilon^X_{ij}(\delta^\dagger) x_{jt} + \Gamma_i' f_t + e_{it} \tag{1.5.18}$$

where $y_{jt}, x_{jt}, \alpha_i, \Gamma_i$ and e_{it} are defined as in (1.3.2); $\Upsilon^\dagger_{ij}(\eta^\dagger)$ is defined in Section 5.3.1. Then, based on the above expression, the IGPC method can be applied in this extension.

1.5.4 Models with additional spatial lags

In model (1.1.1), y_{1it} is affected by its own spatial lag only; likewise for y_{2it}. In this section, I enrich model (1.1.1) with additional spatial lags in both y_{1it} and y_{2it} equations as follows:

$$\begin{aligned} y_{1it} &= \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{1ij} y_{1jt} + \gamma_1 y_{2it} + \rho_3 \sum_{j=1}^{N} w_{3ij} y_{2jt} + x_{1it}'\beta_1 + \lambda_i' f_t + e_{1it} \\ y_{2it} &= \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{2ij} y_{2jt} + \gamma_2 y_{1it} + \rho_4 \sum_{j=1}^{N} w_{4ij} y_{1jt} + x_{2it}'\beta_2 + \psi_i' f_t + e_{2it} \end{aligned} \tag{1.5.19}$$

where y_{1it} is affected by the spatial lag of y_{2it}, and y_{2it} is affected by the spatial lag of y_{1it}, with ρ_3, ρ_4 being the additional parameters measuring the magnitudes of the spatial effects, and $W_3 = (w_{3ij})_{N\times N}$ and $W_4 = (w_{4ij})_{N\times N}$ being additional weights matrices. I analyze the above model using both the QML and IGPC approaches as follows.

Extension 4 using QML approach

I assume that the explanatory variables follow (1.2.1). Then, the extended model combining (1.5.19) and (1.2.1) can be rewritten as:

$$D^{\dagger}(\delta) z_t = \mu + L f_t + \epsilon_t \tag{1.5.20}$$

in the same framework as (1.2.4) but with a different transformation matrix $D^{\dagger}(\delta)$ whose (i, j)th subblock is defined below:

$$D^{\dagger}_{ij}(\delta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 & -\beta_1' & 0 \\ -\gamma_2 & 1 & 0 & -\beta_2' \\ 0 & 0 & I_{k_1} & 0 \\ 0 & 0 & 0 & I_{k_2} \end{bmatrix} & \text{if} \quad i = j \\ \begin{bmatrix} -\rho_1 w_{1ij} & -\rho_3 w_{3ij} & 0 & 0 \\ -\rho_4 w_{4ij} & -\rho_2 w_{2ij} & 0 & 0 \\ 0 & 0 & 0_{k_1} & 0 \\ 0 & 0 & 0 & 0_{k_2} \end{bmatrix} & \text{if} \quad i \neq j \end{cases} \tag{1.5.21}$$

Again, I implement the QML method for this extended model, and the inferential analysis of the QMLE can be derived in a similar way as for (1.2.2). In order to derive the inferential theory of the QMLE, the key is to study the determinant and the inverse matrix of the transformation matrix $D^{\dagger}(\delta)$. Unlike (1.2.7), I define a new $2N \times 2N$ matrix $\Upsilon^{\dagger}(\eta)$ whose (i, j)th block, a 2×2 matrix, is equal to:

$$\Upsilon^{\dagger}_{ij}(\eta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{bmatrix} & \text{if } \ i = j \\ \begin{bmatrix} -\rho_1 w_{1ij} & -\rho_3 w_{3ij} \\ -\rho_4 w_{4ij} & -\rho_2 w_{2ij} \end{bmatrix} & \text{if } \ i \neq j \end{cases} \tag{1.5.22}$$

A preliminary step needed to conduct inferential analysis is to study the determinant and the inverse of the new transformation matrix $\Upsilon^{\dagger}_{ij}(\eta)$. With mathematical calculation, it can be verified that $\det(D^{\dagger}(\delta)) = \det(\Upsilon^{\dagger}(\eta))$. Let $V^{\dagger}(\delta)$ denote the inverse matrix of $D^{\dagger}(\delta)$, which is an $N\bar{k} \times N\bar{k}$ matrix. Its (i, j)th subblock, denoted by $V^{\dagger}_{ij}$, a $\bar{k} \times \bar{k}$ matrix, has the same expression as in Lemma A.2 of the supplementary material but with a different $F^{\dagger}_{ij}$, which now is the (i, j)th 2×2 block of the inverse matrix of $\Upsilon^{\dagger}(\eta)$.[19]

Extension 4 using IGPC approach

In this section, I apply the IGPC method to the extended model (1.5.19) without specifying the model for the explanatory variables. With the notation $\Upsilon^{\dagger}(\eta)$ defined in (1.5.22), I can

[19]In this extended model, Lemma A.3 needs modification due to the new $D^{\dagger}(\delta)$ and $\Upsilon^{\dagger}(\eta)$.

rewrite (1.5.19) as in (1.3.2) (with a new $\Upsilon^\dagger(\eta)$):

$$\sum_{j=1}^{N} \Upsilon^\dagger_{ij}(\eta) y_{jt} = \alpha_i + x'_{it}\beta + \Gamma'_i f_t + e_{it} \tag{1.5.23}$$

Based on the above expression, the IGPC method can be implemented.

Remark 1.5.1. This paper considers a system of simultaneous equations with two dependent variables y_1 and y_2. It can be generalized to the multiple dependent variables case, i.e., $y_1, y_2, \ldots, y_p$. Both the estimation and the corresponding inferential analysis could be studied in a similar way but would require more mathematical calculation. This generalization warrants further study.

Remark 1.5.2. In this paper, I assume that the diagonal elements of the weights matrix w_{1ii} and w_{2ii} are zero for all i. This is a standard assumption implemented in the spatial modeling literature. However, in practice, there are cases where the diagonal elements of weights matrix are not all zero, for example, in an input-output matrix (in the production network by sector level as in Ozdagli and Weber (2016)). Then, one can slightly modify the definition of the transformation matrices $D(\delta)$ and $\Upsilon(\eta)$ and still apply the QML and IGPC estimation methods. The corresponding inferential theory needs modification according to the changes involved in $D(\delta)$ and $\Upsilon(\eta)$. The asymptotic analysis could be conducted as this paper with the same consistency and convergence rate but changes are needed for the limiting distribution. To avoid replication of the analysis in this paper, this is left for future research.

Remark 1.5.3. This paper considers a static case in the sense that the dependent variable y_t does not depend on its previous observation y_{t-1}. However, in practice, there might be cases when dynamic effects exist. In the trade and macroeconomics examples mentioned in the introduction, this year's GDP growth or trade growth might be affected by the previous year's values. Thus, it is potentially useful to study the dynamic case, where there are extra dynamic lags on the right-hand side of model (1.1.1). Taking the y_{1it} equation, for example, there would be a dynamic lag $y_{1i,t-1}$ on the right-hand side. The dynamic case of model (1.1.1) combines four effects: spatial effects, simultaneous effects, common shock effects and dynamic effects. Such a dynamic model would be useful for economic forecasting. Jointly modeling the first three effects is already difficult; the extra dynamic effect would make the analysis even more challenging. The dynamic case is studied in a work-in-progress paper.

1.6 Applications

In this section, I apply model (1.1.1) to explore the relationship between trade and GDP over time and across countries, taking into account spatial effects and global common shock effects.

In the literature on international trade, it has been difficult to establish a robust relationship between trade and GDP, due to the endogeneity issue between them. Many studies try to examine such a relationship using an IV approach. For instance, Frankel and Romer (1999) use a geographic instrument for trade and find a positive effect of trade on GDP in a cross-country setting. Such instruments are implemented by Irwin and Tervio (2002) and extended by Noguer and Siscart (2005), Felbermayr and Groschl (2013), Ortega and Peri

(2014). Despite the importance of geographic instruments, as discussed in Winters and Masters (2013), they are time-invariant and thus preclude the use of panel data to analyze the effects of trade. Then, Feyrer (2009) makes progress by proposing a time-varying geographic instrument for trade. Based on a panel data model with simple additive individual and time fixed effects,[20] he finds that trade has a significant, positive impact on GDP, with an elasticity of approximately one-half. However, as mentioned in Feyrer (2009), his IV estimates are nearly identical to ordinary least squares estimates, indicating that his instrument is doubtful. Regardless of the contribution on the endogeneity issue, common shocks are not captured well by his model setup.

Instead, I apply framework (1.1.1) to model trade and GDP in a simultaneous equation system and incorporate common shock effects using a factor structure. Such a factor structure can be regarded as a form of interactive fixed effects, which provides a flexible way to control for potential serial and cross-sectional correlations. In addition, my model considers the spatial effect (i.e., spillover effects across countries for both trade and GDP) implied by gravity theory as noted in Helpman (1987) and Anderson and van Wincoop (2003).[21] In this application, the model is specified as follows:

$$Trade_{it} = \alpha_{1i} + \rho_1 \sum_{j=1}^{N} w_{ij} Trade_{jt} + \gamma_1 GDP_{it} + \beta_1 x_{1it} + \lambda_i' f_t + e_{1it}$$

[20] Feyrer (2009) uses the real GDP per capita from the Penn World Tables, from 1950 to 1995 and cross 62 countries, with all estimation conducted on a panel with observations every 5 years.

[21] Similar spatial effects among firms due to cultural and social networks among firms as well as regional trade agreements have been studied in Baltagi et al. (2008), Lawless (2009), Rauch and Trindade (2002) and Defever et al. (2015).

Table 1.7: Estimation results using IGPC for the case without x (** significant at 1%)

ρ_1	ρ_2	γ_1	γ_2
0.9597**	0.7751**	0.6326**	0.1074**
(0.0146)	(0.0195)	(0.0298)	(0.0047)

Table 1.8: Estimation results using IGPC for the case with x (** significant at 1%)

ρ_1	ρ_2	γ_1	γ_2	β_1	β_2
0.9903**	0.5061**	0.5952**	0.0933**	0.1956**	0.9041**
(0.0332)	(0.0043)	(0.0853)	(0.0336)	(0.0118)	(0.0207)

$$GDP_{it} = \alpha_{2i} + \rho_2 \sum_{j=1}^{N} w_{ij} GDP_{jt} + \gamma_2 Trade_{it} + \beta_2 x_{2it} + \psi_i' f_t + e_{2it}$$

where $Trade_{it}$ and GDP_{it} are the log of total trade volume (export plus import) and the log of GDP, respectively, for country i in year t; weight w_{ij} is computed as $\frac{TotalTrade_{ij}}{\sum_{j=1}^{N} TotalTrade_{ij}}$ with $TotalTrade_{ij}$ being the total trade volume between country i and j.[22]

I investigate the above model in two ways, without explanatory variables and using population as an explanatory variable (i.e., $x_{1it} = x_{2it} = x_{it}$ denotes the log of population for country i in year t). Frankel and Romer (1999) adopt the same explanatory variable to control for country size.[23]

Without explanatory variables, I find one common factor based on the information criterion in (1.4.2). Given that $r = 1$, I estimate the model using the IGPC method; the results

[22]Data source: Trade data and the weighting matrix (i.e. bilateral trade data) comes from IMF Directions of Trade Statistics. GDP and population data is obtained from Penn World Table. The sample period is from 1961 to 2013, with total 61 countries. In this case, $N = 61$ and $T = 53$. Weights are constructed using the bilateral trade data of the base year 1960, to avoid potential reversal causality. Both trade and GDP data are inflation adjusted.

[23]In both cases of without and with explanatory variables, the model is identified, by Assumption G$'$.

are presented in Table 1.7.[24] First, I find that trade and GDP are positively and significantly affected by each other. Specifically, the elasticity of trade with respect to GDP (γ_1) is approximately 0.6, while the elasticity of GDP with respect to trade (γ_2) is much smaller, approximately 0.1. By comparison, Feyrer (2009) identifies an elasticity of approximately one-half of GDP with respect to trade using an IV approach, which is much larger. The result in Feyrer (2009) might be less convincing, since the instrument is probably inappropriate, the spatial effect is not captured, and only additive individual and time fixed effects are controlled for. On the contrary, my model captures general interactive fixed effects through a factor structure. Moreover, his panel runs from 1950 to 1995 with observations every 5 years, while the application here uses annual data from 1961 to 2013. Second, I find that the trade volume of a country is positively affected by the trading parties' trade volumes; likewise for GDP. Specifically, the trade volume of a country can increase by almost 1% if the average trade volume of the trading parties increases by 1%; similarly, the GDP of a country can increase by 0.77% if the average GDP of the trading parties increases by 1%.

With population included as a control variable, the estimation results and corresponding findings are similar. Again, the information criterion in (1.4.2) implies that there exists one common factor. The estimation results obtained by IGPC based on $r = 1$ are provided in Table 1.8. Similar estimated results are found for coefficients $(\rho_1, \rho_2, \gamma_1, \gamma_2)$. Additionally, the estimates of β_1 and β_2 indicate that country size in terms of population has positive impacts on both trade and GDP, with a larger impact on GDP: a 1% increase in population raises total trade volume by 0.2% whereas it increases GDP by 0.9%. By comparison, Frankel

[24]Based on the IGPC results, I did panel unit root check for the errors by various tests, and overwhelming evidence rejects the hypothesis that the errors contain unit roots. Similarly for the case with control variable.

and Romer (1999) also find a positive impact of population on GDP but the magnitude is smaller: a 1% increase in population increases GDP by approximately 0.35% based on a cross-country study and an IV approach in a single equation setting.[25]

1.7 Conclusion

In this chapter, I consider a simultaneous spatial panel data model, jointly modeling three important effects: spatial effects, common shock effects and simultaneous effects. Under joint modeling, there are many incidental parameters. Moreover, I take into account cross-sectional heteroskedasticity, which gives rise to additional incidental parameters. To estimate the model, I propose two different approaches, the QML method and the IGPC method. For each approach, I derive its identification condition and develop a full inferential theory for its estimators, including consistency, convergence rates and limiting distributions. The estimators from both methods are consistent. There is a trade-off between the model specification of the explanatory variable x and the asymptotic properties of the estimators for the two approaches. The QML method requires the model specification of x, but the gain is that its limiting distribution is unbiased (i.e., centered at zero) and more efficient (less variance than that of IGPC); the IGPC method does not require the specification of x, but the cost is that its limiting distribution contains a bias term and less efficient. Based on the limiting distribution of the IGPC estimator, the bias-corrected IGPC estimator is obtained. Then, I investigate the finite sample performance of both methods using Monte Carlo simulations. I find that both methods perform well and that the simulation results corroborate the infer-

[25]The data used in Frankel and Romer (1999) are based on year 1985 only. In addition, Frankel and Romer (1999) use trade share (i.e., trade as a percentage of GDP) instead of trade itself, to study how openness affects GDP.

ential theories I derived in this paper. I also consider some extensions of the model. Finally, I apply the model to analyze the relationship between trade and GDP over time and cross countries, taking into account spatial effects and global common shock effects.

Chapter 2

Efficient estimation of heterogenous coefficients in panel data models with common shocks

Kunpeng Li and Lina Lu

2.1 Introduction

It has been long recognized and well documented in the literature that a small number of factors can explain a large fraction of the comovement of financial, macroeconomic and sectorial variables, for example, Ross (1976), Sargent and Sims (1977), Geweke (1977) and Stock and Watson (1998). Based on this fact, recent econometric literature places particular focus on panel data models with common shocks. These models specify that the dependent variable and explanatory variables both have a factor structure. A typical example can be written as

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta_i + \lambda_i' f_t + \epsilon_{it}, \\ x_{it} &= \nu_i + \gamma_i' f_t + v_{it}, \qquad i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T. \end{aligned} \tag{2.1.1}$$

where y_{it} denotes the dependent variable; x_{it} denotes a $k \times 1$ vector of explanatory variables; and f_t is an $r \times 1$ vector of unknown factors, which represents the unobserved economic shocks. The factor loadings γ_i and λ_i capture the heterogeneous responses to the shocks. A salient feature of this paper is that the coefficients of x_{it} are assumed to be individual-dependent.

Due to the presence of factor f_t, the error term of the y equation (i.e., $\lambda_i' f_t + \epsilon_{it}$) is correlated with the explanatory variables. The usual estimation methods, such as ordinary least squares method, are not applicable. The instrumental variables (IV) method appears to be an intuitive way to address this issue, but the validity of IV is difficult to justify in practice. A remarkable result from recent studies is that, even without IV, model (2.1.1) can still be consistently estimated. The related literature includes Pesaran (2006), Bai (2009), Song (2013), Bai and Li (2014b) and Moon and Weidner (2015), among others.

Bai (2009) proposes the iterated principal components (PC) method to estimate a model with homogeneous coefficients. His analysis has been reexamined and extended by perturbation theory in Moon and Weidner (2015). Both studies find that a bias arises from cross-sectional heteroscedasticity. Bai and Li (2014b) therefore consider the quasi maximum likelihood method to eliminate this bias from the estimator. All these studies focus on the case of homogeneous coefficient. If the underlying coefficients are heterogeneous, misspecification of homogeneity would lead to inconsistent estimation (see the simulation of Kapetanios et al. (2011)).

There are several studies on the estimation of heterogeneous coefficients.[1] Pesaran (2006) proposes the common correlated effect (CCE) estimation method to estimate the heterogeneous coefficients (2.1.1). The intuition of his method is approximating the unknown projection space of the factors f_t by the space spanned by the cross-sectional average of the observations $(y_{it}, x_{it}')'$. To this end, some rank condition is needed. Song (2013) alternatively considers the iterated principal components method, which extends the analysis of Bai (2009) to the case of heterogeneous coefficients. In this chapter, we propose a new method to estimate (2.1.1). Our estimation method is motivated by both Pesaran's and Song's methods having their limitations in estimating the heterogeneous coefficients for some particular data setups. The CCE estimator has a reputation for computational simplicity and excellent finite sample properties. However, we note that in some cases rank condition alone is not enough for a good approximation. When good approximation breaks down, the CCE

[1] In empirical examples, these β_i are unknown and there might exist common values. There is another literature to consider the case of unknown group membership of β_i, for inference, see Bonhomme and Manresa (2015), Bai and Ando (2015b) and Bai and Ando (2016).

estimator would perform poorly. With Song's method, although his theory is beautiful, the minimizer of the objective function is not easily obtained, especially for the data with heavy cross-sectional heteroscedasticity. As far as we know, there is no good way to address this issue. The limitations of the CCE method and the iterated principal components method are manifested by simulations in Section 2.7.

Our estimation method is a two-step method. In the first step, we use the quasi maximum likelihood (QML) method to estimate a pure factor model. Next, the heterogeneous coefficients are estimated by using relations implied by the model and replacing the parameters with their QML estimates. The proposed estimation method aims to strike a balance between efficiency and computational economy. We note that in model (2.1.1) the computational burden cannot be ignored due to a huge number of βs being estimated, especially when N is large. This problem is made worse because we can only compute β_i ($i = 1, 2, \ldots, N$) sequentially, instead of all β_i simultaneously by matrix algebra. As a result, the iterated computation method, which requires updating β_i one by one in each iteration, may not be attractive because of the heavy computational burden. Our estimation method overcomes this problem by using the iterated computation method to estimate a pure factor model, delaying the estimation of β_i to the second step. Nevertheless, as we will show, the two-step estimators are asymptotically efficient.

The rest of this chapter is organized as follows. Section 2.2 illustrates the idea of our estimation. Section 2.3 presents some theoretical results of the factor models, in which the covariance matrix of idiosyncratic errors are block-diagonal. These results are very useful for the subsequent analysis. Section 2.4 presents the asymptotic properties of the proposed estimator. Section 2.5 extends our method to the case with zero restrictions on the loadings in the y equation. We show that when zero restrictions are present, the loadings contain information for β. We propose a minimum distance estimator to achieve the efficiency. Section 2.6 extends the model to nonzero restrictions. Section 2.7 conducts extensive simulations to investigate the finite sample properties of the proposed estimator and provides some comparisons with the competitors. Section 2.8 concludes. Throughout the paper, the norm of a vector or matrix is that of Frobenius; that is, $\|A\| = [\text{tr}(A'A)]^{1/2}$ for matrix A. In addition, we use $\dot{v}_t$ to denote $v_t - \frac{1}{T}\sum_{s=1}^{T} v_s$ for any column vector v_t and M_{wv} to denote $\frac{1}{T}\sum_{t=1}^{T} \dot{w}_t \dot{v}_t'$ for any vectors w_t and v_t.

2.2 Key idea of the estimation

To illustrate the idea of our estimation, first substitute the second equation of model (2.1.1) into the first one. Then

$$\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \nu_i \end{bmatrix} + \begin{bmatrix} \beta_i'\gamma_i' + \lambda_i' \\ \gamma_i' \end{bmatrix} f_t + \begin{bmatrix} \beta_i' v_{it} + \epsilon_{it} \\ v_{it} \end{bmatrix}.$$

Let $z_{it} = (y_{it}, x_{it}')'$, $\mu_i = (\alpha_i, \nu_i')'$, $u_{it} = (\beta_i' v_{it} + \epsilon_{it}, v_{it}')'$ and Λ_i' be the factor loadings matrix before f_t in the above equation. Now we have

$$z_{it} = \mu_i + \Lambda_i' f_t + u_{it}. \tag{2.2.1}$$

Let Ω_i be the covariance matrix of v_{it} and $\sigma_{\epsilon i}^2$ the variance of ϵ_{it}. Throughout the paper, we assume that ϵ_{it} is independent of v_{js} for all i, j, t, s. This assumption is crucial to the models with common shocks and is maintained by all the related studies; for example, Pesaran

(2006), Bai (2009), Bai and Li (2014b) and Moon and Weidner (2015). The covariance of u_{it}, denoted by Σ_{ii}, now is

$$\Sigma_{ii} = \begin{bmatrix} \Sigma_{i,11} & \Sigma_{i,12} \\ \Sigma_{i,21} & \Sigma_{i,22} \end{bmatrix} = \begin{bmatrix} \beta_i'\Omega_i\beta_i + \sigma_{\epsilon i}^2 & \beta_i'\Omega_i \\ \Omega_i\beta_i & \Omega_i \end{bmatrix}. \tag{2.2.2}$$

This leads to

$$\Sigma_{i,22}\beta_i = \Sigma_{i,21}. \tag{2.2.3}$$

Suppose that we have obtained a consistent estimator of Σ_{ii}, β_i is then estimated by

$$\hat{\beta}_i = \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21} \tag{2.2.4}$$

We call the above estimator CoVariance estimator, denoted by $\hat{\beta}_i^{CV}$ since the estimation for β_i only involves the covariance of u_{it}.

The remaining problem is to consistently estimate Σ_{ii}. A striking feature of the model (2.2.1) is that the variance matrix of its idiosyncratic errors is block-diagonal. So we need to extend the usual factor analysis to accommodate this feature.

2.3 Factor models

Let $i = 1, 2, \ldots, N, t = 1, 2, \ldots, T$. Consider the following factor models

$$z_{it} = \mu_i + \Lambda_i' f_t + u_{it}, \tag{2.3.1}$$

where z_{it} is a $\bar{K}\times 1$ vector of observations with $\bar{K}=k+1$; u_{it} is a $\bar{K}\times 1$ vector of error terms; Λ_i is an $r\times\bar{K}$ loading matrix; and f_t is an $r\times 1$ vector of factors. Let $z_t=(z'_{1t},z'_{2t},\ldots,z'_{Nt})'$, $\mu=(\mu'_1,\mu'_2,\ldots,\mu'_N)'$, $\Lambda=(\Lambda_1,\Lambda_2,\ldots,\Lambda_N)'$ and $u_t=(u'_{1t},u'_{2t},\ldots,u'_{Nt})'$, then we can rewrite (2.3.1) as

$$z_t=\mu+\Lambda f_t+u_t. \tag{2.3.2}$$

Without loss of generality, we assume that $\bar{f}=T^{-1}\sum_{t=1}^{T}f_t=0$ throughout the paper since the model can be rewritten as $z_t=\mu+\Lambda\bar{f}+\Lambda(f_t-\bar{f})+u_t=\mu^*+\Lambda f_t^*+u_t$ with $\mu^*=\mu+\Lambda\bar{f}$ and $f_t^*=f_t-\bar{f}$. To analyze (2.3.2), we make the following assumptions:

Assumption A: The factor f_t is a sequence of constant. Let $M_{ff}=T^{-1}\sum_{t=1}^{T}\dot{f}_t\dot{f}_t'$ with $\dot{f}_t=f_t-T^{-1}\sum_{t=1}^{T}f_t$. We assume that $\overline{M}_{ff}=\lim\limits_{T\to\infty}M_{ff}$ is a strictly positive definite matrix.

Assumption B: The idiosyncratic error term u_{it} is assumed such that

B.1 u_{it} is independent and identically distributed $(i.i.d)$ over t and uncorrelated over i with $E(u_{it})=0$ and $E(\|u_{it}^4\|)\leq\infty$ for all $i=1,\cdots,N$ and $t=1,\cdots,T$. Let Σ_{ii} be the variance of u_{it} and $\Psi=\text{diag}(\Sigma_{11},\Sigma_{22},\ldots,\Sigma_{NN})$ be the variance of u_t.

B.2 f_t is independent of u_{js} for all (j,t,s).

Assumption C: There exists a positive constant C sufficiently large such that

C.1 $\|\Lambda_i\|\leq C$ for all $i=1,\cdots,N$.

C.2 $C^{-1}\leq\tau_{min}(\Sigma_{ii})\leq\tau_{max}(\Sigma_{ii})\leq C$ for all $i=1,\cdots,N$, where $\tau_{min}(\cdot)$ and $\tau_{max}(\cdot)$ denote the smallest and largest eigenvalues of its argument, respectively.

C.3 There exists an $r\times r$ positive matrix Q such that $Q=\lim\limits_{N\to\infty}N^{-1}\Lambda'\Psi^{-1}\Lambda$, where $\Lambda=(\Lambda_1,\Lambda_2,\ldots,\Lambda_N)'$ and Ψ is the variance of $u_t=(u'_{1t},u'_{2t},\ldots,u'_{Nt})'$.

Assumption D: The variances Σ_{ii} for all i are estimated in a compact set; that is, all the eigenvalues of $\hat{\Sigma}_{ii}$ are in an interval $[C^{-1}, C]$ for sufficiently large positive constant C.

Assumptions A-D are usually made in the context of factor analysis; for example, Bai and Li (2012, 2014b). Readers are referred to Bai and Li (2012) for the related discussions on these assumptions.

2.3.1 Estimation

The objective function used to estimate (2.3.2) is

$$\ln \mathscr{L}(\theta) = -\frac{1}{2N} \ln |\Sigma_{zz}| - \frac{1}{2N} \text{tr}[M_{zz} \Sigma_{zz}^{-1}] \tag{2.3.3}$$

where $\theta = (\Lambda, \Psi, M_{ff})$ and $\Sigma_{zz} = \Lambda M_{ff} \Lambda' + \Psi$; $M_{zz} = \frac{1}{T} \sum_{t=1}^{T} \dot{z}_t \dot{z}_t'$ is the data matrix where $\dot{z}_t = z_t - \frac{1}{T} \sum_{s=1}^{T} z_s$. Suppose that f_t is random and follows $N(0, M_{ff})$, the above objective function is the corresponding likelihood function after concentrating out the intercept μ. Although the factors f_t are assumed to be fixed constants, we still use the above objective function and call the maximizer $\hat{\theta} = (\hat{\Lambda}, \hat{\Psi}, \hat{M}_{ff})$, defined by

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmax}} \ \ln \mathscr{L}(\theta),$$

the quasi maximum likelihood estimator, or the QMLE, where Θ is the parameter space specified by Assumption D.

It is known in factor analysis that the loadings and factors can only be identified up to a rotation. To see this, let $\hat{\theta} = (\hat{\Lambda}, \hat{\Psi}, \hat{M}_{ff})$ be the maximizer of (2.3.3), then $\hat{\theta}^{\dagger} =$

$(\hat{\Lambda}\hat{M}_{ff}^{1/2}, \hat{\Psi}, I_r)$ is also a qualified maximizer. From this perspective, it is no loss of generality to normalize that

$$M_{ff} = \frac{1}{T}\sum_{t=1}^{T} f_t f_t' = I_r.$$

Under this normalization, Σ_{zz} is simplified as $\Sigma_{zz} = \Lambda\Lambda' + \Psi$.

Maximizing the objective function (2.3.3) with respect to Λ and Ψ gives the following two first order conditions.

$$\hat{\Lambda}'\hat{\Psi}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = 0 \tag{2.3.4}$$

$$\text{Bdiag}(M_{zz} - \hat{\Sigma}_{zz}) = 0 \tag{2.3.5}$$

where $\text{Bdiag}(\cdot)$ is the block-diagonal operator, which puts the element of its argument to zero if the counterpart of Ψ is nonzero, otherwise unspecified. $\hat{\Lambda}$ and $\hat{\Psi}$ denote the QMLE and $\hat{\Sigma}_{zz} = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$.

2.3.2 Asymptotic properties of the QMLE

This section presents the asymptotic results of the QMLE for (2.3.3). Since we only impose $M_{ff} = I_r$ in (2.3.2), the loadings and factors still cannot be fully identified. We adopt the treatment of Bai (2003), in which the rotational matrix appears in the asymptotic representation. This treatment has two advantages in the present context. First, it simplifies our analysis. Second, it clarifies that the estimation and inferential theory of β is invariant to the rotational matrix. Alternatively, we can impose some additional restrictions to uniquely fix the rotational matrix; see Bai and Li (2012) for full identification strategies. The following theorem, which serves as the base for the subsequent analysis, gives the asymptotic representations of the MLE.

Theorem 2.3.1. *Under Assumptions A-D, as $N, T \to \infty$, we have*

$$\hat{\Lambda}_i - R'\Lambda_i = R'\frac{1}{T}\sum_{t=1}^{T} f_t u'_{jt} + o_p(T^{-1/2})$$
$$\hat{\Sigma}_{ii} - \Sigma_{ii} = \frac{1}{T}\sum_{t=1}^{T}(u_{it}u'_{it} - \Sigma_{ii}) + o_p(T^{-1/2})$$

where $R = \Lambda'\hat{\Psi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}$.

Remark 2.3.1. Notice that the rotational matrix R only enters in the asymptotic representation of $\hat{\Lambda}_i$. This is consistent with only loadings and factors having rotational indeterminacy and idiosyncratic errors not having such a problem.

Remark 2.3.2. By the above theorem, we immediately have $\hat{\Lambda}_i - R'\Lambda_i = O_p(T^{-1/2})$ and $\hat{\Sigma}_{ii} - \Sigma_{ii} = O_p(T^{-1/2})$. These two results continue to hold when N is fixed since the model falls within the scope of traditional factor analysis. But the asymptotic representations will be more complicated when N is finite. An implication of this result is that the covariance estimator $\hat{\beta}_i^{CV}$ is consistent even when N is finite.

2.4 Asymptotic results for the covariance estimator

Now we use the results in Theorem 2.3.1 to derive the asymptotic representation of $\hat{\beta}_i^{CV}$. Notice $\hat{\beta}_i^{CV} = (\hat{\Sigma}_{i,22})^{-1}\hat{\Sigma}_{i,21}$ and $\beta_i = (\Sigma_{i,22})^{-1}\Sigma_{i,21}$. Given $\hat{\Sigma}_{ii} = \Sigma_{ii} + o_p(1)$ by Theorem 2.3.1, the consistency of $\hat{\beta}_i$ is immediately obtained by the continuous mapping theorem.

Furthermore, by Theorem 2.3.1,

$$\hat{\Sigma}_{ii} - \Sigma_{ii} = \frac{1}{T}\sum_{t=1}^{T}(u_{it}u_{it}' - \Sigma_{ii}) + O_p(T^{-1}).$$

Then it follows

$$\hat{\Sigma}_{i,21} - \Sigma_{i,21} = \frac{1}{T}\sum_{t=1}^{T}[v_{it}(\epsilon_{it} + v_{it}'\beta_i) - \Omega_i\beta_i] + O_p(T^{-1}); \tag{2.4.1}$$

$$\hat{\Sigma}_{i,22} - \Sigma_{i,22} = \frac{1}{T}\sum_{t=1}^{T}[v_{it}v_{it}' - \Omega_i] + O_p(T^{-1}). \tag{2.4.2}$$

Notice that

$$\begin{aligned}\hat{\beta}_i^{CV} - \beta_i &= (\hat{\Sigma}_{i,22})^{-1}\hat{\Sigma}_{i,21} - \Sigma_{i,22}^{-1}\Sigma_{i,21} \\ &= (\hat{\Sigma}_{i,22})^{-1}\Big[(\hat{\Sigma}_{i,21} - \Sigma_{i,21}) - (\hat{\Sigma}_{i,22} - \Sigma_{i,22})\Sigma_{i,22}^{-1}\Sigma_{i,21}\Big]\end{aligned} \tag{2.4.3}$$

Substituting (2.4.1) and (2.4.2) into (2.4.3) and noting that $\hat{\Sigma}_{i,22} \xrightarrow{p} \Omega_i$ and $\beta_i = \Sigma_{i,22}^{-1}\Sigma_{i,21}$, we have the following theorem on $\hat{\beta}_i^{CV}$.

Theorem 2.4.1. *Under Assumptions A-D, when $N, T \to \infty$, we have*

$$\sqrt{T}(\hat{\beta}_i^{CV} - \beta_i) = \Omega_i^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{it}\epsilon_{it}\Big) + o_p(1) \tag{2.4.4}$$

Remark 2.4.1. The above asymptotic result implies that our estimator is asymptotically efficient. To see this, suppose that the factors f_t are observed, then the generalized least squares (GLS) estimator has the asymptotic representation:

$$\sqrt{T}(\hat{\beta}_i^{GLS} - \beta_i) = \Omega_i^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} v_{it}\epsilon_{it}\Big) + o_p(1), \tag{2.4.5}$$

which is the same as that of Theorem 2.4.1, implying the asymptotic efficiency of the CV estimator.

Remark 2.4.2. Although the asymptotic result of $\hat{\beta}_i^{CV}$ is derived under Assumption B, we point out that the proposed method works in a very general setup given the results of Bai and Li (2016), which show that the quasi maximum likelihood method can be used to estimate approximate factor models (Chamberlain and Rothschild (1983)). More specifically, let $\Sigma_{ii,t}$ be the variance of u_{it}, where the covariance matrix has an additional superscript t to indicate that it is time-varying. Partition $\Sigma_{ii,t}$ as

$$\Sigma_{ii,t} = \begin{bmatrix} \Sigma_{ii,t,11} & \Sigma_{ii,t,12} \\ \Sigma_{ii,t,21} & \Sigma_{ii,t,22} \end{bmatrix}.$$

Under the assumption that ϵ_{it} is independent of v_{it}, we have $\Sigma_{ii,t,22}\beta_i = \Sigma_{ii,t,21}$ for all t, which implies that

$$\Big(\frac{1}{T}\sum_{t=1}^{T}\Sigma_{ii,t,22}\Big)\beta_i = \frac{1}{T}\sum_{t=1}^{T}\Sigma_{ii,t,21}.$$

To consistently estimate β_i, it suffices to consistently estimate $\frac{1}{T}\sum_{t=1}^{T}\Sigma_{ii,t}$. As shown in Bai and Li (2016), if the underlying covariance is time-varying but misspecified to be time-invariant in the estimation, the resulting estimator of the covariance is a consistent estimator for the average underlying covariance over time, that is, $\frac{1}{T}\sum_{t=1}^{T}\Sigma_{ii,t}$ happens to be estimated by the QMLE.

Remark 2.4.3. For the basic model, the CCE estimator of Pesaran (2006) and the iterated PC estimator of Song (2013) have the same asymptotic representations as in Theorem 2.4.1 and hence are asymptotically efficient. However, different methods require different conditions for the asymptotic theory. Except for the rank condition, the CCE estimator potentially requires N be large, otherwise the average error over the cross section cannot be negligible. The PC estimator is derived under the cross-sectional homoscedasticity. If heteroscedasticity is present, a large N is needed to ensure the consistency. For the CV estimator, the consistency can be maintained for a fixed N even in the presence of the cross-sectional heteroscedasticity. So the CV estimator requires the least restrictive condition for the consistency.

Remark 2.4.4. With slight modification, our method can be used to estimate the homogeneous coefficient. Suppose $\beta_i \equiv \beta$ for all i. Now we have $\Sigma_{i,22}\beta = \Sigma_{i,21}$ for all i, which leads to

$$\Big(\sum_{i=1}^{N} \Sigma_{i,22}\Big)\beta = \sum_{i=1}^{N} \Sigma_{i,21}.$$

So a consistent estimator for β is

$$\hat{\beta} = \big(\sum_{i=1}^{N} \hat{\Sigma}_{i,22}\big)^{-1}\big(\sum_{i=1}^{N} \hat{\Sigma}_{i,21}\big). \tag{2.4.6}$$

The asymptotic properties of $\hat{\beta}$ will not be pursued in this paper. In section 2.7, we conduct a small simulation to examine its finite sample performance.

Corollary 2.4.1. *Under the assumptions of Theorem 2.4.1, we have*

$$\sqrt{T}(\hat{\beta}_i^{CV} - \beta_i) \xrightarrow{d} N\big(0, \sigma_{\epsilon i}^2 \Omega_i^{-1}\big),$$

where $\sigma_{\epsilon i}^2$ is the variance of ϵ_{it} and Ω_i is the variance of v_{it}. The variance $\sigma_{\epsilon i}^2 \Omega_i^{-1}$ can be consistently estimated by $\hat{\sigma}_{\epsilon i}^2 \hat{\Sigma}_{i,22}^{-1}$, where $\hat{\sigma}_{\epsilon i}^2 = \hat{\Sigma}_{i,11} - \hat{\beta}_i^{CV\prime} \hat{\Sigma}_{i,22} \hat{\beta}_i^{CV}$.

2.5 Models with zero restrictions

In this section, we consider the following restricted model:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta_i + \psi_i' g_t + \epsilon_{it} \\ x_{it} &= \nu_i + \gamma_i^{g\prime} g_t + \gamma_i^{h\prime} h_t + v_{it} \end{aligned} \tag{2.5.1}$$

where the dimensions of g_t and h_t are $r_1 \times 1$ and $r_2 \times 1$, respectively. A salient feature of model (2.5.1) is that the explanatory variables include more factors than the error of the y equation. This specification aims to accommodate that both endogenous and exogenous shocks exist in the economic system. Endogenous shocks such as unexpected monetary supply would directly affect all economic variables. Exogenous shocks such as oil prices would first affect the energy-related industries and then gradually affect other economic variables. In model (2.5.1), g_t denotes the endogenous shocks that directly affect y and x, and h_t denotes the exogenous shocks that affect first x then y[2].

[2] Another way to see this point is as follows. Notice that the x equation can always be written as

$$x_{it} = \nu_i + (\gamma_i^{g\prime} + \gamma_i^{h\prime} H'G(G'G)^{-1}) g_t + \gamma_i^{h\prime}(h_t - H'G(G'G)^{-1} g_t) + v_{it} = \nu_i + \gamma_i^{*g\prime} g_t + \gamma_i^{h\prime} h_t^* + v_{it}.$$

The y equation of (2.5.1) can be written as

$$y_{it} = \alpha_i + x'_{it}\beta_i + \psi'_i g_t + \phi'_i h_t + \epsilon_{it}$$

with $\phi_i = 0$ for all i. Let $f_t = (g'_t, h'_t)'$, $\lambda_i = (\psi'_i, \phi'_i)'$ and $\gamma_i = (\gamma_i^{g\prime}, \gamma_i^{h\prime})'$, we have the same representation as (2.1.1). From this perspective, model (2.5.1) can be viewed as a restricted version of model (2.1.1). This implies that the two-step method proposed in Section 2.4 is applicable to (2.5.1). However, this estimation method is not efficient. Consider the ideal case that g_t is observable. To eliminate the endogenous ingredient $\psi'_i g_t$, we post-multiply $M_G = I - G(G'G)^{-1}G'$ on both sides of the y equation. The remaining part of x_{it} includes v_{it} and $\gamma_i^{h\prime}(h_t - H'G(G'G)^{-1}g_t)$, which both provide the information for β. However, as shown in Theorem 2.4.1, only the variations of v_{it} are used to signal β_i in $\hat{\beta}_i^{CV}$. Therefore, partial information is discarded and the two-step method in Section 2.4 is inefficient.

The preceding discussion provides some insights on the improvement of efficiency. To efficiently estimate model (2.5.1), we need to use information contained in the common components of x_{it}. Rewrite model (2.5.1) as

$$\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \nu_i \end{bmatrix} + \begin{bmatrix} \beta'_i\gamma_i^{g\prime} + \psi'_i & \beta'_i\gamma_i^{h\prime} \\ \gamma_i^{g\prime} & \gamma_i^{h\prime} \end{bmatrix} \begin{bmatrix} g_t \\ h_t \end{bmatrix} + \begin{bmatrix} \beta'_i v_{it} + \epsilon_{it} \\ v_{it} \end{bmatrix} \tag{2.5.2}$$

We use Λ'_i to denote the loadings matrix before $f_t = (g'_t, h'_t)'$. The symbols μ_i, z_{it} and u_{it} are defined the same as in the previous section. We then have the same equation as (2.2.1).

In the last equation, g_t is uncorrelated with h_t^*. Given this expression, it is no loss of generality to assume that h_t is uncorrelated with g_t. Now we see that g_t causes the endogeneity problem but h_t does not. So we say that g_t represents endogenous shocks and h_t represents exogenous shocks.

Further partition the loadings matrix Λ_i into four blocks,

$$\Lambda_i = \begin{bmatrix} \Lambda_{i,11} & \Lambda_{i,12} \\ \Lambda_{i,21} & \Lambda_{i,22} \end{bmatrix} = \begin{bmatrix} \psi_i + \gamma_i^g \beta_i & \gamma_i^g \\ \gamma_i^h \beta_i & \gamma_i^h \end{bmatrix}. \tag{2.5.3}$$

So we have $\Lambda_{i,22}\beta_i = \Lambda_{i,21}$. This result together with (2.2.3) leads to

$$\begin{bmatrix} \Lambda_{i,22} \\ \Sigma_{i,22} \end{bmatrix} \beta_i = \begin{bmatrix} \Lambda_{i,21} \\ \Sigma_{i,21} \end{bmatrix} \tag{2.5.4}$$

Given the above structural relationship, a routine to estimate β_i is replacing $\Lambda_{i,22}, \Lambda_{i,21}, \Sigma_{i,22}$ and $\Sigma_{i,21}$ with their QMLE and minimizing the distance on the both sides of the equation with some weighting matrix. While this method is intuitive, it is not correct since $\hat{\Lambda}_{i,22}$ and $\hat{\Lambda}_{i,21}$ are not consistent estimators of $\Lambda_{i,22}$ and $\Lambda_{i,21}$, as shown in Theorem 2.3.1. Let $\Lambda_i^* = R'\Lambda_i$ represent the underlying parameters that the QMLE corresponds to, where R is the rotation matrix defined in Theorem 2.3.1. Then

$$\Lambda_i^{*\prime} = \begin{bmatrix} \Lambda_{i,11}^{*\prime} & \Lambda_{i,21}^{*\prime} \\ \Lambda_{i,12}^{*\prime} & \Lambda_{i,22}^{*\prime} \end{bmatrix} = \Lambda_i' R = \begin{bmatrix} \Lambda_{i,11}' & \Lambda_{i,21}' \\ \Lambda_{i,12}' & \Lambda_{i,22}' \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} \beta_i' \gamma_i^{g\prime} + \psi_i' & \beta_i' \gamma_i^{h\prime} \\ \gamma_i^{g\prime} & \gamma_i^{h\prime} \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

implying

$$\Lambda_{i,21}^* = (R_{12}'\gamma_i^g + R_{22}'\gamma_i^h)\beta_i + R_{12}'\psi_i \tag{2.5.5}$$

$$\Lambda_{i,22}^* = R_{12}'\gamma_i^g + R_{22}'\gamma_i^h \tag{2.5.6}$$

From (2.5.5) and (2.5.6), we see that unless $\psi_i = 0$, $\Lambda_{i,22}^*\beta_i = \Lambda_{i,21}^*$ does not hold. But when $\psi_i = 0$, we see from (2.5.1) that the model is free of the endogeneity problem and the ordinary least squares method is applicable. The preceding analysis indicates that the existence of the rotational indeterminacy for loadings impedes the use of the underlying relation $\Lambda_{i,22}\beta_i = \Lambda_{i,21}$ in the estimation of β_i.

Although this result is a little disappointing, we now show that with some transformation, $\Lambda_{i,22}\beta_i = \Lambda_{i,21}$ can still be used to estimate β_i. First by $\Lambda_i^{*\prime} = \Lambda_i' R$,

$$\Lambda_{i,11}^* = (R_{11}'\gamma_i^g + R_{21}'\gamma_i^h)\beta_i + R_{11}'\psi_i \tag{2.5.7}$$

$$\Lambda_{i,12}^* = R_{11}'\gamma_i^g + R_{21}'\gamma_i^h \tag{2.5.8}$$

By the expressions (2.5.5)-(2.5.8), we have the following equation:

$$(\Lambda_{i,21}^* - \Lambda_{i,22}^*\beta_i) = R_{12}'R_{11}'^{-1}(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i) = V(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i) \tag{2.5.9}$$

where $V = R_{12}'R_{11}'^{-1}$, an $r_2 \times r_1$ rotational matrix. The preceding equation can be written as

$$(\Lambda_{i,22}^* - V\Lambda_{i,12}^*)\beta_i = \Lambda_{i,21}^* - V\Lambda_{i,11}^* \tag{2.5.10}$$

Given the above result, together with (2.2.3), we have

$$\begin{bmatrix} \Lambda_{i,22}^* - V\Lambda_{i,12}^* \\ \Sigma_{i,22} \end{bmatrix} \beta_i = \begin{bmatrix} \Lambda_{i,21}^* - V\Lambda_{i,11}^* \\ \Sigma_{i,21} \end{bmatrix} \tag{2.5.11}$$

If V is known, then we can replace $\Lambda_{i,11}^*, \Lambda_{i,12}^*, \Lambda_{i,21}^*, \Lambda_{i,22}^*$ with the corresponding estimates, and β_i is efficiently estimated. Although V is unknown, it can be consistently estimated by (2.5.9) since β_i can be consistently (albeit not efficiently) estimated by $\hat{\beta}_i^{CV} = \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}$. Given the above analysis, we propose the following estimation procedure:

1. *Use the maximum likelihood method to obtain the estimates* $\hat{\Sigma}_{ii}, \hat{\Lambda}_i, \hat{f}_t$ *for all* i *and* t.

2. *Calculate* $\hat{\beta}_i^{CV} = \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}$ *and*

$$\hat{V} = \Big[\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}-\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]\Big[\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11}-\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]^{-1}.$$

3. *Calculate* $\hat{\beta}_i = (\hat{\Delta}_i' W_i^{-1}\hat{\Delta}_i)^{-1}\hat{\Delta}_i' W_i^{-1}\hat{\delta}_i$, *where* W_i *is a predetermined weighting matrix that is specified below, and*

$$\hat{\Delta}_i = \begin{bmatrix} \hat{\Lambda}_{i,22} - \hat{V}\hat{\Lambda}_{i,12} \\ \hat{\Sigma}_{i,22} \end{bmatrix}, \qquad \hat{\delta}_i = \begin{bmatrix} \hat{\Lambda}_{i,21} - \hat{V}\hat{\Lambda}_{i,11} \\ \hat{\Sigma}_{i,21} \end{bmatrix} \tag{2.5.12}$$

where we call the resulting estimator the Loading-coVariance estimators, denoted by $\hat{\beta}_i^{LV}$.

Remark 2.5.1. We can iterate the second and third steps by using the updated estimator of β_i to calculate $\hat{V}$. We call the estimator resulting from this iterating procedure the *Iterated-LV* estimator, denoted by $\hat{\beta}_i^{ILV}$. The iterated estimator has the same asymptotic representation as the LV estimator, but better finite sample performance; see the simulation results in Section 2.7.

2.5.1 The optimal weighting matrix

To carry out the estimation procedure, we need to specify the weighting matrix W_i. It can be shown that the theoretically optimal weighting matrix is

$$W_i^{opt} = \begin{bmatrix} R'_{22\cdot1}M_{hh\cdot g}^{-1}R_{22\cdot1} & 0_{r_2\times k} \\ 0_{k\times r_2} & \Sigma_{i,22} \end{bmatrix},$$

where $R_{22\cdot1} = R_{22} - R_{21}R_{11}^{-1}R_{12}$ and $M_{hh\cdot g} = M_{hh} - M_{hg}M_{gg}^{-1}M_{gh}$. This weighting matrix can be consistently estimated by

$$\hat{W}_i = \begin{bmatrix} \Big[\big(\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}'_t\big) - \big(\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{\eta}'_t\big)\big(\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t\hat{\eta}'_t\big)^{-1}\big(\frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t\hat{h}'_t\big)\Big]^{-1} & 0_{r_2\times k} \\ 0_{k\times r_2} & \hat{\Sigma}_{i,22} \end{bmatrix} \quad (2.5.13)$$

with $\hat{\eta}_t = \hat{g}_t + \hat{V}'\hat{h}_t$, where $\hat{g}_t$ and $\hat{h}_t$ are given by

$$\begin{bmatrix} \hat{g}_t \\ \hat{h}_t \end{bmatrix} = \Big(\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}'_i\Big)^{-1}\Big(\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}z_{it}\Big).$$

2.5.2 The asymptotic result

The following theorem gives the asymptotic representation of the LV estimator with some remarks following.

Theorem 2.5.1. *Under Assumptions A-D, when $N, T \to \infty$, we have*

$$\sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) = \big(\gamma_i^{h\prime}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})\gamma_i^h + \Omega_i\big)^{-1} \times \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\gamma_i^{h\prime}\big(\dot{h}_t - M_{hg}M_{gg}^{-1}\dot{g}_t\big) + v_{it}\Big]\epsilon_{it} + o_p(1)$$

Given Theorem 2.5.1, we have the following corollary:

Corollary 2.5.1. *Under the assumptions of Theorem 2.5.1, we have*

$$\sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) \xrightarrow{d} N\big(0, \sigma_{\epsilon i}^2(\gamma_i^{h\prime}\overline{M}_{hh\cdot g}\gamma_i^h + \Omega_i)^{-1}\big).$$

where $\overline{M}_{hh\cdot g} = \operatorname*{plim}_{T\to\infty}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})$. The above asymptotic result can be presented alternatively as

$$\sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) \xrightarrow{d} N\big(0, \sigma_{\epsilon i}^2\big[\operatorname*{plim}_{T\to\infty}\frac{1}{T}X_i' M_{\overline{G}} X_i\big]^{-1}\big).$$

with $\overline{G} = (\mathbf{1}_T, G)$, where $\mathbf{1}_T$ is a T-dimensional vector with all the elements equal to 1.

Remark 2.5.2. Consider the "y" equation, which can be written as

$$Y_i = \alpha_i \mathbf{1}_T + X_i\beta_i + G\psi_i + E_i \tag{2.5.14}$$

where $Y_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$, $X_i = (x_{i1}, x_{i2}, \dots, x_{iT})'$, and E_i is defined similarly as Y_i. If the factors g_t are observable, the infeasible GLS estimator for β_i is

$$\hat{\beta}_i^{GLS} = (X_i' M_{\overline{G}} X_i)^{-1}(X_i' M_{\overline{G}} Y_i).$$

By (2.5.14), we have

$$\hat{\beta}_i^{GLS} - \beta_i = (X_i' M_{\overline{G}} X_i)^{-1} (X_i' M_{\overline{G}} E_i).$$

Notice $\text{var}(E_i) = \sigma_{\epsilon i}^2 I_T$. Thus the limiting distribution of $\hat{\beta}_i^{GLS} - \beta_i$ conditional on X_i is

$$\sqrt{T}(\hat{\beta}_i^{GLS} - \beta_i) \xrightarrow{d} N\big(0, \sigma_{\epsilon i}^2 \big[\operatorname*{plim}_{T\to\infty} \frac{1}{T} X_i' M_{\overline{G}} X_i\big]^{-1}\big).$$

the same as that of Corollary (2.5.1). This means that the LV estimator $\hat{\beta}_i^{LV}$ is asymptotically efficient.

Remark 2.5.3. Consider the following model, in which zero restrictions exist in both the x equation and the y equation:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}' \beta_i + \psi_i' g_t + \epsilon_{it} \\ x_{it} &= \nu_i + \gamma_i^{h\prime} h_t + v_{it} \end{aligned} \tag{2.5.15}$$

where g_t and h_t are assumed to be correlated. Model (2.5.15) is a special case of (2.5.1) in view that γ_i^g is restricted to zero. So the loading-covariance two-step method can be directly applied to (2.5.15). We note that the LV estimator is efficient even in the presence of additional zero restrictions $\gamma_i^g = 0$. To see this point, notice that Λ_i in model (2.5.15) is

$$\Lambda_i = \begin{bmatrix} \Lambda_{i,11} & \Lambda_{i,12} \\ \Lambda_{i,21} & \Lambda_{i,22} \end{bmatrix} = \begin{bmatrix} \psi_i & 0 \\ \gamma_i^h \beta_i & \gamma_i^h \end{bmatrix}.$$

The coefficient β_i can only be estimated by the relations of $\Lambda_{i,21}$ and $\Lambda_{i,22}$, which is the same as Model (2.5.1). By the same arguments, we conclude that the model

$$y_{it} = \alpha_i + x_{it}'\beta_i + \psi_i' g_t + \phi_i' h_t + \epsilon_{it},$$

$$x_{it} = \nu_i + \gamma_i^{h\prime} h_t + v_{it}.$$

is efficiently estimated by the CV method.

Remark 2.5.4. If the underlying coefficients are identical, we can also use the information contained in the loadings to improve the efficiency. Let

$$\hat{g}_i(V,\beta) = \begin{bmatrix} \hat{\Lambda}_{i,22} - V\hat{\Lambda}_{i,12} \\ \hat{\Sigma}_{i,22} \end{bmatrix} \beta - \begin{bmatrix} \hat{\Lambda}_{i,21} - V\hat{\Lambda}_{i,11} \\ \hat{\Sigma}_{i,21} \end{bmatrix}.$$

Given equation (2.5.11) (notice that now $\beta_i \equiv \beta$ for all i) we can consistently estimate β by

$$(\hat{\beta}^{LV}, \hat{V}) = \underset{\beta, V}{\operatorname{argmin}} \sum_{i=1}^{N} \hat{g}_i(V,\beta)' \hat{W}_i^{-1} \hat{g}_i(V,\beta). \tag{2.5.16}$$

where $\hat{W}_i$ is defined in (2.5.13). Notice that if Λ is identified, we can estimate β by (2.5.4), replacing the unknown parameters with their estimates. So the additional estimation of V can be regarded as the cost we pay for the rotational indeterminacy. The finite sample properties of the above LV estimator will be investigated in Section 2.7.

2.6 Discussions on models with time-invariant regressors

In some applications, it is of interest to include some time-invariant variables, such as gender, race, education, and so forth. In this section, we address this concern. Consider the following model with time-invariant variables:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}'\beta_i + \psi_i' g_t + \phi_i' h_t + \epsilon_{it} \\ x_{it} &= \nu_i + \gamma_i^{g\prime} g_t + \gamma_i^{h\prime} h_t + v_{it} \end{aligned} \tag{2.6.1}$$

where ϕ_i are observable and represent the time-invariant regressors. Model (2.6.1) specifies that the coefficients of ϕ_i are time-varying. We believe that this is a sensible way to make the model flexible enough. Now we show that our estimation idea can be used to estimate (2.6.1). As in the previous section, rewrite model (2.6.1) as

$$\begin{bmatrix} y_{it} \\ x_{it} \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \nu_i \end{bmatrix} + \begin{bmatrix} \beta_i'\gamma_i^{g\prime} + \psi_i' & \beta_i'\gamma_i^{h\prime} + \phi_i' \\ \gamma_i^{g\prime} & \gamma_i^{h\prime} \end{bmatrix} \begin{bmatrix} g_t \\ h_t \end{bmatrix} + \begin{bmatrix} \beta_i' v_{it} + \epsilon_{it} \\ v_{it} \end{bmatrix} \tag{2.6.2}$$

Let Λ_i' be the loadings matrix before $f_t = (g_t', h_t')'$ and partition it into four blocks, we have

$$\Lambda_i = \begin{bmatrix} \Lambda_{i,11} & \Lambda_{i,12} \\ \Lambda_{i,21} & \Lambda_{i,22} \end{bmatrix} = \begin{bmatrix} \psi_i + \gamma_i^g \beta_i & \gamma_i^g \\ \phi_i + \gamma_i^h \beta_i & \gamma_i^h \end{bmatrix} \tag{2.6.3}$$

Let $\Lambda_i^* = R'\Lambda_i$ be the underlying parameters that the estimators correspond to. So we have

$$\Lambda_i^{*\prime} = \begin{bmatrix} \Lambda_{i,11}^{*\prime} & \Lambda_{i,21}^{*\prime} \\ \Lambda_{i,12}^{*\prime} & \Lambda_{i,22}^{*\prime} \end{bmatrix} = \Lambda_i' R = \begin{bmatrix} \Lambda_{i,11}' & \Lambda_{i,21}' \\ \Lambda_{i,12}' & \Lambda_{i,22}' \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

This leads to

$$\Lambda_{i,11}^* = (R_{11}'\gamma_i^g + R_{21}'\gamma_i^h)\beta_i + R_{11}'\psi_i + R_{21}'\phi_i, \qquad \Lambda_{i,12}^* = R_{11}'\gamma_i^g + R_{21}'\gamma_i^h \tag{2.6.4}$$

$$\Lambda_{i,21}^* = (R_{12}'\gamma_i^g + R_{22}'\gamma_i^h)\beta_i + R_{12}'\psi_i + R_{22}'\phi_i, \qquad \Lambda_{i,22}^* = R_{12}'\gamma_i^g + R_{22}'\gamma_i^h \tag{2.6.5}$$

From (2.6.4)−(2.6.5), we have

$$R_{12}'R_{11}'^{-1}(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i) + R_{22\cdot1}'\phi_i = (\Lambda_{i,21}^* - \Lambda_{i,22}^*\beta_i) \tag{2.6.6}$$

where $R_{22\cdot1} = R_{22} - R_{21}R_{11}^{-1}R_{12}$. Given (2.6.6) together with $\Sigma_{i,22}\beta_i = \Sigma_{i,21}$, we have

$$\begin{bmatrix} \Lambda_{i,22}^* - V\Lambda_{i,12}^* \\ \Sigma_{i,22} \end{bmatrix} \beta_i = \begin{bmatrix} \Lambda_{i,21}^* - V\Lambda_{i,11}^* - R_{22\cdot1}'\phi_i \\ \Sigma_{i,21} \end{bmatrix} \tag{2.6.7}$$

where $V = R_{12}'R_{11}'^{-1}$. If V and $R_{22\cdot1}$ are known, we can use (2.6.7) to efficiently estimate β_i. Similarly as in the previous section, we can use $\hat{\beta}_i^{CV}$ to get a preliminary estimators for V and $R_{22\cdot1}$. This leads to the following estimation procedures:

1. *Use the maximum likelihood method to obtain the estimates $\hat{\Sigma}_{ii}, \hat{\Lambda}_i$ and $\hat{f}_t$ for all i and t.*

2. *Calculate* $\hat{\beta}_i^{CV} = \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}$ *and* $\hat{V}$ *and* $\hat{R}_{22\cdot 1}$ *by*

$$[\hat{V}, \hat{R}'_{22\cdot 1}] = \Big[\sum_{i=1}^{N}(\hat{\Lambda}_{i,21} - \hat{\Lambda}_{i,22}\hat{\beta}_i^{CV})\Xi_i\Big]\Big[\sum_{i=1}^{N}\Xi_i\Xi_i'\Big]^{-1}$$

where $\Xi_i = [(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})', \phi_i']'$.

3. *Calculate* $\hat{\beta}_i^{LV} = (\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\Delta}_i)^{-1}\hat{\Delta}_i'\hat{W}_i^{-1}\hat{\gamma}_i$, *where*

$$\hat{\Delta}_i = \begin{bmatrix} \hat{\Lambda}_{i,22} - \hat{V}\hat{\Lambda}_{i,12} \\ \\ \hat{\Sigma}_{i,22} \end{bmatrix}, \qquad \hat{\gamma}_i = \begin{bmatrix} \hat{\Lambda}_{i,21} - \hat{V}\hat{\Lambda}_{i,11} - \hat{R}'_{22\cdot 1}\phi_i \\ \\ \hat{\Sigma}_{i,21} \end{bmatrix}$$

and $\hat{W}_i$ *is the predetermined weighting matrix, which is the same as (2.5.13).*

Similarly we can iterate Steps 2 and 3 by replacing $\hat{\beta}_i^{CV}$ with the updated LV estimator. This leads to the iterated LV estimator. Under the same conditions of Theorem (2.5.1), we can show

$$\sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) = \big(\gamma_i^{h\prime}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})\gamma_i^h + \Omega_i\big)^{-1} \times \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Big[\gamma_i^{h\prime}\big(\dot{h}_t - M_{hg}M_{gg}^{-1}\dot{g}_t\big) + v_{it}\Big]\epsilon_{it} + o_p(1)$$

The above asymptotic result can be interpreted in a similar way as in Remark 2.5.2. So the LV estimator is asymptotically efficient.

2.7 Finite sample properties

In this section, we run Monte Carlo simulations to investigate the finite sample properties of the proposed estimators. The model considered in the simulation consists of one explanatory variable ($K = 1$) and two factors ($r = 2$), which can be presented as

$$y_{it} = \alpha_i + x_{it}\beta_i + \psi_i g_t + \phi_i h_t + \epsilon_{it},$$
$$x_{it} = \mu_i + \gamma_i^g g_t + \gamma_i^h h_t + v_{it}, \tag{2.7.1}$$

where g_t and h_t are both scalars. We consider the following different specifications on the models (M), loadings (L), errors (E) and coefficients(C):

M1: ψ_i and ϕ_i are random variables for all i;

M2: ϕ_i is zero for all i and ψ_i is random variable.

L1: ψ_i and ϕ_i (if not zero) are generated according to $\psi_i = 2+N(0,1)$ and $\phi_i = 1+N(0,1)$; similarly γ_i^g and γ_i^h are generated by $\gamma_i^g = 1 + N(0,1)$ and $\gamma_i^h = 2 + N(0,1)$.

L2: ψ_i and ϕ_i (if not zero) are generated from $N(0,1)$; γ_i^g and γ_i^h are generated according to $\gamma_i^g = \psi_i + N(0,1)$ and $\gamma_i^h = \phi_i + N(0,1)$.

E1: Let Ξ be a $N(K+1)$ dimensional vector with all its elements being 1. Let $\Upsilon = \text{diag}(\Upsilon_1, \Upsilon_2, \ldots, \Upsilon_N)$ be an $N(K+1) \times N(K+1)$ block diagonal matrix, where $\Upsilon_i = \text{diag}(1, (M_i'M_i)^{-1/2}M_i)$ with M_i being a $K \times K$ standard normal random matrix. Then u_t is generated according to $u_t = \sqrt{\text{diag}(\Xi)}\Upsilon\varepsilon_t$, where ε_t is an $N(K+1) \times 1$ vector with all its elements being i.i.d from $N(0,1)$.

E2: u_t is generated as in **E1** except that

$$\Xi_i = 0.1 + \frac{\eta_i}{1-\eta_i}\iota_i'\iota_i, \qquad i = 1, 2, \cdots, N(K+1)$$

where ι_i' is the ith row of Λ, and η_i is drawn independently from $U[u, 1-u]$ with $u = 0.1$.

C1: $\beta_i = 1 + N(0, 0.04)$ for all i.

C2: $\beta_i = 1$ for all i.

Remark 2.7.1. Two specifications in M denotes the two models considered in the paper. M1 corresponds to the basic model, and M2 corresponds to the model with zero restrictions. We consider two different sets of loadings. In L1 all the loadings have the same mean, but in L2 only the loadings corresponding to the same individual share the same component. Both specifications lead to the correlated loadings, but as will be seen below, the CCE estimator performs quite differently in the two setups. We also consider the cross-sectional heteroscedasticity and homoscedasticity in the simulation, which correspond to E1 and E2, respectively. When generating heteroscedasticity, we add 0.1 to the expression, avoiding the variance being too close to zero. Our approach to generating the idiosyncratic errors is similar to Doz et al. (2012) and Bai and Li (2014b). We also consider two specifications for the coefficients. While we mainly focus on the performance of the estimation of heterogeneous coefficients, we also use simulations to examine the finite sample properties of the two estimators proposed in Remarks 2.4.4 and 2.5.4.

The other parameters including $g_t, h_t, \alpha_i, \nu_i$ are all generated independently from $N(0,1)$. To evaluate the performance of estimators, we use the average of the root mean square error (RMSE) to measure the goodness-of-fit, which is calculated by

$$\sqrt{\frac{1}{N\mathcal{S}}\sum_{s=1}^{\mathcal{S}}\sum_{i=1}^{N}(\hat{\beta}_i^{(s)} - \beta_i)^2},$$

where $\hat{\beta}_i^{(s)}$ is the estimator of the ith unit in the sth experiment, and β_i is the underlying true value. $\mathcal{S}$ is the number of repetitions, which is set to 1000 in the simulation.

2.7.1 Determining the number of factors

We now discuss the determination of the number of factors, which is a relevant issue in the factor-analysis-based method. In the basic model, determining the number of factors is relatively easier. In the first step, we estimate a pure factor model. So the existing determination methods, such as Bai and Ng (2002), Onatski (2009) and Ahn and Horenstein (2013), can be used. Although these methods do work well in the present setup, to be consistent with the theory established in Section 2.3, we instead consider the following MLE-based information criterion in the simulation

$$\hat{r} = \underset{0 \le m \le r_{\max}}{\operatorname{argmin}} \, IC(m) \tag{2.7.2}$$

where

$$IC(m) = \frac{1}{N\bar{K}} \ln |\hat{\Lambda}^m \hat{\Lambda}^{m\prime} + \hat{\Psi}^m| + m \frac{N\bar{K} + T}{NT\bar{K}} \ln \min(N\bar{K}, T).$$

where $\hat{\Lambda}^m$ and $\hat{\Psi}^m$ are the respective estimator of Λ and Ψ when the number of factors is set to m and $\bar{K} = K + 1$. For the model with zero restrictions, we need to determine the factor numbers in the y equation and the x equation, respectively. Following Bai and Li (2014b), we consider a two-step method to determine them. First, we use (2.7.2) to obtain the total number $r = r_1 + r_2$, denoted by $\hat{r}$, and the associated CV estimator $\hat{\beta}_i^{\hat{r}}$; we then use (2.7.2) again to determine the factor number of the residual matrix $\mathscr{R} = (\mathscr{R}_{it})$ with $\mathscr{R}_{it} = \dot{y}_{it} - \dot{x}_{it}' \hat{\beta}_i^{\hat{r}}$, which we use $\hat{r}_1$ to denote. Then $\hat{r}_2 = \hat{r} - \hat{r}_1$. In the simulation, we set $r_{\max} = 3$.

In practice, the basic model and the model with zero restrictions cannot be differentiated. We therefore suggest estimating the two models in a unified way. More specifically, for a given data set, we calculate r and r_1. If $\hat{r} = \hat{r}_1$, we turn to the basic model; if $\hat{r} > \hat{r}_1$, we turn to the model with zero restrictions.

Table 2.1 reports the percentages that the number of factors is correctly estimated by (2.7.2) based on 1000 repetitions. From the table, we see that the number of factors can be correctly estimated with very high probability. This result is robust to all combinations of listed specifications on loadings, errors and models.

Table 2.1: The percentage of correctly estimating the number of factors

		M1				M2			
	T	50	100	150	200	50	100	150	200
		L1+E1				L1+E1			
	50	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		L1+E2				L1+E2			
	50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		L2+E1				L2+E1			
	50	99.8	100.0	100.0	100.0	99.9	100.0	100.0	100.0
N	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		L2+E2				L2+E2			
	50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	150	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

2.7.2 Finite sample properties of several estimators

In this section, we examine the performance of the CV and LV estimators. For the purpose of comparison, we also calculate Pesaran's CCE estimator, Song's PC estimator, and the infeasible GLS estimator. The infeasible GLS estimator, which is calculated by assuming that the factors are observed, serves as the benchmark for comparison. Since the previous subsection has confirmed that the number of factors can be correctly estimated with high probability, we assume that the number of factors is known in this subsection.

Tables 2.2-2.3 report the performance of the CCE, PC, CV and infeasible GLS (denoted by INF) estimators under different loading and error choices in the basic model. In summary, we see that the CCE estimator performs well under L1, but poorly under L2; the PC estimator performs well under E1, but poorly under E2; the CV estimator performs well under all setups.

Table 2.2: The performance of the four estimators in the basic model

		L1+E1				L2+E1			
N	T	CCE	PC	CV	INF	CCE	PC	CV	INF
50	50	0.1517	0.1596	0.1537	0.1501	0.3980	0.1603	0.1533	0.1492
100	50	0.1499	0.1538	0.1512	0.1494	0.3985	0.1543	0.1508	0.1489
150	50	0.1491	0.1519	0.1500	0.1489	0.3961	0.1526	0.1503	0.1492
50	100	0.1052	0.1087	0.1049	0.1024	0.3868	0.1095	0.1051	0.1026
100	100	0.1034	0.1058	0.1040	0.1029	0.3855	0.1060	0.1037	0.1025
150	100	0.1029	0.1046	0.1033	0.1025	0.3863	0.1049	0.1034	0.1026
50	150	0.0857	0.0878	0.0848	0.0830	0.3819	0.0883	0.0847	0.0828
100	150	0.0839	0.0855	0.0841	0.0832	0.3826	0.0858	0.0841	0.0832
150	150	0.0834	0.0846	0.0836	0.0831	0.3819	0.0848	0.0836	0.0830
50	200	0.0749	0.0760	0.0733	0.0717	0.3832	0.0763	0.0732	0.0716
100	200	0.0723	0.0737	0.0723	0.0715	0.3815	0.0741	0.0726	0.0718
150	200	0.0719	0.0729	0.0720	0.0716	0.3813	0.0731	0.0722	0.0717

First consider the different loading choices. Under L1, the performance of the CCE estimator is considerably good and very close to that of the CV estimator. The performance of these two estimators is only slightly inferior to the infeasible GLS estimator regardless of homoscedasticity or heteroscedasticity. However, under L2 the performance of the CCE estimator is poor. Not only does it have a large average RMSE, but it also exhibits a slowly

Table 2.3: The performance of the four estimators in the basic model

		L1+E2				L2+E2			
N	T	CCE	PC	CV	INF	CCE	PC	CV	INF
50	50	0.3505	3.4677	0.3667	0.3581	0.4079	2.2194	0.2456	0.2377
100	50	0.3426	2.7550	0.3592	0.3545	0.4084	1.6894	0.2390	0.2362
150	50	0.3470	2.6504	0.3569	0.3543	0.4128	1.2141	0.2382	0.2363
50	100	0.2515	2.8863	0.2494	0.2427	0.3870	2.0866	0.1672	0.1630
100	100	0.2380	2.5816	0.2430	0.2399	0.3856	1.5579	0.1630	0.1616
150	100	0.2417	2.6489	0.2447	0.2430	0.3864	0.9734	0.1644	0.1630
50	150	0.2141	2.9851	0.2008	0.1956	0.3773	1.9264	0.1333	0.1302
100	150	0.2029	2.7919	0.1996	0.1977	0.3804	1.4195	0.1340	0.1326
150	150	0.1973	2.4904	0.1988	0.1973	0.3791	1.0475	0.1319	0.1310
50	200	0.1944	3.5289	0.1763	0.1718	0.3769	1.8067	0.1168	0.1141
100	200	0.1781	3.0194	0.1715	0.1694	0.3787	1.1939	0.1142	0.1131
150	200	0.1726	2.4151	0.1717	0.1705	0.3771	0.8777	0.1128	0.1122

decreasing rate for the average RMSE. In contrast, the CV estimator performs closely with the infeasible GLS estimator. The average RMSE of the CV estimator decreases almost at the same speed with that of the infeasible estimator.

The reason for the different performance of the CCE estimator under different loading sets is that the space spanned by $\tilde{z}_t = \frac{1}{N}\sum_{i=1}^{N} \dot{z}_{it}$ with $\dot{z}_{it} = (\dot{y}_{it}, \dot{x}_{it}')'$ provides a good approximation to the space spanned by f_t under L1, but a poor approximation under L2. To see this point more clearly, consider (2.2.1), which can be written as $\dot{z}_{it} = \Lambda_i' f_t + \dot{u}_{it}$. Taking the average over i, we have $\tilde{z}_t = \tilde{\Lambda}' f_t + \tilde{u}_t$, where $\tilde{\Lambda}$ and $\tilde{u}_t$ are defined similarly to $\tilde{z}_t$. With some transformation, we have $f_t = (\tilde{\Lambda}\tilde{\Lambda}')^{-1}\tilde{\Lambda}(\tilde{z}_t - \tilde{u}_t)$. So a good approximation requires two conditions. First, $\tilde{z}_t$ dominates $\tilde{u}_t$ so that $\tilde{u}_t$ is negligible. Second, $\tilde{\Lambda}\tilde{\Lambda}'$ is invertible when N goes to infinity. The loadings in L1 satisfy these two conditions, but the loadings in L2 violate the first one. In fact, the terms $\tilde{\Lambda}' f_t$ and $\tilde{u}_t$ are of the same magnitude under L2. So a good approximation fails. There are cases in which the second condition breaks down. For example, if all rows of Λ share the same mean, then $\tilde{\Lambda}$ is of rank one asymptotically, which in turn leads to $\tilde{\Lambda}'\tilde{\Lambda}$ being singular asymptotically. The simulation results confirm that the CCE estimator performs poorly in this case.

Consider then the different choices of the errors. Table 2.3 shows that the PC estimator performs poorly in the presence of cross-sectional heteroscedasticity (E2). In addition, we find that the performance of the PC estimator is improved marginally under E1, but significantly under E2, when N becomes larger. According to the theory of Song (2013), the PC estimate is $\sqrt{T}$-consistent, implying that the performance of the PC estimator should be closely related to T and loosely related to N. This theoretical result is supported by Table 2.2 but contradicted in Table 2.3. We think that the underlying reason is due to the computation problem of the minimizer of the objective function in the iterated PC method, as mentioned in Section 2.1. The extent of this problem depends on the strength of heteroscedasticity. In our simulation, we generate heavy heteroscedasticity, which magnifies the computational problem of the iterated PC method. [3]

Tables 2.4-2.7 report the simulation results for the models with zero restrictions and heterogeneous coefficients. Overall, these tables reaffirm the result that the CCE estimator performs poorly under L2, and the PC estimator performs poorly under E2. Besides this result, there are several additional points worth noting. First, the CCE and CV estimators are inefficient. Under the L1+E1 setup, even when N and T are large, say $N = 150, T = 200$,

[3] In the case of a homogeneous coefficient, this computational problem does not exist. First, as shown in the next subsection, the PC estimator generally has a better convergence under a homogeneous coefficient. Second, as pointed out in Moon and Weidner (2015), the objective function of the PC method can be written into a trace form, which only depends on β. So we can first use the method suggested by Bai (2009) to obtain a preliminary estimator, and then turn to the Newton-Raphson algorithm to get a better estimator.

Table 2.4: The performance of the six estimators under M2+L1+E1

N	T	CCE	PC	CV	LV	ILV	INF
50	50	0.1486	0.0811	0.1527	0.0891	0.0822	0.0790
100	50	0.1483	0.0797	0.1503	0.0868	0.0808	0.0787
150	50	0.1488	0.0792	0.1501	0.0862	0.0803	0.0785
50	100	0.1023	0.0560	0.1046	0.0588	0.0564	0.0546
100	100	0.1026	0.0552	0.1039	0.0575	0.0555	0.0545
150	100	0.1024	0.0549	0.1032	0.0571	0.0552	0.0545
50	150	0.0831	0.0454	0.0849	0.0470	0.0456	0.0443
100	150	0.0831	0.0449	0.0840	0.0463	0.0450	0.0443
150	150	0.0828	0.0445	0.0834	0.0457	0.0447	0.0442
50	200	0.0718	0.0391	0.0732	0.0404	0.0392	0.0382
100	200	0.0717	0.0387	0.0725	0.0396	0.0388	0.0382
150	200	0.0715	0.0384	0.0720	0.0392	0.0385	0.0381

Table 2.5: The performance of the six estimators under M2+L2+E1

N	T	CCE	PC	CV	LV	ILV	INF
50	50	0.2716	0.1231	0.1533	0.1215	0.1210	0.1193
100	50	0.2673	0.1218	0.1512	0.1210	0.1209	0.1200
150	50	0.2674	0.1205	0.1504	0.1201	0.1200	0.1194
50	100	0.2532	0.0849	0.1047	0.0838	0.0836	0.0825
100	100	0.2563	0.0835	0.1034	0.0830	0.0829	0.0823
150	100	0.2562	0.0833	0.1033	0.0829	0.0829	0.0825
50	150	0.2469	0.0691	0.0849	0.0681	0.0680	0.0672
100	150	0.2500	0.0683	0.0845	0.0679	0.0678	0.0674
150	150	0.2476	0.0676	0.0836	0.0673	0.0673	0.0670
50	200	0.2475	0.0595	0.0732	0.0588	0.0587	0.0580
100	200	0.2474	0.0586	0.0725	0.0582	0.0582	0.0578
150	200	0.2476	0.0584	0.0720	0.0581	0.0581	0.0579

Table 2.6: The performance of the six estimators under under M2+L1+E2

N	T	CCE	PC	CV	LV	ILV	INF
50	50	0.2794	0.7402	0.3002	0.2293	0.2172	0.2103
100	50	0.2905	0.2507	0.3020	0.2223	0.2130	0.2081
150	50	0.2980	0.3511	0.3053	0.2282	0.2201	0.2159
50	100	0.2017	0.5204	0.2100	0.1531	0.1495	0.1462
100	100	0.1993	0.1610	0.2081	0.1517	0.1487	0.1468
150	100	0.2057	0.1871	0.2112	0.1524	0.1496	0.1481
50	150	0.1665	0.4558	0.1727	0.1220	0.1198	0.1170
100	150	0.1645	0.3249	0.1675	0.1196	0.1180	0.1166
150	150	0.1641	0.1282	0.1669	0.1202	0.1184	0.1174
50	200	0.1463	0.3222	0.1461	0.1064	0.1048	0.1027
100	200	0.1462	0.1510	0.1484	0.1050	0.1039	0.1027
150	200	0.1447	0.1128	0.1472	0.1043	0.1032	0.1023

the average RMSEs of these two estimators are considerably larger than the remaining four estimators. This is not surprising since the two estimation methods do not use the information contained in the zero restrictions; see the discussion in Section 2.5. Second, several iterations over the LV estimator indeed improve the finite sample performance, especially when N and T are small or moderate. In all combinations of N and T, the ILV estimator outperforms the LV one. Third, under homoscedasticity, the PC, LV and ILV estimators are seen to be efficient since their performance is very close to that of the infeasible GLS estimator, especially when N and T are large.

Table 2.7: The performance of the six estimators under under M2+L2+E2

N	T	CCE	PC	CV	LV	ILV	INF
50	50	0.2891	1.2307	0.1940	0.1606	0.1600	0.1554
100	50	0.2913	0.7183	0.1910	0.1570	0.1567	0.1545
150	50	0.2894	0.4762	0.1879	0.1567	0.1567	0.1557
50	100	0.2710	0.9264	0.1310	0.1091	0.1080	0.1062
100	100	0.2748	0.6029	0.1306	0.1097	0.1097	0.1086
150	100	0.2720	0.4254	0.1297	0.1079	0.1078	0.1070
50	150	0.2567	0.7998	0.1057	0.0895	0.0882	0.0865
100	150	0.2615	0.5410	0.1061	0.0894	0.0890	0.0880
150	150	0.2654	0.3370	0.1057	0.0887	0.0887	0.0881
50	200	0.2593	0.7218	0.0900	0.0754	0.0748	0.0734
100	200	0.2603	0.5082	0.0901	0.0766	0.0766	0.0759
150	200	0.2566	0.3009	0.0898	0.0749	0.0749	0.0742

2.7.3 Homogeneous coefficient

In this subsection, we investigate the finite sample properties of the CV and LV estimators suggested in (2.4.6) and (2.5.16). We also compute the iterated PC estimator of Bai (2009) and the QML estimator of Bai and Li (2014b) for comparison. For simplicity, only the setup "L2+E2" is considered. Table 2.8 presents the simulation results. Overall, we see that the CV (LV) estimation method gives a consistent estimation for the homogeneous coefficient. Additionally, we see that the performance of the CV(LV) estimator is superior to that of the iterated PC estimator, but inferior to that of the QML estimator. This result is consistent with the two-step method partially taking the cross-sectional heteroscedasticity into account, the iterated PC method not accounting for the cross-sectional heteroscedasticity, and the QML method fully taking the cross-sectional heteroscedasticity into account.

Table 2.8: The performance of the CV(LV), PC and QML estimators under L2+E2+C2

		CV(LV)		PC		QML	
N	T	Bias	RMSE	Bias	RMSE	Bias	RMSE
		M1					
50	50	0.0004	0.0216	-0.0002	0.0259	-0.0004	0.0102
100	50	0.0004	0.0151	0.0000	0.0159	0.0002	0.0066
150	50	0.0007	0.0118	0.0008	0.0121	0.0004	0.0052
50	100	0.0006	0.0146	0.0005	0.0194	-0.0000	0.0071
100	100	0.0000	0.0108	-0.0000	0.0117	0.0002	0.0047
150	100	-0.0003	0.0081	-0.0003	0.0086	-0.0002	0.0036
50	150	-0.0000	0.0122	0.0005	0.0181	-0.0000	0.0052
100	150	0.0004	0.0084	0.0002	0.0101	0.0001	0.0037
150	150	-0.0000	0.0067	-0.0002	0.0072	0.0000	0.0031
50	200	0.0008	0.0105	0.0006	0.0173	-0.0001	0.0047
100	200	0.0001	0.0073	0.0002	0.0089	0.0000	0.0033
150	200	0.0000	0.0060	0.0001	0.0065	0.0000	0.0025
		M2					
50	50	0.0003	0.0140	0.0088	0.0224	-0.0001	0.0053
100	50	-0.0002	0.0097	0.0023	0.0111	0.0000	0.0037
150	50	-0.0000	0.0080	0.0012	0.0085	0.0000	0.0030
50	100	0.0003	0.0098	0.0077	0.0185	-0.0001	0.0043
100	100	-0.0001	0.0068	0.0022	0.0086	-0.0001	0.0026
150	100	-0.0001	0.0057	0.0008	0.0063	0.0000	0.0022
50	150	0.0001	0.0075	0.0071	0.0172	0.0002	0.0029
100	150	0.0002	0.0053	0.0025	0.0079	0.0000	0.0021
150	150	0.0000	0.0044	0.0010	0.0052	-0.0001	0.0017
50	200	0.0001	0.0066	0.0075	0.0166	0.0000	0.0026
100	200	0.0001	0.0047	0.0023	0.0071	-0.0000	0.0018
150	200	0.0001	0.0039	0.0010	0.0046	0.0000	0.0015

2.8 Conclusion

This chapter considers the estimation of heterogeneous coefficients in panel data models with common shocks. We propose a two-step method to estimate heterogeneous coefficients, in which the QML method is first used to estimate the loadings and variances of the idiosyncratic errors in a pure factor model, and heterogeneous coefficients are then estimated based on the estimates and structural relations implied by the model. Asymptotic properties of the proposed estimators including the asymptotic representations and limiting distributions are investigated and provided.

In addition, we extend our method to the models with zero restrictions on the partial loadings in the y equation. We point out that efficiency can be gained by using the information contained in the loadings. The asymptotic representation and limiting distribution of the new two-step estimator are studied. We also consider the model with time-invariant regressors.

The proposed estimators are asymptotically efficient in the sense that they have the same limiting distributions as the infeasible GLS estimators. Monte Carlo simulations confirm our theoretical results and show encouraging evidence that the two-step estimators perform robustly in all data setups.

Chapter 3

Quasi Maximum Likelihood Analysis of High Dimensional Constrained Factor Models

Qi Li, Kunpeng Li and Lina Lu

3.1 Introduction

With the rapid development of data collection, storing and processing techniques in computer science, econometricians and statisticians now face large dimensional data setups more often than ever before. A challenge along with the appearances of large data is how to extract useful information from data, or put differently, how to effectively conduct dimension reduction on data. Factor models are proved to be an effective way to perform this task. Over the last three decades, the literature has witnessed wide applications of factor models in many economics disciplines. In finance, Connor and Korajczyk (1986, 1988) and Fan et al. (2015) use factor models to measure the risk and performance of large portfolios. In macroeconomics, Geweke (1977) and Sargent and Sims (1977) use dynamic factor models to identify the source of primitive shocks. In labor economics, Heckman et al. (2006) use factor models to capture unobservable personal abilities. In international economics, Kose et al. (2003) use multilevel factor models to separate global business circles, regional business circles and country-specific business circles. Large dimensional factor models are also used in a variety of ways to deal with strong correlations, see e.g., Fan et al. (2011) and Fan et al. (2013), among others.

A standard factor model can be written as

$$z_t = Lf_t + e_t, \quad t = 1, 2, \ldots, T,$$

where $z_t = (z_{1t}, \ldots, z_{Nt})'$ is a vector of N variables at time t, L is an $N \times r$ loadings matrix, f_t is an r-dimensional vector of factors and e_t is an N-dimensional vector of idiosyncratic errors. The traditional (classical) factor analysis assumes that N is fixed and T is large.

This assumption runs counter to usual shape of large dimensional data sets, in which N is usually comparable to or even greater than T (Stock and Watson (2002a)). Recent literature contributes a lot to the asymptotic theory with N comparable to or even greater than T. Bai and Ng (2002) propose several information criterions to determine the number of factors in a large-N and large-T environment. Under a similar setup to Bai and Ng (2002), Stock and Watson (2002a) prove that the principal components (PC) estimates are consistent in approximate factor models of Chamberlain and Rothschild (1983). Bai (2003) moves forwards along the work of Stock and Watson (2002a) and gives the asymptotic representations of the PC estimates of loadings, factors and common components. Doz et al. (2012) consider the maximum likelihood (ML) method and prove the average consistency of the maximum likelihood estimates (MLE). Bai and Li (2012, 2016) use five different identification strategies to eliminate the rotational indeterminacy from asymptotics and give limiting distributions of the MLE. Fan et al. (2016) propose a new projected principal component method to more accurately estimate the unobserved latent factors.

A potential problem in high dimensional factor models is that too many parameters are estimated within the model, which makes it difficult to analyze and interpret the economic implications of the estimates. However, if the space of the loading matrix is spanned by a low dimension matrix, this problem can be much ameliorated. In this paper, following Tsai and Tsay (2010), we address this problem by considering the following constrained factor model

$$z_t = M\Lambda f_t + e_t,$$

where M is a *known* $N \times k$ matrix with rank k and Λ is a $k \times r$ *unknown* loadings matrix with rank r. We assume $r < k \leq C$ for some generic constant C. In the above specification, M consists of the bases of the loading matrix. The underlying true loadings are a weighted average of these bases associated with the weights matrix Λ, which are the parameters of interests. The number of loading parameters now is kr instead of Nr. So the number of parameters is greatly reduced.

Our work is closely related to Tsai and Tsay (2010) who were the first to consider constrained factor models. This chapter differs from Tsai and Tsay (2010) in several dimensions. First, although Tsai and Tsay (2010) propose using PC and ML methods to estimate constrained factor models, their asymptotic analysis focuses only on the PC method. They obtain convergence rates of the PC estimates. As a comparison, we investigate asymptotics of the ML method and derive the convergence rates and limiting distributions of the MLE. Given the limiting distributions, one can easily construct $(1-\alpha)$-confidence intervals if prediction is the target of interest, or use t-test or F-test to conduct statistical inferences on the underlying parameter values if hypothesis testing is the purpose. Second, Tsai and Tsay (2010) consider the setup that k is large (but still smaller than N). In this paper, we instead assume that k is fixed[1]. In our viewpoints, assuming a fixed k is of practical and theoretical interests. In some typical examples, the parameter k is interpreted as the number of groups or categories, according to which the variables are classified (see Tsai and Tsay (2010)). This value is usually not large in real data. Therefore, a fixed-k assumption is adopted in this paper. Furthermore, in constrained factor models, a large k leads to a larger number of parameters being estimated. The estimation accuracy is reversely linked with k for a

[1]Our analysis can be extended to the case of a large k. But for this case, deriving the limiting distribution of the MLE is very challenging since the matrix Λ is high-dimensional.

given sample size. When k is large, the benefit of constrained factor models against standard factor models becomes weak, which makes constrained factor models less attractive in practice. Third, an importantly related issue in constrained factor models is on conducting valid model specification check on the presence of matrix M. Tsai and Tsay (2010) consider the traditional likelihood ratio test to perform this task. But the traditional likelihood ratio test is designed under fixed-N and large-T setup, which conflicts to large-N and large-T scenarios. In this chapter, we propose new statistics for testing model specifications that are applicable to the large-N and large-T setups.

The rest of this chapter is organized as follows. Section 3.2 provides more empirical examples of the constrained factor model. Section 3.3 introduces the model and lists the assumptions needed for the subsequent analysis. Section 3.4 delivers the consistency and limiting distribution results of the MLE. Section 3.5 considers testing issues within constrained factor models. Section 3.6 considers a partially constrained factor model and presents the asymptotic properties of the MLE for this model. Section 3.7 presents the Expectation-Maximization (EM) algorithm to implement the QML estimation. Section 3.8 conducts Monte Carlo simulations to investigate the finite sample performance of the MLE and to study the empirical size and power of the proposed model specification test. In Section 3.9, we relax Assumption B to allow for the idiosyncratic errors to have a more general weakly dependence structure. Section 3.10 concludes the paper. All technical contents are delegated to several appendices.

3.2 Motivating Applications

The well-known equilibrium arbitrary pricing theory (APT) implies that the observed assets returns can be expressed into a linear factor structure, see Ross (1976), Connor and Korajczyk (1988) among others. This motivates the use of the following factor model

$$r_{it} = \sum_{j=1}^{r} l_{ij} f_{jt} + e_{it}$$

to study the performance of portfolios, where r_{it} is the excess return of the ith security at time t, f_{jt} denotes the jth risk premium at time t and l_{ij} the beta coefficient of the jth risk premium for security i. However, as pointed out by Rosenberg (1974), the common movements among the assets returns may be related with the individual characteristics. Such characteristics include capitalization and book-to-price ratios as suggested in Fama and French (1993), momentum as in Carhart (1997), own-volatility as in Goyal and Santa-Clara (2003). Let x_{ip} denote the observed pth characteristic of the ith security. Rosenberg (1974) considers the specification

$$l_{ij} = \sum_{p=1}^{k} x_{ip}\lambda_{pj} + v_{ij}, \qquad \text{or} \qquad L = M\Lambda + V,$$

where $M = (x_{ip})_{N\times k}$ is the observed characteristics matrix. Rosenberg's specification is very close to the one studied in this paper. With a slight modification, the analysis in this paper can easily be extended to cover the Rosenberg's model.

A limitation of Rosenberg's specification is that the factor betas are assumed to be linear functions of the observed characteristics, which is overly restrictive in practice. To accommodate this concern, Connor and Linton (2007) and Connor and Linton (2012) consider the following nonparametric specification

$$l_{ij} = g_j(x_{ij}).$$

where $g_j(\cdot)$ is an unknown smooth function. Connor and Linton (2012) apply their model to a real dataset and indeed find that the factor betas are nonlinear functions of the characteristics. However, an undesirable feature in these two papers is that the estimation of the model involves an iterative procedure between the factors and unknown functions, which is formidable to many applied researches. To address this issue, we instead consider using a series of polynomial functions to approximate the unknown function $g_j(\cdot)$. More specifically, we consider approximating the function $g_j(\cdot)$ by all the polynomial functions with power less than q, i.e.,

$$g_j(x) \approx \lambda_{j0} + \lambda_{j1}x + \cdots + \lambda_{jq}x^q. \tag{3.2.1}$$

Given this, the model now can be written as $L = M\Lambda$ with

$$M = \begin{bmatrix} 1 & x_{11} & x_{11}^2 & \cdots & x_{11}^q & \cdots & \cdots & x_{1r} & x_{1r}^2 & \cdots & x_{1r}^q \\ 1 & x_{21} & x_{21}^2 & \cdots & x_{21}^q & \cdots & \cdots & x_{2r} & x_{2r}^2 & \cdots & x_{2r}^q \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N1}^2 & \cdots & x_{N1}^q & \cdots & \cdots & x_{Nr} & x_{Nr}^2 & \cdots & x_{Nr}^q \end{bmatrix}$$

and

$$\Lambda = \begin{bmatrix} \lambda_{10} & \lambda_{11} & \cdots & \lambda_{1q} & 0 & \cdots & 0 & \cdots & \cdots & 0 & \cdots & 0 \\ \lambda_{20} & 0 & \cdots & 0 & \lambda_{21} & \cdots & \lambda_{2q} & \cdots & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ \lambda_{r0} & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \cdots & \lambda_{r1} & \cdots & \lambda_{rq} \end{bmatrix}'.$$

The above model can be viewed as a special case of the constrained factor model with some zero restrictions imposed on Λ. The model considered here maintains the nonlinear function feature of Connor and Linton (2007) and Connor and Linton (2012) but the computational burden has been much reduced. A primary issue related with our method is whether the approximation (3.2.1) is good enough. This work can be partially addressed by the W statistic proposed in Section 3.5.

Constrained factor models have other applications. Tsai and Tsay (2010) apply constrained factor models to analyze stock returns where the stocks can be classified into different sectors. They specify the constraint matrix M consisting of orthogonal and binary vectors. In another application, they implement constrained factor models to study the interest-rate yield curve, where the columns of the matrix M are specified to denote the level, slope and curvature feature of interest rates. Matteson et al. (2011) use constrained factor models to forecast the hourly emergency medical service call arrival rates by specifying the constraints on the factor loadings based on the prior information of the pattern of the call arrivals. Similar approach is adopted in Zhou and Matteson (2015) to model the ambulance demand by incorporating covariate information as constraints on the factor loadings.

Remark 3.2.1. In practice, the loading matrix L is unknown but its elements might have common values (i.e., zeros; in this case M consists of orthogonal and binary vectors). Once we know the locations of these common values, we can impose restrictions on L, in other words, construct the constraint matrix M and then apply the framework of constrained factor model. Researchers might consider least absolute shrinkage and selection operator (LASSO) or group LASSO to identify these locations of common values in L. For inference, see Tibshirani (1996) and Zou (2006).

3.3 Constrained Factor Models

Let N denote the number of variables and T the sample size in the time dimension. We consider the following constrained factor model

$$z_t = M\Lambda f_t + e_t, \tag{3.3.1}$$

where $z_t = (z_{1t}, z_{2t}, \dots, z_{Nt})'$ is an N-dimensional vector of explanatory variables at time t; M is a specified $N \times k$ (known) matrix with rank k; Λ is the $k \times r$ loading matrix of rank r; $f_t = (f_{1t}, f_{2t}, \dots, f_{rt})'$ is a vector of r latent common factors; e_t is an N-dimensional vector of idiosyncratic disturbances and is independent of f_t. Throughout the paper, we assume $k > r$. If $k \leq r$, we can simply consider the linear regression $z_t = Mf_t^* + e_t$ with $f_t^* = \Lambda f_t$. The model effectively becomes a factor model with k (when $k \leq r$) factors.

Our analysis is based on similar assumptions used in standard factor models, see Bai and Li (2012) for the asymptotic analysis of the MLE for standard high dimensional factor models. The symbol C appearing in the following assumptions denotes a generic positive constant. Our assumptions include:

Assumption A: $\{f_t\}$ is a sequence of fixed constants with $\bar{f} = \sum_{t=1}^{T} f_t = 0$. Let

$M_{ff} = \frac{1}{T}\sum_{t=1}^{T} f_t f_t'$ be the sample variance of f_t. There exists an $\overline{M}_{ff} > 0$ (positive definite) such that $\overline{M}_{ff} = \lim_{T\to\infty} M_{ff}$.

Assumption B: The idiosyncratic error term e_{it} is independent across the i index and the t index with $E(e_t) = 0$, $E(e_t e_t') = \Sigma_{ee} = \text{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2)$ and $E(e_{it}^8) \leq C$ for all i and t, where $e_t = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$ is the N-dimensional vector of idiosyncratic errors at time t.

Assumption C: The underlying values of parameters satisfy that

C.1 $\|\Lambda\| \leq C$ and $\|m_j\| \leq C$ for all j, where m_j is the transpose of the jth row of M.

C.2 $C^{-2} \leq \sigma_j^2 \leq C^2$ for all j, where $\sigma_j^2 = E(e_{jt}^2)$ is defined in Assumption B.

C.3 Let $P = \Lambda' M' \Sigma_{ee}^{-1} M \Lambda / N$, $R = M' \Sigma_{ee}^{-1} M / N$. We assume that $P_\infty = \lim_{N\to\infty} P$ and $R_\infty = \lim_{N\to\infty} R$ exist. In addition, $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \sigma_i^{-4}(m_i \otimes m_i)(m_i' \otimes m_i') = V_\infty$ exists. Here P_∞, R_∞ and V_∞ are some positive definite matrices.

Assumption D: The estimator of σ_j^2 for $j = 1, ..., N$ takes value in a compact set: $[C^{-2}, C^2]$. Furthermore, M_{ff} is restricted to be in a set consisting of all semi-positive definite matrices with all elements bounded in the interval $[-C, C]$, where C is a large positive constant.

Assumption A requires that factors are sequences of fixed constants. The random factors can be dealt with in a similar way under some suitable moment conditions. Assumption B is commonly imposed in classical factor models. It can be relaxed to allow for cross-sectional and temporal heteroskedasticities and correlations, see extension in Section 3.9. Assumption C requires that underlying values of parameters are in a compact set, which is standard in econometric literature. Assumption D requires that some parameter estimates take values in a compact set. This assumption is often made when dealing with highly nonlinear objective function, see Jennrich (1969). Our objective function is highly nonlinear.

Similar to the case of a standard factor model, a constrained factor model has an identification problem. To see this, for any invertible $r \times r$ matrix B, we have

$$\Lambda f_t = \Lambda B \cdot B^{-1} f_t = \Lambda^* f_t^*.$$

with $\Lambda^* = \Lambda B$ and $f_t^* = B^{-1} f_t$. To sperate (Λ, f_t) from (Λ^*, f_t^*), we impose the following identification condition.

Identification condition (abbreviated by IC hereafter):

IC1 $\Lambda' \big(\frac{1}{N} M' \Sigma_{ee}^{-1} M \big) \Lambda = P$, where P is a diagonal matrix whose diagonal elements are distinct and arranged in a descending order.

IC2 $M_{ff} = \frac{1}{T} \sum_{t=1}^{T} f_t f_t' = I_r$.

Our identification strategy is similar to IC3 in Bai and Li (2012). It is known that this identification strategy identifies the loadings and factors up to a column sign, see Bai and Li (2012) for a detailed discussion on this issue. To eliminate such a problem in our theoretical analysis, we follow Bai and Li (2012) to treat as part of the identification condition that the estimators and the underlying values of loadings matrix have the same column signs. In practice, the sign problem causes no troubles in empirical analysis.

We use the following discrepancy function between M_{zz} and Σ_{zz} as our objective function

$$\mathcal{L}(\theta) = -\frac{1}{2N} \ln |\Sigma_{zz}| - \frac{1}{2N} \mathrm{tr}[M_{zz} \Sigma_{zz}^{-1}], \tag{3.3.2}$$

where $\theta = (\Lambda, \Sigma_{ee})$, $M_{zz} = T^{-1}\sum_{t=1}^{T} z_t z_t'$ and $\Sigma_{zz} = M\Lambda\Lambda' M' + \Sigma_{ee}$. This discrepancy function has the same form as a likelihood function when f_t are independently and normally distributed with mean zero and variance I_r, see Bai and Li (2012) for details. In the current paper, the factors are assumed to be fixed constants in Assumption A, the above discrepancy function is therefore not a likelihood function. Nevertheless, we still call the maximizer of the above function as a quasi MLE or MLE for simplicity. Specifically, the MLE $\hat{\theta} = (\hat{\Lambda}, \hat{\Sigma}_{ee})$ is defined as

$$\hat{\theta} = \underset{\theta \in \Theta}{\text{argmax}}\ \mathcal{L}(\theta),$$

where Θ is the parameters space such that any interior point of it satisfies Assumption D and the identification condition IC. The input parameters include Λ and Σ_{ee}. In a constrained factor model, we only need to estimate kr loadings instead of Nr loadings (the number of parameters in a standard factor model). Therefore, the number of parameters is greatly reduced. Taking derivatives with respect to Λ and Σ_{ee}, we obtain the following first order conditions:

$$\hat{\Lambda}' M' \hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1} M = 0; \tag{3.3.3}$$

$$\text{diag}(\hat{\Sigma}_{zz}^{-1}) = \text{diag}(\hat{\Sigma}_{zz}^{-1} M_{zz} \hat{\Sigma}_{zz}^{-1}), \tag{3.3.4}$$

where $\hat{\Lambda}$ and $\hat{\Sigma}_{ee}$ denote MLE of Λ and Σ_{ee}, respectively, and $\hat{\Sigma}_{zz} = M\hat{\Lambda}\hat{\Lambda}' M' + \hat{\Sigma}_{ee}$. We note that the above two first order conditions are only used in deriving the asymptotic properties of the MLE. One does not need to solve the above nonlinear equations to obtain the MLE. Instead, we can implement the EM algorithm to compute the MLE. Details are given in

Section 3.7.

3.4 Asymptotic properties of the MLE

In this section, we investigate the asymptotic properties of the MLE. The following proposition shows that the MLE is consistent.

Proposition 3.4.1 (Consistency)**.** *Let $\hat{\theta} = (\hat{\Lambda}, \hat{\Sigma}_{ee})$ be the MLE that maximizes (3.3.2). Then under Assumptions A-D, together with IC, when $N, T \to \infty$, we have*

$$\hat{\Lambda} - \Lambda \xrightarrow{p} 0; \qquad \frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2 \xrightarrow{p} 0.$$

In high dimensional factor analysis, the loadings and variances of idiosyncratic errors are high-dimensional. The consistencies have to be defined under some chosen norms, see Stock and Watson (2002a), Bai (2003), Doz et al. (2012) and Bai and Li (2012, 2016). In constrained factor models, due to the presence of matrix M, the loading matrix Λ is low-dimensional. So its consistency is defined in the elementwise sense. But for the variances of idiosyncratic errors, they are still high-dimensional. Their consistency is therefore defined by $\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2$, which can be written as $\frac{1}{N}\|\hat{\Sigma}_{ee} - \Sigma_{ee}\|^2$. So the chosen norm is the Frobenius norm adjusted with the matrix dimension.

Given the consistency results, we have the following theorem on convergence rates of the MLE.

Theorem 3.4.1 (Convergence rates)**.** *Under the assumptions of Proposition 3.4.1, we have*

$$\hat{\Lambda} - \Lambda = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right), \qquad \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\sigma}_i^2 - \sigma_i^2\right)^2 = O_p\left(\frac{1}{T}\right).$$

According to Theorem 3.4.1, the convergence rate of $\hat{\Lambda}$ is $\min(\sqrt{NT}, T)$, which is faster than the $\sqrt{T}$-convergence rate of estimated loadings in standard factor models. This result is plausible since in a constrained factor model, we use NT observations to estimate kr loadings. This is in contrast with a standard factor model, where we use NT observations to estimate Nr loadings.

Remark 3.4.1. As we can see, the convergence rate of $\hat{\Lambda}$ involves $O_p(\frac{1}{T})$, which reflects the fact that the factors are unknown and need to be estimated. This is similar to the case of standard factor model.

To present the asymptotic representation of the MLE, we introduce some notation. Let

$$\mathbb{D}_1 = \begin{bmatrix} 2D_r^+ \\ \mathcal{D}[(P \otimes I_r) + (I_r \otimes P)K_r] \end{bmatrix}, \quad \mathbb{D}_2 = \begin{bmatrix} 2D_r^+ \\ 0_{\frac{1}{2}r(r-1)\times r^2} \end{bmatrix}, \quad \mathbb{D}_3 = \begin{bmatrix} 0_{\frac{1}{2}r(r+1)\times r^2} \\ \mathcal{D} \end{bmatrix},$$

and

$$\mathbb{B}_1 = K_{kr}[(P^{-1}\Lambda') \otimes \Lambda] + R^{-1} \otimes I_r - K_{kr}(I_r \otimes \Lambda)\mathbb{D}_1^{-1}\mathbb{D}_2[(P^{-1}\Lambda') \otimes I_r],$$

$$\mathbb{B}_2 = K_{kr}(I_r \otimes \Lambda)\mathbb{D}^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)', \quad \Delta = \mathbb{B}_2 \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^6}(m_i \otimes m_i)(\kappa_{i,4} - \sigma_i^4),$$

where $P = \frac{1}{N}\Lambda' M' \Sigma_{ee}^{-1} M \Lambda$, $R = \frac{1}{N} M' \Sigma_{ee}^{-1} M$, $\kappa_{i,4} = E(e_{it}^4)$, m_i is the transpose of the ith row of matrix M, K_{uv} is the commutation matrix such that for any $u \times v$ matrix B,

$K_{uv}\text{vec}(B) = \text{vec}(B')$; and K_r is defined to be K_{rr}. $D_r^+ = (D_r' D_r)^{-1} D_r'$ is the Moore-Penrose inverse matrix of the r-dimensional duplication matrix D_r, $\mathcal{D}$ is the matrix such that $\text{veck}(B) = \mathcal{D}\text{vec}(B)$ for any $r \times r$ matrix B, where $\text{veck}(B)$ is the operation which stacks the elements below the diagonal of the matrix B into a vector. Given matrix P, we can easily calculate the matrix $\mathbb{D}_1$ and its inverse. For example, let $P = \text{diag}(1, 2, 3)$ ($r = 3$ in this case), then

$$\mathbb{D}_1 = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 3 & 0 \end{bmatrix}, \quad \mathbb{D}_1^{-1} = \begin{bmatrix} 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1.5 & 0 & 0 & 0 & 0 & -0.5 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & -1 \\ 0 & 0 & -0.5 & 0 & 0 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 \end{bmatrix}.$$

Now we state the asymptotic result of $\hat{\Lambda}$.

Theorem 3.4.2 (Asymptotic representation). *Under assumptions of Theorem 3.4.1, we have*

$$\text{vec}(\hat{\Lambda}' - \Lambda') = \mathbb{B}_1 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (m_i \otimes f_t) e_{it} - \mathbb{B}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^4} (m_i \otimes m_i)(e_{it}^2 - \sigma_i^2)$$
$$+ \frac{1}{T}\Delta + O_p\Big(\frac{1}{N\sqrt{T}}\Big) + O_p\Big(\frac{1}{\sqrt{NT}}\Big) + O_p\Big(\frac{1}{T^{3/2}}\Big), \tag{3.4.1}$$

where the symbols $\mathbb{B}_1$*,* $\mathbb{B}_2$ *and* Δ *are defined above Theorem 3.4.2.*

The first two terms on the right hand side of (3.4.1) are $O_p(\frac{1}{\sqrt{NT}})$ since their variances are $O(\frac{1}{NT})$ and the third term is $O(\frac{1}{T})$. The first three terms dominates the remaining terms. Theorem 3.4.2 reaffirms the convergence rates asserted in Theorem 3.4.1 and sharpens the results by explicitly giving the concrete expressions of the $O_p(\frac{1}{\sqrt{NT}})$ and $O_p(\frac{1}{T})$ terms. Given Theorem 3.4.2, invoking a Central Limit Theorem, we have the following theorem.

Theorem 3.4.3 (Limiting distribution)**.** *Under assumptions of Theorem 3.4.1, as* $N, T \to \infty, N/T^2 \to 0$*, we have*

$$\sqrt{NT}\Big[\text{vec}(\hat{\Lambda}' - \Lambda') - \frac{1}{T}\Delta\Big] \xrightarrow{d} N(0, \Omega),$$

where $\Omega = \lim_{N\to\infty} \Omega_N$ *with*

$$\Omega_N = \mathbb{B}_1(R \otimes I_r)\mathbb{B}_1' + \mathbb{B}_2\Big[\frac{1}{N}\sum_{i=1}^{N} \frac{\kappa_{i,4} - \sigma_i^4}{\sigma_i^8}(m_i m_i') \otimes (m_i m_i')\Big]\mathbb{B}_2'.$$

Theorem 3.4.3 shows that the MLE $\hat{\Lambda}$ has a non-negligible bias. This is in contrast to a result of Bai and Li (2012) who show that, in a high-dimensional standard factor model, the MLE is asymptotically centered around zero. Another interesting result is that the limiting variance of the MLE $\hat{\Lambda}$ depends on the kurtosis of e_{jt}. Given Theorem 3.4.3, when e_{it} is normally distributed, we have $\kappa_{i,4} = 3\sigma_i^4$, the asymptotic variance can be simplified as the next corollary shows.

Corollary 3.4.1. *Under assumptions of Theorem 3.4.3, with normality of e_{it}, we have*

$$\sqrt{NT}\Big[\text{vec}(\hat{\Lambda}'-\Lambda')-\frac{1}{NT}\mathbb{B}_2\sum_{i=1}^{N}\frac{1}{\sigma_i^2}(m_i\otimes m_i)\Big]\xrightarrow{d}N\Big(0,\mathbb{B}_{1,\infty}(R_\infty\otimes I_r)\mathbb{B}_{1,\infty}'+2\mathbb{B}_{2,\infty}V_\infty\mathbb{B}_{2,\infty}'\Big),$$

where R_∞ and V_∞ are defined in Assumption C.3, $\mathbb{B}_{1,\infty}$ and $\mathbb{B}_{2,\infty}$ are almost the same as $\mathbb{B}_1$ and $\mathbb{B}_2$ except that P and R are replaced by P_∞ and R_∞. Furthermore, if $N/T\to 0$, we have

$$\sqrt{NT}\text{vec}(\hat{\Lambda}'-\Lambda')\xrightarrow{d}N\Big(0,\mathbb{B}_{1,\infty}(R_\infty\otimes I_r)\mathbb{B}_{1,\infty}'+2\mathbb{B}_{2,\infty}V_\infty\mathbb{B}_{2,\infty}'\Big).$$

Remark 3.4.2. To estimate the bias and the limiting variance, we use some plug-in methods. Specifically, the bias is estimated by

$$\hat{\Delta}=\hat{\mathbb{B}}_2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^6}(\hat{\kappa}_{i,4}-\hat{\sigma}_i^4)(m_i\otimes m_i),$$

and the limiting variance is estimated by

$$\hat{\Omega}=\hat{\mathbb{B}}_1(\hat{R}\otimes I_r)\hat{\mathbb{B}}_1'+\hat{\mathbb{B}}_2\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\kappa}_{i,4}-\hat{\sigma}_i^4}{\hat{\sigma}_i^8}(m_im_i')\otimes(m_im_i')\Big]\hat{\mathbb{B}}_2',$$

where

$$\hat{\mathbb{B}}_1=K_{kr}[(\hat{P}^{-1}\hat{\Lambda}')\otimes\hat{\Lambda}]+\hat{R}^{-1}\otimes I_r-K_{kr}(I_r\otimes\hat{\Lambda})\hat{\mathbb{D}}_1^{-1}\mathbb{D}_2[(\hat{P}^{-1}\hat{\Lambda}')\otimes I_r],$$

$$\hat{\mathbb{B}}_2=K_{kr}(I_r\otimes\hat{\Lambda})\hat{\mathbb{D}}^{-1}\mathbb{D}_3(\hat{\Lambda}\otimes\hat{\Lambda})'.$$

Here $\hat{\Lambda}$ and $\hat{\sigma}_i^2$ are the MLE; $\hat{R}=\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M$ and $\hat{P}=\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}$; $\hat{\mathbb{D}}_1$ is almost the same as $\mathbb{D}_1$ except that P is replaced by $\hat{P}$; $\hat{\kappa}_{i,4} = \frac{1}{T}\sum_{t=1}^{T}\hat{e}_{it}^4$ with $\hat{e}_{it} = z_{it} - m_i'\hat{\Lambda}\hat{f}_t$ and $\hat{f}_t = (\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}z_t$.

Remark 3.4.3. Theorem 3.4.3 is derived under a full identification of loading matrix Λ. An alternative approach to investigate the asymptotics, as adopted in Bai (2003), is that one only imposes the condition $M_{ff} = I_r$. Since in this case the original identification conditions (IC) are not met, the loading matrix Λ is not fully identified. But one can still deliver the asymptotic theory based on $\hat{\Lambda}' - \mathcal{R}\Lambda'$, where $\mathcal{R}$ is a rotational matrix. According to (C.1.18) in Appendix C, together with Lemma C.2.3 (e), (f) and Lemma C.2.5 (a), we have

$$\hat{\Lambda}' - \mathcal{R}\Lambda' = \mathcal{R}\frac{1}{T}\sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R_N^{-1} + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{T^{3/2}}),$$

where $\mathcal{R}$ is the rotational matrix defined by

$$\mathcal{R} = \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda + \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t'$$

with $\hat{P}_N = \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}$.

Given the above result, we have that under $N, T \to \infty, N/T^2 \to 0$,

$$\sqrt{NT}\text{vec}(\hat{\Lambda}' - \mathcal{R}\Lambda') \xrightarrow{d} N(0, R_\infty^{-1} \otimes \overline{\mathcal{R}}\overline{\mathcal{R}}'),$$

where $\overline{\mathcal{R}} = \underset{N,T\to\infty}{\text{plim}}\ \mathcal{R}$.

Theorem 3.4.4. *Under Assumptions A-D, as $N, T \to \infty$, we have*

$$\sqrt{T}(\hat{\sigma}_i^2 - \sigma_i^2) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2) + o_p(1).$$

Given this result, we have

$$\sqrt{T}(\hat{\sigma}_i^2 - \sigma_i^2) \xrightarrow{d} N(0, \kappa_{i,4} - \sigma_i^4),$$

where $\kappa_{i,4} = E(e_{it}^4)$ is the kurtosis of e_{it}.

We emphasize that the limiting result for $\hat{\sigma}_i^2$ is independent with the identification conditions. In addition, the above limiting result is the same as that in a standard high-dimensional factor model (see, e.g., Theorem 5.4 of Bai and Li (2012)).

We finally consider the estimation of factors. Following Bai and Li (2012), we estimate the factors by the generalized least squares (GLS) method. More specifically, the GLS estimator of f_t is

$$\hat{f}_t = (\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda})^{-1} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} z_t,$$

where $\hat{\Lambda}$ and $\hat{\Sigma}_{ee}$ are the respective MLEs of Λ and Σ_{ee}. The asymptotic representation and limiting distribution of $\hat{f}_t$ are provided in the following theorem.

Theorem 3.4.5. *Under assumptions of Theorem 3.4.1, we have*

$$\hat{f}_t - f_t = P^{-1} \frac{1}{N} \Lambda' M' \Sigma_{ee}^{-1} e_t + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right),$$

where $P = \frac{1}{N}\Lambda' M' \Sigma_{ee}^{-1} M\Lambda$. *Then as* $N, T \to \infty$ *and* $N/T^2 \to 0$, *we have*

$$\sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} N(0, P_\infty^{-1}),$$

where $P_\infty = \lim_{N\to\infty} P$ *is defined in Assumption C.3.*

The above theorem indicates that the asymptotic properties of the GLS estimator for factors in the current model are the same as that in standard high-dimensional factor models[2]. However, the derivation of the above theorem is actually easier due to the faster convergence rate of estimated loadings.

3.5 Testing

The limiting distribution of the MLE in Theorem 3.4.3 allows one to test whether the loading matrix Λ is equal to some known matrix. Consider the following hypothesis:

$$H_{\Lambda,0} : \Lambda = \Lambda^o, \qquad H_{\Lambda,1} : \Lambda \neq \Lambda^o.$$

A Wald statistic for this hypothesis testing is

$$W_\Lambda = NT\Big[\text{vec}(\hat{\Lambda}' - \Lambda^{o\prime}) - \frac{1}{T}\hat{\Delta}\Big]'\hat{\Omega}^{-1}\Big[\text{vec}(\hat{\Lambda}' - \Lambda^{o\prime}) - \frac{1}{T}\hat{\Delta}\Big],$$

where the symbols $\hat{\Delta}$ and $\hat{\Omega}$ are given in Remark 3.4.2. The following theorem, which is a direct result of Theorem 3.4.3, gives the limiting distribution of W_Λ.

[2]For the asymptotic results of the GLS estimator in standard high dimensional factor models, see Theorem 6.1 of Bai and Li (2012).

Theorem 3.5.1. *Under Assumptions A-D, together with IC, as $N, T \to \infty$ and $N/T^2 \to 0$, under $H_{\Lambda,0}$, we have*

$$W_\Lambda \xrightarrow{d} \chi^2_{kr},$$

where χ^2_{kr} denotes a chi-square distribution with degrees of freedom equal to kr.

An important issue related with the constrained factor model is that whether specification (3.3.1) is appropriate in a general factor model. Therefore, in practice one is likely to be interested in testing the correctness of the decomposition of loadings matrix $L = M\Lambda$. For a given M, the corresponding null and alternative hypotheses are

$$H_0: \quad L = M\Lambda \quad \text{for} \quad \text{some} \quad \Lambda,$$

$$H_1: \quad L \neq M\Lambda \quad \text{for} \quad \text{all} \quad \Lambda.$$

In traditional (low-dimensional) factor analysis, testing restrictions on loadings can be conducted by using the likelihood ratio (LR) principle. Because the number of parameters is finite, the number of imposed restrictions is finite too. By standard arguments, one can show that, under the null hypothesis, the LR statistic has an asymptotic χ^2 distribution with the degrees of freedom equal to the number of restrictions. In the high-dimensional setting, the number of parameters increases with the sample size. The number of restrictions possibly increases with the sample size as well. This is the case in our specification test in constrained factor models. As can be seen that under H_0, the number of restrictions for $L = M\Lambda$ is $(N - k)r$, which proportionally increases with the number of cross sectional units. As a result, the limiting distribution of the traditional LR test would have divergent degrees of freedom, an undesirable feature which can make the test unstable. This motives us to design a new test independent of N.

To gain an insight of our test, notice that the estimator $M\hat{\Lambda}$[3] under IC and H_0 should be very close to $\hat{L}$, the MLE of L from a standard factor model ($z_t = Lf_t + e_t$) under the identification condition that $M_{ff} = I_r$ and $\frac{1}{N}L'\Sigma_{ee}^{-1}L$ is diagonal. However, under H_1, the two estimates will not be close to each other. Based on the above analysis, we construct the following test statistic

$$W = \sqrt{NT^2}\text{tr}\left[\frac{1}{N}(M\hat{\Lambda} - \hat{L})'\widetilde{\Sigma}_{ee}^{-1}(M\hat{\Lambda} - \hat{L}) - \frac{1}{T}I_r\right],$$

where $\widetilde{\Sigma}_{ee}$ is an estimator of Σ_{ee} under the alternative hypothesis.

Theorem 3.5.2. *Under the same assumptions of Proposition 1.2.1 and $N/T^2 \to 0$, under H_0, we have*

$$W \xrightarrow{d} N(0, 2r).$$

Remark 3.5.1. As pointed out in Section 2, the identification condition has a sign problem. This problem should be carefully treated in the two statistics (W_Λ and W) in implementations, otherwise it may lead to an erroneous rejection of the null hypothesis. To eliminate this problem, when calculating W_Λ, we first compute the inter product of each column of $\hat{\Lambda}$ and the counterpart of Λ^o. If the value is negative, we multiple -1 on this column of $\hat{\Lambda}$. As regard to W, both $\hat{L}$ and $M\hat{\Lambda}$ have the sign problem, but we can use a similar procedure to

[3] An alternative estimator is $M\hat{\Lambda}^\dagger$, where $\hat{\Lambda}^\dagger$ is the bias-corrected estimator for Λ. It can be shown that the difference of the two statistics (which are based on $\hat{\Lambda}^\dagger$ and $\hat{\Lambda}$) is asymptotically negligible under $N/T^2 \to 0$.

deal with it. That is, for each column of $\hat{L}$, we calculate the inner product of this column and its counterpart of $M\hat{\Lambda}$. If the inner product is negative, we multiple -1 on this column of $\hat{L}$. After this treatment, the sign problem concomitant with the identification condition is removed.

Remark 3.5.2. Although we use the symbol W to denote the proposed statistic in the paper, our W statistic differs from the conventional Wald test. There are some key features that are different between our W test and the Wald test. First, the Wald test only involves estimators from an unconstrained model. In contrast, we use estimators from both constrained and unconstrained models to construct the W statistic. Second, the Wald test has an asymptotic χ^2 distribution with the value of degrees of freedom equal to the number of restrictions. But our W statistic has an asymptotic normal distribution, which is free of degree of freedom. For the same reasons, our W statistic is also different from a conventional Lagrange multiplier test.

Remark 3.5.3. In section 3.8, we run simulations to investigate the size and power performance of our proposed W test, in comparison with the traditional LR test (denoted as W_{LR}, which has an asymptotic χ^2 distribution with degree of freedom $(N-K)r$). We find that the performance of W_{LR} is dominated by W in terms of size and power. Furthermore, we also consider the normalized LR test (denoted as W_{LRN}, which equals $\frac{W_{LR}-(N-k)r}{\sqrt{2(N-k)r}}$ and has an asymptotic standard normal distribution). In finite sample simulations, the performance of W_{LRN} is similar to W_{LR}, and our W still outperforms W_{LRN} in terms of size and power.

3.6 Partially Constrained Factor Models

In this section, we consider the following partially constrained factor model

$$z_t = M\Lambda f_t + \Gamma g_t + e_t \triangleq \Phi h_t + e_t, \tag{3.6.1}$$

where $\Phi = [M\Lambda, \Gamma]$, $h_t = (f_t', g_t')'$ is an r-dimensional vector, f_t is an r_1-dimensional vector and g_t an r_2-dimensional vector with $r_1 + r_2 = r$. Again we study the ML estimation on model (3.6.1).

To analyze the MLE, we make the following assumptions.

Assumption A$'$. The factors $\{h_t\}$ satisfy the conditions in Assumption A.

Assumption C$'$. There exists a positive constant C such that $\|\phi_i\| < C$ for all i, where ϕ_i is the transpose of the ith row of Φ. Let $\mathcal{H} = \frac{1}{N}\Phi'\Sigma_{ee}^{-1}\Phi$, we assume $\overline{\mathcal{H}} = \lim_{N\to\infty} \mathcal{H} > 0$.

Identification condition, IC$'$. The identification conditions considered here are similar to those in the pure constrained factor model. More specifically, we require that $M_{hh} = \frac{1}{T}\sum_{t=1}^{T} h_t h_t' = I_r$ and $\mathcal{H}$ is a diagonal matrix with all its diagonal elements distinct and arranged in a descending order.

Let $\Sigma_{zz} = \Phi\Phi' + \Sigma_{ee}$ and $\theta = (\Lambda, \Gamma, \Sigma_{ee})$. The MLE is defined as

$$\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmax}}\ \mathcal{L}(\theta),$$

where

$$\mathcal{L}(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}\text{tr}[M_{zz}\Sigma_{zz}^{-1}].$$

Here Θ is the parameter space specified by Assumption D and the identification condition IC$'$. In Appendix C, we show that the first order condition for Λ can be written as

$$\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1} M = 0. \tag{3.6.2}$$

The first order condition for Γ can be written as

$$\hat{\Gamma}' \hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = 0. \tag{3.6.3}$$

The first order condition for Σ_{ee} can be written as

$$\text{diag}\Big[(M_{zz} - \hat{\Sigma}_{zz}) - M\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) - (M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\Big] = 0. \tag{3.6.4}$$

Before we present the asymptotic results for the MLE, we first introduce some notation

$$\mathbb{B}_1^\star = R^{-1} \otimes I_{r_1} + K_{kr_1}[(P^{-1}\Lambda') \otimes \Lambda] - K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}^{-1}E_1\Lambda') \otimes E_1],$$

$$\mathbb{B}_2^\star = K_{kr_1}[P^{-1} \otimes \psi] - K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}^{-1}E_1) \otimes E_2],$$

$$\mathbb{B}_3^\star = -K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}^{-1}E_2) \otimes E_1],$$

$$\mathbb{B}_4^\star = -K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}^{-1}E_2) \otimes E_2], \qquad \mathbb{B}_5^\star = -K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_3,$$

$$\Delta^\star = K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_3\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^6}(\phi_i \otimes \phi_i)(\kappa_{i,4} - \sigma_i^4) + \text{vec}(r_1\mathcal{H} - E_2E_2')\right],$$

where $E_1 = [I_{r_1}, 0_{r_1 \times r_2}]'$, $E_2 = [0_{r_2 \times r_1}, I_{r_2}]'$, $\psi = (M'\Sigma_{ee}^{-1}M)^{-1}M'\Sigma_{ee}^{-1}\Gamma$, $\Psi = [\Lambda, \psi]$ and $\mathcal{H}$ is defined in Assumption C$'$. The symbols $\kappa_{i,4}$, K_{mn}, P, R, $\mathbb{D}_1$, $\mathbb{D}_2$ and $\mathbb{D}_3$ are defined the same as in Section 3.4.

Let γ_i be the transpose of the ith row of Γ. The following theorem states the asymptotic representations for the MLE. The consistency and convergence rates are implied by the theorem.

Theorem 3.6.1. *Under Assumptions A', B, C' and D, when $N, T \to \infty$, we have, for all i,*

$$\hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2) + O_p\left(\frac{1}{T}\right).$$

In addition, if IC' is imposed, we have, for all i,

$$\hat{\gamma}_i - \gamma_i = \frac{1}{T}\sum_{t=1}^{T} g_t e_{it} + O_p\left(\frac{1}{T}\right)$$

and

$$\begin{aligned}
\text{vec}(\hat{\Lambda}' - \Lambda') &= \mathbb{B}_1^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it} + \mathbb{B}_2^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\Lambda' m_i \otimes g_t)e_{it} \\
&+ \mathbb{B}_3^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i \otimes f_t)e_{it} + \mathbb{B}_4^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i \otimes g_t)e_{it} \\
&+ \mathbb{B}_5^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}(\phi_i \otimes \phi_i)(e_{it}^2 - \sigma_i^2) + \frac{1}{T}\Delta^\star \\
&+ O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right),
\end{aligned}$$

where $\mathbb{B}_1^\star, \ldots, \mathbb{B}_5^\star$ and $\Delta^\star$ are defined above this theorem.

Given the above theorem, we have the following distribution results for the MLE.

Corollary 3.6.1. *Under Assumptions A', B, C' and D, when $N, T \to \infty$, we have, for all i,*

$$\sqrt{T}(\hat{\sigma}_i^2 - \sigma_i^2) \xrightarrow{d} N(0, \kappa_{i,4} - \sigma_i^4).$$

In addition, if IC' is imposed, we have, for all i,

$$\sqrt{T}(\hat{\gamma}_i - \gamma_i) \xrightarrow{d} N(0, \sigma_i^2 I_{r_2}).$$

If $N/T^2 \to 0$ is further imposed, we have

$$\sqrt{NT}\Big[\text{vec}(\hat{\Lambda}' - \Lambda') - \frac{1}{T}\Delta^\star\Big] \xrightarrow{d} N(0, \Omega^\star),$$

where $\Omega^\star = \lim_{N\to\infty} \Omega_N^\star$ with

$$\begin{aligned}\Omega_N^\star &= \mathbb{B}_1^\star(R \otimes I_{r_1})\mathbb{B}_1^{\star\prime} + \mathbb{B}_2^\star(P \otimes I_{r_1})\mathbb{B}_2^{\star\prime} + \mathbb{B}_3^\star(Q \otimes I_{r_1})\mathbb{B}_3^{\star\prime} + \mathbb{B}_4^\star(Q \otimes I_{r_2})\mathbb{B}_4^{\star\prime} \\ &\quad + \mathbb{B}_1^\star(S \otimes I_{r_1})\mathbb{B}_3^{\star\prime} + \mathbb{B}_3^\star(S' \otimes I_{r_1})\mathbb{B}_1^{\star\prime} + \mathbb{B}_5^\star\Big[\frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^8}(\phi_i\phi_i') \otimes (\phi_i\phi_i')(\kappa_{i,4} - \sigma_i^4)\Big]\mathbb{B}_5^{\star\prime},\end{aligned}$$

where $Q = \Gamma'\Sigma_{ee}^{-1}\Gamma/N$ and $S = M'\Sigma_{ee}^{-1}\Gamma/N$.

The approach to estimate the factors in partially constrained factor models is similar as before. Given the MLE $\hat{\Lambda}, \hat{\Gamma}$ and $\hat{\Sigma}_{ee}$, the GLS estimator of h_t is

$$\hat{h}_t = (\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\hat{\Phi})^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}z_t,$$

where $\hat{\Phi} = (M\hat{\Lambda}, \hat{\Gamma})$. Using the similar arguments as in the proof of Theorem 3.4.5, we have the following asymptotic representation and limiting distribution results on $\hat{h}_t$.

Theorem 3.6.2. *Under Assumptions A′, B, C′ and D, together with IC′, we have, for all t,*

$$\hat{h}_t - h_t = \mathcal{H}^{-1}\frac{1}{N}\Phi'\Sigma_{ee}^{-1}e_t + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right),$$

where $\mathcal{H} = \frac{1}{N}\Phi'\Sigma_{ee}^{-1}\Phi$. *Then as* $N, T \to \infty$ *and* $N/T^2 \to 0$, *we have*

$$\sqrt{N}(\hat{h}_t - h_t) \xrightarrow{d} N(0, \bar{\mathcal{H}}^{-1}),$$

where $\bar{\mathcal{H}} = \lim_{N\to\infty} \mathcal{H}$ *is defined in Assumption C′.*

3.7 EM algorithm

The ML estimation can be easily implemented via the EM algorithm. The iterating formulas for a purely constrained factor model and a partially constrained one are different. We present them separately.

3.7.1 EM algorithm for the pure constrained factor model

Let $\theta^{(k)} = (\Lambda^{(k)}, \Sigma_{ee}^{(k)})$ denote the estimate at the kth iteration. The EM algorithm updates and calculates $\theta^{(k+1)} = (\Lambda^{(k+1)}, \Sigma_{ee}^{(k+1)})$ by

$$\Lambda^{(k+1)} = (M'\Sigma_{ee}^{(k)-1}M)^{-1}\left[M'\Sigma_{ee}^{(k)-1}\frac{1}{T}\sum_{t=1}^{T}E(z_t f_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(f_t f_t'|Z,\theta^{(k)})\right]^{-1},$$

$$\text{diag}(\Sigma_{ee}^{(k+1)}) = \text{diag}\Bigg\{M_{zz} - \frac{2}{T}\sum_{t=1}^{T}E(z_t f_t'|Z,\theta^{(k)})\Lambda^{(k+1)\prime}M'$$

$$+ M\Lambda^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)})\Lambda^{(k+1)\prime}M'\Bigg\},$$

where $\Sigma_{zz}^{(k)} = M\Lambda^{(k)}\Lambda^{(k)\prime}M' + \Sigma_{ee}^{(k)}$ and

$$\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)}) = \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M_{zz}(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)} + I_r - \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)},$$
$$\frac{1}{T}\sum_{t=1}^{T}E(z_tf_t'|Z,\theta^{(k)}) = M_{zz}(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)}.$$

The above iteration continues until $\|\theta^{(k+1)} - \theta^{(k)}\|$ is smaller than a preset tolerance. For the initial values, the PC estimates proposed in Tsai and Tsay (2010) are recommended. When iterations are terminated, the estimates, denoted by $(\Lambda^\dagger, \Sigma_{ee}^\dagger)$, need to be further normalized to satisfy the identification conditions in Section 3.3. The normalization can be conducted as follows. Let $V^\dagger$ be the orthogonal matrix consisting of the eigenvectors of the matrix $\frac{1}{N}\Lambda^{\dagger\prime}M'(\Sigma_{ee}^\dagger)^{-1}M\Lambda^\dagger$ with the corresponding eigenvalues arranged in a descending order. Let $\hat{\Lambda} = \Lambda^\dagger V^\dagger$ and $\hat{\Sigma}_{ee} = \Sigma_{ee}^\dagger$. Then $\hat{\theta} = (\hat{\Lambda}, \hat{\Sigma}_{ee})$ is the MLE that satisfies IC.

Bai and Li (2012) show that the iterating formulas of the EM algorithm approach to the first order conditions of the likelihood function as the iteration tends to infinity. Using their arguments, one can show similar results in constrained factor models. Since the proof is almost the same as in Bai and Li (2012), we omit it for sake of space.

3.7.2 EM algorithm for the partially constrained factor model

Let $\theta^{(k)} = (\Lambda^{(k)}, \Gamma^{(k)}, \Sigma_{ee}^{(k)})$ denote the estimate at the kth iteration. The EM algorithm updates and calculates $\theta^{(k+1)} = (\Lambda^{(k+1)}, \Gamma^{(k+1)}, \Sigma_{ee}^{(k+1)})$ by

$$\begin{aligned}
\Lambda^{(k+1)} &= (M'\Sigma_{ee}^{(k)-1}M)^{-1}\left[M'\Sigma_{ee}^{(k)-1}\frac{1}{T}\sum_{t=1}^{T}E(z_tf_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)})\right]^{-1} \\
&\quad - (M'\Sigma_{ee}^{(k)-1}M)^{-1}\left[M'\Sigma_{ee}^{(k)-1}\Gamma^{(k)}\frac{1}{T}\sum_{t=1}^{T}E(g_tf_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)})\right]^{-1}, \\
\Gamma^{(k+1)} &= \left[\frac{1}{T}\sum_{t=1}^{T}E(z_tg_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(g_tg_t'|Z,\theta^{(k)})\right]^{-1} \\
&\quad - \left[M\Lambda^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(f_tg_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(g_tg_t'|Z,\theta^{(k)})\right]^{-1}, \\
\operatorname{diag}(\Sigma_{ee}^{(k+1)}) &= \operatorname{diag}\left\{M_{zz} - \frac{2}{T}\sum_{t=1}^{T}E(z_tf_t'|Z,\theta^{(k)})\Lambda^{(k+1)\prime}M' - \frac{2}{T}\sum_{t=1}^{T}E(z_tg_t'|Z,\theta^{(k)})\Gamma^{(k+1)\prime}\right. \\
&\quad + M\Lambda^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)})\Lambda^{(k+1)\prime}M' + \Gamma^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(g_tg_t'|Z,\theta^{(k)})\Gamma^{(k+1)\prime} \\
&\quad \left. + 2M\Lambda^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(f_tg_t'|Z,\theta^{(k)})\Gamma^{(k+1)\prime}\right\},
\end{aligned}$$

where $\Sigma_{zz}^{(k)} = M\Lambda^{(k)}\Lambda^{(k)\prime}M' + \Gamma^{(k)}\Gamma^{(k)\prime} + \Sigma_{ee}^{(k)}$ and

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)}) &= \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M_{zz}(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)} + I_{r_1} - \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)}, \\
\frac{1}{T}\sum_{t=1}^{T}E(f_tg_t'|Z,\theta^{(k)}) &= \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M_{zz}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)} - \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)}, \\
\frac{1}{T}\sum_{t=1}^{T}E(g_tg_t'|Z,\theta^{(k)}) &= \Gamma^{(k)\prime}(\Sigma_{zz}^{(k)})^{-1}M_{zz}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)} + I_{r_2} - \Gamma^{(k)\prime}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)}, \\
\frac{1}{T}\sum_{t=1}^{T}E(z_tf_t'|Z,\theta^{(k)}) &= M_{zz}(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)},
\end{aligned}$$

$$\frac{1}{T}\sum_{t=1}^{T} E(z_t g_t' | Z, \theta^{(k)}) = M_{zz}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)}.$$

Likewise, we use the PC estimates as the starting values, and iterate the above formulas until $\|\theta^{(k+1)} - \theta^{(k)}\|$ is smaller than a preset tolerance. Let $\theta^\diamond = (\Lambda^\diamond, \Gamma^\diamond, \Sigma_{ee}^\diamond)$ be the estimates of the last iteration. Again we need rotate $\theta^\diamond$ to satisfy the IC$''$. Let $V^\diamond$ be the orthogonal matrix consisting of the eigenvectors of the matrix $\frac{1}{N}\Phi^{\diamond\prime}(\Sigma_{ee}^\diamond)^{-1}\Phi^\diamond$ with the corresponding eigenvalues arranged in a descending order, where $\Phi^\diamond = (M\Lambda^\diamond, \Gamma^\diamond)$. Let $\Phi^\diamond V^\diamond$ and split $\Phi^\triangle$ into $\Phi^\triangle = (\Phi_1^\triangle, \Phi_2^\triangle)$, where $\Phi_1^\triangle$ is made up with the left r_1 columns and $\Phi_2^\triangle$ the remaining r_2 columns. Then calculate $\hat{\Lambda} = (M'M)^{-1}M'\Phi_1^\triangle$, and simply let $\hat{\Gamma} = \Phi_2^\triangle$ and $\hat{\Sigma}_{ee} = \Sigma_{ee}^\diamond$. Then $\hat{\theta} = (\hat{\Lambda}, \hat{\Gamma}, \hat{\Sigma}_{ee})$ is the MLE that satisfies IC$''$.

Again, we can show that the limit of the iterated EM solutions satisfy the first order conditions (3.6.2), (3.6.3) and (3.6.4). The proof is similar to the pure constrained factor model case and therefore skipped here.

3.8 Simulation results

In this section, we run simulations to investigate the finite sample performance of the MLE, as well as the empirical size and power of the W test.

3.8.1 Finite sample performance of the MLE

We first conduct simulations to investigate the finite sample properties of the MLE and compare it with the PC estimates proposed by Tsai and Tsay (2010).

In the literature on high dimensional factor models, researchers usually use a generalized

R^2 or a trace ratio to measure the goodness-of-fit, e.g., Stock and Watson (2002a), Doz et al. (2012) and Bai and Li (2012). These measures are invariant to the rotational indeterminacy and therefore effective to perform the measure task. However, in constrained factor models, such measures are not suitable since the estimates have faster convergence rates, which often leads to a high value of the generalized R^2 or the trace ratio. For this reason, we instead consider an alternative measure by rotating the underlying values to satisfy the identification condition and investigating the precision of $\hat{\Lambda} - \Lambda$ for rotated values. We calculate the mean absolute deviation (MAD) and the root mean square error (RMSE) based on the rotated underlying values. We also calculate the root asymptotic variance (RAvar) to check the convergence rate of $\hat{\Lambda}$ presented in Theorem 3.4.1. The calculation formulas based on S simulations are as follows

$$\text{MAD} = \frac{1}{S}\sum_{s=1}^{S}\Big(\frac{1}{kr}\sum_{p=1}^{k}\sum_{i=1}^{r}|\hat{\Lambda}_{pi}^{s} - \Lambda_{pi}^{s}|\Big),$$

$$\text{RMSE} = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\Big(\frac{1}{kr}\sum_{p=1}^{k}\sum_{i=1}^{r}(\hat{\Lambda}_{pi}^{s} - \Lambda_{pi}^{s})^2\Big)},$$

$$\text{RAvar} = \sqrt{NT} \times \sqrt{\frac{1}{S}\sum_{s=1}^{S}\Big(\frac{1}{kr}\sum_{p=1}^{k}\sum_{i=1}^{r}(\hat{\Lambda}_{pi}^{bc,s} - \Lambda_{pi}^{s})^2\Big)},$$

where $\hat{\Lambda}_{pi}^{s}$ and $\hat{\Lambda}_{pi}^{bc,s}$ are the MLE and biased-corrected MLE in the sth simulation, respectively.

Data are generated according to $z_t = M\Lambda f_t + e_t$, where all elements of M are drawn independently from $U[0,1]$ and all elements of Λ and F independently from $N(0,1)$. The idiosyncratic errors e_{it} are generated according to $e_{it} = \sigma_i \epsilon_{it}$ with σ_i^2 being the ith diagonal element of $(M\Lambda\Lambda' M')$ multiplying $\frac{b_i}{1-b_i}$, where $b_i = 0.2 + 0.6U_i$ and $U_i \sim U[0,1]$. The component ϵ_{it} is generated from the three distributions: the normal distribution, student's distribution with 5 degrees of freedom and chi-squared distribution with 2 degrees of freedom. For the latter two distributions, we normalize the random variable to have zero mean and init variance. For the values of k and r, we consider two cases: $(k, r) = (3, 1)$ and $(k, r) = (8, 3)$.

Throughout this section, we assume that the number of common factors is known. There are a number of methods at hand to determine the number of factors, for example, the information criterion method by Bai and Ng (2002), the largest eigenvalue-ratios method by Ahn and Horenstein (2013) and the eigenvalue empirical distribution method by Onatski (2010). If the number of factors is unknown, one can choose any of the method mentioned above to estimate it. Tables 3.1 and 3.2 present the performance of the MLE and the PC estimate for normal errors under the sample sizes of $N = 30, 50, 100, 150$ and $T = 30, 50, 100$. The results under student-t errors and chi-square errors are almost the same as those for normal errors and are given in Tables C.51-C.54 in Appendix C to save space. All these results are obtained based on 1000 repetitions.

From Tables 3.1 and 3.2, we can see that both MAD and RMSE of the MLE are much smaller than those of PC estimates for all (N, T) combinations, implying that the MLE performs better than the PC estimate. Regarding the RAvar[4], we see that the MLE has almost constant RAvars when the time dimension T or the cross section dimension N increases, implying that the convergence rate of the MLE is $\sqrt{NT}$. This simulation result is consistent with our theoretical results in Section 3.4. In addition, it seems that the PC estimate also has $\sqrt{NT}$ convergence rate from simulations. Finally, we note that the RMSEs of the MLE

[4]Since we do not know whether the PC estimate is biased, and if biased, what is the bias formula. Hence, we cannot calculate RAvar for the PC estimate.

Table 3.1: $k = 3$, $r = 1$, and $\epsilon_{it} \sim N(0, 1)$.

$\Lambda_{3\times1}$		MLE			PC		
N	T	MAD	RMSE	RAvar	MAD	RMSE	RAvar
30	30	0.0440	0.0716	2.2301	0.0943	0.1386	N/A
50	30	0.0349	0.0540	1.9887	0.0654	0.0934	N/A
100	30	0.0262	0.0417	2.0504	0.0474	0.0677	N/A
150	30	0.0216	0.0340	2.1741	0.0410	0.0582	N/A
30	50	0.0333	0.0533	2.1936	0.0787	0.1145	N/A
50	50	0.0237	0.0368	1.9426	0.0546	0.0800	N/A
100	50	0.0190	0.0306	1.9194	0.0375	0.0541	N/A
150	50	0.0159	0.0255	2.0863	0.0293	0.0417	N/A
30	100	0.0232	0.0374	2.1425	0.0674	0.0964	N/A
50	100	0.0172	0.0263	1.8314	0.0443	0.0611	N/A
100	100	0.0105	0.0168	1.7473	0.0253	0.0358	N/A
150	100	0.0102	0.0165	1.8668	0.0200	0.0288	N/A

Table 3.2: $k = 8$, $r = 3$, and $\epsilon_{it} \sim N(0, 1)$.

$\Lambda_{8\times3}$		MLE			PC		
N	T	MAD	RMSE	RAvar	MAD	RMSE	RAvar
30	30	0.3498	0.5006	15.2632	0.5655	0.8071	N/A
50	30	0.2307	0.3310	13.6988	0.3744	0.5363	N/A
100	30	0.1537	0.2307	12.5998	0.2224	0.3131	N/A
150	30	0.1245	0.1881	11.7159	0.1735	0.2452	N/A
30	50	0.2637	0.3744	14.4701	0.5130	0.7521	N/A
50	50	0.1794	0.2689	13.1269	0.3184	0.4679	N/A
100	50	0.1082	0.1578	12.1691	0.1763	0.2545	N/A
150	50	0.0860	0.1291	12.3152	0.1382	0.2091	N/A
30	100	0.1846	0.2698	15.5540	0.4570	0.6882	N/A
50	100	0.1213	0.1937	13.3273	0.2622	0.4064	N/A
100	100	0.0774	0.1258	11.9418	0.1440	0.2157	N/A
150	100	0.0620	0.1021	12.9696	0.1033	0.1633	N/A

Table 3.3: The empirical size of the test statistic W for $(k, r) = (3, 1)$

Empirical size of W										
$\epsilon_{it} \sim$		$N(0,1)$			t_5			$\chi^2(2)$		
N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%
30	30	3.6%	7.4%	13.5%	3.8%	8.5%	12.9%	2.7%	8.0%	13.3%
50	30	4.4%	11.5%	16.6%	3.9%	9.5%	16.3%	5.4%	10.5%	16.1%
100	30	6.7%	14.2%	20.5%	6.5%	13.9%	20.1%	5.5%	12.9%	21.1%
150	30	9.2%	18.4%	24.8%	8.1%	18.6%	27.1%	8.2%	20.3%	29.0%
30	50	1.7%	5.9%	11.3%	1.3%	5.8%	12.7%	1.7%	6.6%	11.6%
50	50	3.1%	6.8%	13.0%	2.6%	6.1%	11.0%	2.0%	7.0%	12.1%
100	50	3.3%	8.0%	15.2%	2.3%	8.3%	14.2%	3.5%	9.7%	15.7%
150	50	4.6%	11.4%	18.1%	3.4%	11.1%	17.3%	2.8%	9.3%	15.8%
30	100	0.6%	4.5%	10.4%	1.4%	4.0%	10.6%	1.0%	4.8%	10.9%
50	100	1.5%	4.2%	10.9%	1.5%	6.1%	9.9%	1.2%	5.8%	11.7%
100	100	1.4%	6.5%	11.6%	0.9%	5.8%	12.6%	1.5%	6.5%	12.4%
150	100	1.6%	5.6%	10.9%	2.0%	7.5%	12.7%	1.9%	5.8%	11.3%
30	150	0.6%	5.0%	10.5%	1.0%	5.0%	9.9%	1.2%	5.8%	10.2%
50	150	1.5%	5.9%	10.4%	1.5%	4.8%	10.2%	1.5%	5.1%	9.6%
100	150	0.7%	6.2%	10.7%	1.2%	5.4%	10.2%	1.5%	5.8%	11.6%
150	150	1.9%	5.9%	9.6%	1.6%	5.0%	11.5%	1.7%	5.2%	10.8%
100	100	1.4%	6.5%	11.6%	0.9%	5.8%	12.6%	1.5%	6.5%	12.4%
200	100	1.3%	6.1%	11.2%	1.4%	6.7%	13.5%	2.2%	7.2%	12.6%
300	100	2.3%	6.5%	12.8%	2.1%	6.8%	12.7%	1.8%	7.9%	12.9%
100	200	1.3%	4.0%	9.4%	1.3%	5.3%	10.8%	1.1%	5.1%	11.3%
200	200	1.4%	5.6%	10.5%	0.9%	4.9%	9.6%	1.4%	6.1%	11.6%
300	200	1.3%	6.1%	8.6%	1.5%	5.4%	11.6%	1.5%	5.9%	11.7%
100	300	0.4%	4.5%	9.5%	1.2%	5.1%	11.8%	1.2%	5.1%	9.2%
200	300	0.9%	6.1%	10.5%	1.3%	4.9%	9.1%	0.8%	6.2%	11.6%
300	300	1.3%	5.2%	10.9%	0.7%	3.9%	8.5%	1.2%	4.4%	9.0%
100	500	0.8%	5.3%	9.8%	0.8%	4.6%	10.9%	1.1%	5.2%	9.7%
200	500	0.9%	5.4%	9.8%	0.5%	5.1%	9.8%	1.0%	5.2%	10.3%
300	500	0.6%	5.3%	10.5%	1.5%	5.9%	9.2%	0.9%	5.0%	9.4%

are smaller than those of the PC estimates, indicating that the MLE is more efficient than the PC estimates.

3.8.2 Empirical size of the W test

In this subsection, we use simulations to study the empirical size of the W statistic. The data generating process is the same as in previous subsection, but with more combinations of (N, T). We investigate the performance of W under three nominal levels 1%, 5% and 10%. The empirical sizes of W for the case $(k, r) = (3, 1)$ are given in Table 3.3, which is obtained from 1000 repetitions.

From the results in Table 3.3, we emphasize the following findings. First, the performance of the W test is considerably good overall. Except for the sample size when T is small, almost all the empirical sizes of the W statistic fall in the interval $[5\%, 10\%]$ under the 5% nominal level. Second, the distribution type of errors has no significant impact on the performance of W. The W statistic performs very similarly under three different error distributions. This is consistent with the theoretical result in Section 3.5. Third, the performance of W is closely linked with time period number T, loosely with the number of units N. For example, when $T = 30$, the W statistic suffers a mildly severe size distortion. But when T grows to 50, the size distortion considerably decreases. As regard to N, we see that the W statistic performs well even when $N = 30$. We conjecture the reason is that when T is small, the variance σ_i^2 are estimated inaccurately, which leads to a poor performance of W.

Tsai and Tsay (2010) propose using a traditional likelihood ratio (LR) statistic to perform model specification testing. In the factor model literature, LR tests are usually considered under the fixed-N, large-T setup, see Lawley and Maxwell (1971). As mentioned in the introduction, when N and T are both large the traditional LR test may not be suitable. For example, the adjusted likelihood ratio test, which is often used with consideration of finite sample performance, may be negative for too large N. According to the simulation results in Table 7 in Tsai and Tsay (2010), the LR test suffers size distortion issue even when N is not large. As a primary competitor to our W statistic, we compare the performance of the W statistic and the LR one (also the normalized LR test as mentioned in Remark 3.5.3) under the current data generating setup. We find that the performance of the W statistic dominates that of the LR tests. Details are given in Appendix C.

3.8.3 Empirical power of the W test

We next study the empirical power of the W test. Data are generated by $z_t = Lf_t + e_t$ with

$$L = M\Lambda + d \cdot \nu,$$

where M, Λ, f_t and e_t are generated in the same way as in Section 3.8.1. The symbol ν is an $N \times r$ noise matrix with its elements drawn from $N(0,1)$ and d is a prespecified constant, which is related with N and T and is used to control the magnitude of deviation from the null hypothesis. In this section, we set it as

$$d = \frac{\alpha}{\sqrt[4]{N}\sqrt{T}}$$

with $\alpha = 0.2, 0.5, 2$ and 5. In classical models, if an estimator is $\sqrt{T}$-consistent, the local power is studied under $\beta = \beta^* + \frac{1}{\sqrt{T}}\alpha$, where β^* denotes the true value. However, this general result cannot be applied to the present context since we renormalize the distance between estimators from the constrained and unconstrained models to accommodate the large number of restrictions imposed in the null hypothesis. Directly deriving the local power of W is challenging. We conjecture that the W statistic can detect local alternatives that approach the null model at a rate of $N^{-1/4}T^{-1/2}$. Simulation results below seem to support our conjecture since the local power converges to some value as N and T grow larger in all choices of α.

Table 3.4 presents the empirical power of the W test for the case $(k, r) = (3, 1)$ under normal errors. It is seen that the W statistic has higher power when α is larger and lower

Table 3.4: The empirical power of the W test for $(k, r) = (3, 1)$

Empirical power of W													
α		0.2			0.5			2			5		
N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
30	30	22.9%	31.4%	37.4%	52.0%	57.5%	61.7%	91.2%	93.1%	93.7%	99.7%	100.0%	100.0%
50	30	31.8%	39.4%	44.9%	58.2%	64.1%	67.5%	94.1%	95.7%	96.4%	100.0%	100.0%	100.0%
100	30	51.4%	59.4%	63.7%	71.4%	77.3%	81.1%	96.2%	98.0%	98.7%	100.0%	100.0%	100.0%
150	30	55.5%	63.9%	68.0%	74.4%	78.9%	81.6%	97.9%	98.9%	99.2%	100.0%	100.0%	100.0%
30	50	22.9%	30.3%	35.2%	51.1%	57.4%	60.7%	89.3%	91.9%	93.6%	99.6%	99.8%	99.8%
50	50	29.2%	36.3%	42.2%	58.2%	63.8%	67.4%	93.7%	95.8%	96.7%	99.8%	99.9%	99.9%
100	50	45.5%	51.7%	56.3%	69.2%	72.7%	76.1%	96.5%	97.7%	98.1%	100.0%	100.0%	100.0%
150	50	51.3%	58.3%	63.4%	70.9%	76.0%	79.2%	97.3%	98.2%	98.5%	100.0%	100.0%	100.0%
30	100	20.5%	25.7%	31.5%	53.6%	60.7%	62.9%	90.0%	92.2%	93.8%	99.5%	99.6%	99.6%
50	100	29.8%	35.6%	41.1%	59.3%	64.2%	67.2%	93.1%	94.7%	95.7%	100.0%	100.0%	100.0%
100	100	37.7%	43.3%	47.5%	65.6%	70.1%	72.3%	94.1%	96.2%	97.3%	99.9%	100.0%	100.0%
150	100	49.8%	55.4%	59.0%	70.1%	74.2%	77.6%	95.5%	96.6%	97.2%	100.0%	100.0%	100.0%
30	150	19.9%	25.4%	29.8%	55.8%	62.1%	64.5%	88.2%	91.2%	92.0%	99.6%	99.8%	99.9%
50	150	28.4%	34.9%	40.8%	58.1%	62.2%	65.3%	90.8%	93.4%	93.8%	99.8%	99.9%	99.9%
100	150	37.7%	44.8%	49.8%	66.5%	69.9%	72.8%	93.1%	95.1%	96.4%	100.0%	100.0%	100.0%
150	150	46.2%	51.1%	55.3%	67.1%	71.0%	74.3%	95.9%	97.0%	97.5%	100.0%	100.0%	100.0%
100	100	40.0%	46.1%	51.5%	65.4%	70.2%	73.3%	93.8%	96.3%	96.9%	100.0%	100.0%	100.0%
200	100	52.5%	57.3%	61.4%	71.6%	74.8%	77.0%	96.6%	97.3%	97.7%	100.0%	100.0%	100.0%
300	100	59.5%	63.7%	68.2%	75.0%	77.7%	80.0%	95.9%	97.1%	97.4%	100.0%	100.0%	100.0%
100	200	39.9%	46.9%	51.9%	66.2%	70.9%	73.2%	93.4%	94.8%	95.6%	99.8%	99.9%	99.9%
200	200	48.5%	54.8%	58.2%	68.4%	72.9%	76.2%	95.9%	97.0%	97.3%	100.0%	100.0%	100.0%
300	200	56.0%	59.9%	63.0%	69.3%	72.8%	75.9%	96.4%	97.4%	98.3%	100.0%	100.0%	100.0%
100	300	41.0%	47.4%	50.2%	67.4%	71.9%	73.4%	93.3%	94.9%	95.4%	100.0%	100.0%	100.0%
200	300	50.6%	55.6%	58.9%	68.7%	72.3%	74.4%	94.7%	95.8%	96.4%	100.0%	100.0%	100.0%
300	300	54.9%	59.0%	63.1%	72.3%	74.9%	77.3%	94.8%	96.8%	97.6%	100.0%	100.0%	100.0%
100	500	39.5%	45.0%	49.0%	65.1%	68.9%	71.2%	94.0%	95.6%	96.6%	99.9%	99.9%	99.9%
200	500	50.4%	54.4%	58.4%	69.4%	72.6%	75.6%	95.4%	97.2%	97.6%	100.0%	100.0%	100.0%
300	500	53.4%	58.3%	61.8%	71.2%	73.2%	75.2%	96.1%	97.4%	97.9%	100.0%	100.0%	100.0%

power when α is smaller. This is an expected result. As α becomes larger, the distance between the null hypothesis and the alternative hypothesis is larger and then we have more chances to differentiate the two hypotheses. Given that the W statistic has considerable power even against the local alternatives that are $N^{-1/4}T^{-1/2}$ away from the null model, we conclude that the W has good performance in terms of empirical power. We also compare empirical powers of the W statistic and the LR test (also the normalized LR test as mentioned in Remark 3.5.3). We find that the performance of the W test is better than that of the LR tests. Details are given in Appendix C.

3.9 Extension

In this section, we relax Assumption B to allow for general weakly dependence idiosyncratic errors. Following Chamberlain and Rothschild (1983) we call a factor model with weak dependence idiosyncratic errors the approximate factor model. Approximate factor models are the primary research interests in a number of studies, e.g., Bai and Ng (2002), Bai (2003) and Bai and Li (2016), among others. To relax Assumption B, we introduce the following assumption to control the heteroskedasticity and weak correlations over cross section and time.

Assumption B″: (weak dependence on errors)

B″.1 $E(e_{it}) = 0$, and $E(e_{it}^8) \leq C$.

B″.2 Let $\mathbb{O}_t = E(e_t e_t')$, $\mathbb{O} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{O}_t$, and $\mathbb{W} = \text{diag}(\mathbb{O})$, which is the diagonal matrix that sets the off-diagonal elements of $\mathbb{O}$ to zero. Specifically, let w_i^2 be the ith diagonal element of $\mathbb{W}$, then $\mathbb{W} = \text{diag}(w_1^2, w_2^2, \ldots, w_N^2)$.

B″.3 For all i, $C^{-2} \leq w_i^2 \leq C^2$;

B″.4 Let $\tau_{ij,t} \equiv E(e_{it}e_{jt})$, assume there exists some positive τ_{ij} such that $|\tau_{ij,t}| \leq \tau_{ij}$ for all t and $\sum_{i=1}^{N}\tau_{ij} \leq C$ for all j.

B″.5 Let $\rho_{i,ts} \equiv E(e_{it}e_{is})$, assume there exists some positive ρ_{ts} such that $|\rho_{i,ts}| \leq \rho_{ts}$ for all i and $\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t}\rho_{ts} \leq C$.

B″.6 Assume $E\left[\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[e_{it}e_{jt} - E(e_{it}e_{jt})\right]\right|^4\right] \leq C$ for all i and all j.

To be consistent with the changes in Assumption B″, we modify Assumptions C and D as follows.

Assumption C″:

C″.1 $\|\Lambda\| \leq C$ and $\|m_j\| \leq C$ for all j, where m_j is the transpose of the jth row of M.

C″.2 Let $\mathbb{P} = \Lambda' M' \mathbb{W}^{-1} M \Lambda / N$, $\mathbb{R} = M' \mathbb{W}^{-1} M / N$. We assume that $\mathbb{P}_\infty = \lim_{N \to \infty} \mathbb{P}$ and $\mathbb{R}_\infty = \lim_{N \to \infty} \mathbb{R}$ exist. Here $\mathbb{P}_\infty$ and $\mathbb{R}_\infty$ are some positive definite matrices.

Assumption D″: The estimator of w_j^2 for $j = 1, ..., N$ takes value in a compact set: $[C^{-2}, C^2]$. Furthermore, M_{ff} is restricted to be in a set consisting of all semi-positive definite matrices with all elements bounded in the interval $[-C, C]$.

For theoretical analysis, we further assume the following two assumptions.

Assumption E″: We assume

E″.1 Let $\delta_{ijts} = E(e_{it} e_{js})$, and we assume $\frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T |\delta_{ijts}| \leq C$.

E″.2 Let $\pi_1 = \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \frac{\delta_{ijts}}{w_i^2 w_j^2} (m_i \otimes f_t)(m_j' \otimes f_s')$, and assume $\lim_{N,T \to \infty} \pi_1 = \pi_{1\infty} > 0$; in other words, the limit of π_1 exits and is positive definite.

E″.3 Let $\pi_2 = \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \frac{\varrho_{ijts}}{w_i^4 w_j^4} (m_i \otimes m_i)(m_j' \otimes m_j')$ with $\varrho_{ijts} = E\Big[(e_{it}^2 - w_i^2)(e_{js}^2 - w_j^2)\Big]$. We assume $\lim_{N,T \to \infty} \pi_2 = \pi_{2\infty} > 0$.

E″.4 Let $\pi_3 = \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \frac{\vartheta_{ijts}}{w_i^2 w_j^4} (m_i \otimes f_t)(m_j' \otimes m_j')$ with $\vartheta_{ijts} = E\Big[e_{it}(e_{js}^2 - w_j^2)\Big]$. We assume $\lim_{N,T \to \infty} \pi_3 = \pi_{3\infty} > 0$.

E″.5 For each i, as $T \to \infty$, $\frac{1}{\sqrt{T}} \sum_{t=1}^T (e_{it}^2 - w_i^2) \xrightarrow{d} N(0, \varpi_{i\infty}^2)$, with $\varpi_{i\infty}^2 = \lim_{T \to \infty} \varpi_i^2$ and $\varpi_i^2 = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E\big[(e_{it}^2 - w_i^2)(e_{is}^2 - w_i^2)\big]$.

Assumption F″: We assume

F″.1 For all j, $E\left[\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \frac{m_i' \Lambda}{w_i^2} \big[e_{it} e_{jt} - E(e_{it} e_{jt})\big] \right\|^2\right] \leq C$.

F″.2 We assume $E\left[\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{t=1}^T \frac{m_i' \Lambda \Lambda' m_i}{w_i^2} \big(e_{it}^2 - w_i^2\big) \right\|^2\right] \leq C$.

F″.3 For all t, $E\left[\left\| \frac{1}{\sqrt{NT}} \sum_{i=1}^N \sum_{s=1}^T \frac{1}{w_i^2} f_s \big[e_{it} e_{is} - E(e_{it} e_{is})\big] \right\|^2\right] \leq C$.

F$''$.4 For all t, $E\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{\sqrt{T}}\sum_{s=1}^{N}f_s\big[e_{it}e_{is}-E(e_{it}e_{is})\big]\right\|^2\right]\leq C.$

F$''$.5 For all t, $E\left[\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{s=1}^{T}\frac{1}{w_i^4}m_i'\Lambda\big(e_{is}^2-w_i^2\big)e_{it}\right\|^2\right]\leq C.$

F$''$.6 We assume $E\left[\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{w_i^4}f_te_{it}\big(e_{is}^2-w_i^2\big)m_i'\right\|^2\right]\leq C.$

Assumption E$''$ is used in deriving the limiting distributions. Assumption F$''$ provides some moment conditions which are needed in inferential analysis.

To remove the rotational indeterminacy, the identification conditions considered here, which are denoted by IC$''$, are the same with those in Section 3.3 except that the matrix Σ_{ee} is replaced with $\mathbb{W}$.

Even that the model allows for general weak dependence among idiosyncratic errors, we still use (3.3.2) as the objective function to estimate the loadings and idiosyncratic variances, with Σ_{ee} replaced by $\mathbb{W}$. Now the parameter is $\theta=(\Lambda,\mathbb{W})$. As shown in Bai and Li (2016), although the objective function is misspecified, the consistency of the estimated loadings can be maintained if some regularity conditions are satisfied.

Let $\hat{\theta}=(\hat{\Lambda},\hat{\mathbb{W}})$ be the maximizer of the objective function. Then we can derive the first order conditions for Λ and $\mathbb{W}$, which are similar to (3.3.3) and (3.3.4), except that $\hat{\Sigma}_{ee}$ should be replaced by $\hat{\mathbb{W}}$. Based on these first order conditions, together with the similar arguments, we develop inferential theories under the weak dependence idiosyncratic errors. The following theorem presents the convergence rates of the MLE. The consistency is implied by the theorem.

Theorem 3.9.1 (Convergence rates)**.** *Under Assumptions A,B$''$, C$''$, D$''$ and F$''$, together* *with IC″, when $N, T \to \infty$, we have*

$$\hat{\Lambda} - \Lambda = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right), \qquad \frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2 = O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N^2}\right).$$

In contrast with the results in Theorem 3.4.1, we see that there is an extra term $O_p(\frac{1}{N})$ in $(\hat{\Lambda} - \Lambda)$ and another extra term $O_p(\frac{1}{N^2})$ in $\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2$ under the weak dependence data structure.

Before we state the asymptotic result of $\hat{\Lambda}$, below we first introduce some symbols.

$$\mathbb{D}_1^{\dagger} = \begin{bmatrix} 2D_r^{+} \\ \mathcal{D}[(\mathbb{P} \otimes I_r) + (I_r \otimes \mathbb{P})K_r] \end{bmatrix},$$

$$\mathbb{B}_1^{\dagger} = K_{kr}[(\mathbb{P}^{-1}\Lambda') \otimes \Lambda] + \mathbb{R}^{-1} \otimes I_r - K_{kr}(I_r \otimes \Lambda)(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_2[(\mathbb{P}^{-1}\Lambda') \otimes I_r],$$

$$\mathbb{B}_2^{\dagger} = K_{kr}(I_r \otimes \Lambda)(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)', \quad \mathbb{B}_3^{\dagger} = K_{kr}(I_r \otimes \Lambda)(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)',$$

$$\mathbb{B}_4^{\dagger} = \Big((\mathbb{R}^{-1}) \otimes (\mathbb{P}^{-1}\Lambda')\Big) - \frac{1}{2}K_{kr}(I_r \otimes \Lambda)(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_2(\mathbb{P} \otimes \mathbb{P})^{-1}(\Lambda \otimes \Lambda)',$$

$$\Delta^{\dagger} = \mathbb{B}_2^{\dagger}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\varpi_i^2}{w_i^6}(m_i \otimes m_i),$$

$$\Pi^{\dagger} = \mathbb{B}_4^{\dagger}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}\frac{\mathbb{O}_{ij}}{w_i^2 w_j^2}(m_j \otimes m_i) - \mathbb{B}_3^{\dagger}\frac{1}{N}\sum_{i=1}^{N}\frac{\varsigma_i}{w_i^4}(m_i \otimes m_i).$$

where $D_r^{+}, \mathcal{D}, K_r, K_{kr}, \mathbb{D}_2$ and $\mathbb{D}_3$ are defined the same as in Theorem 4.2; $\mathbb{P}$ and $\mathbb{R}$ are defined in Assumption C″; $\mathbb{O}_{ij}$ is the (i,j)th entry of matrix $\mathbb{O}$; $\varsigma_i = \frac{1}{N}m_i'\Lambda\mathbb{P}^{-1}\Lambda' M'\mathbb{W}^{-1}(\mathbb{O} - \mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}^{-1}\Lambda' m_i - 2m_i'\Lambda\mathbb{G}_N\Lambda' M'\mathbb{W}^{-1}(\mathbb{O} - \mathbb{W})_i$ where $\mathbb{G}_N = N\mathbb{G}$ with $\mathbb{G} = (I_r + \Lambda' M'\mathbb{W}^{-1}M\Lambda)^{-1}$ and $(\mathbb{O} - \mathbb{W})_i$ is the ith column of $(\mathbb{O} - \mathbb{W})$; $\varpi_i^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\big[(e_{it}^2 - w_i^2)(e_{is}^2 - w_i^2)\big]$ is defined in Assumption E″.5; both ς_i and ϖ_i^2 are scalars. Then we provide the asymptotic representation of $\hat{\Lambda}$ in the following theorem.

Theorem 3.9.2 (Asymptotic representation for $\hat{\Lambda}$)**.** *Under assumptions of Theorem 3.9.1,*

$$\operatorname{vec}(\hat{\Lambda}' - \Lambda') = \mathbb{B}_1^{\dagger} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{w_i^2} (m_i \otimes f_t) e_{it} - \mathbb{B}_2^{\dagger} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{w_i^4} (m_i \otimes m_i)(e_{it}^2 - w_i^2)$$
$$+ \frac{1}{T}\Delta^{\dagger} + \frac{1}{N}\Pi^{\dagger} + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right) + O_p\left(\frac{1}{N^2}\right), \tag{3.9.1}$$

where the symbols $\mathbb{B}_1^{\dagger}, \mathbb{B}_2^{\dagger}, \Delta^{\dagger}$ and $\Pi^{\dagger}$ are defined in the preceding paragraph.

Given the above theorem, we have the following corollary.

Corollary 3.9.1 (Limiting distribution for $\hat{\Lambda}$)**.** *Under assumptions of Theorem 3.9.1 and Assumption E$''$, as $N, T \to \infty, N/T^2 \to 0$ and $T/N^3 \to 0$, we have*

$$\sqrt{NT}\Big[\operatorname{vec}(\hat{\Lambda}' - \Lambda') - \frac{1}{T}\Delta^{\dagger} - \frac{1}{N}\Pi^{\dagger}\Big] \xrightarrow{d} N(0, \Xi),$$

where $\Xi = \lim_{N\to\infty} \Xi_{NT}$, and

$$\Xi_{NT} = \mathbb{B}_1^{\dagger}\pi_1\mathbb{B}_1^{\dagger\prime} + \mathbb{B}_2^{\dagger}\pi_2\mathbb{B}_2^{\dagger\prime} - \mathbb{B}_1^{\dagger}\pi_3\mathbb{B}_2^{\dagger\prime} - \mathbb{B}_2^{\dagger}\pi_3'\mathbb{B}_1^{\dagger\prime}$$

where $\mathbb{B}_1^{\dagger}$ and $\mathbb{B}_2^{\dagger}$ are defined the same as in Theorem 3.9.2; the symbols π_1, π_2 and π_3 are defined in Assumption E$''$. Furthermore, by Assumption E$''$.2, E$''$.3 and E$''$.4, we have

$$\Xi = \mathbb{B}_1^{\dagger}\pi_{1\infty}\mathbb{B}_1^{\dagger\prime} + \mathbb{B}_2^{\dagger}\pi_{2\infty}\mathbb{B}_2^{\dagger\prime} - \mathbb{B}_1^{\dagger}\pi_{3\infty}\mathbb{B}_2^{\dagger\prime} - \mathbb{B}_2^{\dagger}\pi_{3\infty}'\mathbb{B}_1^{\dagger\prime}.$$

where the symbols $\pi_{1\infty}, \pi_{2\infty}$ *and* $\pi_{3\infty}$ *are defined in Assumption* E''.

we also have the following theorem for w_i^2.

Theorem 3.9.3 (Asymptotic properties for $\hat{w}_i^2$). *Under assumptions of Theorem 3.9.1,*

$$\hat{w}_i^2 - w_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right).$$

As $N, T \to \infty$ *and* $T/N^2 \to 0$, *we have*

$$\sqrt{T}(\hat{w}_i^2 - w_i^2) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(e_{it}^2 - w_i^2) + o_p(1).$$

Furthermore, by Assumption $E''.5$, *we have*

$$\sqrt{T}(\hat{w}_i^2 - w_i^2) \xrightarrow{d} N(0, \varpi_{i\infty}^2),$$

where $\varpi_{i\infty}^2$ *is defined in Assumption* $E''.5$.

This limiting result is the same as that in the unconstrained approximate factor model, see Bai and Li (2016).

3.10 Conclusion

This chapter considers the ML estimation of large dimensional constrained factor models in which both cross sectional units (N) and time periods (T) are large but the number of loadings is fixed. We investigate the asymptotic properties of the MLE including consistency, convergence rates, asymptotic representations and limiting distributions. We show that the

MLE for the loadings in a constrained factor model converges much faster than that in a standard factor model. In addition, we also find that the MLE has a non-negligible bias asymptotically and some bias corrections are needed when conducting inference. A W statistic is proposed to conduct model specification check in a constrained factor model versus a standard factor model. The test is valid for a large N and a large T setup. We also analyze partially constrained factor models where only partial factor loadings are constrained. We run simulations to investigate the finite sample performance of the MLE and the proposed W test. The simulation results are encouraging and show that the MLE outperform the PC estimates and the proposed W test has good empirical sizes and powers. Monte Carlo simulations show that our proposed MLE has better finite sample performances than that of PC estimates. In addition, we consider the extension of a general weak dependence structure on idiosyncratic errors and we study MLE asymptotic properties of the resulting approximate factor models.

Bibliography

Ahn, S. C. and A. R. Horenstein (2013), "Eigenvalue ratio test for the number of factors." *Econometrica*, 81(3), 1203–1227.

Ahn, S.G., Y.H. Lee, and P. Schmidt (2013), "Panel data models with multiple time-varying effects." *Journal of Econometrics*, 174, 1–14.

Allers, M.A. and J.P. Elhorst (2011), "A simultaneous equations model of fiscal policy interactions." *Journal of Regional Science*, 51(2), 271–291.

Anderson, J.E. and E. van Wincoop (2003), "Gravity with gravitas: a solution to the border puzzle." *American Economic Review*, 93(1), 170–192.

Anderson, T.W. (2003), *Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Hoboken, New Jersey.

Anselin, L. (1988), *Spatial econometrics: methods and models*. Kluwer Academic Publishers, Netherlands.

Arezki, R., B. Candelon, and A.N.R. Sy (2011), "Sovereign rating news and financial markets spillovers: evidence from the european debt crisis." Manuscript.

Bai, J. (2003), "Inferential theory for factor models of large dimensions." *Econometrica*, 71(1), 135–171.

Bai, J. (2009), "Panel data models with interactive fixed effects." *Econometrica*, 77(4), 1229–1279.

Bai, J. and T. Ando (2015a), "Asset pricing with a general multifactor structure." *Journal of Financial Econometrics*, 13(3), 556–604.

Bai, J. and T. Ando (2015b), "A simple new test for slope homogeneity in panel data models with interactive effects." *Economic Letters*, 136, 112–117.

Bai, J. and T. Ando (2016), "Panel data models with grouped factor structure under unknown group membership." *Journal of Applied Econometrics*, 31(1), 163–191.

Bai, J. and K. Li (2012), "Statistical analysis of factor models of high dimension." *The Annals of Statistics*, 40(1), 436–465.

Bai, J. and K. Li (2014a), "Spatial panel data models with common shocks." Manuscript.

Bai, J. and K. Li (2014b), "Theory and methods of panel data models with interactive effects." *The Annals of Statistics*, 42(1), 142–170.

Bai, J. and K. Li (2015), "Dynamic spatial panel data models with common shocks." Manuscript.

Bai, J. and K. Li (2016), "Maximum likelihood estimation and inference of approximate factor mdoels of high dimension." *Review of Economics and Statistics*, 98(2), 298–309.

Bai, J. and S. Ng (2002), "Determining the number of factors in approximate factor models." *Econometrica*, 70(1), 191–221.

Bai, J. and S. Ng (2013), "Principal components estimation and identification of static factors." *Journal of Econometrics*, 176, 18–29.

Balestra, P. and J. Varadharajan-Krishnakumar (1987), "Full information estimations of a system of simultaneous equations with error component structure." *Econometric Theory*, 3(2), 223–246.

Baltagi, B.H. (1981), "Simultaneous equations with error components." *Journal of Econometrics*, 17(2), 603–624.

Baltagi, B.H. and G. Bresson (2011), "Maximum likelihood estimation and lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in paris." *Journal of Urban Economics*, 69(1), 24–42.

Baltagi, B.H. and Y. Deng (2015), "Ec3sls estimator for a simultaneous system of spatial autoregressive equations with random effects." *Econometric Reviews*, 34(6-10), 659–694.

Baltagi, B.H., P. Egger, and M. Pfaffermayr (2008), "Estimating regional trade agreement effects on fdi in an interdependent world." *Journal of Econometrics*, 145(1), 194–208.

Baltagi, B.H. and D. Li (2006), "Prediction in the panel data model with spatial correlation: the case of liquor." *Spatial Economic Analysis*, 1, 175–185.

Baltagi, B.H. and Q. Li (1992), "A note on the estimation of simultaneous equations with error components." *Econometric Theory*, 8(1), 113–119.

Bartlett, M.S. (1950), "Test of significance in factor analysis." *British Journal of Psychology (Statistics Section)*, 3, 77–85.

Bonhomme, S. and E. Manresa (2015), "Grouped patterns of heterogeneity in panel data." *Econometrica*, 83(3), 1147–1184.

Bradford, C.J. and N. Chakwin (1993), "Alternative explanations of the trade-output correlation in east asian economies." OECD Development Centre Technical Paper No.87.

Carhart, M. M. (1997), "On persistence of mutual fund performance." *Journal of Finance*, 52, 57–82.

Castagnetti, C., E. Rossi, and L. Trapani (2015), "Inference on factor structures in heterogeneous panels." *Journal of Econometrics*, 184, 145–157.

Chamberlain, G. (1977a), "Education, income and ability revisited." In *Latent Variables in Socio-Economic Models* (D.J. Aigner and editors A.S. Goldberger, eds.), 143–161, North-Holland, Amsterdam.

Chamberlain, G. (1977b), "An instrumental variable interpretation of identification in variance-components and mimic models." In *Kinometrics: Determinants of Social-Economic Success within and between Families* (editor P.Taubman, ed.), 235–254, North-Holland, Amsterdam.

Chamberlain, G. and Z. Griliches (1975), "Unobservables with a variance-components structure: ability, schooling and the economic success of brothers." *International Economic Review*, 16, 422–450.

Chamberlain, G. and M. Rothschild (1983), "Arbitrage, factor structure, and mean-variance analysis on large asset markets." *Econometrica*, 51(5), 1281–1304.

Chen, C.R., T.L. Steiner, and A.M. Whyte (2006), "Does stock option-based executive compensation induce risk-taking? an analysis of the banking industry." *Journal of Banking and Finance*, 30, 915–945.

Cohen-Cole, E., X. Liu, and Y. Zenou (2013), "Multivariate choice and identification of social interactions." CEPR Discussion Papers 9159.

Connor, G. and R.A. Korajczyk (1986), "Performance measurement with the arbitrage pricing theory : A new framework for analysis." *Journal of Financial Economics*, 15(3), 373–394.

Connor, G. and R.A. Korajczyk (1988), "Risk and return in an equilibrium apt: Application of a new test methodology." *Journal of Financial Economics*, 21, 255–289.

Connor, G. and O. Linton (2007), "Semiparametric estimation of a characteristic-based factor model of stock returns." *Journal of Empirical Finance*, 14, 694–717.

Connor, M. Hagmann, G. and O. Linton (2012), "Efficient semiparametric estimation of the fama-french model and extensions." *Econometrica*, 80(2), 713–754.

Cornwell, C., P. Schmidt, and D. Wyhowski (1992), "Simultaneous equations and panel data." *Journal of Econometrics*, 51, 151–181.

Defever, F., B. Heid, and M. Larch (2015), "Spatial exporters." *Journal of International Economics*, 95(1), 145–156.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977), "Maximum likelihood from incomplete data via the em algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.

Dewachter, H., R. Houssa, and P. Toffano (2012), "Spatial propagation of macroeconomic shocks in europe." *Review of World Economics*, 148(2), 377–402.

Doz, C., D. Giannone, and L. Reichlin (2012), "A quasi maximum likelihood approach for large approximate dynamic factor models." *Review of Economics and Statistics*, 94(4), 1014–1024.

Druska, V. and W.C. Horrace (2004), "Generalized moments estimation for spatial panel data: Indonesian rice farming." *American Journal of Agricultural Economics*, 86, 185–198.

Eberhardt, M., C. Helmers, and H. Strauss (2013), "Do spillovers matter when estimating private returns to r&d?" *Review of Economics and Statistics*, 95(2), 436–448.

Egger, P. and M. Pfaffermayr (2004), "Distance, trade and fdi: a hausman-taylor sur approach." *Journal of Applied Econometrics*, 19, 227–246.

Egger, P., M. Pfaffermayr, and H. Winner (2005), "An unbalanced spatial panel data approach to us state tax competition." *Economic Letters*, 88, 329–335.

Fama, E.F. and K.R. French (1993), "Common risk factors in the returns to stocks and bonds." *Journal of Financial Economics*, 33, 3–56.

Fan, J., Y. Liao, and M. Mincheva (2011), "High dimensional covariance matrix estimation in approximate factor models." *The Annals of Statistics*, 39(6), 3320–3356.

Fan, J., Y. Liao, and M. Mincheva (2013), "Large covariance estimation by thresholding principal orthogonal complements." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4), 603–680.

Fan, J., Liao Y., and Wang W. (2016), "Projected principal component analysis in factor models." *The Annals of Statistics*, 44(1), 219–254.

Fan, J., Liao Y., and Shi X. (2015), "Risks of large portfolios." *Journal of Econometrics*, 186(2), 367–387.

Felbermayr, G. and J. Groschl (2013), "Natural disasters and the effect of trade on income: a new panel approach." *European Economic Review*, 58, 18–30.

Feyrer, J. (2009), "Trade and income –exploiting time series in geography." NBER, working paper 14910.

Frankel, J. and D. Romer (1999), "Does trade cause growth?" *American Economic Review*, 89(3), 379–399.

Frazier, C. and K.M. Kockelman (2005), "Spatial econometric models for panel data: incorporating spatial and temporal data." *Transportation Research Record: Journal of the Transportation Research Board*, 1902(2005), 80–90.

Gebremariam, G.H., Gebremedhin T.G., and P.V. Schaeffer (2011), "Employment, income and migration in appalachia: a spatial simultaneous equations approach." *Journal of Regional Science*, 51(1), 102–120.

Geweke, John. (1977), "The dynamic factor analysis of economic time series." In *Latent Variables in Socio-Economic Models* (Dennis J. Aigner and eds. Arthur S. Goldberger, eds.), North-Holland, Amsterdam.

Goldsmith-Pinkham, P. and W.G. Imbens (2013), "Social networks and the identification of peer effects." *Journal of Business and Economic Statistics*, 31(3), 253–264.

Goyal, A. and P. Santa-Clara (2003), "Idiosyncratic risk matters." *Journal of Finance*, 58, 975–1008.

Griliches, Z. (1979), "Sibling models and data in economics: beginning of a survey." *Journal of Political Economy*, 87(Supplement2), S37–S64.

Hauptmeier, S., F. Mittermaier, and J. Rincke (2012), "Fiscal competition over taxes and public inputs." *Regional Science and Urban Economics*, 42(3), 407–419.

Heckman, J.J., J. Stixrud, and S. Urzua (2006), "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior." *Journal of Labor Economics*, 24(3), 411–482.

Helpman, E. (1987), "Imperfect competition and international trade: evidence from fourteen industrial countries." *Journal of the Japanese and International Economies*, 1, 62–81.

Helpman, E. (1988), "Growth, technological progress, and trade." NBER, reprint no. 1145.

Ho, C. and D. Hite (2008), "The benefit of environmental improvement in the southeastern united states: Evidence from a simultaneous model of cancer mortality, toxic chemical releases and house values." *Papers in Regional Science*, 87(4), 589–604.

Irwin, D.A. and M. Tervio (2002), "Does trade raise income? evidence from the twentieth century." *Journal of International Economics*, 58(1), 1–18.

Jain, N., K. Girotra, and S. Netessine (2013), "Managing global sourcing: inventory performance." *Management Science*, 60(5), 1202–1222.

Jeanty, P.W., M. Partridge, and E. Irwin (2010), "Estimation of a spatial simultaneous equation model of population migration and housing price dynamics." *Regional Science and Urban Economics*, 40, 343–352.

Jennrich, R. I. (1969), "Asymptotic properties of non-linear least squares estimators." *The Annals of Mathematical Statistics*, 40(2), 633–643.

Kapetanios, G., M. H. Pesaran, and T. Yamagata (2011), "Panels with non-stationary multifactor error structures." *Journal of Econometrics*, 160(2), 326–348.

Kelejian, H. H. and I. R. Prucha (1998), "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances." *The Journal of Real Estate Finance and Economics*, 17(1), 99–121.

Kelejian, H. H. and I. R. Prucha (1999), “A generalized moments estimator for the autoregressive parameter in a spatial model.” *International Economic Review*, 40(2), 509–533.

Kelejian, H. H. and I. R. Prucha (2004), “Estimation of simultaneous systems of spatially interrelated cross sectional equations.” *Journal of Econometrics*, 118, 27–50.

Kelejian, H. H. and I. R. Prucha (2010), “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances.” *Journal of Econometrics*, 157(1), 53–67.

Keller, W. and C. Shiue (2007), “The origin of spatial interaction.” *Journal of Econometrics*, 140(1), 304–332.

Kneip, A., R.C. Sickles, and W. Song (2012), “A new panel data treatment for heterogeneity in time trends.” *Econometric Theory*, 28, 590–628.

Korte, J. and S. Steffen (2014), “Zero risk contagion -banks' sovereign exposure and sovereign risk spillovers.” Manuscript.

Kose, A., C. Otrok, and C. Whiteman (2003), “International business cycles: world region and country specific factors.” *American Economic Review*, 93(4), 1216–1239.

Kuersteiner, G.M. and I.R. Prucha (2015), “Dynamic spatial panel models: networks, common shocks, and sequential exogeneity.” CESIFO working paper no. 5445.

Lawless, M. (2009), “Firm export dynamics and the geography of trade.” *Journal of International Economics*, 77(2), 245–254.

Lawley, D. N. and A. E. Maxwell (1971), *Factor analysis as a statitical method*. Ameriacan Elsevier, New York.

Lee, L. (2004), "Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models." *Econometrica*, 72(6), 1899–1925.

Lee, L. and J. Yu (2010a), "Estimation of spatial autoregressive panel data models with fixed effects." *Journal of Econometrics*, 154(2), 165–185.

Lee, L. and J. Yu (2010b), "A spatial dynamic panel data model with both time and individual fixed effects." *Econometric Theory*, 26, 564–597.

Lin, X. and L. F. Lee (2010), "Gmm estimation of spatial autoregressive models with unknown heteroskedasticity." *Journal of Econometrics*, 157(1), 34–52.

Liu, C. and D. B. Rubin (1994), "The ecme algorithm: a simple extension of em and ecm with faster monotone convergence." *Biometrika*, 81(4), 633–648.

Liu, X. (2014), "Identification and efficient estimation of simultaneous equations network models." *Journal of Business and Economic Statistics*, 32(4), 516–536.

Matteson, D.S., M.W. McLean, D.B. Woodard, and S.G. Henderson (2011), "Forecasting emergency medical service call arrival rates." *Annals of Applied Statistics*, 5, 1379–1406.

McLachlan, G. J. and T. Krishnan (1997), *The EM Algorithm and Extensions*. Wiley, New York.

Moon, H., M. Shum, and M. Weidner (2014), "Interactive fixed effects in the blp random coefficients demand model." R&R Journal of Econometrics.

Moon, H. and M. Weidner (2015), "Linear regression for panel with unknown number of factors as interactive fixed effect." *Econometrica*, 83(4), 1543–1579.

Moon, H. and M. Weidner (2017), "Dynamic linear panel regression models with interactive fixed effects." *Econometric Theory*, 33(1), 158–195.

Newey, W. and D. McFadden (1994), "Large sample estimation and hypothesis testing." In *Handbook of Econometrics* (R.F. Engle and D. McFadden (eds.), eds.), North-Holland, Amsterdam.

Noguer, M. and M. Siscart (2005), "Trade raises income a precise and robust result." *Journal of International Economics*, 65(2), 447–460.

Onatski, A. (2009), "Testing hypotheses about the number of factors in large factor models." *Econometrica*, 77(5), 1447–1479.

Onatski, A. (2010), "Determining the number of factors from empirical distribution of eigenvalues." *Review of Economics and Statistics*, 92(4), 1004–1016.

Ortega, F. and G. Peri (2014), "Openness and income: The roles of trade and migration." *Journal of International Economics*, 92, 231–251.

Ozdagli, A. and M. Weber (2016), "Monetary policy through production networks: evidence from the stock market." Manuscript.

Pesaran, M. H. (2006), "Estimation and inference in large heterogeneous panels with a multifactor error structure." *Econometrica*, 74(4), 967–1012.

Pesaran, M. H. and E. Tosetti (2011), “Large panels with spatial correlations and common factors.” *Journal of Econometrics*, 161, 182–202.

Rauch, J.E. and V. Trindade (2002), “Ethnic chinese networks in international trade.” *Review of Economics and Statistics*, 84(1), 116–130.

Rodrik, D. (1995), “Getting interventions right: how south korea and taiwan grew rich.” *Economic Policy*, 20, 53–97.

Rosenberg, B. (1974), “Extra-market components of covariance in security prices.” *Journal of Financial and Quantitative Analysis*, 9, 263–274.

Ross, S. A. (1976), “The arbitrage theory of capital asset pricing.” *Journal of Economic Theory*, 13(3), 341–360.

Sargent, T. J. and C. A. Sims (1977), “Business cycle modeling without pretending to have too much a priori economic theory.” In *New methods in business cycle research* (eds. C. Sims et al., ed.), Federal Reserve Bank of Minneapolis, Minneapolis.

Serlenga, L. and Y. Shin (2007), “Gravity models of intra-eu trade: application of the ccep-ht estimation in heterogeneous panels with unobserved common time-specific factors.” *Journal of Applied Econometrics*, 22, 361–381.

Song, M. (2013), “Asymptotic theory for dynamic heterogeneous panels with cross-sectional dependence and its applications.” Manuscipt.

Stock, J. H. and M. W. Watson (1998), “Diffusion indexes.” NBER, working paper, No. w6702.

Stock, J.H. and M.W. Watson (2002a), "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association*, 97, 1167–1179.

Stock, J.H. and M.W. Watson (2002b), "Macroeconomic forecasting using diffusion indexes." *Journal of Business and Economic Statistics*, 20, 147–162.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society, Series B (Methodological)*, 58, 267–288.

Tsai, H. and R. Tsay (2010), "Constrained factor models." *Journal of the American Statistical Association*, 105(492), 1593–1605.

Wang, L., K. Li, and Z. Wang (2014), "Quasi maximum likelihood estimation for simultaneous spatial autoregressive models." Manuscript.

Winters, L.A. (2004), "Trade liberalisation and economic performance: an overview." *The Economic Journal*, 114, F4–F21.

Winters, L.A. and A. Masters (2013), "Openness and growth: still an open question?" *Journal of International Development*, 25, 1061–1070.

Wu, C. (1981), "Asymptotic theory of nonlinear least squares estimation." *The Annals of Statistics*, 9(3), 501–513.

Wu, J. and H. Lin (2010), "The effect of the conservation reserve program on land values." *Land Economics*, 86, 1–21.

Yang, K. and L. Lee (2017), "Identification and qml estimation of multivariate and simultaneous spatial autoregressive models." *Journal of Econometrics*, 196(1), 196–214.

Yu, J., R. de Jong, and L. Lee (2008), "Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large." *Journal of Econometrics*, 146(1), 118–134.

Yu, J. and L. Lee (2010), "Estimation of unit root spatial dynamic panel data models." *Econometric Theory*, 26, 1332–1362.

Zhou, Z. and D.S. Matteson (2015), "Temporal and spatio-temporal models for ambulance demand." In *Healthcare Data Analytics* (H.Yang and eds. E.K. Lee, eds.), Wiley Series in Operations Research and Management Science, Hoboken, New Jersey.

Zou, H. (2006), "The adaptive lasso and its oracle properties." *Journal of the American Statistical Association*, 101, 1418–1429.

Appendix A

Appendix for Chapter 1

A.1 Notation I

In this Appendix, I define the important notation used in this paper. Table A.11 includes the definitions of important matrices.

Table A.11: Some important symbols used in the paper

$Y_{1t} = (y_{11t}, \ldots, y_{1Nt})'$	$Y_{2t} = (y_{21t}, \ldots, y_{2Nt})'$
$\ddot{Y}_{1t} = (\ddot{y}_{11t}, \ldots, \ddot{y}_{1Nt})'$	$\ddot{Y}_{2t} = (\ddot{y}_{21t}, \ldots, \ddot{y}_{2Nt})'$
$\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$	$\Psi = (\psi_1, \psi_2, \ldots, \psi_N)'$
$\Phi_{1p} = (\phi_{1p}, \phi_{2p}, \ldots, \phi_{Np})'$	$\Phi_{2p} = (\phi_{1p}, \phi_{2p}, \ldots, \phi_{Np})'$
$e_{1t} = (e_{11t}, \ldots, e_{1Nt})'$	$e_{2t} = (e_{21t}, \ldots, e_{2Nt})'$
$V_{1t} = (v_{11t}, \ldots, v_{1Nt})'$	$V_{2t} = (v_{21t}, \ldots, v_{2Nt})'$
$\Sigma_{1ee} = \text{diag}(\sigma_{11}^2, \ldots, \sigma_{1N}^2)$	$\Sigma_{2ee} = \text{diag}(\sigma_{21}^2, \ldots, \sigma_{2N}^2)$
$\Delta_{11} = \text{diag}(\sigma_{11}^2 + \beta_1'\Sigma_{11v1}\beta_1, \ldots, \sigma_{1N}^2 + \beta_1'\Sigma_{NNv1}\beta_1)$ $\Delta_{22} = \text{diag}(\sigma_{21}^2 + \beta_2'\Sigma_{11v2}\beta_2, \ldots, \sigma_{2N}^2 + \beta_2'\Sigma_{NNv2}\beta_2)$	
$P_1 = I_N - \rho_1 W_1$	$P_2 = I_N - \rho_2 W_2$
$B_{12} = (I_N - \gamma_1\gamma_2 P_1^{-1} P_2^{-1})^{-1}$	$B_{21} = (I_N - \gamma_1\gamma_2 P_2^{-1} P_1^{-1})^{-1}$
$\mathbb{G}_1 = W_1 B_{12} P_1^{-1}$	$\mathbb{G}_2 = W_2 B_{21} P_2^{-1}$
$\mathbb{G}_3 = \gamma_2 B_{21} P_2^{-1} P_1^{-1}$	$\mathbb{G}_4 = \gamma_1 B_{12} P_1^{-1} P_2^{-1}$

Table A.12 presents the definitions of scalars $a_1, a_2, b_1, b_2, c_1, c_2$ and matrices M_i $(i = 1, 2, 3, 4)$ used in Assumption G.

Table A.12: The definition of $a_1, a_2, b_1, b_2, c_1, c_2$ as in Section 1.2

$a_1 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$[\mathbb{G}_{1,ij}]^2 + [(W_1\mathbb{G}_4)_{ij}]^2$
$a_2 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$[(W_2\mathbb{G}_3)_{ij}]^2 + [\mathbb{G}_{2,ij}]^2$
$b_1 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$\mathbb{G}_{1,ij}\mathbb{G}_{3,ij} + (W_1\mathbb{G}_4)_{ij}(B_{21}P_2^{-1})_{ij}$
$b_2 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$(W_2\mathbb{G}_3)_{ij}(B_{12}P_1^{-1})_{ij} + \mathbb{G}_{2,ij}\mathbb{G}_{4,ij}$
$c_1 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$[\mathbb{G}_{3,ij}]^2 + [(B_{21}P_2^{-1})_{ij}]^2$
$c_2 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}$	$[(B_{12}P_1^{-1})_{ij}]^2 + [\mathbb{G}_{4,ij}]^2$
$M_1 = -W_1\mathbb{G}_4\Sigma_{2ee} + \mathbb{G}_1\Sigma_{1ee}\mathbb{G}_3'W_2'(\rho_2^{\dagger} - \rho_2) + \mathbb{G}_1\Sigma_{1ee}P_1^{-1\prime}B_{12}'(\gamma_2^{\dagger} - \gamma_2)$	
$M_2 = -\Sigma_{1ee}\mathbb{G}_3W_2' + W_1\mathbb{G}_4\Sigma_{2ee}\mathbb{G}_2'(\rho_1^{\dagger} - \rho_1) + B_{21}P_2^{-1}\Sigma_{2ee}\mathbb{G}_2'(\gamma_1^{\dagger} - \gamma_1)$	
$M_3 = -B_{21}P_2^{-1}\Sigma_{2ee} + \mathbb{G}_3\Sigma_{1ee}\mathbb{G}_3'W_2'(\rho_2^{\dagger} - \rho_2) + \mathbb{G}_3\Sigma_{1ee}P_1^{-1\prime}B_{12}'(\gamma_2^{\dagger} - \gamma_2)$	
$M_4 = -\Sigma_{1ee}P_1^{-1\prime}B_{12}' + W_1\mathbb{G}_4\Sigma_{2ee}\mathbb{G}_4'(\rho_1^{\dagger} - \rho_1) + B_{21}P_2^{-1}\Sigma_{2ee}\mathbb{G}_4'(\gamma_1^{\dagger} - \gamma_1)$	

A.2 Notation II

In this Appendix, I first introduce the definitions of $\mathbb{D}_b, \mathbb{D}_c$ and ζ, which are involved in Assumption G$'$, and then the definitions of the $\mathbb{D}_\eta, \Phi, \vartheta$ and $\mathbb{D}_\beta$ matrices involved in Theorem 1.3.2. All these matrices are used in the IGPC approach.

Some matrices in Assumption G$'$: $\mathbb{D}_b, \mathbb{D}_c$ and ζ.

For $p = 1, \ldots, k_1$, let $\widetilde{X}_p = (\widetilde{X}_{1p}, \ldots, \widetilde{X}_{Tp})$ be a $2N \times T$ matrix, where $\widetilde{X}_{tp} = (\dot{x}_{11t,p}, 0, \dot{x}_{12t,p}, 0, \ldots, \dot{x}_{1Nt,p}, 0)'$, with $\dot{x}_{1it,p}$ being the de-meaned version of $x_{1it,p}$ defined as $\dot{x}_{1it,p} = x_{1it,p} - \frac{1}{T}\sum_{s=1}^{T} x_{1is,p}$. For $p = k_1+1, \ldots, k$ with $k = k_1+k_2$, let $\widetilde{X}_p = (\widetilde{X}_{1p}, \ldots, \widetilde{X}_{Tp})$ be a $2N \times T$ matrix,where $\widetilde{X}_{tp} = (0, \dot{x}_{21t,(p-k_1)}, 0, \dot{x}_{22t,(p-k_1)}, \ldots, 0, \dot{x}_{2Nt,(p-k_1)})'$ with $\dot{x}_{2it,p}$ is defined as $\dot{x}_{1it,p}$. For $p = (k+1), \ldots, (k+4)$, let $\widetilde{X}_p = Q_{p-k}\Big(\sum_{q=1}^{k}\widetilde{X}_q\tilde{\beta}_q\Big)$, where $\tilde{\beta}_l = \beta_{1l}$ for $l = 1, \ldots, k_1$ and $\tilde{\beta}_l = \beta_{2,l-k_1}$ for $l = k_1+1, \ldots, k$. Each Q_l $(l = 1, \ldots, 4)$ is a $2N \times 2N$ matrix, defined as in Theorem 1.3.2, with its specification presented in Table A.17 in Appendix A.1.

Tables A.13 and A.14 list the definitions of Ω and ε, which are contained in Theorem 1.2.2.

Table A.13: The detailed definition of Ω

$\Omega_{11} = \frac{1}{N}\mathrm{tr}\left[\Sigma_{1ee}^{-1}\left(\mathbb{G}_1\Delta_{11}\mathbb{G}_1' + W_1\mathbb{G}_4\Delta_{22}\mathbb{G}_4'W_1'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^N(\mathbb{G}_{1,ii})^2\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_1^2\right]$	
$\Omega_{13} = \frac{1}{N}\mathrm{tr}\left[\Sigma_{1ee}^{-1}\left(\mathbb{G}_1\Delta_{11}\mathbb{G}_3' + W_1\mathbb{G}_4\Delta_{22}P_2^{-1\prime}B_{21}'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^N \mathbb{G}_{1,ii}\mathbb{G}_{3,ii}\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_3\mathbb{G}_1\right]$	
$\Omega_{12} = \frac{1}{N}\mathrm{tr}\left[W_2\mathbb{G}_3W_1\mathbb{G}_4\right]$	$\Omega_{14} = \frac{1}{N}\mathrm{tr}\left[B_{12}P_1^{-1}W_1\mathbb{G}_4\right]$
$\Omega_{15} = \beta_1'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{1,ii}\Sigma_{iiv1}/\sigma_{1i}^2\right]$	$\Omega_{16} = 0_{1\times k_2}$
$\Omega_{22} = \frac{1}{N}\mathrm{tr}\left[\Sigma_{2ee}^{-1}\left(W_2\mathbb{G}_3\Delta_{11}\mathbb{G}_3'W_2' + \mathbb{G}_2\Delta_{22}\mathbb{G}_2'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^N(\mathbb{G}_{2,ii})^2\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_2^2\right]$	
$\Omega_{24} = \frac{1}{N}\mathrm{tr}\left[\Sigma_{2ee}^{-1}\left(W_2\mathbb{G}_3\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_2\Delta_{22}\mathbb{G}_4'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^N \mathbb{G}_{2,ii}\mathbb{G}_{4,ii}\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_4\mathbb{G}_2\right]$	
$\Omega_{21} = \frac{1}{N}\mathrm{tr}\left[W_1\mathbb{G}_4W_2\mathbb{G}_3\right]$	$\Omega_{23} = \frac{1}{N}\mathrm{tr}\left[B_{21}P_2^{-1}W_2\mathbb{G}_3\right]$
$\Omega_{25} = 0_{1\times k_1}$	$\Omega_{26} = \beta_2'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{2,ii}\Sigma_{iiv2}/\sigma_{2i}^2\right]$
$\Omega_{31} = \frac{1}{N}\mathrm{tr}\left[\Sigma_{1ee}^{-1\prime}\left(\mathbb{G}_3\Delta_{11}\mathbb{G}_1' + B_{21}P_2^{-1}\Delta_{22}\mathbb{G}_4'W_1'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^N \mathbb{G}_{1,ii}\mathbb{G}_{3,ii}\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_1\mathbb{G}_3\right]$	
$\Omega_{33} = \frac{1}{N}\mathrm{tr}\left[\Sigma_{1ee}^{-1\prime}\left(\mathbb{G}_3\Delta_{11}\mathbb{G}_3' + B_{21}P_2^{-1}\Delta_{22}P_2^{-1\prime}B_{21}'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^N(\mathbb{G}_{3,ii})^2\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_3^2\right]$	
$\Omega_{32} = \frac{1}{N}\mathrm{tr}\left[W_2\mathbb{G}_3B_{21}P_2^{-1}\right]$	$\Omega_{34} = \frac{1}{N}\mathrm{tr}\left[B_{12}P_1^{-1}B_{21}P_2^{-1}\right]$
$\Omega_{35} = \beta_1'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{3,ii}\Sigma_{iiv1}/\sigma_{1i}^2\right]$	$\Omega_{36} = 0_{1\times k_2}$
$\Omega_{42} = \frac{1}{N}\mathrm{tr}\left[\Sigma_{2ee}^{-1}\left(W_2\mathbb{G}_3\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_2\Delta_{22}\mathbb{G}_4'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^N \mathbb{G}_{2,ii}\mathbb{G}_{4,ii}\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_2\mathbb{G}_4\right]$	
$\Omega_{44} = \frac{1}{N}\mathrm{tr}\left[\Sigma_{2ee}^{-1}\left(B_{12}P_1^{-1}\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_4\Delta_{22}\mathbb{G}_4'\right)\right] - \left[\frac{2}{N}\sum_{i=1}^N(\mathbb{G}_{4,ii})^2\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_4^2\right]$	
$\Omega_{41} = \frac{1}{N}\mathrm{tr}\left[W_1\mathbb{G}_4B_{12}P_1^{-1}\right]$	$\Omega_{43} = \frac{1}{N}\mathrm{tr}\left[B_{21}P_2^{-1}B_{12}P_1^{-1}\right]$
$\Omega_{45} = 0_{1\times k_1}$	$\Omega_{46} = \beta_2'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{4,ii}\Sigma_{iiv2}/\sigma_{2i}^2\right]$
$\Omega_{51} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{1,ii}\Sigma_{iiv1}/\sigma_{1i}^2\right]\beta_1$	$\Omega_{52} = 0_{k_1\times 1}$
$\Omega_{53} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{3,ii}\Sigma_{iiv1}/\sigma_{1i}^2\right]\beta_1$	$\Omega_{54} = 0_{k_1\times 1}$
$\Omega_{55} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iiv1}/\sigma_{1i}^2$	$\Omega_{56} = 0_{k_1\times k_2}$
$\Omega_{62} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{2,ii}\Sigma_{iiv2}/\sigma_{2i}^2\right]\beta_2$	$\Omega_{61} = 0_{k_2\times 1}$
$\Omega_{64} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{4,ii}\Sigma_{iiv2}/\sigma_{2i}^2\right]\beta_2$	$\Omega_{63} = 0_{k_2\times 1}$
$\Omega_{66} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iiv2}/\sigma_{2i}^2$	$\Omega_{65} = 0_{k_2\times k_1}$

Table A.14: The detailed definition of $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_6)'$

$\varepsilon_1 = \frac{1}{NT}\sum_{t=1}^T \beta_1' V_{1t}' \mathbb{G}_1' \Sigma_{1ee}^{-1} e_{1t} + \frac{1}{NT}\sum_{t=1}^T \beta_2' V_{2t}' \mathbb{G}_4' W_1' \Sigma_{1ee}^{-1} e_{1t}$
$+\frac{1}{NT}\sum_{t=1}^T e_{2t}' \mathbb{G}_4' W_1' \Sigma_{1ee}^{-1} e_{1t} + \frac{1}{NT}\sum_{t=1}^T e_{1t}' \mathbb{G}_1^{o\prime} \Sigma_{1ee}^{-1} e_{1t}$
$\varepsilon_2 = \frac{1}{NT}\sum_{t=1}^T \beta_2' V_{2t}' \mathbb{G}_2' \Sigma_{2ee}^{-1} e_{2t} + \frac{1}{NT}\sum_{t=1}^T \beta_1' V_{1t}' \mathbb{G}_3' W_2' \Sigma_{2ee}^{-1} e_{2t}$
$+\frac{1}{NT}\sum_{t=1}^T e_{1t}' \mathbb{G}_3' W_2' \Sigma_{2ee}^{-1} e_{2t} + \frac{1}{NT}\sum_{t=1}^T e_{2t}' \mathbb{G}_2^{o\prime} \Sigma_{2ee}^{-1} e_{2t}$
$\varepsilon_3 = \frac{1}{NT}\sum_{t=1}^T \beta_2' V_{2t}' P_2^{-1\prime} B_{21}' \Sigma_{1ee}^{-1} e_{1t} + \frac{1}{NT}\sum_{t=1}^T \beta_1' V_{1t}' \mathbb{G}_3' \Sigma_{1ee}^{-1} e_{1t}$
$+\frac{1}{NT}\sum_{t=1}^T e_{2t}' P_2^{-1\prime} B_{21}' \Sigma_{1ee}^{-1} e_{1t} + \frac{1}{NT}\sum_{t=1}^T e_{1t}' \mathbb{G}_3^{o\prime} \Sigma_{1ee}^{-1} e_{1t}$
$\varepsilon_4 = \frac{1}{NT}\sum_{t=1}^T \beta_1' V_{1t}' P_1^{-1\prime} B_{12}' \Sigma_{2ee}^{-1} e_{2t} + \frac{1}{NT}\sum_{t=1}^T \beta_2' V_{2t}' \mathbb{G}_4' \Sigma_{2ee}^{-1} e_{2t}$
$+\frac{1}{NT}\sum_{t=1}^T e_{1t}' P_1^{-1\prime} B_{12}' \Sigma_{2ee}^{-1} e_{2t} + \frac{1}{NT}\sum_{t=1}^T e_{2t}' \mathbb{G}_4^{o\prime} \Sigma_{2ee}^{-1} e_{2t}$
$\varepsilon_5 = \frac{1}{NT}\sum_{t=1}^T V_{1t}' \Sigma_{1ee}^{-1} e_{1t}$
$\varepsilon_6 = \frac{1}{NT}\sum_{t=1}^T V_{2t}' \Sigma_{2ee}^{-1} e_{2t}$

where $\mathbb{G}_p^o = \mathbb{G}_p - \mathbb{G}_p^d$, with $\mathbb{G}_p^d$ being a diagonal matrix whose diagonal elements equal the diagonal elements of $\mathbb{G}_p$ for $p = 1, 2, 3, 4$.

Tables A.16 and A.15 provide the definitions of $\tilde{\Omega}$ and $\tilde{\varepsilon}$, which are involved in Remark 2.15.

Table A.15: The detailed definition of $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots, \tilde{\varepsilon}_6)'$

$\tilde{\varepsilon}_1 = \frac{1}{NT\sigma_1^2}\sum_{t=1}^T \beta_1' V_{1t}' \mathbb{G}_1' e_{1t} + \frac{1}{NT\sigma_1^2}\sum_{t=1}^T \beta_2' V_{2t}' \mathbb{G}_4' W_1' e_{1t}$
$+\frac{1}{NT\sigma_1^2}\sum_{t=1}^T e_{2t}' \mathbb{G}_4' W_1' e_{1t} + \frac{1}{NT\sigma_1^2}\left[\sum_{t=1}^T e_{1t}' \mathbb{G}_1^{o\prime} e_{1t} + \omega_1\right]$
$\tilde{\varepsilon}_2 = \frac{1}{NT\sigma_2^2}\sum_{t=1}^T \beta_2' V_{2t}' \mathbb{G}_2' e_{2t} + \frac{1}{NT\sigma_2^2}\sum_{t=1}^T \beta_1' V_{1t}' \mathbb{G}_3' W_2' e_{2t}$
$+\frac{1}{NT\sigma_2^2}\sum_{t=1}^T e_{1t}' \mathbb{G}_3' W_2' e_{2t} + \frac{1}{NT\sigma_2^2}\left[\sum_{t=1}^T e_{2t}' \mathbb{G}_2^{o\prime} e_{2t} + \omega_2\right]$
$\tilde{\varepsilon}_3 = \frac{1}{NT\sigma_1^2}\sum_{t=1}^T \beta_2' V_{2t}' P_2^{-1\prime} B_{21}' e_{1t} + \frac{1}{NT\sigma_1^2}\sum_{t=1}^T \beta_1' V_{1t}' \mathbb{G}_3' e_{1t}$
$+\frac{1}{NT\sigma_1^2}\sum_{t=1}^T e_{2t}' P_2^{-1\prime} B_{21}' e_{1t} + \frac{1}{NT\sigma_1^2}\left[\sum_{t=1}^T e_{1t}' \mathbb{G}_3^{o\prime} e_{1t} + \omega_3\right]$
$\tilde{\varepsilon}_4 = \frac{1}{NT\sigma_2^2}\sum_{t=1}^T \beta_1' V_{1t}' P_1^{-1\prime} B_{12}' e_{2t} + \frac{1}{NT\sigma_2^2}\sum_{t=1}^T \beta_2' V_{2t}' \mathbb{G}_4' e_{2t}$
$+\frac{1}{NT\sigma_2^2}\sum_{t=1}^T e_{1t}' P_1^{-1\prime} B_{12}' e_{2t} + \frac{1}{NT\sigma_2^2}\left[\sum_{t=1}^T e_{2t}' \mathbb{G}_4^{o\prime} e_{2t} + \omega_4\right]$
$\tilde{\varepsilon}_5 = \frac{1}{NT}\sum_{t=1}^T V_{1t}' \Sigma_{1ee}^{-1} e_{1t}$
$\tilde{\varepsilon}_6 = \frac{1}{NT}\sum_{t=1}^T V_{2t}' \Sigma_{2ee}^{-1} e_{2t}$

where
$\omega_1 = \sum_{i=1}^N \sum_{t=1}^T \left[\mathbb{G}_{1,ii} - \frac{1}{N}\mathrm{tr}(\mathbb{G}_1)\right](e_{1it}^2 - \sigma_1^2)$, $\omega_2 = \sum_{i=1}^N \sum_{t=1}^T \left[\mathbb{G}_{2,ii} - \frac{1}{N}\mathrm{tr}(\mathbb{G}_2)\right](e_{2it}^2 - \sigma_2^2)$
$\omega_3 = \sum_{i=1}^N \sum_{t=1}^T \left[\mathbb{G}_{3,ii} - \frac{1}{N}\mathrm{tr}(\mathbb{G}_3)\right](e_{1it}^2 - \sigma_1^2)$, $\omega_4 = \sum_{i=1}^N \sum_{t=1}^T \left[\mathbb{G}_{4,ii} - \frac{1}{N}\mathrm{tr}(\mathbb{G}_4)\right](e_{2it}^2 - \sigma_2^2)$

Table A.16: The detailed definition of $\tilde{\Omega}$

$\tilde{\Omega}_{11} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\left(\mathbb{G}_1\Delta_{11}\mathbb{G}_1' + W_1\mathbb{G}_4\Delta_{22}\mathbb{G}_4'W_1'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_1)\right]^2 + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_1^2\right]$	
$\tilde{\Omega}_{13} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\left(\mathbb{G}_1\Delta_{11}\mathbb{G}_3' + W_1\mathbb{G}_4\Delta_{22}P_2^{-1\prime}B_{21}'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_1)\mathrm{tr}(\mathbb{G}_3)\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_3\mathbb{G}_1\right]$	
$\tilde{\Omega}_{12} = \frac{1}{N}\mathrm{tr}\left[W_2\mathbb{G}_3W_1\mathbb{G}_4\right]$	$\tilde{\Omega}_{14} = \frac{1}{N}\mathrm{tr}\left[B_{12}P_1^{-1}W_1\mathbb{G}_4\right]$
$\tilde{\Omega}_{15} = \beta_1'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{1,ii}\Sigma_{iiv1}/\sigma_1^2\right]$	$\tilde{\Omega}_{16} = 0_{1\times k_2}$
$\tilde{\Omega}_{22} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_2^2}\left(W_2\mathbb{G}_3\Delta_{11}\mathbb{G}_3'W_2' + \mathbb{G}_2\Delta_{22}\mathbb{G}_2'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_2)\right]^2 + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_2^2\right]$	
$\tilde{\Omega}_{24} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_2^2}\left(W_2\mathbb{G}_3\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_2\Delta_{22}\mathbb{G}_4'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_2)\mathrm{tr}(\mathbb{G}_4)\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_4\mathbb{G}_2\right]$	
$\tilde{\Omega}_{21} = \frac{1}{N}\mathrm{tr}\left[W_1\mathbb{G}_4W_2\mathbb{G}_3\right]$	$\tilde{\Omega}_{23} = \frac{1}{N}\mathrm{tr}\left[B_{21}P_2^{-1}W_2\mathbb{G}_3\right]$
$\tilde{\Omega}_{25} = 0_{1\times k_1}$	$\tilde{\Omega}_{26} = \beta_2'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{2,ii}\Sigma_{iiv2}/\sigma_2^2\right]$
$\tilde{\Omega}_{31} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\left(\mathbb{G}_3\Delta_{11}\mathbb{G}_1' + B_{21}P_2^{-1}\Delta_{22}\mathbb{G}_4'W_1'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_1)\mathrm{tr}(\mathbb{G}_3)\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_1\mathbb{G}_3\right]$	
$\tilde{\Omega}_{33} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_1^2}\left(\mathbb{G}_3\Delta_{11}\mathbb{G}_3' + B_{21}P_2^{-1}\Delta_{22}P_2^{-1\prime}B_{21}'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_3)\right]^2 + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_3^2\right]$	
$\tilde{\Omega}_{32} = \frac{1}{N}\mathrm{tr}\left[W_2\mathbb{G}_3B_{21}P_2^{-1}\right]$	$\tilde{\Omega}_{34} = \frac{1}{N}\mathrm{tr}\left[B_{12}P_1^{-1}B_{21}P_2^{-1}\right]$
$\tilde{\Omega}_{35} = \beta_1'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{3,ii}\Sigma_{iiv1}/\sigma_1^2\right]$	$\tilde{\Omega}_{36} = 0_{1\times k_2}$
$\tilde{\Omega}_{42} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_2^2}\left(W_2\mathbb{G}_3\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_2\Delta_{22}\mathbb{G}_4'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_2)\mathrm{tr}(\mathbb{G}_4)\right] + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_2\mathbb{G}_4\right]$	
$\tilde{\Omega}_{44} = \frac{1}{N}\mathrm{tr}\left[\frac{1}{\sigma_2^2}\left(B_{12}P_1^{-1}\Delta_{11}P_1^{-1\prime}B_{12}' + \mathbb{G}_4\Delta_{22}\mathbb{G}_4'\right)\right] - \frac{2}{N^2}\left[\mathrm{tr}(\mathbb{G}_4)\right]^2 + \frac{1}{N}\mathrm{tr}\left[\mathbb{G}_4^2\right]$	
$\tilde{\Omega}_{41} = \frac{1}{N}\mathrm{tr}\left[W_1\mathbb{G}_4B_{12}P_1^{-1}\right]$	$\tilde{\Omega}_{43} = \frac{1}{N}\mathrm{tr}\left[B_{21}P_2^{-1}B_{12}P_1^{-1}\right]$
$\tilde{\Omega}_{45} = 0_{1\times k_1}$	$\tilde{\Omega}_{46} = \beta_2'\left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{4,ii}\Sigma_{iiv2}/\sigma_2^2\right]$
$\tilde{\Omega}_{51} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{1,ii}\Sigma_{iiv1}/\sigma_1^2\right]\beta_1$	$\tilde{\Omega}_{52} = 0_{k_1\times 1}$
$\tilde{\Omega}_{53} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{3,ii}\Sigma_{iiv1}/\sigma_1^2\right]\beta_1$	$\tilde{\Omega}_{54} = 0_{k_1\times 1}$
$\tilde{\Omega}_{55} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iiv1}/\sigma_1^2$	$\tilde{\Omega}_{56} = 0_{k_1\times k_2}$
$\tilde{\Omega}_{62} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{2,ii}\Sigma_{iiv2}/\sigma_2^2\right]\beta_2$	$\tilde{\Omega}_{61} = 0_{k_2\times 1}$
$\tilde{\Omega}_{64} = \left[\frac{1}{N}\sum_{i=1}^N \mathbb{G}_{4,ii}\Sigma_{iiv2}/\sigma_2^2\right]\beta_2$	$\tilde{\Omega}_{63} = 0_{k_2\times 1}$
$\tilde{\Omega}_{66} = \frac{1}{N}\sum_{i=1}^N \Sigma_{iiv2}/\sigma_2^2$	$\tilde{\Omega}_{65} = 0_{k_2\times k_1}$

Table A.17 provides the specifications of matrices Q_l ($l = 1, \ldots, 4$), which are involved in the IGPC approach as in Assumption G$'$ and Theorem 1.3.2. Each Q_l is a $2N \times 2N$ matrix whose (i, j)th 2×2 subblock $Q_{l,ij}$ is defined as below.

Table A.17: The specification of $Q_{l,ij}$ ($l = 1, \ldots, 4$) involved in the IGPC approach

$Q_{1,ij} = \begin{bmatrix} \mathbb{G}_{1,ij} & (W_1\mathbb{G}_4)_{ij} \\ 0 & 0 \end{bmatrix}$,	$Q_{2,ij} = \begin{bmatrix} 0 & 0 \\ (W_2\mathbb{G}_3)_{ij} & \mathbb{G}_{2,ij} \end{bmatrix}$
$Q_{3,ij} = \begin{bmatrix} \mathbb{G}_{3,ij} & (B_{21}P_2^{-1})_{ij} \\ 0 & 0 \end{bmatrix}$,	$Q_{4,ij} = \begin{bmatrix} 0 & 0 \\ (B_{12}P_1^{-1})_{ij} & \mathbb{G}_{4,ij} \end{bmatrix}$

where matrices $\mathbb{G}_l$ for $l = 1, 2, \ldots, 4$ are defined in Table A.11.

Further, $\mathbb{D}_b$ is a $k \times k$ matrix defined as:

$$\mathbb{D}_b = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}(\widetilde{X}_1' \ddot{M} \widetilde{X}_1 M_F) & \mathrm{tr}(\widetilde{X}_1' \ddot{M} \widetilde{X}_2 M_F) & \dots & \mathrm{tr}(\widetilde{X}_1' \ddot{M} \widetilde{X}_k M_F) \\ \mathrm{tr}(\widetilde{X}_2' \ddot{M} \widetilde{X}_1 M_F) & \mathrm{tr}(\widetilde{X}_2' \ddot{M} \widetilde{X}_2 M_F) & \dots & \mathrm{tr}(\widetilde{X}_2' \ddot{M} \widetilde{X}_k M_F) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\widetilde{X}_k' \ddot{M} \widetilde{X}_1 M_F) & \mathrm{tr}(\widetilde{X}_k' \ddot{M} \widetilde{X}_2 M_F) & \dots & \mathrm{tr}(\widetilde{X}_k' \ddot{M} \widetilde{X}_k M_F) \end{bmatrix}$$

$\mathbb{D}_c$ is a 4×4 matrix defined as:

$$\mathbb{D}_c = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}(\widetilde{X}_{k+1}' \ddot{M} \widetilde{X}_{k+1} M_F) & \mathrm{tr}(\widetilde{X}_{k+1}' \ddot{M} \widetilde{X}_{k+2} M_F) & \dots & \mathrm{tr}(\widetilde{X}_{k+1}' \ddot{M} \widetilde{X}_{k+4} M_F) \\ \mathrm{tr}(\widetilde{X}_{k+2}' \ddot{M} \widetilde{X}_{k+1} M_F) & \mathrm{tr}(\widetilde{X}_{k+2}' \ddot{M} \widetilde{X}_{k+2} M_F) & \dots & \mathrm{tr}(\widetilde{X}_{k+2}' \ddot{M} \widetilde{X}_{k+4} M_F) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\widetilde{X}_{k+4}' \ddot{M} \widetilde{X}_{k+1} M_F) & \mathrm{tr}(\widetilde{X}_{k+4}' \ddot{M} \widetilde{X}_{k+2} M_F) & \dots & \mathrm{tr}(\widetilde{X}_{k+4}' \ddot{M} \widetilde{X}_{k+4} M_F) \end{bmatrix}$$

ζ is a $k \times 4$ matrix defined as:

$$\zeta = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}(\widetilde{X}_1' \ddot{M} \widetilde{X}_{k+1} M_F) & \mathrm{tr}(\widetilde{X}_1' \ddot{M} \widetilde{X}_{k+2} M_F) & \dots & \mathrm{tr}(\widetilde{X}_1' \ddot{M} \widetilde{X}_{k+4} M_F) \\ \mathrm{tr}(\widetilde{X}_2' \ddot{M} \widetilde{X}_{k+1} M_F) & \mathrm{tr}(\widetilde{X}_2' \ddot{M} \widetilde{X}_{k+2} M_F) & \dots & \mathrm{tr}(\widetilde{X}_2' \ddot{M} \widetilde{X}_{k+4} M_F) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(\widetilde{X}_k' \ddot{M} \widetilde{X}_{k+1} M_F) & \mathrm{tr}(\widetilde{X}_k' \ddot{M} \widetilde{X}_{k+2} M_F) & \dots & \mathrm{tr}(\widetilde{X}_k' \ddot{M} \widetilde{X}_{k+4} M_F) \end{bmatrix}$$

where $M_F = I_T - F(F'F)^{-1}F'$, and $\ddot{M} = \Sigma_{ee}^{-1} - \frac{1}{N}\Sigma_{ee}^{-1}\Gamma\Gamma'\Sigma_{ee}^{-1}$ with Σ_{ee} as defined in Assumption C.

Some matrices in Theorem 1.3.2: $\mathbb{D}_\eta, \Phi, \vartheta$ and $\mathbb{D}_\beta$. First, $\mathbb{D}_\eta$ is defined as:

$$\mathbb{D}_\eta = \frac{1}{NT}\begin{bmatrix} \mathrm{tr}[\ddot{Y}_1'\ddot{M}_\Lambda \ddot{Y}_1 M_F] & \mathrm{tr}[\ddot{Y}_1'\ddot{M}_{\Lambda\Psi} \ddot{Y}_2 M_F] & \mathrm{tr}[\ddot{Y}_1'\ddot{M}_\Lambda \dot{Y}_2 M_F] & \mathrm{tr}[\ddot{Y}_1'\ddot{M}_{\Lambda\Psi} \dot{Y}_1 M_F] \\ \mathrm{tr}[\ddot{Y}_2'\ddot{M}_{\Psi\Lambda} \ddot{Y}_1 M_F] & \mathrm{tr}[\ddot{Y}_2'\ddot{M}_{\Psi} \ddot{Y}_2 M_F] & \mathrm{tr}[\ddot{Y}_2'\ddot{M}_{\Psi\Lambda} \dot{Y}_2 M_F] & \mathrm{tr}[\ddot{Y}_2'\ddot{M}_{\Psi} \dot{Y}_1 M_F] \\ \mathrm{tr}[\dot{Y}_2'\ddot{M}_\Lambda \ddot{Y}_1 M_F] & \mathrm{tr}[\dot{Y}_2'\ddot{M}_{\Lambda\Psi} \ddot{Y}_2 M_F] & \mathrm{tr}[\dot{Y}_2'\ddot{M}_\Lambda \dot{Y}_2 M_F] & \mathrm{tr}[\dot{Y}_2'\ddot{M}_{\Lambda\Psi} \dot{Y}_1 M_F] \\ \mathrm{tr}[\dot{Y}_1'\ddot{M}_{\Psi\Lambda} \ddot{Y}_1 M_F] & \mathrm{tr}[\dot{Y}_1'\ddot{M}_{\Psi} \ddot{Y}_2 M_F] & \mathrm{tr}[\dot{Y}_1'\ddot{M}_{\Psi\Lambda} \dot{Y}_2 M_F] & \mathrm{tr}[\dot{Y}_1'\ddot{M}_{\Psi} \dot{Y}_1 M_F] \end{bmatrix}$$

where $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$; $\Psi = (\psi_1, \psi_2, \ldots, \psi_N)'$; $\ddot{M}_\Lambda = \Sigma_{1ee}^{-1} - \frac{1}{N}\Sigma_{1ee}^{-1}\Lambda\Lambda'\Sigma_{1ee}^{-1}$; $\ddot{M}_\Psi = \Sigma_{2ee}^{-1} - \frac{1}{N}\Sigma_{2ee}^{-1}\Psi\Psi'\Sigma_{2ee}^{-1}$; $\ddot{M}_{\Lambda\Psi} = -\frac{1}{N}\Sigma_{1ee}^{-1}\Lambda\Psi'\Sigma_{2ee}^{-1}$; and $\ddot{M}_{\Psi\Lambda} = -\frac{1}{N}\Sigma_{2ee}^{-1}\Psi\Lambda'\Sigma_{1ee}^{-1}$, with both Σ_{1ee} and Σ_{2ee} being $N \times N$ matrices defined as $\Sigma_{1ee} = \mathrm{diag}(\sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{1N}^2)$ and $\Sigma_{2ee} = \mathrm{diag}(\sigma_{21}^2, \sigma_{22}^2, \ldots, \sigma_{2N}^2)$. In addition, $\ddot{Y}_1$ is an $N \times T$ matrix whose (i,t)th entry is $\ddot{y}_{1it}$; the $N \times T$ matrices $\ddot{Y}_2$, $\dot{Y}_1$, $\dot{Y}_2$ are defined similarly.

Then, Φ is a 4×4 symmetric matrix whose diagonal elements and upper diagonal elements are denoted by Φ_{ij} defined as follows:

$$\Phi_{pp} = -\Big[\frac{2}{N}\sum_{i=1}^{N}(\mathbb{G}_{p,ii})^2\Big] + \frac{1}{N}\mathrm{tr}\Big[\mathbb{G}_p^2\Big], \qquad \text{for} \quad p = 1,2,3,4$$

$$\Phi_{12} = \frac{1}{N}\mathrm{tr}[W_2\mathbb{G}_3 W_1 \mathbb{G}_4], \qquad \Phi_{13} = -\frac{2}{N}\sum_{i=1}^{N}(\mathbb{G}_{1,ii}\mathbb{G}_{3,ii}) + \frac{1}{N}\mathrm{tr}[\mathbb{G}_3\mathbb{G}_1]$$

$$\Phi_{14} = \frac{1}{N}\mathrm{tr}[B_{12}P_1^{-1}W_1\mathbb{G}_4], \qquad \Phi_{24} = -\frac{2}{N}\sum_{i=1}^{N}(\mathbb{G}_{2,ii}\mathbb{G}_{4,ii}) + \frac{1}{N}\mathrm{tr}[\mathbb{G}_4\mathbb{G}_2]$$

$$\Phi_{23} = \frac{1}{N}\mathrm{tr}[B_{21}P_2^{-1}W_2\mathbb{G}_3], \qquad \Phi_{34} = \frac{1}{N}\mathrm{tr}[B_{12}P_1^{-1}B_{21}P_2^{-1}]$$

where those $N \times N$ matrices $\mathbb{G}_p$ for $p = 1,2,3,4$ are defined in Table A1 in Notation Appendix I, with its (i,j)th entry being denoted by $\mathbb{G}_{p,ij}$; those $N \times N$ matrices P_1, P_2, B_{12}, B_{21} are defined in Assumption F.

The $4 \times k$ matrix ϑ is defined as:

$$\vartheta = [\vartheta_a \quad \vartheta_b]$$

with ϑ_a being a $4 \times k_1$ matrix:

$$\vartheta_a = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}[\ddot{Y}_1' \ddot{M}_\Lambda \dot{X}_{11} M_F] & \mathrm{tr}[\ddot{Y}_1' \ddot{M}_\Lambda \dot{X}_{12} M_F] & \dots & \mathrm{tr}[\ddot{Y}_1' \ddot{M}_\Lambda \dot{X}_{1k_1} M_F] \\ \mathrm{tr}[\ddot{Y}_2' \ddot{M}_{\Psi\Lambda} \dot{X}_{11} M_F] & \mathrm{tr}[\ddot{Y}_2' \ddot{M}_{\Psi\Lambda} \dot{X}_{12} M_F] & \dots & \mathrm{tr}[\ddot{Y}_2' \ddot{M}_{\Psi\Lambda} \dot{X}_{1k_1} M_F] \\ \mathrm{tr}[\dot{Y}_2' \ddot{M}_\Lambda \dot{X}_{11} M_F] & \mathrm{tr}[\dot{Y}_2' \ddot{M}_\Lambda \dot{X}_{12} M_F] & \dots & \mathrm{tr}[\dot{Y}_2' \ddot{M}_\Lambda \dot{X}_{1k_1} M_F] \\ \mathrm{tr}[\dot{Y}_1' \ddot{M}_{\Psi\Lambda} \dot{X}_{11} M_F] & \mathrm{tr}[\dot{Y}_1' \ddot{M}_{\Psi\Lambda} \dot{X}_{12} M_F] & \dots & \mathrm{tr}[\dot{Y}_1' \ddot{M}_{\Psi\Lambda} \dot{X}_{1k_1} M_F] \end{bmatrix}$$

and ϑ_b being a $4 \times k_2$ matrix:

$$\vartheta_b = \frac{1}{NT} \begin{bmatrix} \mathrm{tr}[\ddot{Y}_1' \ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\ddot{Y}_1' \ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\ddot{Y}_1' \ddot{M}_{\Lambda\Psi} \dot{X}_{2K_2} M_F] \\ \mathrm{tr}[\ddot{Y}_2' \ddot{M}_\Psi \dot{X}_{21} M_F] & \mathrm{tr}[\ddot{Y}_2' \ddot{M}_\Psi \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\ddot{Y}_2' \ddot{M}_\Psi \dot{X}_{2k_2} M_F] \\ \mathrm{tr}[\dot{Y}_2' \ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{Y}_2' \ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\dot{Y}_2' \ddot{M}_{\Lambda\Psi} \dot{X}_{2k_2} M_F] \\ \mathrm{tr}[\dot{Y}_1' \ddot{M}_\Psi \dot{X}_{21} M_F] & \mathrm{tr}[\dot{Y}_1' \ddot{M}_\Psi \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\dot{Y}_1' \ddot{M}_\Psi \dot{X}_{2k_2} M_F] \end{bmatrix}$$

where $\dot{X}_{1p}$ is a $N \times T$ matrix whose (i,t)th entry is $\dot{x}_{1it,p}$ for $p = 1, 2, \dots, k_1$; $\dot{X}_{2q}$ is defined similarly for $q = 1, 2, \dots, k_2$.

The $k \times k$ matrix $\mathbb{D}_\beta$ is defined as

$$\mathbb{D}_\beta = \begin{bmatrix} \mathbb{D}_\beta^a & \mathbb{D}_\beta^b \\ \mathbb{D}_\beta^{b\prime} & \mathbb{D}_\beta^c \end{bmatrix}$$

where the $k_1 \times k_1$ matrix $\mathbb{D}^a_\beta$ is defined as

$$\mathbb{D}^a_\beta = \frac{1}{NT}\begin{bmatrix} \mathrm{tr}[\dot{X}'_{11}\ddot{M}_\Lambda \dot{X}_{11} M_F] & \mathrm{tr}[\dot{X}'_{11}\ddot{M}_\Lambda \dot{X}_{12} M_F] & \dots & \mathrm{tr}[\dot{X}'_{11}\ddot{M}_\Lambda \dot{X}_{1k_1} M_F] \\ \mathrm{tr}[\dot{X}'_{12}\ddot{M}_\Lambda \dot{X}_{11} M_F] & \mathrm{tr}[\dot{X}'_{12}\ddot{M}_\Lambda \dot{X}_{12} M_F] & \dots & \mathrm{tr}[\dot{X}'_{12}\ddot{M}_\Lambda \dot{X}_{1k_1} M_F] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}[\dot{X}'_{1k_1}\ddot{M}_\Lambda \dot{X}_{11} M_F] & \mathrm{tr}[\dot{X}'_{1k_1}\ddot{M}_\Lambda \dot{X}_{12} M_F] & \dots & \mathrm{tr}[\dot{X}'_{1k_1}\ddot{M}_\Lambda \dot{X}_{1k_1} M_F] \end{bmatrix}$$

and the $k_1 \times k_2$ matrix $\mathbb{D}^b_\beta$ is defined as

$$\mathbb{D}^b_\beta = \frac{1}{NT}\begin{bmatrix} \mathrm{tr}[\dot{X}'_{11}\ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}'_{11}\ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\dot{X}'_{11}\ddot{M}_{\Lambda\Psi} \dot{X}_{2k_2} M_F] \\ \mathrm{tr}[\dot{X}'_{12}\ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}'_{12}\ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\dot{X}'_{12}\ddot{M}_{\Lambda\Psi} \dot{X}_{2k_2} M_F] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}[\dot{X}'_{1k_1}\ddot{M}_{\Lambda\Psi} \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}'_{1k_1}\ddot{M}_{\Lambda\Psi} \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\dot{X}'_{1k_1}\ddot{M}_{\Lambda\Psi} \dot{X}_{2k_2} M_F] \end{bmatrix}$$

and the $k_2 \times k_2$ matrix $\mathbb{D}^c_\beta$ is defined as

$$\mathbb{D}^c_\beta = \frac{1}{NT}\begin{bmatrix} \mathrm{tr}[\dot{X}'_{21}\ddot{M}_\Psi \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}'_{21}\ddot{M}_\Psi \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\dot{X}'_{21}\ddot{M}_\Psi \dot{X}_{2k_2} M_F] \\ \mathrm{tr}[\dot{X}'_{22}\ddot{M}_\Psi \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}'_{22}\ddot{M}_\Psi \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\dot{X}'_{22}\ddot{M}_\Psi \dot{X}_{2k_2} M_F] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}[\dot{X}'_{2k_2}\ddot{M}_\Psi \dot{X}_{21} M_F] & \mathrm{tr}[\dot{X}'_{2k_2}\ddot{M}_\Psi \dot{X}_{22} M_F] & \dots & \mathrm{tr}[\dot{X}'_{2k_2}\ddot{M}_\Psi \dot{X}_{2k_2} M_F] \end{bmatrix}$$

A.3 Some proofs

This appendix only includes the proof of consistency in Proposition 1.2.1 of QML approach. The proofs of other propositions and theorems in this paper are provided in the Supplementary Material. The symbols introduced in Table A.11 and the following table will be used throughout proofs.

Table A8: More symbols

$\hat{H} = (\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L})^{-1}$	$\hat{H}_N = N \cdot \hat{H}$
$\hat{G} = (I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L})^{-1}$	$\hat{G}_N = N \cdot \hat{G}$
$\hat{J} = \Upsilon(\eta) - \Upsilon(\hat{\eta})$	$S_N = \Upsilon(\eta)^{-1}$

From $(A+B)^{-1} = A^{-1} - A^{-1}B(A+B)^{-1}$, I have $\hat{H} = \hat{G}(I_r - \hat{G})^{-1}$ and $\hat{H} + \hat{G} = \hat{H}\hat{G} = \hat{G}\hat{H}$. From $\Sigma_{zz} = LL' + \Sigma_{\epsilon\epsilon}$, I have

$$\Sigma_{zz}^{-1} = \Sigma_{\epsilon\epsilon}^{-1} - \Sigma_{\epsilon\epsilon}^{-1}L(I_r + L'\Sigma_{\epsilon\epsilon}^{-1}L)^{-1}L'\Sigma_{\epsilon\epsilon}^{-1} \tag{A.3.1}$$

The above formulas will be used frequently throughout the appendix.

While in the main text, I use $(\delta, L, \Sigma_{\epsilon\epsilon})$ to denote the true value of the coefficients. For proving consistency, I shall use a superscript "*" to denote the true values of parameters; the variables without "*" denote the input variables of the likelihood function. This notation is only used in Appendix A. Proofs of all the following lemmas are provided in the Supplementary Material.

Lemma A.1. *Let $\eta = (\rho_1, \rho_2, \gamma_1, \gamma_2)$ and $\Upsilon(\eta)$ be a $2N \times 2N$ matrix, with its (i,j)th block,* *a* 2×2 *matrix, equal to*

$$\Upsilon_{ij}(\eta) = \begin{cases} \begin{bmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{bmatrix} & \text{if } i = j \\ \begin{bmatrix} -\rho_1 w_{1ij} & 0 \\ 0 & -\rho_2 w_{2ij} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{A.1}$$

Then I have $\det(D(\delta)) = \det(\Upsilon(\eta))$.

Lemma A.2. *Let* $V(\delta)$ *be the inverse matrix of* $D(\delta)$*, then its* (i,j)*th block, a* $(k+2)\times(k+2)$ *matrix, denoted as* $V_{ij}(\delta)$ *has a closed form, which is equal to*

$$V_{ij}(\delta) = \begin{cases} \begin{bmatrix} F_{ii} & F_{ii}\beta' \\ 0 & I_k \end{bmatrix} & \text{if } i = j \\ \begin{bmatrix} F_{ij} & F_{ij}\beta' \\ 0 & 0_{k\times k} \end{bmatrix} & \text{if } i \neq j \end{cases} \tag{A.2}$$

where $\beta = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix}$ *and* F_{ij} *is the* (i,j)*th* 2×2 *block of the inverse matrix of* $\Upsilon(\eta)$.

Lemma A.3. *Let* $R = \Upsilon(\eta)(\Upsilon(\eta^*))^{-1}$*, which is a* $2N \times 2N$ *matrix, then I can specify its* (i,j)*th block, a* 2×2 *matrix, denoted as* R_{ij}*, as following. For* $i = j$,

$$R_{ii} = I_2 - \begin{bmatrix} (\rho_1-\rho_1^*)\mathbb{G}_{1,ii} + (\gamma_1-\gamma_1^*)\mathbb{G}_{3,ii} & (\rho_1-\rho_1^*)(W_1\mathbb{G}_4)_{ii} + (\gamma_1-\gamma_1^*)(B_{21}P_2^{-1})_{ii} \\ (\rho_2-\rho_2^*)(W_2\mathbb{G}_3)_{ii} + (\gamma_2-\gamma_2^*)(B_{12}P_1^{-1})_{ii} & (\rho_2-\rho_2^*)\mathbb{G}_{2,ii} + (\gamma_2-\gamma_2^*)\mathbb{G}_{4,ii} \end{bmatrix}$$

For $i \neq j$,

$$R_{ij} = -\begin{bmatrix} (\rho_1-\rho_1^*)\mathbb{G}_{1,ij} + (\gamma_1-\gamma_1^*)\mathbb{G}_{3,ij} & (\rho_1-\rho_1^*)(W_1\mathbb{G}_4)_{ij} + (\gamma_1-\gamma_1^*)(B_{21}P_2^{-1})_{ij} \\ (\rho_2-\rho_2^*)(W_2\mathbb{G}_3)_{ij} + (\gamma_2-\gamma_2^*)(B_{12}P_1^{-1})_{ij} & (\rho_2-\rho_2^*)\mathbb{G}_{2,ij} + (\gamma_2-\gamma_2^*)\mathbb{G}_{4,ij} \end{bmatrix}$$

where W_1 and W_2 are the weights matrices defined in Assumption E, and $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_3, \mathbb{G}_4, P_1, P_2, B_{12}$ and B_{21} are defined in Table 1.

Furthermore, let $\mathbb{D} = DD^{*-1}$ with $D^* = D(\delta^*)$, I have

$$\mathbb{D}_{ij} = \begin{cases} \begin{bmatrix} R_{ii} & R_{ii}\beta^{*\prime} - \beta' \\ 0 & I_k \end{bmatrix} & \text{if } i = j \\ \begin{bmatrix} R_{ij} & R_{ij}\beta^{*\prime} \\ 0 & 0_{k\times k} \end{bmatrix} & \text{if } i \neq j \end{cases}$$

where $\beta = \begin{bmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{bmatrix}$ and $\beta^* = \begin{bmatrix} \beta_1^* & 0 \\ 0 & \beta_2^* \end{bmatrix}$ as defined in Lemma A.2; $\mathbb{D}_{ij}$ is the (i,j)th $(k+2)\times(k+2)$ subblock of $\mathbb{D}$ and R_{ij} is the (i,j)th 2×2 subblock of R defined as above.

Lemma A.4. *Let $(\rho_1, \rho_2, \gamma_1, \gamma_2, \beta_1', \beta_2')' \in \mathcal{A}_1\times\mathcal{A}_2\times\mathcal{A}_3\times\mathcal{A}_4\times\mathcal{A}_5\times\mathcal{A}_6$, where $\mathcal{A}_l$ is a compact set for all $l = 1, \dots, 6$. Under Assumptions A-F, uniformly on $\mathcal{A}_1\times\mathcal{A}_2\times\mathcal{A}_3\times\mathcal{A}_4\times\mathcal{A}_5\times\mathcal{A}_6$, I have*

(a) $\quad \Big\| \sum_{j=1}^{N} (\varphi_i^* + \varphi_i^*\beta^*)R_{ij}' - \phi_i^*\beta \Big\| \leq C, \quad$ *for all i;*

(b) $\quad \frac{1}{N}\sum_{i=1}^{N} \Big\| \frac{1}{T}\sum_{t=1}^{T} \Big(\sum_{j=1}^{N} R_{ij}(e_{jt} + \beta^{*\prime}v_{jt}) - \beta' v_{it} \Big) f_t' \Big\|^2 = O_p(T^{-1}),$

(c) $\dfrac{1}{N}\sum_{i=1}^{N}\Big|\dfrac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{pit}^{2}-E(\tilde{e}_{pit}^{2})]\Big|^{2}=O_p(T^{-1}),\quad for\ p=1,2,3,4;$

(d) $\dfrac{1}{N}\sum_{i=1}^{N}\Big|\dfrac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{pit}e_{qit}-E(\tilde{e}_{pit}e_{qit})]\Big|^{2}=O_p(T^{-1}),\quad for\ (p,q)=(1,1),(3,1),(2,2),(4,2);$

(e) $\dfrac{1}{N}\sum_{i=1}^{N}\Big\|\dfrac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{pit}v_{qit}-E(\tilde{e}_{pit}v_{qit})]\Big\|^{2}=O_p(T^{-1}),\quad for\ (p,q)=(1,1),(3,1),(2,2),(4,2);$

(f) $\dfrac{1}{N}\sum_{i=1}^{N}\Big|\dfrac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{pit}\tilde{e}_{qit}-E(\tilde{e}_{pit}\tilde{e}_{qit})]\Big|^{2}=O_p(T^{-1}),\quad for\ (p,q)=(1,3),(2,4);$

(g) $\dfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\dfrac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{pit}\tilde{e}_{qjt}-E(\tilde{e}_{pit}\tilde{e}_{qjt})]\Big|^{2}=O_p(T^{-1}),\quad for\ p,q=1,2,3,4;$

(h) $\dfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\dfrac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{pit}e_{qjt}-E(\tilde{e}_{pit}e_{qjt})]\Big|^{2}=O_p(T^{-1}),\quad for\ p=1,2,3,4\ and\ q=1,2;$

(i) $\dfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\dfrac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{pit}v_{qjt}-E(\tilde{e}_{pit}v_{qjt})]\Big|^{2}=O_p(T^{-1}),\quad for\ p=1,2,3,4\ and\ q=1,2;$

(j) $\dfrac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\dfrac{1}{T}\sum_{t=1}^{T}[\tilde{e}_{pit}v_{qjtl}-E(\tilde{e}_{pit}v_{qjtl})]\Big|^{2}=O_p(T^{-1}),$

$$for\ p=1,2,3,4,\ q=1,2\ and\ l=1,2,\ldots,k_q;$$

where β,β^* *are the same as defined in Lemma A.3;* $\varphi_i^*=(\lambda_i^*,\psi_i^*)$, $\phi_i^*=(\phi_{1i}^*,\phi_{2i}^*)$; $\tilde{e}_{1it}=\sum_{j=1}^{N}\left[\mathbb{G}_{1,ij}(e_{1jt}+\beta_1^{*\prime}v_{1jt})+(W_1\mathbb{G}_4)_{ij}(e_{2jt}+\beta_2^{*\prime}v_{2jt})\right]$, $\tilde{e}_{2it}=\sum_{j=1}^{N}\left[(W_2\mathbb{G}_3)_{ij}(e_{1jt}+\beta_1^{*\prime}v_{1jt})+\mathbb{G}_{2,ij}(e_{2jt}+\beta_2^{*\prime}v_{2jt})\right]$, $\tilde{e}_{3it}=\sum_{j=1}^{N}\left[\mathbb{G}_{3,ij}(e_{1jt}+\beta_1^{*\prime}v_{1jt})+(B_{21}P_2^{-1})_{ij}(e_{2jt}+\beta_2^{*\prime}v_{2jt})\right]$, *and* $\tilde{e}_{4it}=\sum_{j=1}^{N}\left[(B_{12}P_1^{-1})_{ij}(e_{1jt}+\beta_1^{*\prime}v_{1jt})+\mathbb{G}_{4,ij}(e_{2jt}+\beta_2^{*\prime}v_{2jt})\right]$, *where the matrices* $\mathbb{G}_1,\mathbb{G}_2,\mathbb{G}_3,\mathbb{G}_4,P_1,P_2,B_{12},B_{21}$ *are defined in Table 1.*

Lemma A.5. *Under Assumptions A-F,*

$$(a)\quad \sup_{\theta\in\Theta}\left|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}L^*\Big(\frac{1}{T}\sum_{t=1}^{T} f_t\epsilon_t'\Big)\mathbb{D}'\Sigma_{zz}^{-1}\Big]\right| = o_p(1)$$

$$(b)\quad \sup_{\theta\in\Theta}\left|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\frac{1}{T}\sum_{t=1}^{T}(\epsilon_t\epsilon_t' - \Sigma_{\epsilon\epsilon}^*)\mathbb{D}'\Sigma_{zz}^{-1}\Big]\right| = o_p(1)$$

$$(c)\quad \sup_{\theta\in\Theta}\left|\frac{1}{N}\mathrm{tr}\Big[\mathbb{D}\bar{\epsilon}\bar{\epsilon}'\mathbb{D}'\Sigma_{zz}^{-1}\Big]\right| = o_p(1)$$

where $\mathbb{D} = DD^{*-1}$ *and* $\Sigma_{zz} = LL' + \Sigma_{\epsilon\epsilon}$.

Lemma A.6. *Let* $P(\eta) = \begin{bmatrix} \rho_1 W_1 & \gamma_1 I_N \\ \gamma_2 I_N & \rho_2 W_2 \end{bmatrix}$, $\Delta(\eta) = I_{2N} - P(\eta)$ *and* $\mathcal{R} = \Delta(\eta)\big(\Delta(\eta^*)\big)^{-1}$. *Then I have*

$$\mathrm{tr}\Big((R\Sigma_{ee}^*R' - \Sigma_{ee})(R\Sigma_{ee}^*R' - \Sigma_{ee})'\Big) = \mathrm{tr}\Big((\mathcal{R}\Omega_{ee}^*\mathcal{R}' - \Omega_{ee})(\mathcal{R}\Omega_{ee}^*\mathcal{R}' - \Omega_{ee})'\Big)$$

where R *is defined in Lemma A.3;* $\Omega_{ee} = \begin{bmatrix} \Omega_{1ee} & 0 \\ 0 & \Omega_{2ee} \end{bmatrix}$ *with* $\Omega_{1ee} = \mathrm{diag}(\sigma_{11}^2, \sigma_{12}^2, \ldots, \sigma_{1N}^2)$ *and* $\Omega_{2ee} = \mathrm{diag}(\sigma_{21}^2, \sigma_{22}^2, \ldots, \sigma_{2N}^2)$*;* Ω_{ee}^* *equals* Ω_{ee} *with parameters evaluated at their true values. In addition,*

$$\mathcal{R} = I_{2N} - \begin{bmatrix} (\rho_1-\rho_1^*)\mathbb{G}_1 + (\gamma_1-\gamma_1^*)\mathbb{G}_3, & (\rho_1-\rho_1^*)W_1\mathbb{G}_4 + (\gamma_1-\gamma_1^*)B_{21}P_2^{-1} \\ (\rho_2-\rho_2^*)W_2\mathbb{G}_3 + (\gamma_2-\gamma_2^*)B_{12}P_1^{-1}, & (\rho_2-\rho_2^*)\mathbb{G}_2 + (\gamma_2-\gamma_2^*)\mathbb{G}_4 \end{bmatrix}$$

where matrices P_1, P_2, B_{12}, B_{21} *and* $\mathbb{G}_l$ $(l = 1, 2, 3, 4)$ *are defined in Table A.11 with parameters evaluated at true values.*

Proof of Proposition 1.2.1: Consider the following centered objective function:

$$\mathcal{L}(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| + \frac{1}{N}\ln|D| - \frac{1}{2N}\mathrm{tr}[DM_{zz}D'\Sigma_{zz}^{-1}] + \frac{1}{2N}\ln|\Sigma_{zz}^*| - \frac{1}{N}\ln|D^*| + \frac{k+2}{2}. \tag{A.3}$$

Note that the term $(\frac{1}{2N}\ln|\Sigma_{zz}^*| - \frac{1}{N}\ln|D^*| + \frac{k+2}{2})$ is a constant as it does not depend on any unknown parameters and is for the purpose of centering. By $D^*\dot{z}_t = \Phi^* f_t + \dot{\epsilon}_t$ and the identification condition $\bar{f} = 0$, I have

$$D^*M_{zz}D^{*\prime} = \Sigma_{zz}^* + L^*\Big(\frac{1}{T}\sum_{t=1}^{T} f_t\epsilon_t'\Big) + \Big(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t f_t'\Big)L^{*\prime} + \frac{1}{T}\sum_{t=1}^{T}(\epsilon_t\epsilon_t' - \Sigma_{\epsilon\epsilon}^*) - \bar{\epsilon}\bar{\epsilon}',$$

where $\Sigma_{zz}^* = L^*L^{*\prime} + \Sigma_{\epsilon\epsilon}^*$ and $\bar{\epsilon} = \frac{1}{T}\sum_{t=1}^{T}\epsilon_t$. Then I get

$$M_{zz} = D^{*-1}\Sigma_{zz}^* D^{*\prime-1} + \mathbb{R}, \tag{A.4}$$

where

$$\mathbb{R} = D^{*-1}L^*\Big(\frac{1}{T}\sum_{t=1}^{T} f_t\epsilon_t'\Big)D^{*\prime-1} + D^{*-1}\Big(\frac{1}{T}\sum_{t=1}^{T}\epsilon_t f_t'\Big)L^{*\prime}D^{*\prime-1} + D^{*-1}\frac{1}{T}\sum_{t=1}^{T}(\epsilon_t\epsilon_t' - \Sigma_{\epsilon\epsilon}^*)D^{*\prime-1} - D^{*-1}\bar{\epsilon}\bar{\epsilon}'D^{*\prime-1}.$$

Substituting (A.4) into (A.3),

$$\mathcal{L}(\theta) = \mathcal{L}_1(\theta) + \mathcal{L}_2(\theta), \tag{A.5}$$

where

$$\begin{aligned}\mathcal{L}_1(\theta) = &- \frac{1}{2N}\ln|\Sigma_{zz}| + \frac{1}{N}\ln|D| - \frac{1}{2N}\mathrm{tr}[DD^{*-1}\Sigma_{zz}^{*}D^{*-1\prime}D'\Sigma_{zz}^{-1}] \\ &+ \frac{1}{2N}\ln|\Sigma_{zz}^{*}| - \frac{1}{N}\ln|D^{*}| + \frac{k+2}{2}\end{aligned}$$

and

$$\mathcal{L}_2(\theta) = -\frac{1}{2N}\mathrm{tr}[D\mathbb{R}D'\Sigma_{zz}^{-1}]. \tag{A.6}$$

Since $\hat{\theta}$ maximizes $\mathcal{L}(\theta)$, I have $\mathcal{L}(\hat{\theta}) \geq \mathcal{L}(\theta^*)$, implying $\mathcal{L}_1(\hat{\theta}) \geq \mathcal{L}_1(\theta^*) + \mathcal{L}_2(\theta^*) - \mathcal{L}_2(\hat{\theta})$. By Lemma A.5, I have $\sup_{\theta\in\Theta}|\mathcal{L}_2(\theta)| = o_p(1)$, and then $|\mathcal{L}_2(\theta^*) - \mathcal{L}_2(\hat{\theta})| \geq -2\sup_{\theta\in\Theta}|\mathcal{L}_2(\theta)| = -|o_p(1)|$. Given this result, together with $\mathcal{L}_1(\theta^*) = 0$, I have

$$\mathcal{L}_1(\hat{\theta}) \geq -|o_p(1)|. \tag{A.7}$$

With notation $\hat{\mathbb{D}} = \hat{D}D^{*-1}$, I rewrite $\mathcal{L}_1(\hat{\theta})$ as

$$\begin{aligned}\mathcal{L}_1(\hat{\theta}) = &- \frac{1}{2N}\ln|\hat{\Sigma}_{zz}| + \frac{1}{N}\ln|\hat{D}| - \frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}\Sigma_{zz}^{*}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] \\ &+ \frac{1}{2N}\ln|\Sigma_{zz}^{*}| - \frac{1}{N}\ln|D^{*}| + \frac{k+2}{2}\end{aligned}$$

With the definition $\hat{\Sigma}_{zz} = \hat{L}\hat{L}' + \hat{\Sigma}_{\epsilon\epsilon}$, I have $|\hat{\Sigma}_{zz}| = |\hat{\Sigma}_{\epsilon\epsilon}| \cdot |I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}|$. Thus,

$$\ln|\hat{\Sigma}_{zz}| = \ln|\hat{\Sigma}_{\epsilon\epsilon}| + \ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}| = \sum_{i=1}^{N}(\ln|\hat{\Sigma}_{iie}| + \ln|\hat{\Sigma}_{iiv}|) + \ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1\prime}\hat{L}|$$

Similarly $\ln|\Sigma_{zz}^{*}| = \sum_{i=1}^{N}(\ln|\Sigma_{iie}^{*}| + \ln|\Sigma_{iiv}^{*}|) + \ln|I_r + L^{*\prime}\Sigma_{\epsilon\epsilon}^{*-1}L^{*}|$. Notice that $|I_r + L^{*\prime}\Sigma_{\epsilon\epsilon}^{*-1}L^{*}| =$

$O_p(N)$, so uniformly on Θ,

$$-\frac{1}{2N}\ln|\hat{\Sigma}_{zz}| + \frac{1}{2N}\ln|\Sigma^*_{zz}| = -\frac{1}{2N}\sum_{i=1}^{N}(\ln|\hat{\Sigma}_{iie}| + \ln|\hat{\Sigma}_{iiv}|) + \frac{1}{2N}\sum_{i=1}^{N}(\ln|\Sigma^*_{iie}| + \ln|\Sigma^*_{iiv}|)$$
$$-\frac{1}{2N}\ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1\prime}\hat{L}| + O_p(\frac{\ln(N)}{N})$$

Next consider the term $\frac{1}{2N}\text{tr}[\hat{\mathbb{D}}\Sigma^*_{zz}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}]$, which can be written as, in view of $\Sigma^*_{zz} = L^*L^{*\prime} + \Sigma^*_{\epsilon\epsilon}$,

$$\frac{1}{2N}\text{tr}[\hat{\mathbb{D}}\Sigma^*_{zz}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] = \frac{1}{2N}\text{tr}[\hat{\mathbb{D}}\Sigma^*_{\epsilon\epsilon}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] + \frac{1}{2N}\text{tr}[\hat{\mathbb{D}}L^*L^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] \triangleq i_1 + i_2, \quad \text{say.}$$

By the Woodbury formula $\hat{\Sigma}_{zz}^{-1} = \hat{\Sigma}_{\epsilon\epsilon}^{-1} - \hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}$, where $\hat{G} = (I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L})^{-1}$, i_1 can be written as

$$i_1 = \frac{1}{2N}\text{tr}[\hat{\mathbb{D}}\Sigma^*_{\epsilon\epsilon}\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}] - \frac{1}{2N}\text{tr}[\hat{\mathbb{D}}\Sigma^*_{\epsilon\epsilon}\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}\hat{G}\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}] \triangleq i_3 - i_4, \quad \text{say}$$

With the definition of $\mathbb{D}$ and calculation, I get

$$i_3 = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\text{tr}\Big(\hat{R}_{ij}\Sigma^*_{jje}\hat{R}'_{ij}\hat{\Sigma}_{iie}^{-1}\Big) + \frac{1}{2N}\sum_{i=1}^{N}\text{tr}\Big((\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')\Sigma^*_{iiv}(\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')'\hat{\Sigma}_{iiv}^{-1}\Big)$$
$$+\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\text{tr}\Big(\hat{R}_{ij}\beta^{*\prime}\Sigma^*_{jjv}\beta^*\hat{R}'_{ij}\hat{\Sigma}_{iie}^{-1}\Big) + \frac{1}{2N}\text{tr}\Big[\sum_{i=1}^{N}\Sigma^*_{iiv}\hat{\Sigma}_{iiv}^{-1}\Big]$$

where $\hat{R}_{ij}$ is the (i,j)-th subblock $\hat{R}$. Now we show $i_4 = o_p(1)$ uniformly on Θ. To see this, by the boundedness of $\hat{\Sigma}_{ii}$ and Σ^*_{ii}, $\hat{\mathbb{D}}\Sigma^*_{\epsilon\epsilon}\hat{\mathbb{D}}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}$ is less than[1] $C_1\hat{\mathbb{D}}\hat{\mathbb{D}}'$ for some C_1, which is

[1]For matrices A and B, I say $A \leq B$ if $B - A$ is a semi-definite positive matrix.

further less than $C_1C_2I_{N(k+2)}$ for some constant C_2, as shown in the proof of Lemma A.5 (c). This result leads to $i_4 \leq C_1C_2\frac{1}{2N}\mathrm{tr}[\hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}\hat{G}] = O_p(N^{-1})$.

Given the above results, together with the fact that $\ln|\hat{D}| - \ln|D^*| = \ln|\hat{D}D^{*-1}| = \ln|\hat{\mathbb{D}}| = \ln|\hat{R}|$, I can rewrite the $\mathcal{L}_1(\theta)$ as

$$\begin{aligned}
\mathcal{L}_1(\hat{\theta}) = &-\frac{1}{2N}\sum_{i=1}^{N}(\ln|\hat{\Sigma}_{iie}| + \ln|\hat{\Sigma}_{iiv}|) + \frac{1}{2N}\sum_{i=1}^{N}(\ln|\Sigma_{iie}^*| + \ln|\Sigma_{iiv}^*|) - \frac{1}{2N}\ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1\prime}\hat{L}| \\
&-\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{tr}\Big(\hat{R}_{ij}\Sigma_{jje}^*\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\Big) - \frac{1}{2N}\sum_{i=1}^{N}\mathrm{tr}\Big((\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')\Sigma_{iiv}^*(\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')'\hat{\Sigma}_{iiv}^{-1}\Big) \\
&-\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\mathrm{tr}\Big(\hat{R}_{ij}\beta^{*\prime}\Sigma_{jjv}^*\beta^*\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\Big) - \frac{1}{2N}\mathrm{tr}\Big[\sum_{i=1}^{N}\Sigma_{iiv}^*\hat{\Sigma}_{iiv}^{-1}\Big] \\
&-\frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}L^*L^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] + \frac{1}{N}\ln|\hat{R}| + \frac{k+2}{2} \geq -|o_p(1)| \qquad \text{(A.8)}
\end{aligned}$$

Notice that the term $-\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{tr}\Big(\hat{R}_{ij}\Sigma_{jje}^*\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\Big) + \frac{1}{N}\ln|\hat{R}|$ is equivalent to

$$-\frac{1}{2N}\mathrm{tr}[\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}] + \frac{1}{2N}\ln|\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}| + \frac{1}{2N}\sum_{i=1}^{N}(\ln|\hat{\Sigma}_{iie}| - \ln|\Sigma_{iie}^*|) \qquad \text{(A.9)}$$

Substituting (A.9) into (A.8), I can rewrite $\mathcal{L}_1(\hat{\theta})$ as

$$\begin{aligned}
\mathcal{L}_1(\hat{\theta}) = &-\Big\{\frac{1}{2N}\mathrm{tr}[\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}] - \frac{1}{2N}\ln|\hat{R}\Sigma_{ee}^*\hat{R}'\hat{\Sigma}_{ee}^{-1}| - 1\Big\} - \Big\{\frac{1}{2N}\mathrm{tr}[\hat{\mathbb{D}}L^*L^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}]\Big\} \\
&-\Big\{\frac{1}{2N}\sum_{i=1}^{N}\mathrm{tr}\Big((\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')\Sigma_{iiv}^*(\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')'\hat{\Sigma}_{iiv}^{-1}\Big)\Big\} - \Big\{\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\mathrm{tr}\Big(\hat{R}_{ij}\beta^{*\prime}\Sigma_{jjv}^*\beta^*\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\Big)\Big\} \\
&-\Big\{\frac{1}{2N}\sum_{i=1}^{N}\Big(\mathrm{tr}[\Sigma_{iiv}^*\hat{\Sigma}_{iiv}^{-1}] - \ln|\Sigma_{iiv}^*\hat{\Sigma}_{iiv}^{-1}| - k\Big)\Big\} - \Big\{\frac{1}{2N}\ln|I_r + \hat{L}'\hat{\Sigma}_{\epsilon\epsilon}^{-1}\hat{L}|\Big\} \geq -|o_p(1)|
\end{aligned}$$

In the above equation, all the expressions in the braces are non-negative, so each expression must be $o_p(1)$. From the first five expressions, I have

$$\frac{1}{2N}\text{tr}[\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1}] - \frac{1}{2N}\ln|\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1}| - 1 = o_p(1) \tag{A.10}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\text{tr}\Big((\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')\Sigma_{iiv}^{*}(\hat{R}_{ii}\beta^{*\prime} - \hat{\beta}')'\hat{\Sigma}_{iiv}^{-1}\Big) = o_p(1) \tag{A.11}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\text{tr}\Big(\hat{R}_{ij}\beta^{*\prime}\Sigma_{jjv}^{*}\beta^{*}\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\Big) = o_p(1) \tag{A.12}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\Big(\text{tr}[\Sigma_{iiv}^{*}\hat{\Sigma}_{iiv}^{-1}] - \ln|\Sigma_{iiv}^{*}\hat{\Sigma}_{iiv}^{-1}| - k\Big) = o_p(1) \tag{A.13}$$

$$\frac{1}{2N}\text{tr}[\hat{\mathbb{D}}L^{*}L^{*\prime}\hat{\mathbb{D}}'\hat{\Sigma}_{zz}^{-1}] = o_p(1) \tag{A.14}$$

First consider (A.12). For $\beta_1^* \neq 0$ and $\beta_2^* \neq 0$, by definition of $\beta^* = (\beta_1', \beta_2')'$ and the boundedness of Σ_{iiv}^* and $\hat{\Sigma}_{iie}$, there exists a positive constant c such that $\beta^{*\prime}\Sigma_{jjv}^{*}\beta^{*} > c$ and $\hat{\Sigma}_{iie}^{-1} > cI_2$. Then together with (A.12), I have

$$o_p(1) = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\text{tr}\Big(\hat{R}_{ij}\beta^{*\prime}\Sigma_{jjv}^{*}\beta^{*}\hat{R}_{ij}'\hat{\Sigma}_{iie}^{-1}\Big) > \frac{c^2}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\text{tr}\Big(\hat{R}_{ij}\hat{R}_{ij}'\Big) > 0$$

which implies that

$$\frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\text{tr}\Big(\hat{R}_{ij}\hat{R}_{ij}'\Big) = o_p(1) \tag{A.15}$$

By the expressions of R_{ij} in Lemma A.3, I have

$$\begin{aligned} o_p(1) =& \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \mathrm{tr}\Big(\hat{R}_{ij}\hat{R}'_{ij}\Big) \\ =& (\hat{\rho}_1 - \rho_1^*)^2 a_1 + 2(\hat{\rho}_1 - \rho_1^*)(\hat{\gamma}_1 - \gamma_1^*) b_1 + (\hat{\gamma}_1 - \gamma_1^*)^2 c_1 \\ & + (\hat{\rho}_2 - \rho_2^*)^2 a_2 + 2(\hat{\rho}_2 - \rho_2^*)(\hat{\gamma}_2 - \gamma_2^*) b_2 + (\hat{\gamma}_2 - \gamma_2^*)^2 c_2 \\ =& (\hat{\eta} - \eta^*)' \mathbb{M}_a (\hat{\eta} - \eta^*) \end{aligned} \tag{A.16}$$

where the 4×4 matrix $\mathbb{M}_a$ is defined as in Assumption G.1, and a_p, b_p, c_p for $(p = 1, 2)$ are all scalars, and their definitions are given in Table A.12. Based on the above equation, together with Assumption (G.1), I have the consistency $\hat{\eta} - \eta^* = o_p(1)$.

Next consider (A.10), which can be written as

$$\frac{1}{2N} \mathrm{tr}[\hat{\Sigma}_{ee}^{-1/2} \hat{R} \Sigma_{ee}^* \hat{R}' \hat{\Sigma}_{ee}^{-1/2}] - \frac{1}{2N} \ln |\hat{\Sigma}_{ee}^{-1/2} \hat{R} \Sigma_{ee}^* \hat{R}' \hat{\Sigma}_{ee}^{-1/2}| - 1 = o_p(1)$$

Let l_i $(i = 1, 2, \ldots, 2N)$ denote the eigenvalues of the $2N \times 2N$ matrix $\hat{\Sigma}_{ee}^{-1/2} \hat{R} \Sigma_{ee}^* \hat{R}' \hat{\Sigma}_{ee}^{-1/2}$. By the boundedness of $\hat{\eta}$ and $\hat{\Sigma}_{iie}$, there exits some large constant C such that $l_i \in [C^{-1}, C]$ for all i. Together with the fact that $x - \ln x - 1 \geq \frac{1}{4C^2}(x-1)^2$ for all $x \in [C^{-1}, C]$, I have

$$\begin{aligned} o_p(1) =& \frac{1}{2N} \mathrm{tr}[\hat{\Sigma}_{ee}^{-1/2} \hat{R} \Sigma_{ee}^* \hat{R}' \hat{\Sigma}_{ee}^{-1/2}] - \frac{1}{2N} \ln |\hat{\Sigma}_{ee}^{-1/2} \hat{R} \Sigma_{ee}^* \hat{R}' \hat{\Sigma}_{ee}^{-1/2}| - 1 \\ =& \frac{1}{2N} \sum_{i=1}^{2N} (l_i - \ln l_i - 1) \geq \frac{1}{4C^2} \frac{1}{2N} \sum_{i=1}^{2N} (l_i - 1)^2 = \frac{1}{4C^2} \frac{1}{2N} \| \hat{\Sigma}_{ee}^{-1/2} \hat{R} \Sigma_{ee}^* \hat{R}' \hat{\Sigma}_{ee}^{-1/2} - I_{2N} \|^2 \end{aligned}$$

implying

$$\frac{1}{2N} \| \hat{\Sigma}_{ee}^{-1/2} \hat{R} \Sigma_{ee}^* \hat{R}' \hat{\Sigma}_{ee}^{-1/2} - I_{2N} \|^2 = o_p(1)$$

which is equivalent to

$$\frac{1}{2N}\mathrm{tr}\Big[(\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_{2N})(\hat{\Sigma}_{ee}^{-1/2}\hat{R}\Sigma_{ee}^{*}\hat{R}'\hat{\Sigma}_{ee}^{-1/2} - I_{2N})'\Big] = o_p(1)$$

can be further written as

$$\frac{1}{2N}\mathrm{tr}\Big[\hat{\Sigma}_{ee}^{-1/2}(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1}(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1/2}\Big] = o_p(1)$$

By the boundedness of $\hat{\Sigma}_{iie}$, there exists come constant c such that $\hat{\Sigma}_{ee}^{-1/2} \geq cI_{2N}$. Then

$$\begin{aligned} o_p(1) =& \frac{1}{2N}\mathrm{tr}\Big[\hat{\Sigma}_{ee}^{-1/2}(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1}(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})\hat{\Sigma}_{ee}^{-1/2}\Big] \\ \geq& c^4\frac{1}{2N}\mathrm{tr}\Big[(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})(\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee})\Big] = c^4\frac{1}{2N}\|\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee}\|^2 > 0 \end{aligned}$$

implying

$$\frac{1}{2N}\|\hat{R}\Sigma_{ee}^{*}\hat{R}' - \hat{\Sigma}_{ee}\|^2 = o_p(1) \tag{A.17}$$

By Lemma A.6, the above equation is equivalent to

$$\frac{1}{2N}\|\hat{\mathcal{R}}\Omega_{ee}^{*}\hat{\mathcal{R}}' - \hat{\Omega}_{ee}\|^2 = o_p(1) \tag{A.18}$$

Let $\mathcal{U} = \hat{\mathcal{R}}\Omega_{ee}^{*}\hat{\mathcal{R}}' - \hat{\Omega}_{ee}$, together with Lemma A.6, I have

$$\mathcal{U} = \Omega_{ee}^{*} - \hat{\Omega}_{ee} + \mathbb{H}$$

where $\mathbb{H} = -\hat{H}\Omega^*_{ee} - \Omega^*_{ee}\hat{H}' + \hat{H}\Omega^*_{ee}\hat{H}'$, and

$$\hat{H} = \begin{bmatrix} (\hat{\rho}_1 - \rho_1^*)\mathbb{G}_1 + (\hat{\gamma}_1 - \gamma_1^*)\mathbb{G}_3, & (\hat{\rho}_1 - \rho_1^*)W_1\mathbb{G}_4 + (\hat{\gamma}_1 - \gamma_1^*)B_{21}P_2^{-1} \\ (\hat{\rho}_2 - \rho_2^*)W_2\mathbb{G}_3 + (\hat{\gamma}_2 - \gamma_2^*)B_{12}P_1^{-1}, & (\hat{\rho}_2 - \rho_2^*)\mathbb{G}_2 + (\hat{\gamma}_2 - \gamma_2^*)\mathbb{G}_4 \end{bmatrix}$$

where matrices P_1, P_2, B_{12}, B_{21} and $\mathbb{G}_l$ $(l = 1, 2, 3, 4)$ are all evaluated at true parameters. Then (A.18) can be further rewritten as

$$o_p(1) = \text{tr}\left[\left(\Omega^*_{ee} - \hat{\Omega}_{ee} + \text{diag}(\mathbb{H})\right)^2\right] + (\hat{\eta} - \eta^*)'\text{tr}(\mathbb{M})(\hat{\eta} - \eta^*) = m_1 + m_2, \text{say}$$

where the 4×4 matrix $\mathbb{M}$ is defined in Assumption G. Note that both m_1 and m_2 are non-negative, therefore I have $m_1 = o_p(1)$ and $m_2 = o_p(1)$. Combining the result $m_2 = o_p(1)$ and Assumption (G.2) implies that $\hat{\eta} - \eta^* = o_p(1)$, which further implies $\mathbb{H} = o_p(1)$. Plugging these results into $m_1 = o_p(1)$, together with the boundedness of variances, I get the average consistency for the variances:

$$\begin{aligned} \frac{1}{N}\sum_{i=1}^{N}(\sigma_{1i}^{*2} - \hat{\sigma}_{1i}^2)^2 = o_p(1) \\ \frac{1}{N}\sum_{i=1}^{N}(\sigma_{2i}^{*2} - \hat{\sigma}_{2i}^2)^2 = o_p(1) \end{aligned} \tag{A.19}$$

The consistency of $\hat{\eta} = (\hat{\rho}_1, \hat{\rho}_2, \gamma_1, \gamma_2)$ implies that the 2×2 subblock $\hat{R}_{ii} \to I_2$ for all i. Plug this result into (A.11), I prove the consistency that $\hat{\beta}_1 \to \beta_1^*$ and $\hat{\beta}_2 \to \beta_2^*$. Now consider (A.13) and (A.14), which are the results corresponding to the pure factor structure part. Using a similar way as in Bai and Li (2014a), I can show from (A.13) that

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{iiv}-\Sigma_{iiv}^{*}\|^{2}=o_{p}(1) \tag{A.20}$$

Combining the results (A.19) and (A.20), I have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}^{*}\|^{2}=o_{p}(1) \tag{A.21}$$

The last claim of Proposition (1.2.1) can be proved from (A.14) together with NC.1–NC.3 and first order condition of L, using a similar approach as in Bai and Li (2014a). This completes the proof for Proposition 1.2.1. □

Appendix B

Appendix for Chapter 2

B.1 Proof of Theorem 2.3.1

Throughout the appendix, we use C to denote a generic finite constant large enough, which need not to be the same at each appearance. In addition, we introduce following notations for ease of exposition.

$$H = (\Lambda'\Psi^{-1}\Lambda)^{-1}; \quad \hat{H} = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}; \quad R = M_{ff}\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}.$$

We first show that $R = O_p(1)$. The following lemma is useful.

Lemma B.11. *Under Assumptions A-D,*

$$(a) \quad R = \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(1)$$

$$(b) \quad R' M_{ff}^{-1} \frac{1}{T} \sum_{t=1}^{T} f_t u_t' \hat{\Psi}^{-1} \hat{\Lambda} \hat{H} = \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$$

$$(c) \quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}(u_t u_t' - \Psi)\Big]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$$

$$(d) \quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}(\hat{\Psi} - \Psi)\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = \|N^{1/2}\hat{H}^{1/2}\|^2 \cdot o_p\Big(\big[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\big]^{1/2}\Big)$$

Proof of Lemma B.11: Consider (a). By the definition of R and $\hat{H}$, we have

$$R = M_{ff}\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1} = M_{ff}(\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}^{1/2})\hat{H}^{1/2}$$

By the Cauchy-Schwarz inequality,

$$\big\|\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}^{1/2}\big\| = \big\|\sum_{i=1}^{N}\Lambda_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}^{1/2}\big\| \leq \Big(\sum_{i=1}^{N}\big\|\Lambda_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i'\hat{H}^{1/2}\|^2\Big)^{1/2}$$

However,

$$\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i'\hat{H}^{1/2}\|^2 = \mathrm{tr}\big[\sum_{i=1}^{N}\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}^{1/2}\big] = \mathrm{tr}\big[\hat{H}^{1/2}\hat{H}^{-1}\hat{H}^{1/2}\big] = r \qquad \text{(B.1.1)}$$

Given (B.1.1), together with the boundedness of $\hat{\Sigma}_{ii}^{-1/2}$ and Λ_i, we have

$$\big\|\Lambda'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}^{1/2}\big\| = O_p(N^{1/2})$$

Then (a) follows.

Consider (b). We first show

$$\frac{1}{T}\sum_{t=1}^{T} u_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = \frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t u_{it}'\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H} = \|N^{1/2}\hat{H}^{1/2}\| \cdot O_p(T^{-1/2}) \tag{B.1.2}$$

By the Cauchy-Schwarz inequality,

$$\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t u_{it}'\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H} \le C\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} f_t u_{it}'\Big\|^2\Big)^{1/2} \times \Big(\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i'\hat{H}^{1/2}\|^2\Big)^{1/2}\|N^{1/2}\hat{H}^{1/2}\|$$

So (B.1.2) follows by (B.1.1). Given (B.1.2) together with result (a), we have (b).

Consider (c), which is equal to

$$\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\Big)\hat{\Sigma}_{jj}^{-1}\hat{\Lambda}_j'\hat{H}.$$

The above expression is bounded in norm by

$$\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i'\hat{H}^{1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\Big\|^2\Big)^{1/2}$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$ by (B.1.1). Then (c) follows.

Consider (d), which is equal to

$$\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}.$$

The above epression is bounded in norm by

$$\|N^{1/2}\hat{H}^{1/2}\|^2 \frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 \cdot \|\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1/2}\|$$

By (B.1.1), we have $\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 = r$, which means $\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\| \leq \sqrt{r}$ for all i. Given this result, together with the boundedness of $\hat{\Sigma}_{ii}^{-1}$, we have that the above expression is bounded by

$$C\sqrt{r}\|N^{1/2}\hat{H}^{1/2}\|^2 \frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\| \cdot \|\hat{\Sigma}_{ii}-\Sigma_{ii}\|$$

which is further bounded by

$$C\sqrt{r}\|N^{1/2}\hat{H}^{1/2}\|^2 \Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)^{1/2},$$

implying (d). □

Proposition B.1.1. *Under Assumptions A-D,*

$$\|N^{1/2}\hat{H}^{1/2}\| = O_p(1), \qquad\qquad R = O_p(1).$$

Proof of Proposition B.1.1: By (2.3.4), we have $\hat{\Lambda}'\hat{\Psi}^{-1}(M_{zz}-\hat{\Sigma}_{zz})\hat{\Psi}^{-1}\hat{\Lambda} = 0$. By

$$M_{zz} = \Lambda M_{ff}\Lambda' + \Lambda\frac{1}{T}\sum_{t=1}^{T}f_t u_t' + \frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Lambda' + \frac{1}{T}\sum_{t=1}^{T}(u_t u_t' - \Psi) + \Psi$$

and $\hat{\Sigma}_{zz} = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$, we have

$$I_r = R'M_{ff}^{-1}R + R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t u_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} + \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T} u_t f_t' M_{ff}^{-1}R$$
$$+\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}(u_t u_t' - \Psi)\Big]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} - \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}(\hat{\Psi} - \Psi)\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} \tag{B.1.3}$$

Consider the right hand side of (B.1.3). By Lemma B.11, the first term is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(1)$ and the 2nd-4th terms are all $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(T^{-1/2})$. The last term is equivalent to

$$\hat{H}^{1/2}\Big(\sum_{i=1}^{N} \hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\big(\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1/2}\big)\hat{\Sigma}_{ii}^{-1/2}\hat{\Lambda}_i\hat{H}^{1/2}\Big)\hat{H}^{1/2}$$

which is bounded by

$$\frac{1}{N}\|N^{1/2}\hat{H}^{1/2}\|^2\big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 \cdot \|\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1/2}\|\big)$$

which is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(N^{-1})$ by $\|\hat{\Sigma}_{ii}^{-1/2}(\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1/2}\| = O_p(1)$ and (B.1.1). So the last term is $\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot O_p(N^{-1})$. However, by the equation (A.10) of Bai and Li (2012), we have

$$\|N^{1/2}\hat{H}^{1/2}\|^2 = \mathrm{tr}(N\hat{H}) = \mathrm{tr}\Big[R'M_{ff}^{-1}\Big(\frac{1}{N}\Lambda'\Psi^{-1}\Lambda\Big)^{-1}M_{ff}^{-1}R\Big] + o_p(1).$$

Given these results, we have that the first term dominates the remaining four terms. If R is stochastically unbounded, the right hand side of (B.1.3) will also be unbounded. However, the left hand side is an identity matrix. A contradiction is obtained. So $R = O_p(1)$, which means $\|N^{1/2}\hat{H}^{1/2} = O_p(1)\|$ by Lemma B.11(a). This completes the proof. □

Lemma B.12. *Under Assumptions A-D with $\|N^{1/2}\hat{H}^{1/2}\| = O_p(1)$, we have*

$$(a) \quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda\frac{1}{T}\sum_{t=1}^{T}f_t u_{jt}'\Big\|^2 = O_p(T^{-1})$$

$$(b) \quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Lambda_j\Big\|^2 = O_p(T^{-1})$$

$$(c) \quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t u_{jt}' - E(u_t u_{jt}')]\Big\|^2 = O_p(T^{-1})$$

$$(d) \quad \frac{1}{N}\sum_{j=1}^{N}\Big\|\hat{H}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj} - \Sigma_{jj})\Big\|^2 = o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj} - \Sigma_{jj}\|^2\Big)$$

Proof of Lemma B.12. Consider (a), which is bounded by

$$\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda\| \cdot \frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t u_{jt}'\Big\|^2$$

By Lemma B.11(a) and Proposition B.1.1, we have $\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda\| = O_p(1)$. So we have (a).

Consider (b), which is bounded by

$$\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Big\|^2\Big(\frac{1}{N}\sum_{j=1}^{N}\|\Lambda_j\|^2\Big)$$

which is $O_p(T^{-1})$ by (B.1.2). Then (b) follows.

Consider (c). The left hand side of (c) is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\Big\|^2\Big)$$

which is $O_p(T^{-1})$ by Proposition B.1.1. Then (c) follows.

Consider (d). The left hand side of (d) is bounded by

$$C\|\hat{H}^{1/2}\|^2\frac{1}{N}\sum_{j=1}^{N}\left(\|\hat{H}^{1/2}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{1/2}\|^2\cdot\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2\right)$$

Since $\sum_{j=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1/2}\|^2 = r$, the above expression is bounded by

$$Cr\|N^{1/2}\hat{H}^{1/2}\|^2\frac{1}{N^2}\sum_{j=1}^{N}\left(\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2\right)$$

which is $\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2\cdot O_p(N^{-1})$ by Proposition B.1.1. Thus we have (d). $\square$

Proposition B.1.2. *Under Assumptions A-D, we have*

$$\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j - R'\Lambda_j\|^2 = O_p(T^{-1}), \qquad \frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}-\Sigma_{jj}\|^2 = O_p(T^{-1}).$$

Proof of Proposition B.1.2: Consider (2.3.4), which is equivalent to

$$(\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})\hat{\Lambda}_j = (\hat{\Lambda}\hat{\Psi}^{-1}\Lambda)M_{ff}\Lambda_j + (\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_t u_{jt}' \tag{B.1.4}$$
$$+\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Lambda_j + \hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t u_{jt}' - E(u_t u_{jt}')] - \hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj})$$

So we have

$$\hat{\Lambda}_j - R'\Lambda_j = R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T}f_t u_{jt}' + \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Lambda_j \tag{B.1.5}$$
$$+\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t u_{jt}' - E(u_t u_{jt}')] - \hat{H}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj}-\Sigma_{jj})$$

where $R' = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}(\hat{\Lambda}'\hat{\Psi}^{-1}\Lambda)M_{ff}$ and $\hat{H} = (\hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda})^{-1}$. We use a_{j1}, a_{j2}, a_{j3} and a_{j4} to denote the right hand side of (B.1.5). By triangular inequality,

$$\|\hat{\Lambda}_j - R'\Lambda_j\| \leq \|a_{j1}\| + \|a_{j2}\| + \|a_{j3}\| + \|a_{j4}\|$$

Then we have

$$\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j - R'\Lambda_j\|^2 \leq 4\frac{1}{N}\sum_{j=1}^{N}\Big(\|a_{j1}\|^2 + \cdots + \|a_{j4}\|^2\Big)$$

Using the results in Lemma B.12, we have

$$\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j - R'\Lambda_j\|^2 = O_p(T^{-1}) + o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj} - \Sigma_{jj}\|^2\Big) \tag{B.1.6}$$

Consider (2.3.5), which can be written as

$$\hat{\Sigma}_{ii} - \Sigma_{ii} = \frac{1}{T}\sum_{t=1}^{T}(u_{it}u_{it}' - \Sigma_{ii}) + \Lambda_i'\Big(\frac{1}{T}\sum_{t=1}^{T}f_t u_{it}'\Big) + \Big(\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'\Big)\Lambda_i \tag{B.1.7}$$
$$-\Lambda_i'R(\hat{\Lambda}_i - R'\Lambda_i) - (\hat{\Lambda}_i - R'\Lambda_i)'R'\Lambda_i - (\hat{\Lambda}_i - R'\Lambda_i)'(\hat{\Lambda}_i - R'\Lambda_i) - \Lambda_i'(RR' - M_{ff})\Lambda_i$$

We use $b_{i1}, b_{i2}, \ldots, b_{i7}$ to denote the seven terms on the right hand side. By the Cauchy-Schwarz inequality, we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 \leq 7\frac{1}{N}\sum_{i=1}^{N}\Big(\|b_{i1}\|^2 + \cdots + \|b_{i7}\|^2\Big) \tag{B.1.8}$$

The first three terms are all $O_p(T^{-1})$. Consider the fourth term, which is bounded in norm by

$$C\|R\|^2\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j - R'\Lambda_j\|^2 = O_p(T^{-1}) + o_p\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj} - \Sigma_{jj}\|^2\Big)$$

by $R = O_p(1)$ and (B.1.6). The fifth is just the transpose of the fourth. The sixth is $O_p(T^{-2}) + o_p\big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\Sigma}_{jj} - \Sigma_{jj}\|^2\big)$, which can be verified by substituting (B.1.5) in it. Consider the last term. Since the last term is bounded in norm by $\|RR' - M_{ff}\|^2\frac{1}{N}\sum_{i=1}^{N}\|\Lambda_i\|^4$, it suffices to consider the term $RR' - M_{ff}$, which we will show to be $O_p(T^{-1/2}) + o_p([\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2})$. For ease of exposition, we use S to denote the last fourth terms of (B.1.3). By Lemma B.11 together with $\|N^{1/2}\hat{H}^{1/2}\| = O_p(1)$, we have

$$S = O_p(T^{-1/2}) + o_p\big(\big[\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\big]^{1/2}\big).$$

Now equation (B.1.3) can be written as $I_r = R'M_{ff}^{-1}R + S$, which is equivalent to $RR' - M_{ff} = -RSR^{-1}M_{ff}$. Since $R = O_p(1)$, if $R \neq o_p(1)$, then $R^{-1} = O_p(1)$. However, R is impossible to be $o_p(1)$ since $I_r = R'M_{ff}^{-1}R + o_p(1)$. So we have

$$RR' - M_{ff} = O_p(T^{-1/2}) + o_p([\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2]^{1/2}), \tag{B.1.9}$$

implying that the last term is $O_p(T^{-1}) + o_p(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2)$. Given the above results, we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = O_p(T^{-1}) + o_p\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\big)$$

which implies $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2 = O_p(T^{-1})$. Substituting this result into (B.1.6), we have the remaining result of the proposition. This completes the proof of this proposition. □

To prove Theorem 2.3.1, we further need the following two lemmas.

Lemma B.13. *Under Assumptions A-D,*

$$(a)\quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

$$(b)\quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t u_{jt}' - E(u_t u_{jt}')] = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

$$(c)\quad \hat{H}\hat{\Lambda}_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj} - \Sigma_{jj}) = O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}) + \|\hat{\Lambda}_j - R'\Lambda_j\| \cdot o_p(1)$$

$$(d)\quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Big[\frac{1}{T}\sum_{t=1}^{T}(u_t u_t' - \Psi)\Big]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = O_p(N^{-1}T^{-1/2}) + O_p(T^{-1})$$

$$(e)\quad \hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}(\hat{\Psi} - \Psi)\hat{\Psi}^{-1}\hat{\Lambda}\hat{H} = O_p(N^{-1/2}T^{-1/2})$$

Proof of Lemma B.13: Consider (a). The left hand side of (a) is equal to

$$\hat{H}\sum_{i=1}^{N}(\hat{\Lambda}_i - R'\Lambda_i)\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t' + \hat{H}R'\sum_{i=1}^{N}\Lambda_i(\hat{\Sigma}_{ii}^{-1} - \Sigma_{ii}^{-1})\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'$$
$$+ \hat{H}R'\sum_{i=1}^{N}\Lambda_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'$$

The first term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'\Big\|^2\Big)^{1/2}$$

which is $O_p(T^{-1})$ by Proposition B.1.1 and B.1.2. The second term is bounded in norm by

$$C^3\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}u_{it}f_t'\Big\|^2\Big)^{1/2}$$

which is also $O_p(T^{-1})$ by Proposition B.1.1 and B.1.2. The third term is $O_p(N^{-1/2}T^{-1/2})$ by Proposition B.1.1. So we have (a).

Consider (b). The left hand side (b) is equal to

$$\begin{aligned}&\hat{H}\sum_{i=1}^{N}(\hat{\Lambda}_i - R'\Lambda_i)\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\\ &+\hat{H}R'\sum_{i=1}^{N}\Lambda_i(\hat{\Sigma}_{ii}^{-1} - \Sigma_{ii}^{-1})\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\\ &+\hat{H}R'\sum_{i=1}^{N}\Lambda_i\Sigma_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')].\end{aligned}$$

The first term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\Big\|^2\Big)^{1/2}$$

which is $O_p(T^{-1})$ by Proposition B.1.1 and B.1.2. The second term is bounded in norm by

$$C^3\|N^{1/2}\hat{H}^{1/2}\|^2\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\Big\|^2\Big)^{1/2}$$

which is also $O_p(T^{-1})$ by Proposition B.1.1 and B.1.2. The third term is $O_p(T^{-1/2}T^{-1/2})$ by Proposition B.1.1. Given these results, we have (b).

Consider (c). The left hand side of (a) is equal to

$$\hat{H}(\hat{\Lambda}_j - R'\Lambda_j)\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj} - \Sigma_{jj}) + \hat{H}R'\Lambda_j\hat{\Sigma}_{jj}^{-1}(\hat{\Sigma}_{jj} - \Sigma_{jj}) \tag{B.1.10}$$

By the boundedness of $\hat{\Sigma}_{jj}, \Sigma_{jj}$ and $\hat{H} = O_p(N^{-1})$ (since $\|N^{1/2}\hat{H}^{1/2}\| = O_p(1)$), we have the first term is $\|\hat{\Lambda}_j - R'\Lambda_j\| \cdot o_p(1)$. Consider the second term of (B.1.10). Substituting (B.1.7) into the second term, we obtain an expression consisting of 7 terms. The first three terms are all $O_p(N^{-1}T^{-1/2})$ by the boundedness of $\hat{\Sigma}_{ii}$ and $\hat{H} = O_p(N^{-1})$. The fourth and fifth terms are both $\|\hat{\Lambda}_j - R\Lambda_j\| \cdot o_p(1)$. The sixth term is equal to

$$\hat{H}R'\Lambda_j\hat{\Sigma}_{jj}^{-1}(\hat{\Lambda}_j - R'\Lambda_j)'(\hat{\Lambda}_j - R'\Lambda_j)$$

which is bounded by

$$\|N^{1/2}\hat{H}^{1/2}\| \cdot \|R\| \cdot \|\hat{\Sigma}_{jj}^{-1}\| \cdot \frac{1}{N}\sum_{j=1}^{N}\|\hat{\Lambda}_j - R'\Lambda_j\|^2$$

By Propositions B.1.1 and B.1.2 and the boundedness of $\hat{\Sigma}_{ii}$, the above expression is $O_p(T^{-1})$. The seventh term is $O_p(N^{-1}T^{-1/2})$ since $S = O_p(T^{-1/2})$. Given these results, we have the second term of (B.1.10) is $O_p(N^{-1}T^{-1/2}) + O_p(T^{-1}) + \|\hat{\Lambda}_j - R'\Lambda_j\| \cdot o_p(1)$. Then (c) follows.

Consider (d). The left hand side of (d) is equal to

$$\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt} - E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Lambda}_j'\hat{H}$$

which is equivalent to

$$\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\Lambda}_i - R'\Lambda_i)\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Lambda}_j'\hat{H}$$
$$\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}' - E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}(\hat{\Lambda}_j - R'\Lambda_j)'\hat{H}$$

$$\hat{H}R' \sum_{i=1}^{N} \sum_{j=1}^{N} \Lambda_i (\hat{\Sigma}_{ii}^{-1} - \Sigma_{ii}^{-1}) \frac{1}{T} \sum_{t=1}^{T} [u_{it}u'_{jt} - E(u_{it}u'_{jt})] \hat{\Sigma}_{jj}^{-1} \Lambda'_j R \hat{H} \tag{B.1.11}$$

$$\hat{H}R' \sum_{i=1}^{N} \sum_{j=1}^{N} \Lambda_i \Sigma_{ii}^{-1} \frac{1}{T} \sum_{t=1}^{T} [u_{it}u'_{jt} - E(u_{it}u'_{jt})] (\hat{\Sigma}_{jj}^{-1} - \Sigma_{jj}^{-1}) \Lambda'_j R \hat{H}$$

$$\hat{H}R' \sum_{i=1}^{N} \sum_{j=1}^{N} \Lambda_i \Sigma_{ii}^{-1} \frac{1}{T} \sum_{t=1}^{T} [u_{it}u'_{jt} - E(u_{it}u'_{jt})] \Sigma_{jj}^{-1} \Lambda'_j R \hat{H}$$

The first term is bounded in norm by

$$C \cdot \|N^{1/2}\hat{H}^{1/2}\|^3 \cdot \big(\frac{1}{N} \sum_{i=1}^{N} \|\hat{\Lambda}_i - R'\Lambda_i\|^2\big)^{1/2} \big(\sum_{j=1}^{N} \|\hat{\Sigma}_{jj}^{-1/2} \hat{\Lambda}_j \hat{H}^{1/2}\|^2\big)^{1/2}$$
$$\times \big(\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \big\|\frac{1}{T} \sum_{t=1}^{T} [u_{it}u'_{jt} - E(u_{it}u'_{jt})]\big\|^2\big)^{1/2} = O_p(T^{-1})$$

by Proposition B.1.1 and (B.1.1). The second term is bounded in norm by

$$C \cdot \|N^{1/2}\hat{H}^{1/2}\|^4 \cdot \|R\| \cdot \big(\frac{1}{N} \sum_{i=1}^{N} \|\Lambda_i\|^2\big)^{1/2} \big(\frac{1}{N} \sum_{j=1}^{N} \|\hat{\Lambda}_j - R'\Lambda_j\|^2\big)^{1/2}$$
$$\times \big(\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \big\|\frac{1}{T} \sum_{t=1}^{T} [u_{it}u'_{jt} - E(u_{it}u'_{jt})]\big\|^2\big)^{1/2} = O_p(T^{-1})$$

by Propositions B.1.1 and B.1.2. The third and fourth terms are both bounded in norm by

$$C \cdot \|N^{1/2}\hat{H}^{1/2}|^4 \cdot \|R\|^2 \cdot \big(\frac{1}{N} \sum_{i=1}^{N} \|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\big)^{1/2} \big(\frac{1}{N} \sum_{j=1}^{N} \|\Lambda_j\|^2\big)^{1/2}$$
$$\times \big(\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \big\|\frac{1}{T} \sum_{t=1}^{T} [u_{it}u'_{jt} - E(u_{it}u'_{jt})]\big\|^2\big)^{1/2} = O_p(T^{-1})$$

by Propositions B.1.1 and B.1.2. The last term is $O_p(N^{-1}T^{-1/2})$. Given these results, we have (d).

Consider (e). The left hand side of (c) is equal to

$$\hat{H}^{1/2}\big(\sum_{i=1}^{N}\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}^{1/2}\big)\hat{H}^{1/2}.$$

The above expression is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|\big).$$

Since $\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 = r$, then $\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\| \leq \sqrt{r}$ uniformly in i. So the above expression is further bounded by

$$C\sqrt{r}\|N^{1/2}\hat{H}^{1/2}\|^2\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|\big).$$

By the Cauchy-Schwarz inequality, the preceding expression is bounded by

$$C\sqrt{r}\|N^{1/2}\hat{H}^{1/2}\|^2\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\big)^{1/2}\big(\frac{1}{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\big)^{1/2},$$

which is $O_p(N^{-1/2}T^{-1/2})$ by Propositions B.1.1 and B.1.2 and (B.1.1). Then (e) follows. □

Lemma B.14. *Under Assumptions A-D, we have*

$$RR' - M_{ff} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Proof of Lemma B.14. Consider (B.1.3). Given the results in Lemma B.13, we have

$$R' M_{ff}^{-1} R = I_r + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Taking inverse on the both sides yields

$$R^{-1} M_{ff} R^{-1\prime} = I_r + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

Pre-multiplying R and post-multiplying R', together with $R = O_p(1)$, we have Lemma B.14.

□

Proof of Theorem 2.3.1: Consider (B.1.5). The last three terms of the right hand side of (B.1.5) are summarized in Lemma B.13(a)-(c). So we have

$$\hat{\Lambda}_j - R'\Lambda_j = R' M_{ff}^{-1} \frac{1}{T} \sum_{t=1}^{T} f_t u_{jt}' + \|\hat{\Lambda}_j - R'\Lambda_j\| \cdot o_p(1) + o_p(T^{-1/2}). \tag{B.1.12}$$

The first term of the right hand side is $O_p(T^{-1/2})$. The second term is of smaller order term than the left hand side and hence negligible. Given this result, we have

$$\hat{\Lambda}_j - R'\Lambda_j = O_p(T^{-1/2}). \tag{B.1.13}$$

Substituting (B.1.13) into (B.1.12), we have

$$\hat{\Lambda}_j - R'\Lambda_j = R' M_{ff}^{-1} \frac{1}{T} \sum_{t=1}^{T} f_t u_{jt}' + o_p(T^{-1/2}).$$

Now consider (B.1.7). Substituting (B.1.5) into (B.1.7), we have

$$
\begin{aligned}
\hat{\Sigma}_{ii} - \Sigma_{ii} = {} & \frac{1}{T}\sum_{t=1}^{T}(u_{it}u_{it}' - \Sigma_{ii}) - (\hat{\Lambda}_i - R'\Lambda_i)'(\hat{\Lambda}_i - R'\Lambda_i) - \Lambda_i'(RR' - M_{ff})\Lambda_i \\
& - \frac{1}{T}\sum_{t=1}^{T} u_{it}f_t'M_{ff}^{-1}(RR' - M_{ff})\Lambda_i - \Lambda_i'(RR' - M_{ff})M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t u_{it}' \\
& - \Lambda_i' R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T} u_t f_t'\Lambda_i - \Lambda_i' R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t u_{it}' - E(u_t u_{it}')] \qquad \text{(B.1.14)} \\
& - \Lambda_i'\frac{1}{T}\sum_{t=1}^{T} f_t u_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}R'\Lambda_i - \frac{1}{T}\sum_{t=1}^{T}[u_{it}u_t' - E(u_{it}u_t')]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}R'\Lambda_i \\
& + \Lambda_i' R\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii}) + (\hat{\Sigma}_{ii} - \Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}R'\Lambda_i
\end{aligned}
$$

The second term is $O_p(T^{-1})$ by (B.1.13). The third term is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma B.14. The fourth and fifth terms are both $O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2})$ by Lemma B.14. The sixth and eighth terms are both $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma B.13(a). The seventh and ninth terms are also $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma B.13(b). The last two terms are of smaller order terms than the left hand side and hence negligible. Given these results, we have

$$
\hat{\Sigma}_{ii} - \Sigma_{ii} = \frac{1}{T}\sum_{t=1}^{T}(u_{it}u_{it}' - \Sigma_{ii}) + o_p(T^{-1/2})
$$

This completes the proof of Theorem 2.3.1. □

B.2 Proofs of the results in Section 2.5

Lemma B.2.1. *Under Assumptions A-D, we have*

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\| = O_p(T^{-1}).$$

Proof of Lemma B.2.1. Notice

$$\begin{aligned}\hat{\beta}_i^{CV}-\beta_i &= \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}-\Sigma_{i,22}^{-1}\Sigma_{i,21}\\ &= \hat{\Sigma}_{i,22}^{-1}\big[(\hat{\Sigma}_{i,21}-\Sigma_{i,21})-(\hat{\Sigma}_{i,22}-\Sigma_{i,22})\Sigma_{i,22}^{-1}\Sigma_{i,21}\big]\end{aligned}$$

By the boundedness of $\hat{\Sigma}_{ii}, \Sigma_{ii}$, we have $\|\hat{\Sigma}_{i,22}^{-1}\| < C$, $\|\Sigma_{i,22}^{-1}\Sigma_{i,21}\| < C$. Then

$$\|\hat{\beta}_i^{CV}-\beta_i\|^2 \le C\|\hat{\Sigma}_{i,21}-\Sigma_{i,21}\|^2 + C\|\hat{\Sigma}_{i,22}-\Sigma_{i,22}\|^2 \le 2C\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2$$

So $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\| = O_p(T^{-1})$ by $\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2 = O_p(T^{-1})$. □

Lemma B.2.2. *Under Assumptions A-D, we have*

$$\begin{aligned}
&(a)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}\hat{\Lambda}_{i,11}'-\Lambda_{i,21}^{*}\Lambda_{i,11}^{*\prime}) = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})\\
&(b)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\Lambda}_{i,11}'-\Lambda_{i,22}^{*}\beta_i\Lambda_{i,11}^{*\prime}) = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})\\
&(c)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}'-\Lambda_{i,21}^{*}\beta_i'\Lambda_{i,12}^{*\prime}) = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})\\
&(d)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}'-\Lambda_{i,22}^{*}\beta_i\beta_i'\Lambda_{i,12}^{*\prime}) = O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})
\end{aligned}$$

where $\Lambda_j^* = R'\Lambda_j$ *and* $\Lambda_{i,pq}^*$ *is defined similarly as* $\Lambda_{i,pq}$.

PROOF OF LEMMA B.2.2. By (B.1.5), we have

$$\begin{aligned}\hat{\Lambda}_{i,11} - \Lambda_{i,11}^* &= \boldsymbol{v}_1 R' M_{ff}^{-1} \frac{1}{T}\sum_{t=1}^T f_t e_{it} + \boldsymbol{v}_1 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^T u_t f_t' \Lambda_i \boldsymbol{w}_1\Big) \\ &+ \boldsymbol{v}_1 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^T [u_t e_{it} - E(u_t e_{it})] - \boldsymbol{v}_1 \hat{H}\hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_1\end{aligned} \tag{B.2.1}$$

$$\begin{aligned}\hat{\Lambda}_{i,21} - \Lambda_{i,21}^* &= \boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{T}\sum_{t=1}^T f_t e_{it} + \boldsymbol{v}_2 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^T u_t f_t' \Lambda_i \boldsymbol{w}_1\Big) \\ &+ \boldsymbol{v}_2 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^T [u_t e_{it} - E(u_t e_{it})] - \boldsymbol{v}_2 \hat{H}\hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_1\end{aligned} \tag{B.2.2}$$

$$\begin{aligned}\hat{\Lambda}_{i,12} - \Lambda_{i,12}^* &= \boldsymbol{v}_1 R' M_{ff}^{-1} \frac{1}{T}\sum_{t=1}^T f_t v_{it}' + \boldsymbol{v}_1 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^T u_t f_t' \Lambda_i \boldsymbol{w}_2\Big) \\ &+ \boldsymbol{v}_1 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^T [u_t v_{it}' - E(u_t v_{it}')] - \boldsymbol{v}_1 \hat{H}\hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_2\end{aligned} \tag{B.2.3}$$

$$\begin{aligned}\hat{\Lambda}_{i,22} - \Lambda_{i,22}^* &= \boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{T}\sum_{t=1}^T f_t v_{it}' + \boldsymbol{v}_2 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^T u_t f_t' \Lambda_i \boldsymbol{w}_2\Big) \\ &+ \boldsymbol{v}_2 \hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^T [u_t v_{it}' - E(u_t v_{it}')] - \boldsymbol{v}_2 \hat{H}\hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_2\end{aligned} \tag{B.2.4}$$

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are defined as $I_r = [\boldsymbol{v}_1, \boldsymbol{v}_2]$ with $\boldsymbol{v}_1$ an $r \times r_1$ matrix and $\boldsymbol{v}_2$ an $r \times r_2$ matrix, respectively. $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are defined as $I_{K+1} = [\boldsymbol{w}_1, \boldsymbol{w}_2]$ with $\boldsymbol{w}_1$ an $(K+1) \times 1$ vector and $\boldsymbol{w}_2$ an $(K+1) \times K$ matrix. In addition, $e_{it} = \epsilon_{it} + v_{it}'\beta_i$.

Consider (a). The left hand side of (a) is equivalent to

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^N (\hat{\Lambda}_{i,21} - \Lambda_{i,21}^*)\Lambda_{i,11}^{*\prime} + \frac{1}{N}\sum_{i=1}^N \Lambda_{i,21}^*(\hat{\Lambda}_{i,11} - \Lambda_{i,11}^*)' \\ &+ \frac{1}{N}\sum_{i=1}^N (\hat{\Lambda}_{i,21} - \Lambda_{i,21}^*)(\hat{\Lambda}_{i,11} - \Lambda_{i,11}^*)' = ii_1 + ii_2 + ii_3 \quad \text{say}\end{aligned}$$

Consider ii_1. By (B.2.2), we have

$$ii_1 = \boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_t e_{it} \Lambda_{i,11}^{*\prime} + \boldsymbol{v}_2 \hat{H} \hat{\Lambda} \hat{\Psi}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [u_t e_{it} - E(u_t e_{it})] \Lambda_{i,11}^{*\prime}$$
$$+ \boldsymbol{v}_2 \hat{H} \hat{\Lambda} \hat{\Psi}^{-1} \frac{1}{T} \sum_{t=1}^{T} u_t f_t' \Big(\frac{1}{N} \sum_{i=1}^{N} \Lambda_i \boldsymbol{w}_1 \Lambda_{i,11}^{*\prime} \Big) - \boldsymbol{v}_2 \frac{1}{N} \sum_{i=1}^{N} \hat{H} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1} (\hat{\Sigma}_{ii} - \Sigma_{ii}) \boldsymbol{w}_1 \Lambda_{i,11}^{*\prime}$$
$$= iii_1 + iii_2 + iii_3 - iii_4$$

Consider iii_1. By $\Lambda_{i,11}^{*\prime} = \Lambda_{i,11}' R_{11} + \Lambda_{i,21}' R_{21}$, we have

$$iii_1 = \boldsymbol{v}_2 R' M_{ff}^{-1} \Big(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_t e_{it} \Lambda_{i,11}' \Big) R_{11} + \boldsymbol{v}_2 R' M_{ff}^{-1} \Big(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_t e_{it} \Lambda_{i,21}' \Big) R_{21}$$

which is $O_p(N^{-1/2} T^{-1/2})$ by $R = O_p(1)$. Consider iii_2, which is equivalent to

$$\boldsymbol{v}_2 \hat{H} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1} \frac{1}{T} \sum_{t=1}^{T} [u_{it} e_{jt}' - E(u_{it} e_{jt})] \Lambda_{j,11}^{*\prime}$$

which is bounded in norm by

$$C \| N^{1/2} \hat{H}^{1/2} \| \cdot \Big[\sum_{i=1}^{N} \| \hat{H}^{1/2} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1/2} \|^2 \Big]^{-1/2} \Big[\frac{1}{N} \sum_{i=1}^{N} \Big\| \frac{1}{NT} \sum_{j=1}^{N} \sum_{t=1}^{T} [u_{it} e_{jt}' - E(u_{it} e_{jt}')] \Lambda_{j,11}^{*\prime} \Big\|^2 \Big]^{1/2}$$

By $\Lambda_{j,11}^{*\prime} = \Lambda_{j,11}' R_{11} + \Lambda_{j,21}' R_{21}$, we have

$$\frac{1}{NT} \sum_{j=1}^{N} \sum_{t=1}^{T} [u_{it} e_{jt} - E(u_{it} e_{jt})] \Lambda_{j,11}^{*\prime} = \frac{1}{NT} \sum_{j=1}^{N} \sum_{t=1}^{T} [u_{it} e_{jt} - E(u_{it} e_{jt})] \Lambda_{j,11}' R_{11}$$
$$+ \frac{1}{NT} \sum_{j=1}^{N} \sum_{t=1}^{T} [u_{it} e_{jt} - E(u_{it} e_{jt})] \Lambda_{j,21}' R_{21} = O_p(N^{-1/2} T^{-1/2}).$$

Given this result, together with $\|N^{1/2}\hat{H}^{1/2}\| = O_p(1)$ and (B.1.1), we have $iii_2 = O_p(N^{-1/2}T^{-1/2})$.

Consider iii_3. Notice that

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_1 \Lambda_{i,11}^{*\prime} = \Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_1 \Lambda_{i,11}'\Big)R_{11} + \Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_1 \Lambda_{i,21}'\Big)R_{21} = O_p(1).$$

Given the above result, together with Lemma B.13(a), we obtain $iii_3 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Consider iii_4, which is equal to

$$\Big[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{v}_2\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_1\Lambda_{i,11}'\Big]R_{11}+\Big[\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{v}_2\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_1\Lambda_{i,21}'\Big]R_{21}. \quad \text{(B.2.5)}$$

Consider the first term of the above expression. Ignore R_{11} and $\boldsymbol{v}_2$, the expression in the bracket is bounded in norm by

$$\frac{1}{N}\|\hat{H}^{1/2}\|\sum_{i=1}^{N}\big(\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|\cdot\|\hat{\Sigma}_{ii}^{-1/2}\|\cdot\|\boldsymbol{w}_1\Lambda_{i,11}'\|\cdot\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|\big)$$

which is further bounded by

$$\frac{1}{N}\|N^{1/2}\hat{H}^{1/2}\|\Big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\cdot\|\hat{\Sigma}_{ii}^{-1/2}\|^2\cdot\|\boldsymbol{w}_1\Lambda_{i,11}'\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)^{1/2}$$

The above expression is $O_p(N^{-1}T^{-1/2})$ by $\|\hat{\Sigma}_{ii}^{-1/2}\| < C, \|\boldsymbol{w}_1\Lambda_{i,11}'\| < C$ and Propositions B.1.1 and B.1.2 as well as (B.1.1). Given this result, together with $R = O_p(1)$, we have the first term of (B.2.5) is $O_p(N^{-1}T^{-1/2})$. The second term can be proved to be $O_p(N^{-1}T^{-1/2})$ similarly as the first term. So we have $iii_4 = O_p(N^{-1}T^{-1/2})$. Summarizing all the results, we have $ii_1 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Term ii_2 can be proved to be $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ similarly as ii_1 and the details are omitted. For term ii_3, notice that it is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_{i,21} - \Lambda^*_{i,21}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_{i,11} - \Lambda^*_{i,11}\|^2\Big)^{1/2}$$

However, we have $\|\hat{\Lambda}_{i,21} - \Lambda^*_{i,21}\| \leq \|\hat{\Lambda}_i - \Lambda^*_i\| = \|\hat{\Lambda}_i - R'\Lambda_i\|$ and $\|\hat{\Lambda}_{i,11} - \Lambda^*_{i,11}\| \leq \|\hat{\Lambda}_i - \Lambda^*_i\| = \|\hat{\Lambda}_i - R'\Lambda_i\|$. Given this result, the preceding expression is bounded by

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2 = O_p(T^{-1})$$

by Proposition B.1.2. Summarizing the results on ii_1, ii_2 and ii_3, we obtain (a).

Consider (b). The left hand side of (b) can be written as

$$\begin{aligned}&\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\Lambda}'_{i,11} - \frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,22}\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11}\Big)\\ &+\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,22}\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11} - \frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,22}\beta_i\Lambda^{*\prime}_{i,11}\Big) = ii_4 + ii_5, \quad \text{say}\end{aligned}$$

Consider ii_4, which is equal to

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22} - \Lambda^*_{i,22})\hat{\beta}_i^{CV}\Lambda^{*\prime}_{i,11} + \frac{1}{N}\sum_{i=1}^{N}\Lambda^*_{i,22}\hat{\beta}_i^{CV}(\hat{\Lambda}_{i,11} - \Lambda^*_{i,11})'\\ &+\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22} - \Lambda^*_{i,22})\hat{\beta}_i^{CV}(\hat{\Lambda}_{i,11} - \Lambda^*_{i,11})' = iii_5 + iii_6 + iii_7, \quad \text{say}\end{aligned}$$

Consider iii_5. By (B.2.4), we have

$$iii_5 = \boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_t v_{it}' \hat{\beta}_i^{CV} \Lambda_{i,11}^{*\prime} + \boldsymbol{v}_2 \hat{H} \hat{\Lambda} \hat{\Psi}^{-1} \frac{1}{T} \sum_{t=1}^{T} u_t f_t' \Big(\frac{1}{N} \sum_{i=1}^{N} \Lambda_i \boldsymbol{w}_2 \hat{\beta}_i^{CV} \Lambda_{i,11}^{*\prime} \Big)$$
$$+ \boldsymbol{v}_2 \hat{H} \hat{\Lambda} \hat{\Psi}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} [u_t v_{it}' - E(u_t v_{it}')] \hat{\beta}_i^{CV} \Lambda_{i,11}^{*\prime}$$
$$- \boldsymbol{v}_2 \hat{H} \frac{1}{N} \sum_{i=1}^{N} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1} (\hat{\Sigma}_{ii} - \Sigma_{ii}) \boldsymbol{w}_2 \hat{\beta}_i^{CV} \Lambda_{i,11}^{*\prime} \qquad \text{(B.2.6)}$$

Consider the first term of (B.2.6), which can be written as

$$\boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_t v_{it}' \beta_i \Lambda_{i,11}^{*\prime} + \boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_t v_{it}' (\hat{\beta}_i^{CV} - \beta_i) \Lambda_{i,11}^{*\prime}. \qquad \text{(B.2.7)}$$

Treating $v_{it}'\beta_i$ as a new e_{it}, the first term of the above expression can be proved to be $O_p(N^{-1/2}T^{-1/2})$ similarly as the iii_1. By $\Lambda_{i,11}^{*\prime} = \Lambda_{i,11}' R_{11} + \Lambda_{i,21}' R_{21}$, the second term is equal to

$$\boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_t v_{it}' (\hat{\beta}_i^{CV} - \beta_i) \Lambda_{i,11}' R_{11} + \boldsymbol{v}_2 R' M_{ff}^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} f_t v_{it}' (\hat{\beta}_i^{CV} - \beta_i) \Lambda_{i,21}' R_{21}.$$

The first term of the above expression is bounded in norm by

$$\|\boldsymbol{v}_2 R'\| \cdot \|M_{ff}^{-1}\| \cdot \|\Lambda_{i,11}\| \cdot \|R_{11}\| \cdot \Big(\frac{1}{N} \sum_{i=1}^{N} \|\hat{\beta}_i^{CV} - \beta_i\|^2 \Big)^{1/2} \Big(\frac{1}{N} \sum_{i=1}^{N} \Big\| \frac{1}{T} \sum_{t=1}^{T} f_t v_{it}' \Big\|^2 \Big)^{1/2},$$

which is $O_p(T^{-1})$ by Lemma B.2.1 and $R = O_p(1)$. The second term can be proved similarly as the first term. Given these results, we have the expression of (B.2.7) is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. So the first term of (B.2.6) is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Consider the second term. First note that

$$\|\hat{\beta}_i^{CV}\| < C, \qquad \forall\, i \leq N \tag{B.2.8}$$

The above result is due to the boundedness of $\hat{\Sigma}_{ii}$ and $\hat{\beta}_i^{CV} = \hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}$. Given this result, we have

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i \boldsymbol{w}_2 \hat{\beta}_i^{CV}\Lambda_{i,11}^{*\prime} = O_p(1).$$

Given the above result, together with Lemma B.13(a), we have the second term of (B.2.6) is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Consider the third term, which is equal to

$$\boldsymbol{v}_2 \frac{1}{NT}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime} \tag{B.2.9}$$

Ignore $\boldsymbol{v}_2$, The above expression can be rewritten as

$$\frac{1}{NT}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}(\hat{\Lambda}_i - R'\Lambda_i)\hat{\Sigma}_{ii}^{-1}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}$$
$$\frac{1}{NT}\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i(\hat{\Sigma}_{ii}^{-1} - \Sigma_{ii}^{-1})\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}$$
$$\frac{1}{NT}\hat{H}R'\sum_{i=1}^{N}\sum_{j=1}^{N}\Lambda_i\Sigma_{ii}^{-1}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}$$

The first term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot \Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}\|^2\Big)^{1/2} \\ \times\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\Big\|^2\Big)^{1/2},$$

By the boundedness of $\hat{\beta}_i^{CV}$ and $\Lambda_{i,11}^{*\prime} = \Lambda_{i,11}'R_{11} + \Lambda_{i,21}'R_{21}$,

$$\frac{1}{N}\sum_{j=1}^{N}\|\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}\|^2 \leq C\frac{1}{N}\sum_{j=1}^{N}\|\Lambda_{j,11}^{*\prime}\|^2 \\ \leq 2C\big(\|R_{11}\|^2\frac{1}{N}\sum_{j=1}^{N}\|\Lambda_{j,11}\|^2 + \|R_{21}\|^2\frac{1}{N}\sum_{j=1}^{N}\|\Lambda_{j,21}\|^2\big)$$

Given this result, we have that the first term is $O_p(T^{-1})$ by Propositions B.1.1 and (B.1.2).

The second term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot \|R\| \cdot \Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}\|^2\Big)^{1/2} \\ \times\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\Big\|^2\Big)^{1/2},$$

which is $O_p(T^{-1})$ by the same arguments. The last term is bounded in norm by

$$\|N^{1/2}\hat{H}^{1/2}\|^2 \cdot \|R\| \cdot \big(\frac{1}{N}\sum_{j=1}^{N}\|\hat{\beta}_j^{CV\prime}\Lambda_{j,11}^{*\prime}\|^2\big)^{1/2} \\ \times\big(\frac{1}{N}\sum_{j=1}^{N}\big\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_i\Sigma_{ii}^{-1}[u_{it}v_{jt}' - E(u_{it}v_{jt}')]\big\|^2\big)^{1/2}$$

which is $O_p(N^{-1/2}T^{-1/2})$. Given these results, we have the third term of (B.2.6) is $O_p(N^{-1/2}T^{-1/2})+$ $O_p(T^{-1})$.

Consider the fourth term. Ignore $\boldsymbol{v}_2$, this term is bounded in norm by

$$C\|\hat{H}^{1/2}\|\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\cdot\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}\Lambda_{i,11}^{*\prime}\|^2\Big)^{1/2}$$

Notice $\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 = r$. So $\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2 \leq r$ for all i. This leads to

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\cdot\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2 \leq r\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2.$$

Given the above result, together with $\hat{H} = O_p(N^{-1})$, we have the fourth term is $O_p(N^{-1/2}T^{-1/2})$. Summarizing all the results, we have $iii_5 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Term iii_6 can be proved to be $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ similarly as iii_5 and the details are omitted. Consider iii_7, which is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_{i,22}-\Lambda_{i,22}^*\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}\|^2\cdot\|\hat{\Lambda}_{i,11}-\Lambda_{i,11}^*\|^2\Big)^{1/2}$$

By the boundedness of $\hat{\beta}_i^{CV}$, together with $\|\hat{\Lambda}_{i,22}-\Lambda_{i,22}^*\| \leq \|\hat{\Lambda}_i-\Lambda_i^*\| = \|\hat{\Lambda}_i - R'\Lambda_i\|$ and $\|\hat{\Lambda}_{i,11}-\Lambda_{i,11}^*\| \leq \|\hat{\Lambda}_i-\Lambda_i^*\| = \|\hat{\Lambda}_i - R'\Lambda_i\|$, we have that the preceding expression is bounded by

$$C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2 = O_p(T^{-1})$$

by Proposition B.1.2. Summarizing the results on iii_5, iii_6 and iii_7, we have

$$ii_4 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}).$$

We proceed to consider ii_5, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda^{*}_{i,22}(\hat{\beta}_i^{CV}-\beta_i)\Lambda^{*\prime}_{i,11}.$$

By $\Lambda^{*}_{i,22} = R'_{12}\Lambda_{i,12} + R'_{22}\Lambda_{i,22}$ and $\Lambda^{*\prime}_{i,11} = \Lambda'_{i,11}R_{11} + \Lambda'_{i,21}R_{21}$, the above expression can be written as

$$\begin{aligned} &R'_{12}\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}(\hat{\beta}_i^{CV}-\beta_i)\Lambda'_{i,11}\Big)R_{11} + R'_{12}\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}(\hat{\beta}_i^{CV}-\beta_i)\Lambda'_{i,21}\Big)R_{21} \\ &R'_{22}\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}(\hat{\beta}_i^{CV}-\beta_i)\Lambda'_{i,11}\Big)R_{11} + +R'_{22}\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}(\hat{\beta}_i^{CV}-\beta_i)\Lambda'_{i,21}\Big)R_{21}\end{aligned}$$

The derivations on the above four terms are almost the same. So we only choose the first one to illustrate. Ignore R'_{12} and R_{11}. By

$$\hat{\beta}_i^{CV}-\beta_i = \hat{\Sigma}^{-1}_{i,22}[(\hat{\Sigma}_{i,21}-\Sigma_{i,21})-(\hat{\Sigma}_{i,22}-\Sigma_{i,22})\beta_i],$$

we can rewrite the expression in the bracket as

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\hat{\Sigma}^{-1}_{i,22}(\hat{\Sigma}_{i,21}-\Sigma_{i,21})\Lambda'_{i,11} - \frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\hat{\Sigma}^{-1}_{i,22}(\hat{\Sigma}_{i,22}-\Sigma_{i,22})\beta_i\Lambda'_{i,11}.$$

Again, the derivations on the above two terms are almost the same. So we only choose the first term to illustrate. This term is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}(\hat{\Sigma}^{-1}_{i,22}-\Sigma^{-1}_{i,22})(\hat{\Sigma}_{i,21}-\Sigma_{i,21})\Lambda'_{i,11} + \frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma^{-1}_{i,22}(\hat{\Sigma}_{i,21}-\Sigma_{i,21})\Lambda'_{i,11}. \quad \text{(B.2.10)}$$

The first term of the above expression is

$$-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\hat{\Sigma}_{i,22}^{-1}(\hat{\Sigma}_{i,22}-\Sigma_{i,22})\Sigma_{i,22}^{-1}(\hat{\Sigma}_{i,21}-\Sigma_{i,21})\Lambda_{i,11}',$$

which is bounded in norm by $C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii}-\Sigma_{ii}\|^2 = O_p(T^{-1})$ by $\|\Lambda_{i,21}\| < C, \|\Lambda_{i,11}\| < C, \|\hat{\Sigma}_{i,22}^{-1}\| < C, \|\Sigma_{i,22}^{-1}\| < C$ as well as $\|\hat{\Sigma}_{i,22}-\Sigma_{i,22}\| \leq \|\hat{\Sigma}_{ii}-\Sigma_{ii}\|$ and $\|\hat{\Sigma}_{i,21}-\Sigma_{i,21}\| \leq \|\hat{\Sigma}_{ii}-\Sigma_{ii}\|$. So the first term of (B.2.10) is $O_p(T^{-1})$. Consider the second term, which, by (B.1.14), be be written as

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{i,12}\Sigma_{i,22}^{-1}[v_{it}e_{it}-E(v_{it}e_{it})]\Lambda_{i,11}' \\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'(\hat{\Lambda}_i-R'\Lambda_i)'(\hat{\Lambda}_i-R'\Lambda_i)\boldsymbol{w}_1\Lambda_{i,11}' \\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'(RR'-M_{ff})\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}' \\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}v_{it}f_t'\Big)M_{ff}^{-1}(RR'-M_{ff})\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}' \\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'(RR'-M_{ff})M_{ff}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big)\Lambda_{i,11}' \\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'R\Big(\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_tf_t'\Big)\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}' \qquad \text{(B.2.11)} \\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_te_{it}-E(u_te_{it})]\Lambda_{i,11}' \\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'\Big(\frac{1}{T}\sum_{t=1}^{T}f_tu_t'\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}\Big)R'\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}' \\
&-\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\frac{1}{T}\sum_{t=1}^{T}[v_{it}u_t'-E(v_{it}u_t')]\hat{\Psi}^{-1}\hat{\Lambda}\hat{H}R'\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}' \\
&+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'R\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_1\Lambda_{i,11}'
\end{aligned}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i'\hat{H}R'\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}'.$$

where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are defined as $I_{K+1}=[\boldsymbol{w}_1,\boldsymbol{w}_2]$ where $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are $(K+1)\times 1$ and $(K+1)\times K$, respectively. The first term is $O_p(N^{-1/2}T^{-1/2})$. The second term is bounded in norm by $C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i-R'\Lambda_i\|^2=O_p(T^{-1})$ by Proposition B.1.2. The third term is $C\|RR'-M_{ff}\|$ which is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ by Lemma B.14. The fourth term is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\|\Lambda_{i,12}\Sigma_{i,22}^{-1}\|\cdot\Big\|\frac{1}{T}\sum_{t=1}^{T}v_{it}f_t'\Big\|\cdot\|\Lambda_i\boldsymbol{w}_1\Lambda_{i,11}'\|\Big)\cdot\|M_{ff}^{-1}\|\cdot\|RR'-M_{ff}\|,$$

which is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$ by Lemma B.14. The fifth term is also $O_p(N^{-1/2}T^{-1})$ $+O_p(T^{-3/2})$ by the similar arguments in the fourth. The sixth and eighth terms are both bounded in norm by

$$C\|R\|\cdot\Big\|\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_tf_t'\Big\|$$

which is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ by Lemma B.13(a). For the seventh term, we temporarily use L_i to denote $\Lambda_{i,12}\Sigma_{i,22}^{-1}\boldsymbol{w}_2'\Lambda_i'$. Notice the left hand side of the seventh term is an $r_1\times r_1$ matrix. So it suffices to show its the (p,q)th element $(p,q=1,2\ldots,r_1)$, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}L_{i,p}R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_te_{it}-E(u_te_{it})]\Lambda_{i,11,q}$$

where $L_{i,p}$ is the pth row of L_i and $\Lambda_{i,11,q}$ is the qth element of $\Lambda_{i,11}$. The above expression can be rewritten as

$$tr\Big[R\hat{H}\hat{\Lambda}'\hat{\Psi}^{-1}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[u_t e_{it} - E(u_t e_{it})]\Lambda_{i,11,q}L_{i,p}\Big)\Big].$$

The expression in the trace operator is bounded in norm by

$$C \cdot \|N^{1/2}\hat{H}^{1/2}\| \cdot \|R\| \cdot \Big(\sum_{j=1}^{N}\|\hat{\Sigma}_{jj}^{-1/2}\hat{\Lambda}_j\hat{H}^{1/2}\|^2\Big)^{1/2}$$
$$\times\Big(\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[u_{jt}e_{it} - E(u_{jt}e_{it})]\Lambda_{i,11,q}L_{i,p}\Big\|^2\Big)^{1/2}$$

which is $O_p(N^{-1/2}T^{-1/2})$. So the seventh term is $O_p(N^{-1/2}T^{-1/2})$. The ninth term can be proved to be $O_p(N^{-1/2}T^{-1/2})$ similarly as the seventh. Consider the tenth term, which is bounded in norm by

$$CN^{-1/2}\|\hat{H}^{1/2}\| \cdot \|R\| \cdot \Big(\sum_{i=1}^{N}\|\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Sigma}_{ii} - \Sigma_{ii}\|^2\Big)^{1/2},$$

which is $O_p(N^{-1}T^{-1/2})$. So the tenth term is $O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1})$. The eleventh term can be proved to be $O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})$ similarly as the tenth. Summarizing all the results, we have the second term of (B.2.10) is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$. This leads to $ii_5 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Given the results on ii_4 and ii_5, we have (b).

Result (c) can be proved similarly as (b) and the details are omitted.

Consider (d). The left hand side of (d) can be written as

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}\Big)$$
$$+\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime} - \frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}\beta_i\beta_i'\Lambda_{i,12}^{*\prime}\Big) = ii_6 + ii_7 \quad \text{say}$$

We first consider ii_6, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}-\Lambda_{i,22}^{*})\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime} + \frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}(\hat{\Lambda}_{i,12}-\Lambda_{i,12}^{*})'$$
$$+\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}-\Lambda_{i,22}^{*})\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}(\hat{\Lambda}_{i,12}-\Lambda_{i,12}^{*})' = iii_8 + iii_9 + iii_{10} \quad \text{say}$$

Consider iii_8, which is equal to

$$\begin{aligned} iii_8 &= \boldsymbol{v}_2 R' M_{ff}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v_{it}'\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime} \\ &+\boldsymbol{v}_2\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}u_t f_t'\Big(\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\boldsymbol{w}_2\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}\Big) \\ &+\boldsymbol{v}_2\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}[u_t v_{it}' - E(u_t v_{it}')]\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime} \\ &-\boldsymbol{v}_2\hat{H}\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\boldsymbol{w}_1^{-}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime} \end{aligned} \tag{B.2.12}$$

Consider the first term. Ignore $\boldsymbol{v}_2 R' M_{ff}^{-1}$, the remaining expression can be written as

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v_{it}'(\hat{\beta}_i^{CV}-\beta_i)(\hat{\beta}_i^{CV}-\beta_i)'\Lambda_{i,12}^{*\prime} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v_{it}'\beta_i\beta_i'\Lambda_{i,12}^{*\prime}$$
$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v_{it}'(\hat{\beta}_i^{CV}-\beta_i)\beta_i'\Lambda_{i,12}^{*\prime} + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_t v_{it}'\beta_i(\hat{\beta}_i^{CV}-\beta_i)'\Lambda_{i,12}^{*\prime}.$$

The first term is bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^4\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t v_{it}'\Big\|^2\Big)^{1/2}$$

However, by the boundedness of $\hat{\beta}_i^{CV}$ and β_i ($\hat{\beta}_i^{CV}$ is bounded due to the boundedness of $\hat{\Sigma}_{ii}$ and $\hat{\beta}_i^{CV}=\hat{\Sigma}_{i,22}^{-1}\hat{\Sigma}_{i,21}$), we have

$$\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^4\leq C\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^2$$

Given the above result, we have the first term is $O_p(T^{-1})$. The second term is $O_p(N^{-1/2}T^{-1/2})$. The third and fourth terms are both bounded in norm by

$$C\Big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t v_{it}'\Big\|^2\Big)^{1/2}\|R\|=O_p(T^{-1}).$$

Given these results, we have the first term of (B.2.12) is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$. The second term is also $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ by Lemma B.13(a) and the fact

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_i\boldsymbol{w}_2\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}=O_p(1).$$

The third term can be proved to be $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ similarly as proving (B.2.9) by replacing $\hat{\beta}_i^{CV}\Lambda_{i,11}^{*\prime}$ with $\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\Lambda_{i,12}^{*\prime}$. The last term of (B.2.12) can be proved to be $O_p(N^{-1/2}T^{-1/2})$ similarly as the last one of (B.2.6). Given these results, we have $iii_8=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$. Term iii_9 is also $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$, which can be proved by the same arguments in deriving iii_8. Term iii_{10} can be shown to be $O_p(T^{-1})$ similarly as

iii_7. Summarizing the results on iii_8, iii_9 and iii_{10}, we have $ii_6 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$.

Consider ii_7, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}(\hat{\beta}_i^{CV}-\beta_i)\beta_i'\Lambda_{i,12}^{*\prime}+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}\beta_i(\hat{\beta}_i^{CV}-\beta_i)'\Lambda_{i,12}^{*\prime}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\Lambda_{i,22}^{*}(\hat{\beta}_i^{CV}-\beta_i)(\hat{\beta}_i^{CV}-\beta_i)'\Lambda_{i,12}^{*\prime}$$

Treating $\beta_i'\Lambda_{i,12}^{*\prime}$ as a new $\Lambda_{i,11}^{*\prime}$, the first term can be proved to be $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ similarly as ii_5. The second term is also $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by the same arguments. The third term is bounded in norm by

$$C\|R\|^2\frac{1}{N}\sum_{i=1}^{N}\|\hat{\beta}_i^{CV}-\beta_i\|^2 = O_p(T^{-1}).$$

Given the above results, we have $ii_7 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. Summarizing the results on ii_6 and ii_7, we have (d).

This completes the proof of Lemma B.2.2. □

Lemma B.2.3. *Under Assumptions A-D, we have*

$$(a)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}\hat{\Lambda}_{i,11}'-\Lambda_{i,11}^{*}\Lambda_{i,11}^{*\prime}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$
$$(b)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV}\hat{\Lambda}_{i,11}'-\Lambda_{i,12}^{*}\beta_i\Lambda_{i,11}^{*\prime}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$
$$(c)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}'-\Lambda_{i,11}^{*}\beta_i'\Lambda_{i,12}^{*\prime}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$
$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}'-\Lambda_{i,12}^{*}\beta_i\beta_i'\Lambda_{i,12}^{*\prime}) = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

PROOF OF LEMMA B.2.3. The proof of Lemma B.2.3 is quite similar as the one of Lemma B.2.2. So we omit it. □

Lemma B.2.4. *Under Assumptions A-D, we have*

$$\hat{V} - V = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

PROOF OF LEMMA B.2.4. By the definitions of $\hat{V}$ and V, i.e.,

$$\hat{V} = \Big[\sum_{i=1}^{N}(\hat{\Lambda}_{i,21} - \hat{\Lambda}_{i,22}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]\Big[\sum_{i=1}^{N}(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]^{-1}$$

and

$$V = \Big[\sum_{i=1}^{N}(\Lambda_{i,21}^* - \Lambda_{i,22}^*\beta_i)(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i)'\Big]\Big[\sum_{i=1}^{N}(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i)(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i)'\Big]^{-1},$$

together with the fact that $\hat{A}\hat{B}^{-1} - AB^{-1} = \big((\hat{A} - A) - AB^{-1}(\hat{B} - B)\big)\hat{B}^{-1}$, we have

$$\hat{V} - V = (J_1 - VJ_2)\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})'\Big]^{-1},$$

where

$$J_1 = \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21} - \hat{\Lambda}_{i,22}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})' - \frac{1}{N}\sum_{i=1}^{N}(\Lambda_{i,21}^* - \Lambda_{i,22}^*\beta_i)(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i)'$$
$$J_2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})(\hat{\Lambda}_{i,11} - \hat{\Lambda}_{i,12}\hat{\beta}_i^{CV})' - \frac{1}{N}\sum_{i=1}^{N}(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i)(\Lambda_{i,11}^* - \Lambda_{i,12}^*\beta_i)'$$

Consider J_1, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}\hat{\Lambda}_{i,11}' - \Lambda_{i,21}^{*}\Lambda_{i,11}^{*\prime}) - \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\Lambda}_{i,11}' - \Lambda_{i,22}^{*}\beta_i\Lambda_{i,11}^{*\prime})$$
$$-\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,21}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,21}^{*}\beta_i'\Lambda_{i,12}^{*\prime}) + \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,22}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,22}^{*}\beta_i\beta_i'\Lambda_{i,12}^{*\prime}).$$

By Lemma B.2.2, we have $J_1 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. Consider J_2, which is equal to

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}\hat{\Lambda}_{i,11}' - \Lambda_{i,11}^{*}\Lambda_{i,11}^{*\prime}) - \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV}\hat{\Lambda}_{i,11}' - \Lambda_{i,12}^{*}\beta_i\Lambda_{i,11}^{*\prime})$$
$$-\frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,11}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,11}^{*}\beta_i'\Lambda_{i,12}^{*\prime}) + \frac{1}{N}\sum_{i=1}^{N}(\hat{\Lambda}_{i,12}\hat{\beta}_i^{CV}\hat{\beta}_i^{CV\prime}\hat{\Lambda}_{i,12}' - \Lambda_{i,12}^{*}\beta_i\beta_i'\Lambda_{i,12}^{*\prime}).$$

By Lemma B.2.3, we have $J_2 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. Given $J_1 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ and $J_2 = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$, together with $V = O_p(1)$, we have $\hat{V} - V = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. □

Lemma B.2.5. *Under Assumptions A-D, we have*

$$(a)\quad \hat{\Lambda}_{i,22} - \hat{V}\hat{\Lambda}_{i,12} - (\Lambda_{i,22}^{*} - V\Lambda_{i,12}^{*}) = R_{22\cdot 1}'\frac{1}{T}\sum_{t=1}^{T}h_t^{\star}v_{it}' + o_p(T^{-1/2})$$
$$(b)\quad \hat{\Lambda}_{i,21} - \hat{V}\hat{\Lambda}_{i,11} - (\Lambda_{i,21}^{*} - V\Lambda_{i,11}^{*}) = R_{22\cdot 1}'\frac{1}{T}\sum_{t=1}^{T}h_t^{\star}e_{it} + o_p(T^{-1/2})$$

where $R_{22\cdot 1} = R_{22} - R_{21}R_{11}^{-1}R_{12}$, $f_t^{\star} = M_{ff}^{-1}f_t \equiv [g_t^{\star\prime}, h_t^{\star\prime}]'$ *and* $e_{it} = \epsilon_{it} + \beta_i'v_{it}$.

Proof of Lemma B.2.5. The left hand side of (a) is equal to

$$(\hat{\Lambda}_{i,22} - \Lambda_{i,22}^{*}) - V(\hat{\Lambda}_{i,12} - \Lambda_{i,12}^{*}) - (\hat{V} - V)\hat{\Lambda}_{i,12} \tag{B.2.13}$$

The last term is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by $\hat{\Lambda}_{i,12} = \Lambda^*_{i,12} + o_p(1)$ and Lemma B.2.4. Substituting (B.2.3) and (B.2.4) into (B.2.13), we can rewrite the first two term of (B.2.13) (denoted by i_1) as

$$i_1 = (\boldsymbol{v}_2 - V\boldsymbol{v}_1)R'M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t v_{it}' + (\boldsymbol{v}_2 - V\boldsymbol{v}_1)\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T} u_t f_t' \Lambda_i \boldsymbol{w}_2\Big)$$
$$+ (\boldsymbol{v}_2 - V\boldsymbol{v}_1)\hat{H}\hat{\Lambda}\hat{\Psi}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_t v_{it}' - E(u_t v_{it}')] - (\boldsymbol{v}_2 - V\boldsymbol{v}_1)\hat{H}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii} - \Sigma_{ii})\boldsymbol{w}_2$$

By Lemma B.13, the last three terms are $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$. However,

$$(\boldsymbol{v}_2 - V\boldsymbol{v}_1)R' = [0_{r_2\times r_1}, R'_{22\cdot 1}].$$

So we have

$$i_1 = [0_{r_2\times r_1}, R'_{22\cdot 1}]M_{ff}^{-1}\frac{1}{T}\sum_{t=1}^{T} f_t v_{it}' + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$$

implies (a).

Result (b) can be proved similarly as (a). The details are omitted. □

Lemma B.2.6. *Under Assumptions A-D, we have, for all i,*

$$\hat{W}_i = W_i + o_p(1).$$

Proof of Lemma B.2.6. Let

$$\hat{W}_{i,11} = \left\{ \left[\frac{1}{T}\sum_{t=1}^{T} \hat{h}_t \hat{h}_t'\right] - \left[\frac{1}{T}\sum_{t=1}^{T} \hat{h}_t \hat{\eta}_t'\right]\left[\frac{1}{T}\sum_{t=1}^{T} \hat{\eta}_t \hat{\eta}_t'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} \hat{\eta}_t \hat{h}_t'\right] \right\}^{-1}$$

We first show that

$$\hat{W}_{i,11} = W_{i,11} + o_p(1) \tag{B.2.14}$$

where

$$W_{i,11} = R_{22\cdot1}' M_{hh}^{\star} R_{22\cdot1} = R_{22\cdot1}'\big(T^{-1}\sum_{t=1}^{T} h_t^{\star} h_t^{\star\prime}\big) R_{22\cdot1} = R_{22\cdot1}'(M_{hh} - M_{hg} M_{gg}^{-1} M_{gh})^{-1} R_{22\cdot1}$$

with $R_{22\cdot1} = R_{22} - R_{21} R_{11}^{-1} R_{12}$. The last equation of the above expression is due to the definition of $f_t^{\star}$, i.e., $f_t^{\star} \equiv (g_t^{\star\prime}, h_t^{\star\prime})' = M_{ff}^{-1} f_t$.

Let $f_t^* = [g_t^{*\prime}, h_t^{*\prime}]' = R^{-1} f_t$ and $\eta_t^* = g_t^* + V' h_t^*$. By $f_t^* = R^{-1} f_t$, we have

$$g_t^* = (R_{11}^{-1} + R_{11}^{-1} R_{12} R_{22\cdot1}^{-1} R_{21} R_{11}^{-1}) g_t - R_{11}^{-1} R_{12} R_{22\cdot1}^{-1} h_t \tag{B.2.15}$$

$$h_t^* = -R_{22\cdot1}^{-1} R_{21} R_{11}^{-1} g_t + R_{22\cdot1}^{-1} h_t \tag{B.2.16}$$

From (B.2.15) and (B.2.16), together with $V = R_{11}^{-1} R_{12}$, we have

$$\eta_t^* = g_t^* + V' h_t^* = R_{11}^{-1} g_t. \tag{B.2.17}$$

Thus,

$$\left\{\left[\frac{1}{T}\sum_{t=1}^{T} h_t^* h_t^{*\prime}\right] - \left[\frac{1}{T}\sum_{t=1}^{T} h_t^* \eta_t^{*\prime}\right]\left[\frac{1}{T}\sum_{t=1}^{T} \eta_t^* \eta_t^{*\prime}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} \eta_t^* h_t^{*\prime}\right]\right\}^{-1} \tag{B.2.18}$$

$$= \left\{R_{22\cdot 1}^{-1}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})R_{22\cdot 1}^{-1\prime}\right\}^{-1} = R_{22\cdot 1}^{\prime}(M_{hh} - M_{hg}M_{gg}^{-1}M_{gh})^{-1}R_{22\cdot 1} = W_{i,11}$$

So to prove (B.2.14), it suffices to prove

$$\frac{1}{T}\sum_{t=1}^{T} \hat{h}_t \hat{h}_t^{\prime} - \frac{1}{T}\sum_{t=1}^{T} h_t^* h_t^{*\prime} = o_p(1), \tag{B.2.19}$$

$$\frac{1}{T}\sum_{t=1}^{T} \hat{h}_t \hat{\eta}_t^{\prime} - \frac{1}{T}\sum_{t=1}^{T} h_t^* \eta_t^{*\prime} = o_p(1), \tag{B.2.20}$$

$$\frac{1}{T}\sum_{t=1}^{T} \hat{\eta}_t \hat{\eta}_t^{\prime} - \frac{1}{T}\sum_{t=1}^{T} \eta_t^* \eta_t^{*\prime} = o_p(1). \tag{B.2.21}$$

Notice that

$$\hat{f}_t = \big(\sum_{i=1}^{N} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1} \hat{\Lambda}_i^{\prime}\big)^{-1}\big(\sum_{i=1}^{N} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1} z_{it}\big).$$

Then we have

$$\begin{aligned}\hat{f}_t - f_t^* = &-\big(\sum_{i=1}^{N} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1} \hat{\Lambda}_i^{\prime}\big)^{-1}\big(\sum_{i=1}^{N} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1}(\hat{\Lambda}_i - R^{\prime}\Lambda_i)^{\prime}\big) f_t^* \\ &+ \big(\sum_{i=1}^{N} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1} \hat{\Lambda}_i^{\prime}\big)^{-1}\big(\sum_{i=1}^{N} \hat{\Lambda}_i \hat{\Sigma}_{ii}^{-1} u_{it}\big).\end{aligned} \tag{B.2.22}$$

where $f_t^* = R^{-1} f_t$. Equation (B.2.22) leads to

$$\frac{1}{T}\sum_{t=1}^{T} (\hat{f}_t - f_t^*) f_t^{*\prime} = o_p(1), \tag{B.2.23}$$

$$\frac{1}{T}\sum_{t=1}^{T}(\hat{f}_t - f_t^*)(\hat{f}_t - f_t^*)' = o_p(1). \tag{B.2.24}$$

To see this, notice the left hand side of (B.2.23) is equal to

$$-\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Lambda}_i - R'\Lambda_i)'\big(\frac{1}{T}\sum_{t=1}^{T}f_t^* f_t^{*\prime}\big) + \hat{H}\big(\frac{1}{T}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}u_{it}f_t^{*\prime}\big),$$

where $\hat{H} = (\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i')^{-1}$. The second term is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma B.13(a). The first term is bounded in norm by

$$C\|N^{1/2}\hat{H}^{1/2}\|\big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\big)^{1/2}\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\big)^{1/2}\big\|\frac{1}{T}\sum_{t=1}^{T}f_t^* f_t^{*\prime}\big\|$$

which is $O_p(T^{-1/2})$ by Proposition B.1.2. So we obtain (B.2.23).

Proceed to consider (B.2.24). The left hand side of (B.2.24) is bounded in norm by

$$2\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Lambda}_i - R'\Lambda_i)'\Big\|^2\big(\frac{1}{T}\sum_{t=1}^{T}\|f_t^*\|^2\big) + 2\frac{1}{T}\sum_{T=1}^{T}\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}u_{it}\Big\|^2.$$

The first term is $O_p(T^{-1/2})$ since

$$\begin{aligned}\Big\|\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Lambda}_i - R'\Lambda_i)'\Big\| \le & C\|N^{1/2}\hat{H}^{1/2}\|\cdot\big(\sum_{i=1}^{N}\|\hat{H}^{1/2}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1/2}\|^2\big)^{1/2} \\ & \times\big(\frac{1}{N}\sum_{i=1}^{N}\|\hat{\Lambda}_i - R'\Lambda_i\|^2\big)^{1/2}\end{aligned}$$

which is $O_p(T^{-1/2})$. Ignore 2, the second term is equal to

$$tr\Big[\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}\frac{1}{T}\sum_{t=1}^{T}[u_{it}u_{jt}'-E(u_{it}u_{jt}')]\hat{\Sigma}_{jj}^{-1}\hat{\Lambda}_j'\hat{H}\Big]+tr\Big[\hat{H}\sum_{i=1}^{N}\hat{\Lambda}_i\hat{\Sigma}_{ii}^{-1}(\hat{\Sigma}_{ii}-\Sigma_{ii})\hat{\Sigma}_{ii}^{-1}\hat{\Lambda}_i\hat{H}\Big]$$

which is $O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1})$ by Lemma B.13(d) and (e). So we have (B.2.24).

Given (B.2.23) and (B.2.24), we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{f}_t\hat{f}_t'-\frac{1}{T}\sum_{t=1}^{T}f_t^*f_t^{*\prime}=o_p(1) \tag{B.2.25}$$

From (B.2.25), we immediately obtain (B.2.19). Now consider (B.2.20). By the definition of $\hat{\eta}_t$,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{\eta}_t' &= \frac{1}{T}\sum_{t=1}^{T}\hat{h}_t(\hat{g}_t+\hat{V}'\hat{h}_t)'=\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{g}_t'+\big(\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t'\big)\hat{V}\\ &=\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{g}_t'+\big(\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t'\big)V+\big(\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{h}_t'\big)(\hat{V}-V)\end{aligned}$$

From (B.2.25), we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{g}_t'=\frac{1}{T}\sum_{t=1}^{T}h_t^*g_t^{*\prime}+o_p(1)$$

Given the above result, together with (B.2.19) and Lemma B.2.4, we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{h}_t\hat{\eta}_t'=\frac{1}{T}\sum_{t=1}^{T}h_t^*g_t^{*\prime}+\big(\frac{1}{T}\sum_{t=1}^{T}h_t^*h_t^{*\prime}\big)V+o_p(1)=\frac{1}{T}\sum_{t=1}^{T}h_t^*\eta_t^{*\prime}+o_p(1).$$

Equation (B.2.21) can be proved similarly as (B.2.20) and the proof is omitted. Given (B.2.19), (B.2.20) and (B.2.21), we have (B.2.14).

Given (B.2.14), in combination with $\hat{\Sigma}_{i,22} = \Sigma_{i,22} + o_p(1)$, we have $\hat{W}_i = W_i + o_p(1)$. This completes the proof. □

Proof of Theorem 2.5.1. The consistency of $\hat{\beta}_i^{LV}$ is implied by the asymptotic expression. So we only focus on the derivation of the asymptotic expression. Notice that $\hat{\beta}_i^{LV}$ is defined by

$$\hat{\beta}_i^{LV} = (\hat{\Delta}_i' \hat{W}_i^{-1} \hat{\Delta}_i)^{-1} (\hat{\Delta}_i' \hat{W}_i^{-1} \hat{\delta}_i).$$

By $\Delta_i \beta_i = \delta_i$, we also have

$$\beta_i = (\Delta_i' \hat{W}_i^{-1} \Delta_i)^{-1} (\Delta_i' \hat{W}_i^{-1} \delta_i),$$

From the two preceding equations, we have

$$\begin{aligned}
\hat{\beta}_i^{LV} - \beta_i &= (\hat{\Delta}_i' \hat{W}_i^{-1} \hat{\Delta}_i)^{-1} \big[(\hat{\Delta}_i' \hat{W}_i^{-1} \hat{\delta}_i - \Delta_i' \hat{W}_i^{-1} \delta_i) - (\hat{\Delta}_i' \hat{W}_i^{-1} \hat{\Delta}_i - \Delta_i' \hat{W}_i^{-1} \Delta_i) \beta_i \big] \\
&= \Big\{ (\hat{\Delta}_i' \hat{W}_i^{-1} \hat{\Delta}_i)^{-1} \big[\Delta_i' \hat{W}_i^{-1} (\hat{\delta}_i - \delta_i) - \Delta_i' \hat{W}_i^{-1} (\hat{\Delta}_i - \Delta_i) \beta_i \big] \Big\} \qquad \text{(B.2.26)} \\
&+ \Big\{ (\hat{\Delta}_i' \hat{W}_i^{-1} \hat{\Delta}_i)^{-1} \big[(\hat{\Delta}_i - \Delta_i) \hat{W}_i^{-1} (\hat{\delta}_i - \delta_i) - (\hat{\Delta}_i - \Delta_i) \hat{W}_i^{-1} (\hat{\Delta}_i - \Delta_i) \beta_i \big] \Big\}
\end{aligned}$$

By Lemma B.2.5 and Theorem 2.3.1, we have

$$\hat{\Delta}_i - \Delta_i = \begin{bmatrix} R_{22\cdot 1}' & 0 \\ 0 & I_K \end{bmatrix} \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} h_t^{\star} v_{it}' \\ v_{it} v_{it}' - E(v_{it} v_{it}') \end{bmatrix} + o_p(T^{-1/2}) \qquad \text{(B.2.27)}$$

and

$$\hat{\delta}_i - \delta_i = \begin{bmatrix} R'_{22\cdot 1} & 0 \\ 0 & I_K \end{bmatrix} \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} h_t^{\star} e_{it} \\ v_{it} e_{it} - E(v_{it} e_{it}) \end{bmatrix} + o_p(T^{-1/2}) \tag{B.2.28}$$

where $e_{it} = \epsilon_{it} + \beta_i' v_{it}$. Equations (B.2.27) and (B.2.28) implies that $\hat{\Delta}_i = \Delta_i + O_p(T^{-1/2})$ and $\hat{\delta}_i = \delta_i + O_p(T^{-1/2})$. Given these results, together with Lemma B.2.6, we have

$$\hat{\Delta}_i' \hat{W}_i^{-1} \hat{\Delta}_i - \Delta_i' W_i^{-1} \Delta_i = o_p(1), \qquad \hat{\Delta}_i' \hat{W}_i^{-1} - \Delta_i' W_i^{-1} = o_p(1),$$

$$(\hat{\Delta}_i - \Delta_i) \hat{W}_i^{-1} (\hat{\delta}_i - \delta_i) = O_p(T^{-1}), \qquad (\hat{\Delta}_i - \Delta_i) \hat{W}_i^{-1} (\hat{\Delta}_i - \Delta_i) = O_p(T^{-1})$$

Then we can simplify the expression of $\hat{\beta}_i^{LV} - \beta_i$ as

$$\hat{\beta}_i^{LV} - \beta_i = (\Delta_i' W_i^{-1} \Delta_i)^{-1} \Delta_i' W_i^{-1} \begin{bmatrix} R'_{22\cdot 1} & 0 \\ 0 & I_K \end{bmatrix} \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} h_t^{\star} \\ v_{it} \end{bmatrix} \epsilon_{it} + o_p(T^{-1/2}) \tag{B.2.29}$$

By definition of W_i, together with

$$\Delta_i = \begin{bmatrix} \Lambda_{i,22}^{*} - V \Lambda_{i,12}^{*} \\ \Sigma_{i,22} \end{bmatrix} = \begin{bmatrix} R'_{22\cdot 1} \Lambda_{i,22} \\ \Sigma_{i,22} \end{bmatrix} = \begin{bmatrix} R'_{22\cdot 1} \gamma_i^h \\ \Sigma_{i,22} \end{bmatrix}$$

we have

$$\begin{aligned} \sqrt{T}(\hat{\beta}_i^{LV} - \beta_i) &= \left(\gamma_i^{h\prime} (M_{hh} - M_{hg} M_{gg}^{-1} M_{gh}) \gamma_i^h + \Omega_i\right)^{-1} \\ &\quad \times \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\gamma_i^{h\prime} \left(h_t - M_{hg} M_{gg}^{-1} g_t\right) + v_{it}\right] \epsilon_{it} + o_p(1) \end{aligned}$$

This completes the proof of Theorem 2.5.1. $\square$

Appendix C

Appendix for Chapter 3

C.1 Proof for Proposition 3.4.1

The following notation will be used in this appendix.

$$\hat{P} = \frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}; \qquad \hat{R} = \frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M; \qquad \hat{G} = (I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1};$$
$$\hat{P}_N = N \cdot \hat{P} = \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}; \quad \hat{R}_N = N \cdot \hat{R} = M'\hat{\Sigma}_{ee}^{-1}M, \quad \hat{G}_N = N \cdot \hat{G}.$$

From $(A+B)^{-1} = A^{-1} - A^{-1}B(A+B)^{-1}$, we have $\hat{P}_N^{-1} = \hat{G}(I - \hat{G})^{-1}$. From $\Sigma_{zz} = M\Lambda\Lambda'M' + \Sigma_{ee}$, we have

$$\Sigma_{zz}^{-1} = \Sigma_{ee}^{-1} - \Sigma_{ee}^{-1}M\Lambda(I_r + \Lambda'M'\Sigma_{ee}^{-1}M\Lambda)^{-1}\Lambda'M'\Sigma_{ee}^{-1}. \tag{C.1.1}$$

It follows that

$$\hat{\Lambda}'M'\hat{\Sigma}_{zz}^{-1} = \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1} - \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1} = \hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}. \tag{C.1.2}$$

We use symbols with superscript "*" to denote the true parameters. Variables without superscript "*" denote the arguments of the likelihood function.

Let $\theta = (\Lambda, \sigma_1^2, \cdots, \sigma_N^2)$ and let Θ be the parameter set such that Λ take values in a compact set and $C^{-2} \leq \sigma_i^2 \leq C^2$ for all $i = 1, ..., N$. We assume $\theta^* = (\Lambda^*, \sigma_1^{*2}, \cdots, \sigma_N^{*2})$ is an interior point of Θ. For simplicity, we write $\theta = (\Lambda, \Sigma_{ee})$ and $\theta^* = (\Lambda^*, \Sigma_{ee}^*)$.

The following lemmas are useful for our analysis

Lemma C.1.1. *Under assumptions of A-D, we have*

$$(a) \quad \sup_{\theta\in\Theta} \frac{1}{NT}\left|\mathrm{tr}\Big[\Lambda^{*\prime} M' \Sigma_{zz}^{-1} \sum_{t=1}^{T} e_t f_t^{*\prime}\Big]\right| \xrightarrow{p} 0;$$

$$(b) \quad \sup_{\theta\in\Theta} \frac{1}{NT}\left|\mathrm{tr}\Big[\sum_{t=1}^{T}(e_t e_t' - \Sigma_{ee}^*)\Sigma_{zz}^{-1}\Big]\right| \xrightarrow{p} 0;$$

where $\theta^* = (\Lambda^*, \Sigma_{ee}^*)$ denotes the true parameters and $\Sigma_{zz} = M\Lambda\Lambda' M' + \Sigma_{ee}$.

Proof of Lemma C.1.1. First, we consider (a). Let m_{ip} be the (i,p)th element of M for $i = 1, \ldots, N, p = 1, \ldots, k$ and $\Lambda = [\lambda_1, \lambda_2, \ldots, \lambda_k]'$. By equation (C.1.1), we have

$$\frac{1}{NT}\Lambda'^* M' \Sigma_{zz}^{-1} \sum_{t=1}^{T} e_t f_t^{*\prime} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big(\sum_{p=1}^{k}\lambda_p^* m_{ip}\Big)\frac{1}{\sigma_i^2} e_{it} f_t^{*\prime} \qquad (C.1.3)$$
$$-\Lambda^{*\prime} M' \Sigma_{ee}^{-1} M\Lambda(I_r + \Lambda' M' \Sigma_{ee}^{-1} M\Lambda)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big(\sum_{p=1}^{k}\lambda_p m_{ip}\Big)\frac{1}{\sigma_i^2} e_{it} f_t^{*\prime}.$$

By the Cauchy-Schwartz inequality, the first term on the right side of (C.1.3) is bounded in norm by

$$\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\|\sum_{p=1}^{k}\lambda_p^* m_{pi}\|^2\Big)^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\|\frac{1}{T}\sum_{t=1}^{T} f_t^* e_{it}\|^2\Big]^{1/2}.$$

The first factor $(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\|\sum_{p=1}^{k}\lambda_p^* m_{pi}\|^2)^{1/2}$ is bounded by the boundedness of σ^{-2} and

$\frac{1}{N}\sum_{i=1}^{N} \| \sum_{p=1}^{k} \lambda_p^* m_{pi}\|^2$ by Assumptions C and D. The second factor does not depend on any unknown parameters, and it is $O_p(T^{-1/2})$ because $E(\frac{1}{N}\sum_{i=1}^{N} \|\frac{1}{T}\sum_{t=1}^{T} f_t^* e_{it}\|^2) = O(T^{-1})$. Therefore, the first part on the right hand side of (C.1.3) is $o_p(1)$ uniformly on θ. For the second part, we rewrite it in terms of P_N as

$$\Lambda^{*\prime} M' \Sigma_{ee}^{-1} M \Lambda P_N^{-1/2} (P_N^{-1} + I_r)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} P_N^{-1/2} \Big(\sum_{p=1}^{k} \lambda_p m_{ip} \Big) \frac{1}{\sigma_i^2} e_{it} f_t^{*\prime}. \tag{C.1.4}$$

The term $\Lambda^{*\prime} M' \Sigma_{ee}^{-1} M \Lambda P_N^{-1/2} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} (\sum_{p=1}^{k} \lambda_p^* m_{ip})(\sum_{p=1}^{k} \lambda_p' m_{ip}) P_N^{-1/2}$ is bounded in norm by

$$C \Big(\sum_{i=1}^{N} \Big\| \sum_{p=1}^{k} \lambda_p^* m_{ip} \Big\|^2 \Big)^{1/2} \Big(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \Big\| \sum_{p=1}^{k} \lambda_p' m_{ip} P_N^{-1/2} \Big\|^2 \Big)^{1/2} = a_1, \quad \text{say.}$$

Notice that

$$\begin{aligned} \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \Big\| P_N^{-1/2} \sum_{p=1}^{k} \lambda_p m_{ip} \Big\|^2 &= \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \Big(\sum_{p=1}^{k} \lambda_p' m_{ip} P_N^{-1} \sum_{q=1}^{k} \lambda_q m_{iq} \Big) \\ &= \mathrm{tr}\Big[P_N^{-1} \Lambda' M' \Sigma_{ee}^{-1} M \Lambda \Big] = \mathrm{tr}[P_N^{-1} P_N] = r. \end{aligned} \tag{C.1.5}$$

We have $a_1 = O_p(N^{1/2})$. As regard to the term $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} P_N^{-1/2} (\sum_{p=1}^{k} \lambda_p m_{ip}) \frac{1}{\sigma_i^2} e_{it} f_t^{*\prime}$, it is bounded in norm by

$$C \frac{1}{\sqrt{N}} \Big(\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \Big\| P_N^{-1/2} \sum_{p=1}^{k} \lambda_p m_{ip} \Big\|^2 \Big)^{1/2} \Big(\frac{1}{N} \sum_{i=1}^{N} \Big\| \frac{1}{T} \sum_{t=1}^{T} f_t^* e_{it} \Big\|^2 \Big)^{1/2} = O_p(N^{-1/2} T^{-1/2})$$

by (C.1.5). In addition, the term $(P_N^{-1} + I_r)^{-1} = O_p(1)$ uniformly on Θ. So the expression in (C.1.4) is $O_p(T^{-1/2})$ uniformly on θ. Then result (a) follows.

Next, we consider (b). By equation (C.1.1), we have

$$
\begin{aligned}
&\operatorname{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\Sigma_{zz}^{-1}\Big]\\
&\quad=\operatorname{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\big(\Sigma_{ee}^{-1}-\Sigma_{ee}^{-1}M\Lambda(I_r+\Lambda'M'\Sigma_{ee}^{-1}M\Lambda)^{-1}\Lambda'M'\Sigma_{ee}^{-1}\big)\Big]\\
&\quad=\operatorname{tr}\big[\frac{1}{NT}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\Sigma_{ee}^{-1}\big]\\
&\qquad-\operatorname{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}\big(\Lambda'M'\Sigma_{ee}^{-1}(e_te_t'-\Sigma_{ee}^*)\Sigma_{ee}^{-1}M\Lambda\big)(I_r+\Lambda'M'\Sigma_{ee}^{-1}M\Lambda)^{-1}\Big].
\end{aligned}
$$

The first term $\operatorname{tr}[\frac{1}{NT}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\Sigma_{ee}^{-1}]=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(e_{it}^2-\sigma_i^{*2})$ is bounded by

$$
\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\big(\frac{1}{T}\sum_{t=1}^{T}e_{it}^2-\sigma_i^{*2}\big)^2\Big)^{1/2},
$$

which is $O_p(T^{-1/2})$ uniformly on θ. The second term can be written as

$$
\operatorname{tr}\Big[\frac{1}{NT}P_N^{-1/2}\Lambda'M'\Sigma_{ee}^{-1}\big[\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\big]\Sigma_{ee}^{-1}M\Lambda P_N^{-1/2}(P_N^{-1}+I_r)^{-1}\Big].
$$

The above term is equal to

$$
\operatorname{tr}\Big[\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2\sigma_j^2}P_N^{-1/2}\sum_{p=1}^{k}\lambda_pm_{ip}\sum_{q=1}^{k}\lambda_q'm_{qj}P_N^{-1/2}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)(P_N^{-1}+I_r)^{-1}\Big].
$$

Since the expression

$$
\frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\sigma_i^2\sigma_j^2}P_N^{-1/2}\sum_{p=1}^{k}\lambda_pm_{ip}\sum_{q=1}^{k}\lambda_q'm_{qj}P_N^{-1/2}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]
$$

is bounded in norm by

$$C^2\Big[\sum_{i=1}^N \frac{1}{\sigma_i^2}\|P_N^{-1/2}\sum_{p=1}^k \lambda_p m_{ip}\|^2\Big]\Big[\frac{1}{N^2}\sum_{i=1}^N\sum_{j=1}^N\Big(\frac{1}{T}\sum_{t=1}^T[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)^2\Big]^{1/2}$$

which is $O_p(T^{-1/2})$ uniformly on θ by (C.1.5). Given $(P_N^{-1}+I_r)^{-1}=O(1)$ uniformly on θ, the second term is $o_p(1)$ uniformly on θ. This proves (b). □

Lemma C.1.2. *Under Assumptions A-D, we have*

$$(a)\quad \Big\|\frac{1}{N}\Lambda^{*\prime}M'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{*-1})M\Lambda^*\Big\| = O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^{*2})^2\Big]^{\frac{1}{2}}\Big);$$

$$(b)\quad \Big\|\frac{1}{N}M'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{*-1})M\Big\| = O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^{*2})^2\Big]^{\frac{1}{2}}\Big).$$

Given the above results, if $N^{-1}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^{*2})^2=o_p(1)$, *we have*

$$(c)\quad \hat{R}_N=O_p(N),\quad \hat{R}=\frac{1}{N}\hat{R}_N=O_p(1);$$

$$(d)\quad \|\hat{R}^{-1/2}\|=O_p(1).$$

where $\hat{R}$ *and* $\hat{R}_N$ *are defined above appendix A.*

Proof of Lemma C.1.2. We first consider (a). The left hand side of (a) can be written as

$$\frac{1}{N}\sum_{i=1}^N(\sum_{p=1}^k\lambda_p^* m_{ip})(\sum_{q=1}^k m_{qi}\lambda_q^{*\prime})\frac{\hat{\sigma}_i^2-\sigma_i^{*2}}{\hat{\sigma}_i^2\sigma_i^2},$$

which is bounded in norm by

$$C^4\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\sum_{p=1}^{k}\lambda_p^* m_{ip}\Big\|^4\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^{*2})^2\Big)^{1/2}.$$

Then result (a) follows because $\|\sum_{p=1}^{k}\lambda_p^* m_{ip}\|^4$ is bounded by Assumption C.

Next, we consider (b). The left hand side of (b) can be written as $\frac{1}{N}\sum_{i=1}^{N} m_i m_i' \frac{\hat{\sigma}_i^2-\sigma_i^{*2}}{\hat{\sigma}_i^2\sigma_i^{*2}}$, where m_i is the transpose of the ith row of M. This term is bounded in norm by

$$C^4\Big(\frac{1}{N}\sum_{i=1}^{N}\|m_i\|^4\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^{*2})^2\Big)^{1/2}.$$

Then result (b) follows because $\frac{1}{N}\sum_{i=1}^{N}\|m_i\|^4$ is bounded by Assumption C.

We now consider (c). From result (b) and result $N^{-1}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^{*2})^2 = o_p(1)$, we have $\hat{R}-\frac{1}{N}M'\Sigma_{ee}^{-1}M = o_p(1)$ which implies $\hat{R}\xrightarrow{p} R > 0$, where R is defined in Assumption C. So $\hat{R} = O_p(1)$ and $\hat{R}_N = N\hat{R} = O_p(N)$. Result (c) follows.

Result (d) is a direct result of $\|\hat{R}^{-1/2}\|^2 = \operatorname{tr}(\hat{R}^{-1}) = O_p(1)$ by $\hat{R}\xrightarrow{p} R > 0$ from result (c).

This completes the proof of Lemma C.1.2. $\square$

Lemma C.1.3. *Under Assumptions A-D, we have*

$$(a)\quad \frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}^{-1} = \|\hat{P}^{-1/2}\|^2\cdot O_p(T^{-1/2});$$

$$(b)\quad \frac{1}{N}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t' = \|\hat{P}^{-1/2}\|\cdot O_p(T^{-1/2});$$

$$(c)\quad \frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}^{-1} = \|\hat{P}_N^{-1}\|\cdot O_p(1);$$

(d) $\dfrac{1}{NT}\sum_{t=1}^{T} f_t e_t' \hat{\Sigma}_{ee}^{-1} M \hat{R}^{-1} = O_p(T^{-1/2});$

(e) $\dfrac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} \dfrac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \Sigma_{ee}]\hat{\Sigma}_{ee}^{-1} M \hat{R}^{-1} = \|\hat{P}^{-1/2}\| \cdot O_p(T^{-1/2});$

(f) $\dfrac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1} M \hat{R}^{-1} = \|\hat{P}^{-1/2}\| \cdot O_p\Big(\Big[\dfrac{1}{N^3}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2\Big]^{\frac{1}{2}}\Big).$

Proof of Lemma C.1.3. We first consider (a). The left hand side can be rewritten as

$$\frac{1}{N^2}\hat{P}^{-1/2}\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{P}^{-1/2}\Big(\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big)\frac{1}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]\Big(\sum_{q=1}^{k} m_{jq}\hat{\lambda}_q'\Big)\hat{P}^{-1/2}\Big]\hat{P}^{-1/2},$$

which is bounded in norm by

$$C^2\|\hat{P}^{-1/2}\|^2\Big[\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|^2\Big]\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]\Big|^2\Big]^{1/2},$$

which is $\|\hat{P}^{-1/2}\|^2 \cdot O_p(T^{-1/2})$ by (C.1.5). Thus, (a) follows.

Next, we consider (b). The left hand side can be rewritten as

$$\frac{1}{\sqrt{N}}\hat{P}^{-1/2}\sum_{i=1}^{N}\hat{P}_N^{-1/2}\frac{1}{\hat{\sigma}_i^2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\frac{1}{T}\sum_{t=1}^{T} e_{it} f_t',$$

which is bounded in norm by

$$C\|\hat{P}^{-1/2}\|\Big(\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} e_{it} f_t'\Big\|^2\Big)^{1/2},$$

which is $\|\hat{P}^{-1/2}\| \cdot O_p(T^{-1/2})$ by (C.1.5). This proves result (b).

To prove result (c), notice that $\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})$ is bounded by $2C^4 I_N$ by $C^{-2} \le \hat{\sigma}_i^2 \le C^2$

and $C^{-2} \leq \sigma_i^2 \leq C^2$. Hence, the left hand side is bounded in norm by

$$\left\| \hat{P}_N^{-1} \hat{\Lambda}' M' \Big(2C^4 I_N \Big) \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} \hat{P}_N^{-1} \right\| = 2C^4 \| \hat{P}_N^{-1} \|.$$

Result (c) then follows.

We now consider (d). The left hand side is equal to

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\hat{\sigma}_i^2} f_t e_{it} m_i' \hat{R},$$

which is bounded in norm by

$$C \| \hat{R} \| \cdot \Big[\frac{1}{N} \sum_{i=1}^{N} \| m_i \|^2 \Big]^{1/2} \Big[\frac{1}{N} \sum_{i=1}^{N} \Big\| \frac{1}{T} \sum_{t=1}^{T} f_t e_{it} \Big\|^2 \Big]^{1/2},$$

which is $O_p(T^{-1/2})$ by Lemma C.1.2(c) and Assumption C. Hence, result (d) follows.

For result (e), the left hand side is equal to

$$\frac{1}{N^{3/2}} \hat{P}^{-1/2} \Big[\sum_{i=1}^{N} \sum_{j=1}^{N} \hat{P}_N^{-1/2} \Big(\sum_{p=1}^{k} \hat{\lambda}_p m_{ip} \Big) \frac{1}{\hat{\sigma}_i^2 \hat{\sigma}_j^2} \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - E(e_{it} e_{jt})] m_j' \Big] \hat{R}^{-1},$$

which is bounded in norm by

$$C^2 \| \hat{P}^{-1/2} \| \cdot \| \hat{R}^{-1} \| \cdot \Big[\sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2} \Big\| \hat{P}_N^{-1/2} \sum_{p=1}^{k} \hat{\lambda}_p m_{ip} \Big\|^2 \Big]^{1/2} \Big[\frac{1}{N} \sum_{j=1}^{N} \| m_j \|^2 \Big]^{1/2}$$

$$\times \Big[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \Big| \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - E(e_{it} e_{jt})] \Big|^2 \Big]^{1/2},$$

which is $\|\hat{P}^{-1/2}\| \cdot O_p(T^{-1/2})$ by (C.1.5) and Lemma C.1.2(c). Thus, result (d) follows.

Finally, we consider (f). The left hand side can be written as

$$\frac{1}{N^{3/2}}\hat{P}^{-1/2}\sum_{i=1}^{N}\hat{P}_N^{-1/2}\Big(\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big)\Big(\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^4}\Big)m_i'\hat{R}^{-1},$$

which is bounded in norm by

$$\frac{1}{N}\cdot\|\hat{P}^{-1/2}\|\cdot\|\hat{R}^{-1}\|\Big[\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{\|m_i\|^2}{\hat{\sigma}_i^4}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{1/2}.$$

By the boundedness of $\|m_i\|$ and $\hat{\sigma}^{-2}$ by Assumptions C and D, we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\|m_i\|^2}{\hat{\sigma}_i^4}(\hat{\sigma}_i^2-\sigma_i^2)^2 \le C\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2.$$

This result, together with (C.1.5) and Lemma C.1.2(c), gives result (f). □

Proof of Proposition 3.4.1. Throughout the proof, we use the following centered objective function

$$L(\theta) = \overline{L}(\theta) + R(\theta),$$

where

$$\overline{L}(\theta) = -\frac{1}{N}\ln|\Sigma_{zz}| - \frac{1}{N}\mathrm{tr}\Big(\Sigma_{zz}^{*}\Sigma_{zz}^{-1}\Big) + 1 + \frac{1}{N}\ln|\Sigma_{zz}^{*}|$$

and

$$R(\theta) = -\frac{1}{N}\mathrm{tr}\Big[(M_{zz}-\Sigma_{zz}^{*})\Sigma_{zz}^{-1}\Big],$$

where $\Sigma_{zz} = M\Lambda\Lambda' M' + \Sigma_{ee}$ and $\Sigma_{zz}^* = M\Lambda^*\Lambda^{*\prime} M' + \Sigma_{ee}^*$. The above objective function differs from the objective function of the main text only by a constant and is convenient for the subsequent analysis. By the definition of M_{zz}, we have

$$R(\theta) = -2\frac{1}{NT}\mathrm{tr}\Big[M\Lambda^* \sum_{t=1}^{T} f_t^* e_t' \Sigma_{zz}^{-1}\Big] - \frac{1}{NT}\mathrm{tr}\Big[\sum_{t=1}^{T}(e_t e_t' - \Sigma_{ee}^*)\Sigma_{zz}^{-1}\Big].$$

By Lemma C.1.1, we have $\sup_\theta |R(\theta)| = o_p(1)$. Since $\hat{\theta}$ maximizes $L(\theta)$, it follows $\overline{L}(\hat{\theta}) + R(\hat{\theta})) \geq \overline{L}(\theta^*) + R(\theta^*)$. This implies that $\overline{L}(\hat{\theta}) \geq \overline{L}(\theta^*) + R(\theta^*) - R(\hat{\theta}) \geq \overline{L}(\theta^*) - 2\sup_{\theta\in\Theta} |R(\theta)| = -|o_p(1)|$, where $\overline{L}(\theta^*)$ is normalized to be zero.

Now consider $\overline{L}(\hat{\theta})$ which is equivalent to

$$\overline{L}(\hat{\theta}) = -\frac{1}{N}\ln|\hat{\Sigma}_{zz}| - \frac{1}{N}\mathrm{tr}(\Sigma_{zz}^*\hat{\Sigma}_{zz}^{-1}) + 1 + \frac{1}{N}\ln|\Sigma_{zz}^*|. \tag{C.1.6}$$

By $\Sigma_{zz} = M\Lambda\Lambda' M' + \Sigma_{ee}$, we have $|\Sigma_{zz}| = |\Sigma_{ee}| \cdot |I_r + \Lambda' M' \Sigma_{ee}^{-1} M\Lambda|$. Similarly, $|\Sigma_{zz}^*| = |\Sigma_{ee}^*| \cdot |I_r + \Lambda^{*'} M' \Sigma_{ee}^{*-1} M\Lambda^*|$. Then equation (C.1.6) can be written as

$$\begin{aligned}
\overline{L}(\hat{\theta}) = & -\frac{1}{N}\ln|\hat{\Sigma}_{ee}| - \frac{1}{N}\ln|I_r + \hat{\Lambda}' M' \Sigma_{ee}^{-1} M\hat{\Lambda}| - \frac{1}{N}\mathrm{tr}[M\Lambda^*\Lambda^{*\prime} M'\hat{\Sigma}_{zz}^{-1}] \\
& - \frac{1}{N}\mathrm{tr}[\Sigma_{ee}^*\hat{\Sigma}_{zz}^{-1}] + \frac{1}{N}\ln|\Sigma_{ee}^*| + \frac{1}{N}\ln|I_r + \Lambda^{*\prime} M'\Sigma_{ee}^{*-1} M\Lambda^*| + 1 \\
= & \left\{-\frac{1}{N}\ln|\hat{\Sigma}_{ee}| + \frac{1}{N}\ln|\Sigma_{ee}^*| - \frac{1}{N}\mathrm{tr}[\Sigma_{ee}^*\hat{\Sigma}_{zz}^{-1}] + 1\right\} \\
& + \left\{-\frac{1}{N}\mathrm{tr}[M\Lambda^*\Lambda^{*\prime} M'\hat{\Sigma}_{zz}^{-1}]\right\} + \left\{-\frac{1}{N}\ln|I_r + \hat{\Lambda}' M'\hat{\Sigma}_{ee}^{-1} M\hat{\Lambda}|\right\} \\
& + \left\{\frac{1}{N}\ln|I_r + \Lambda^{*\prime} M'\Sigma_{ee}^{*-1} M\Lambda^*|\right\}.
\end{aligned}$$

Notice that

$$\frac{1}{N}\text{tr}[\Sigma_{ee}^{*}\hat{\Sigma}_{zz}^{-1}] = \frac{1}{N}\text{tr}[\Sigma_{ee}^{*}\hat{\Sigma}_{ee}^{-1}] - \frac{1}{N}\text{tr}[\Sigma_{ee}^{*}\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}] = \frac{1}{N}\text{tr}[\Sigma_{ee}^{*}\hat{\Sigma}_{ee}^{-1}] + o_p(1)$$

by

$$0 < \frac{1}{N}\text{tr}[\Sigma_{ee}^{*}\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}] \leq C\frac{1}{N}\text{tr}[\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}] \leq C\frac{r}{N},$$

where we use the fact that there exists a constant C such that $\Sigma_{ee}^{*}\hat{\Sigma}_{ee}^{-1} \leq C \cdot I_N$ due to the boundedness of $\hat{\sigma}_i^2$ and σ_i^{*2}.

Given the above result, together with $\frac{1}{N}\ln|I_r + \Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^{*}| = O(\ln N/N)$, we can further write $\overline{L}(\hat{\theta})$ as

$$\begin{aligned}\overline{L}(\hat{\theta}) = &-\left\{\frac{1}{N}\ln|\hat{\Sigma}_{ee}| - \frac{1}{N}\ln|\Sigma_{ee}^{*}| + \frac{1}{N}\text{tr}[\Sigma_{ee}^{*}\hat{\Sigma}_{ee}^{-1}] - 1\right\} \\ &- \left\{\frac{1}{N}\text{tr}[M\Lambda^{*}\Lambda^{*\prime}M'\hat{\Sigma}_{zz}^{-1}]\right\} - \left\{\frac{1}{N}\ln|I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}|\right\} + o_p(1).\end{aligned}$$

The above three expressions in the big curly bracket are all non-negative. Together with $\overline{L}(\hat{\theta}) \geq -2|o_p(1)|$, we have that each expression is $o_p(1)$, that is,

$$\frac{1}{N}\ln|\hat{\Sigma}_{ee}| - \frac{1}{N}\ln|\Sigma_{ee}^{*}| + \frac{1}{N}\text{tr}[\Sigma_{ee}^{*}\hat{\Sigma}_{ee}^{-1}] - 1 \xrightarrow{p} 0, \tag{C.1.7}$$

$$\frac{1}{N}\text{tr}[M\Lambda^{*}\Lambda^{*\prime}M'\hat{\Sigma}_{zz}^{-1}] \xrightarrow{p} 0. \tag{C.1.8}$$

Equation (C.1.7) is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}(\ln\hat{\sigma}_i^2 - \ln\sigma_i^{*2} + \frac{\sigma_i^{*2}}{\hat{\sigma}_i^2} - 1) \xrightarrow{p} 0.$$

Consider the function $g(x) = \ln x + \frac{\sigma_i^{*2}}{x} - \ln \sigma_i^{*2} - 1$. Given that $0 < C^{-2} \leq \sigma_i^2 \leq C^2 < \infty$ for $C > 1$, for any $x \in [C^{-2}, C^2]$, we can find a constant d (for example, let $d = \frac{1}{4C^4}$) such that $g(x) \geq d(x - \sigma_i^{*2})^2$. It follows

$$o_p(1) = \frac{1}{N}\sum_{i=1}^{N}(\ln \hat{\sigma}_i^2 + \frac{\sigma_i^{*2}}{\hat{\sigma}_i^2} - 1 - \ln \sigma_i^{*2}) \geq d\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^{*2})^2.$$

The above argument implies

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^{*2})^2 \xrightarrow{p} 0. \tag{C.1.9}$$

This proves the first result of Proposition 3.4.1.

Next, we consider (C.1.8), which is equivalent to

$$\frac{1}{N}\text{tr}(M\Lambda^*\Lambda'^* M'\hat{\Sigma}_{zz}^{-1}) = \frac{1}{N}\text{tr}\Big[\Lambda'^* M'\big(\hat{\Sigma}_{ee}^{-1} - \hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\big)M\Lambda^*\Big].$$

By $(I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1} = (\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1} - (\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}(I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}$, the preceding expression can be alternatively written as

$$\begin{aligned}&\frac{1}{N}\text{tr}(M\Lambda^*\Lambda'^* M'\hat{\Sigma}_{zz}^{-1})\\ =&\frac{1}{N}\text{tr}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^* - \Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big]\\ &+\frac{1}{N}\text{tr}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}(I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big]\end{aligned}$$

Both terms on the right hand side are non-negative. By (C.1.8), it follows that

$$\frac{1}{N}\text{tr}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^* - \Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big] \xrightarrow{p} 0, \tag{C.1.10}$$

$$\frac{1}{N}\text{tr}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}(I_r+\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big] \xrightarrow{p} 0. \quad \text{(C.1.11)}$$

By (C.1.9) and Lemma C.1.2(a), we know $\frac{1}{N}\text{tr}(\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*)$ converges to a positive constant. Then (C.1.10) implies that $\frac{1}{N}\text{tr}(\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*)$ converges to the same positive constant. Together with (C.1.11), we have $(I_r+\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1} = o_p(1)$, i.e. $\hat{G} = o_p(1)$. Furthermore, from $\hat{P}_N^{-1} = \hat{G}(I-\hat{G})^{-1}$, we have $\hat{P}_N^{-1} = o_p(1)$. We obtain the following results

$$\hat{G} = o_p(1); \qquad \hat{P}_N^{-1} = o_p(1). \quad \text{(C.1.12)}$$

Consider (C.1.10) again. The matrix on the left-hand side is finite dimensional ($r\times r$) and is semi-positive definite, so its trace is $o_p(1)$ if and only if every entry is $o_p(1)$. Thus, we have

$$\frac{1}{N}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^* - \Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big] \xrightarrow{p} 0. \quad \text{(C.1.13)}$$

Let $A \equiv (\hat{\Lambda}-\Lambda^*)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}$. Then $I_r - A = \Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}$. So equation (C.1.13) simplifies to

$$\frac{1}{N}\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^* - (I_r-A)\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(I_r-A)' \xrightarrow{p} 0.$$

By Lemma C.1.2(a) and (C.1.9), we know $\frac{1}{N}\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^* = \frac{1}{N}\Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^* + o_p(1)$. Thus,

$$\frac{1}{N}\Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^* - (I_r-A)\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(I_r-A)' \xrightarrow{p} 0. \quad \text{(C.1.14)}$$

By Assumption C.3, the expression $\frac{1}{N}\Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^*$ is positive definite in the limit, so the second term is of full rank in the limit which implies that $(I_r - A)$ is of full rank in the limit.

Alternatively, equation (C.1.13) can be rewritten as

$$\frac{1}{N}(\hat{\Lambda}-\Lambda^*)'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda^*) - A\Big(\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\Big)A' \xrightarrow{p} 0. \tag{C.1.15}$$

We now make use of the first-order conditions to proceed the proof. The first-order condition (3.3.3) post-multiplied by $\hat{\Lambda}$ implies

$$\hat{\Lambda}'M'\hat{\Sigma}_{zz}^{-1}(M_{zz}-\hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}M\hat{\Lambda} = 0.$$

By (C.1.2), the above equation can be simplified as

$$\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz}-\hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} = 0,$$

which is equivalent to

$$\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} = -\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}$$

$$+\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\frac{1}{T}\sum_{t=1}^{T}f_t^*e_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}$$

$$+\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t^{*\prime}\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}.$$

With notations of $\hat{P}$ and A, we have

$$I_r = (I_r-A)'(I_r-A) + \frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}^{-1}$$

$$+(I_r - A)'\frac{1}{NT}\sum_{t=1}^{T} f_t^* e_t' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} \hat{P}^{-1} + \frac{1}{N}\hat{P}^{-1}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}(I_r - A) \quad \text{(C.1.16)}$$

$$-\frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} \hat{P}^{-1} = i_1 + i_2 + \cdots + i_5, \quad \text{say}$$

Term i_2 is $\|\hat{P}^{-1/2}\|^2 \cdot O_p(T^{-1/2})$ by Lemma C.1.3(a). Term i_3 is $\|I - A\| \cdot \|\hat{P}^{-1/2}\| \cdot O_p(T^{-1/2})$ by Lemma C.1.3(b). Term i_4 is the transpose of i_3 and therefore has the same convergence rate as i_3. The last term is $o_p(1)$ by Lemma C.1.3(c) and (C.1.12). Given these results, we have

$$I_r = (I - A)'(I - A) + \|\hat{P}^{-1/2}\|^2 O_p(T^{-1/2}) + \|I - A\| \cdot \|\hat{P}^{-1/2}\| \cdot O_p(T^{-1/2}) + o_p(1). \quad \text{(C.1.17)}$$

Moreover, by the definition of $\hat{P}$, equation (C.1.14) yields

$$\left(\frac{1}{N}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda}\right)^{-1} = (I_r - A)'\left(\frac{1}{N}\Lambda^{*\prime} M' \Sigma_{ee}^{*-1} M \Lambda^*\right)^{-1}(I_r - A) + o_p(\|I_r - A\|^2).$$

This implies that

$$\|\hat{P}^{-1/2}\|^2 = \mathrm{tr}(\hat{P}^{-1}) = \mathrm{tr}\Big[(I_r - A)'\Big(\frac{1}{N}\Lambda^{*\prime} M' \Sigma_{ee}^{*-1} M \Lambda^*\Big)^{-1}(I_r - A) + o_p(\|I_r - A\|^2)\Big].$$

The right hand side is at most $O_p[(A^2) \vee 1]$, implying that $\|\hat{P}^{-1/2}\| = O_p(A \vee 1)$, where $a \vee b$ denotes the maximum of a and b. So together with (C.1.17), we obtain $A = O_p(1)$. To see this, notice that the left hand side of equation (C.1.17) is bounded. Hence, if $A \neq O_p(1)$, then A is stochastically unbounded, the right hand side of (C.1.17) is dominated by $A'A$ in view of $\|\hat{P}^{-1/2}\| = O_p(A)$, but $A'A$ diverges. Then a contradiction arises. Thus, $A = O_p(1)$, which in turn implies that $\|\hat{P}^{-1/2}\| = O_p(1)$, or equivalently $\|\hat{P}^{-1}\| = O_p(1)$.

Now we sharpen the result to $A = o_p(1)$. From equation (C.1.17), $\|\hat{P}^{-1/2}\| = O_p(1)$ and $A = O_p(1)$, we have

$$(I_r - A)'(I_r - A) - I_r \xrightarrow{p} 0.$$

And from (C.1.14),

$$\frac{1}{N}\Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^* - (I_r - A)\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(I_r - A)' = o_p(1).$$

By the identification condition, $\frac{1}{N}\Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^*$ and $\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}$ are both diagonal with distinct diagonal elements. Applying Lemma A.1 of the supplement of Bai and Li (2012) to the preceding two equations, we have that $I_r - A$ converges in probability to a diagonal matrix with diagonal elements either 1 or -1. By correctly choosing the column signs, the case -1 is precluded. Therefore, we have $I_r - A \xrightarrow{p} I_r$, or equivalently $A = o_p(1)$.

Next, we consider the first-order condition on Λ (equation (3.3.3)). By (C.1.2), we can simplify equation (3.3.3) as

$$\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M = 0.$$

Using the expression of M_{zz}, we can write the preceding equation as

$$\hat{\Lambda}' - \Lambda^{*\prime} = -A'\Lambda^{*\prime} + (I - A)'\frac{1}{T}\sum_{t=1}^{T} f_t^* e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} + \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}\Lambda^{*\prime} \quad \text{(C.1.18)}$$
$$+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \Sigma_{ee}^*]\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} - \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}.$$

By $A = o_p(1)$ and Lemma C.1.3 (d), we have that the first two terms are $o_p(1)$. By $\|\hat{P}^{-1}\| = O_p(1)$ and Lemma C.1.3 (b), the third term is $o_p(1)$. By $\|\hat{P}^{-1}\| = O_p(1)$ and Lemma C.1.3 (e), the fourth term is $o_p(1)$. By $\|\hat{P}^{-1}\| = O_p(1)$ and Lemma C.1.3 (f), the last term is $o_p(1)$. Given the above result, we have $\hat{\Lambda}' - \Lambda^{*\prime} = o_p(1)$, which implies that $\hat{\Lambda} \xrightarrow{p} \Lambda^{*\prime}$. This completes the proof of Proposition 3.4.1. □

Corollary C.1.1. *Under Assumptions A-D,*

$$(a) \quad \frac{1}{N}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} - \frac{1}{N}\Lambda^{*\prime} M' \Sigma_{ee}^{*-1} M \Lambda^* = o_p(1);$$

$$(b) \quad \hat{P}_N = O_p(N),\ \hat{P} = O_p(1),\ \hat{G} = O_p(N^{-1}),\ \hat{G}_N = O_p(1);$$

$$(c) \quad \frac{1}{N}(\hat{\Lambda} - \Lambda)' M' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} = o_p(1).$$

Proof of Corollary C.1.1. Result (a) follows from equation (C.1.14) and $A = (\hat{\Lambda} - \Lambda)' M' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} \hat{P}_N^{-1} = o_p(1)$.

For part (b), by Assumption C.3, $N^{-1}\Lambda^{*\prime} M' \Sigma_{ee}^{*-1} M \Lambda^* \to P_\infty > 0$. This result, together with result (a) of this corollary, implies $\hat{P} = O_p(1)$ and therefore $\hat{P}_N = O_p(N)$. From $\hat{G} = (I_r + \hat{P}_N)^{-1}$, we have $\hat{G} = O_p(N^{-1})$ and hence $\hat{G}_N = O_p(1)$.

Result (c) follows from $\hat{P} = O_p(N)$ and $A = o_p(1)$. □

C.2 Proofs of Theorems 3.4.1, 3.4.2 and 3.4.5

Hereafter, for notational simplicity, we drop "*" from the symbols of underlying true values. The following lemmas are used in the proofs of Theorems 3.4.1 and 3.4.2.

Lemma C.2.1. *Under Assumptions A-D,*

$$(a)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}=O_p(T^{-1/2});$$

$$(b)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'=O_p(T^{-1/2});$$

$$(c)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}=\frac{1}{\sqrt{N}}O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{\frac{1}{2}}\Big);$$

$$(d)\quad \frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}=O_p(T^{-1/2});$$

$$(e)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_t'-\Sigma_{ee}]\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}=O_p(T^{-1/2});$$

$$(f)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}=\frac{1}{\sqrt{N}}O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{\frac{1}{2}}\Big).$$

Proof of Lemma C.2.1. First, we consider (a). The left hand side is equal to

$$\hat{P}^{-1}\frac{1}{N^2}\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\Big)\frac{1}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big(\sum_{q=1}^{k}m_{jq}\hat{\lambda}_q'\Big)\Big]\hat{P}^{-1},$$

which is bounded in norm by

$$C^2\|\hat{P}^{-1}\|^2\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\Big\|^2\Big]\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big|^2\Big]^{1/2}.$$

Moreover, by Corollary C.1.1(a), we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\Big\|^2=\mathrm{tr}[\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}]\xrightarrow{p}\mathrm{tr}[\frac{1}{N}\Lambda'M'\Sigma_{ee}^{-1}M\Lambda]=\mathrm{tr}(P).\qquad(\text{C.2.1})$$

By

$$E\left[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big|^2\right]=O(T^{-1}),$$

together with Corollary C.1.1(b) and (C.2.1), we obtain (a).

Next, we consider (b). The left hand side can be written as

$$\hat{P}^{-1}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\big(\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\big)\frac{1}{T}\sum_{t=1}^{T}e_{it}f_t',$$

which is bounded in norm by

$$C\|\hat{P}^{-1}\|\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}e_{it}f_t'\Big\|^2\Big]^{1/2},$$

which is $O_p(T^{-1/2})$ by (C.2.1). Thus, (b) follows.

For part (c), the left hand side can be written as

$$\hat{P}_N^{-1/2}\Big[\sum_{i=1}^{N}\hat{P}_N^{-1/2}\Big(\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big)\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^4}\Big(\sum_{q=1}^{k}m_{iq}\hat{\lambda}_q'\Big)\hat{P}_N^{-1/2}\Big]\hat{P}_N^{-1/2},$$

which is bounded in norm by

$$C^2\|\hat{P}_N^{-1/2}\|^2\cdot\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\hat{P}_N^{-1/2}\Big(\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big)\Big\|^2(\hat{\sigma}_i^2-\sigma_i^2). \tag{C.2.2}$$

Since

$$\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|^2=r$$

by (C.1.5), this gives

$$\frac{1}{\hat{\sigma}_i}\Big\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\| \leq \sqrt{r}.$$

Hence, expression in (C.2.2) is bounded by

$$C^2\sqrt{r}\|\hat{P}_N^{-1/2}\|^2\cdot\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i}\Big\|\hat{P}_N^{-1/2}\Big(\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big)\Big\|(\hat{\sigma}_i^2-\sigma_i^2),$$

which is further bounded by

$$C^2\sqrt{r}\|\hat{P}_N^{-1/2}\|^2\cdot\Big[\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|^2\Big]^{1/2}\Big[\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{1/2}.$$

Then result (c) follows by noticing that $\hat{P}_N = O_p(N)$.

The proofs of the remaining three parts are similar to those of the first three. The details are therefore omitted. □

Lemma C.2.2. *Under Assumptions A-D,*

$$A \equiv (\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1} = O_p(T^{-1/2}) + O_p(\|\hat{\Lambda}-\Lambda\|^2) + O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{\frac{1}{2}}\Big).$$

Proof of Lemma C.2.2. Consider equation (C.1.16) in the proof of Proposition 3.4.1, we had shown $A = o_p(1)$. So term AA' is of a smaller order and hence negligible. With Lemma C.2.2 (a), (b) and (c), equation (C.1.16) can be simplified as

$$A + A' = O_p(T^{-1/2}) + o_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{\frac{1}{2}}\Big). \quad \text{(C.2.3)}$$

By the identification condition, we know both $\Lambda'(\frac{1}{N}M'\Sigma_{ee}^{-1}M)\Lambda$ and $\hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda}$ are diagonal matrices, which implies

$$\text{Ndg}\left\{\Lambda'(\frac{1}{N}M'\Sigma_{ee}^{-1}M)\Lambda - \hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda}\right\} = 0,$$

where Ndg denotes the operator which sets the diagonal elements of its input to zeros. By adding and subtracting terms,

$$\text{Ndg}\left\{(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda} + \hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda}-\Lambda) \right. \tag{C.2.4}$$

$$\left. -(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda}-\Lambda) + \Lambda'\Big[\frac{1}{N}M'(\hat{\Sigma}_{ee}^{-1} - \Sigma_{ee}^{-1})M\Big]\Lambda\right\} = 0.$$

By Lemma C.1.2 (b), $\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M = \frac{1}{N}M'\Sigma_{ee}^{-1}M + o_p(1) = R + o_p(1)$, where the last equation is due to Assumption C.3. So term $(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda}-\Lambda) = O_p(\|\hat{\Lambda}-\Lambda\|^2)$. Given this result, together with Lemma C.1.2(a), we have

$$\text{Ndg}\left\{(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda} + \hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda}-\Lambda)\right\} \tag{C.2.5}$$

$$= O_p(\|\hat{\Lambda}-\Lambda\|^2) + O_p([\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2]^{1/2}).$$

Notice that $(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda} = (\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda}\hat{P}^{-1}\hat{P} = A\hat{P}$, where the last inequality is due to the definition of A. By $\hat{P} = P + o_p(1)$ from Corollary C.1.1 (a), we have

$$(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda} = AP + o_p(A).$$

According to the preceding result, we can rewrite (C.2.5) as

$$\text{Ndg}\{AP + PA'\} = O_p(\|\hat{\Lambda} - \Lambda\|^2) + O_p([\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2]^{1/2}), \tag{C.2.6}$$

where $o_p(A)$ is discarded since it has an smaller order than other terms.

Now equation (C.2.3) has $\frac{1}{2}r(r+1)$ restrictions and equation (C.2.6) has $\frac{1}{2}r(r-1)$ restrictions, the $r \times r$ matrix A can be uniquely determined. Solving this linear equation system, we have

$$A = O_p(T^{-1/2}) + O_p(\|\hat{\Lambda} - \Lambda\|^2) + O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2\Big]^{\frac{1}{2}}\Big).$$

This completes the proof. □

Proof of Theorem 3.4.1. We first consider the first order condition (3.3.4), which can be written as

$$\text{diag}\Big\{(M_{zz} - \hat{\Sigma}_{zz}) - (M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'M' - M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\Big\} = 0,$$

where "diag" denotes the diagonal operator and $\hat{G} = (I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}$. By

$$M_{zz} = M\Lambda\Lambda'M' + \Sigma_{ee} + M\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_t' + \frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Lambda'M' + \frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \Sigma_{ee}),$$

with some algebra manipulations, we can further write the preceding equation as

$$\hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2 + 2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}$$

$$-2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i-2m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})] \tag{C.2.7}$$

$$+m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i-2m_i'(\hat{\Lambda}-\Lambda)\hat{\Lambda}'m_i+2m_i'(\hat{\Lambda}-\Lambda)\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i$$

$$+2m_i'\Lambda(\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i+2\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i.$$

By $\hat{G}\hat{P}_N=\hat{P}_N\hat{G}=I_N-\hat{G}$, we have $\hat{G}=(I_N-\hat{G})\hat{P}_N^{-1}=\hat{P}_N^{-1}(I_N-\hat{G})$. Then, the third term on right hand side (ignoring the factor 2) is equal to

$$m_i'\hat{\Lambda}(I_N-\hat{G})\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_{it}=m_i'\hat{\Lambda}(I_N-\hat{G})(I-A)'\frac{1}{T}\sum_{t=1}^{T}f_te_{it} \tag{C.2.8}$$

and the sum of the seventh and eighth terms is equal to $-2m_i'(\hat{\Lambda}-\Lambda)\hat{G}\hat{\Lambda}'m_i$. Define

$$\ddot{\psi}=\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1};\qquad \ddot{\phi}=\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}.$$

Now consider the sum of the fourth and ninth terms. By $\hat{G}=\hat{P}_N^{-1}(I_N-\hat{G})$, together with the definitions of $\ddot{\psi}$, this term is equal to

$$\begin{aligned}
&-2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i+2m_i'\Lambda(\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i\\
&=-2m_i'\Lambda\ddot{\psi}(I_N-\hat{G})\hat{\Lambda}'m_i+2m_i'\Lambda A(I_N-\hat{G})\hat{\Lambda}'m_i\\
&=2m_i'\Lambda\ddot{\psi}\hat{G}\hat{\Lambda}'m_i-2m_i'\Lambda A\hat{G}\hat{\Lambda}'m_i-2m_i'\Lambda\ddot{\psi}(\hat{\Lambda}-\Lambda)'m_i+2m_i'\Lambda A(\hat{\Lambda}-\Lambda)'m_i\\
&\quad+m_i'\Lambda(A+A'-\ddot{\psi}-\ddot{\psi}')\Lambda'm_i.
\end{aligned}$$

Also, by (C.1.16), we have

$$A' + A = A'A + \ddot{\phi} + (I_r - A)'\ddot{\psi} + \ddot{\psi}'(I_r - A) - \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1},$$

or equivalently

$$A' + A - \ddot{\psi} - \ddot{\psi}' = A'A + \ddot{\phi} - A'\ddot{\psi} - \ddot{\psi}'A - \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}.$$

Thus, it follows that

$$\begin{aligned} & -2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i + 2m_i'\Lambda(\hat{\Lambda} - \Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i \qquad \text{(C.2.9)} \\ &= 2m_i'\Lambda\ddot{\psi}\hat{G}\hat{\Lambda}'m_i - 2m_i'\Lambda A\hat{G}\hat{\Lambda}'m_i - 2m_i'\Lambda\ddot{\psi}(\hat{\Lambda} - \Lambda)'m_i + 2m_i'\Lambda A(\hat{\Lambda} - \Lambda)'m_i - m_i'\Lambda A'A\Lambda'm_i \\ & \quad -m_i'\Lambda\ddot{\phi}\Lambda'm_i + 2m_i'\Lambda A'\ddot{\psi}\Lambda'm_i + m_i'\Lambda\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}\Lambda'm_i. \end{aligned}$$

Using (C.2.8) and (C.2.9), we can rewrite (C.2.7) as

$$\begin{aligned} \hat{\sigma}_i^2 - \sigma_i^2 &= \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2) - 2m_i'(\hat{\Lambda} - \Lambda)\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}\hat{G}\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} \qquad \text{(C.2.10)} \\ & + 2m_i'\hat{\Lambda}A'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{G}A'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\Lambda\ddot{\psi}\hat{G}\hat{\Lambda}'m_i \\ & - 2m_i'\Lambda A\hat{G}\hat{\Lambda}'m_i - 2m_i'\Lambda\ddot{\psi}(\hat{\Lambda} - \Lambda)'m_i + 2m_i'\Lambda A(\hat{\Lambda} - \Lambda)'m_i \\ & + m_i'\Lambda A'A\Lambda'm_i - 2m_i'\Lambda A'\ddot{\psi}\Lambda'm_i - 2m_i'(\hat{\Lambda} - \Lambda)\hat{G}\hat{\Lambda}'m_i + 2\frac{\hat{\sigma}_i^2 - \sigma_i^2}{\hat{\sigma}_i^2}m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i \\ & + m_i'\Lambda\ddot{\phi}\Lambda'm_i - m_i'\Lambda\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}\Lambda'm_i \\ & - 2m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] + m_i'(\hat{\Lambda} - \Lambda)(\hat{\Lambda} - \Lambda)'m_i \end{aligned}$$

$$= a_{i,1} + a_{i,2} + \cdots + a_{i,17}, \quad \text{say.}$$

By the Cauchy-Schwartz inequality, we have

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2 \le 17\frac{1}{N}\sum_{i=1}^{N}(\|a_{i,1}\|^2 + \cdots + \|a_{i,17}\|^2).$$

The first term $N^{-1}\sum_{i=1}^{N}\|a_{1i}\|^2 = O_p(T^{-1})$ by

$$E\left[\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2)\right|^2\right] = O(T^{-1}).$$

The second term is bounded in norm by

$$4C^2\|\hat{\Lambda} - \Lambda\|^2\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\right\|^2 = o_p(T^{-1})$$

by $\hat{\Lambda} - \Lambda = o_p(1)$ and

$$E\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\right\|^2\right] = O(T^{-1}).$$

Similarly, one can show that the 3rd, 4th, 5th, 6th, 8th, 11th and 14th terms are all $o_p(T^{-1})$.

The 7th term is bounded in norm by

$$(4\|\Lambda\|^2 \cdot \|\hat{\Lambda}\|^2 \cdot \|\hat{G}\|^2 \cdot \|A\|^2)\frac{1}{N}\sum_{i=1}^{N}\|m_i\|^4,$$

which is $O_p(N^{-2}T^{-1}) + O_p(N^{-2}) \cdot O_p(\|\hat{\Lambda} - \Lambda\|^4) + O_p(N^{-2}) \cdot O_p[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)]$ by $\hat{G} = O_p(N^{-1})$, $\hat{\Lambda} = \Lambda + o_p(1)$ and Lemma C.2.2. This result can be simplified to $\frac{1}{N}\sum_{i=1}^{N}\|a_{i,7}\|^2 =$

$o_p(T^{-1})+o_p(\|\hat{\Lambda}-\Lambda\|^2)$ since $O_p(N^{-2})\cdot O_p[\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2)]$ is of smaller order than $\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2)^2$. Similar to the 7th term, the 9th and 10th terms are both of the order $o_p(T^{-1})+o_p(\|\hat{\Lambda}-\Lambda\|^2)$. The 12th term is $o_p(\|\hat{\Lambda}-\Lambda\|^2)$ by $\hat{G}=O_p(N^{-1})$. The 13th term is of smaller order term than $\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2)$ and therefore negligible. The 15th term is $o_p(\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2))$ by Lemma C.2.1 (f). The 16th term is $O_p(T^{-1})$. The last term is $O_p(\|\hat{\Lambda}-\Lambda\|^4)$. Given the above results, we have

$$\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2)^2=O_p(T^{-1})+o_p(\|\hat{\Lambda}-\Lambda\|^2). \tag{C.2.11}$$

Next, we derive bounds for $\|\hat{\Lambda}-\Lambda\|^2$. By equation (C.1.18), together with Lemma C.2.1(b), (d), (e) and (f) and Lemma C.2.2, we have

$$\hat{\Lambda}-\Lambda=O_p(T^{-1/2})+O_p([\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2)^2]^{1/2}). \tag{C.2.12}$$

Substituting equation (C.2.12) into (C.2.11), we have $\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2)^2=O_p(T^{-1})$. This proves the second result of Theorem 3.4.1. □

To prove the first result of Theorem 3.4.1, we need the following lemmas.

Lemma C.2.3. *Under Assumptions A-D, we have*

$$(a)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}$$
$$=O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2});$$

$$(b)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T e_tf_t'=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1});$$

$(c)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1} = O_p(N^{-1}T^{-1/2});$

$(d)\quad \frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1});$

$$(e)\quad \hat{P}_N^{-1}\hat{\Lambda}'\Big(M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_t'-\Sigma_{ee}]\hat{\Sigma}_{ee}^{-1}M\Big)\hat{R}_N^{-1}$$

$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});$$

$(f)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} = O_p(N^{-1}T^{-1/2}).$

Proof of Lemma C.2.3. We first consider (a). We rewrite it as

$$\hat{P}^{-1}\hat{\Lambda}'\Big(\frac{1}{N^2}M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\Big)\hat{\Lambda}\hat{P}^{-1}.$$

Since we already know that $\|\hat{P}^{-1}\| = O_p(1)$ and $\|\hat{\Lambda}'\| = O_p(1)$, we only need to consider the term in the big parenthesis, which is

$$\begin{aligned}
&\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_im_j'\frac{1}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})] \\
&= \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_im_j'\Big(\frac{1}{\hat{\sigma}_i^2}-\frac{1}{\sigma_i^2}\Big)\Big(\frac{1}{\hat{\sigma}_j^2}-\frac{1}{\sigma_j^2}\Big)\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})] \\
&\quad +\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_im_j'\frac{1}{\sigma_i^2}\Big(\frac{1}{\hat{\sigma}_j^2}-\frac{1}{\sigma_j^2}\Big)\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})] \\
&\quad +\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_im_j'\frac{1}{\sigma_j^2}\Big(\frac{1}{\hat{\sigma}_i^2}-\frac{1}{\sigma_i^2}\Big)\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})] \\
&\quad +\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_im_j'\frac{1}{\sigma_i^2\sigma_j^2}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})].
\end{aligned}$$

By the Cauchy-Schwarz inequality, one can show the first term is bounded in norm by

$$C^8\Big(\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big)\Big(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big\|^2\Big)^{1/2},$$

which is $O_p(T^{-3/2})$ by the second part of Theorem 3.4.1. The second term equals to

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}m_i m_j'\frac{1}{\sigma_i^2}\Big(\frac{1}{\hat{\sigma}_j^2}-\frac{1}{\sigma_j^2}\Big)\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]$$

$$=\frac{1}{N}\sum_{j=1}^{N}m_j'\Big(\frac{1}{\hat{\sigma}_j^2}-\frac{1}{\sigma_j^2}\Big)\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}m_i[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big),$$

which is bounded in norm by

$$C^4\Big[\frac{1}{N}\sum_{j=1}^{N}(\hat{\sigma}_j^2-\sigma_j^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{j=1}^{N}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}m_i[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)^2\Big]^{1/2},$$

which is $O_p(N^{-1/2}T^{-1})$. Similarly, the third term is also $O_p(N^{-1/2}T^{-1})$. The last term is $O_p(N^{-1}T^{-1/2})$. Hence result (a) follows.

Next, we consider (b). The left hand side of (b) is equivalent to

$$\hat{P}^{-1}\hat{\Lambda}'\Big(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Big).$$

Similarly to (a), it suffices to consider the term inside the parenthesis, which is

$$\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t'=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}m_i\frac{1}{T}\sum_{t=1}^{T}e_{it}f_t'$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}m_i f_t' e_{it}+\frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{\hat{\sigma}_i^2}-\frac{1}{\sigma_i^2}\Big)\frac{1}{T}\sum_{t=1}^{T}m_i f_t' e_{it}.$$

The first term is $O_p(N^{-1/2}T^{-1/2})$. The second term is bounded in norm by

$$C^4\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big\|^2\Big]^{1/2},$$

which is $O_p(T^{-1})$ by the second part of Theorem 3.4.1. Hence. result (b) follows.

For part (c), the left hand side of (c) is equivalent to

$$\hat{P}^{-1}\hat{\Lambda}'\Big(\frac{1}{N^2}M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\Big)\hat{\Lambda}\hat{P}^{-1}.$$

It suffices to consider the expression in the parenthesis:

$$\frac{1}{N^2}\sum_{i=1}^{N}m_im_i'\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^4}\le\frac{1}{N}\Big(\frac{1}{N}\sum_{i=1}^{N}\|m_i\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\|m_i'\|^2\frac{(\hat{\sigma}_i^2-\sigma_i^2)^2}{\hat{\sigma}_i^8}\Big)^{1/2},$$

which is $O_p(N^{-1}T^{-1/2})$ by the second part of Theorem 3.4.1. This proves result (c). The proofs of results (d), (e) and (f) are similar to those of (a), (b) and (c). The details are therefore omitted. □

Lemma C.2.4. *Under Assumptions A-D,*

$$A\equiv(\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}=O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T}\right)+O_p(\|\hat{\Lambda}-\Lambda\|^2).$$

Proof of Lemma C.2.4. Consider equation (C.1.16). Using the results in Lemma C.2.3 and the fact that $A'A$ has an order smaller than that of A and is therefore negligible, we have

$$A+A'=O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T}\right). \tag{C.2.13}$$

Now consider the term $\frac{1}{N}\Lambda' M'(\hat{\Sigma}_{ee}^{-1} - \Sigma_{ee}^{-1})M\Lambda$, which can be written as

$$\begin{aligned}\frac{1}{N}\Lambda' M'(\hat{\Sigma}_{ee}^{-1} - \Sigma_{ee}^{-1})M\Lambda &= -\Lambda'\Big[\frac{1}{N}\sum_{i=1}^{N} m_i m_i' \frac{\hat{\sigma}_i^2 - \sigma_i^2}{\hat{\sigma}_i^2\sigma_i^2}\Big]\Lambda \\ &= -\Lambda'\Big[\frac{1}{N}\sum_{i=1}^{N} m_i m_i' \frac{\hat{\sigma}_i^2 - \sigma_i^2}{\sigma_i^4}\Big]\Lambda + \Lambda'\Big[\frac{1}{N}\sum_{i=1}^{N} m_i m_i' \frac{(\hat{\sigma}_i^2 - \sigma_i^2)^2}{\hat{\sigma}_i^2\sigma_i^4}\Big]\Lambda.\end{aligned} \tag{C.2.14}$$

The norm of the second expression on the right hand side of (C.2.14) is bounded by

$$C\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2 = O_p(T^{-1}),$$

by the boundedness of $m_i, \hat{\sigma}_i^2, \sigma_i^2$ by Assumptions C and D. Substituting (C.2.10) into the first expression on the right hand side of (C.2.14) and using the same arguments as we did at before (C.2.11), one can show that the first expression is $O_p(\frac{1}{\sqrt{NT}}) + o_p(\frac{1}{T})$. Hence, we have

$$\frac{1}{N}\Lambda' M'(\hat{\Sigma}_{ee}^{-1} - \Sigma_{ee}^{-1})M\Lambda = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right). \tag{C.2.15}$$

Now consider (C.2.4). Using the same arguments as in the derivation of (C.2.6) except that the result for $\frac{1}{N}\Lambda' M'(\hat{\Sigma}_{ee}^{-1} - \Sigma_{ee}^{-1})M\Lambda$ is given by (C.2.15) instead of $o_p([\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2]^{1/2})$, we have

$$\text{Ndg}\{AP + PA'\} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right) + O_p(\|\hat{\Lambda} - \Lambda\|^2). \tag{C.2.16}$$

Solving the equation system (C.2.13) and (C.2.16), we have

$$A = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right) + O_p(\|\hat{\Lambda} - \Lambda\|^2),$$

as asserted in this lemma. This proves Lemma C.2.4. □

Proof of Theorem 3.4.1 (continued). Using the results in Lemma C.2.3 and Lemma C.2.4 and noticing that $\|\hat{\Lambda}-\Lambda\|^2$ is of smaller order than $\hat{\Lambda}-\Lambda$ and therefore negligible, we have from (C.1.18)

$$\hat{\Lambda}-\Lambda = O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T}\right),$$

as asserted by the first result of Theorem 3.4.1. This completes the proof of Theorem 3.4.1.

Corollary C.2.1. *Under Assumptions A-D,*

$$A \equiv (\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1} = O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T}\right).$$

Corollary C.2.1 is a direct result of Lemma C.2.4 and Theorem 3.4.1.

Lemma C.2.5. *Under Assumptions A-D,*

$$(a)\quad \frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} = \frac{1}{T}\sum_{t=1}^{T} f_t e_t'\Sigma_{ee}^{-1}MR_N^{-1} + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});$$

$$(b)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t' = P_N^{-1}\Lambda'M'\Sigma_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t' + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});$$

$$(c)\quad \frac{1}{N}M'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})M = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}m_i m_i'(e_{it}^2-\sigma_i^2) + \frac{1}{NT}\sum_{i=1}^{N} m_i m_i'\frac{\kappa_{i,4}-\sigma_i^4}{\sigma_i^4}$$
$$+ O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).$$

Proof of Lemma C.2.5. Equation (C.2.10) can be written as

$$\hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2) + \mathcal{R}_i, \tag{C.2.17}$$

where

$$\mathcal{R}_i = -2m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] + \mathcal{S}_i$$

with

$$\begin{aligned}
\mathcal{S}_i = & -2m_i'(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}\hat{G}\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}A'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{G}A'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} \\
& + 2m_i'\Lambda\ddot{\psi}\hat{G}\hat{\Lambda}'m_i - 2m_i'\Lambda A\hat{G}\hat{\Lambda}'m_i - 2m_i'\Lambda\ddot{\psi}(\hat{\Lambda}-\Lambda)'m_i + 2m_i'\Lambda A(\hat{\Lambda}-\Lambda)'m_i \\
& + m_i'\Lambda A'A\Lambda'm_i - 2m_i'\Lambda A'\ddot{\psi}\Lambda'm_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{G}\hat{\Lambda}'m_i + 2\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i \\
& + m_i'\Lambda\ddot{\phi}\Lambda'm_i - m_i'\Lambda\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}\Lambda'm_i + m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i.
\end{aligned}$$

Given that $\ddot{\psi} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma C.2.3 (b), $\hat{\Lambda}-\Lambda = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Theorem 3.4.1, $A = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Corollary C.2.1, by the same arguments in the derivation of (C.2.10), we have

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_i^2 = O_p(N^{-1}T^{-2}) + O_p(N^{-2}T^{-1}) + O_p(T^{-3}). \tag{C.2.18}$$

We now consider

$$\frac{1}{N}\sum_{i=1}^{N}\Big|m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})]\Big|^2,$$

which is bounded in norm by

$$C^2\|\hat{\Lambda}\|^4 \cdot \|\hat{G}_N\|^2 \cdot \frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it} - E(e_te_{it})]\Big|^2.$$

Since $\hat{\Lambda} = \Lambda + o_p(1)$ and $\hat{G}_N = O_p(1)$, it suffices to consider the term

$$\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it} - E(e_te_{it})]\Big|^2,$$

which, by the Cauchy-Schwarz inequality, is bounded by

$$2\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{NT}\sum_{j=1}^{N}\frac{1}{\sigma_j^2}m_j\sum_{t=1}^{T}[e_{jt}e_{it} - E(e_{jt}e_{it})]\Big|^2$$

$$+2\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{NT}\sum_{j=1}^{N}\frac{\hat{\sigma}_j^2 - \sigma_j^2}{\hat{\sigma}_j^2\sigma_j^2}m_j\sum_{t=1}^{T}[e_{jt}e_{it} - E(e_{jt}e_{it})]\Big|^2.$$

The first expression is $O_p(N^{-1}T^{-1})$. The second expression is bounded by

$$C^{10}\Big[\frac{1}{N}\sum_{j=1}^{N}(\hat{\sigma}_j^2 - \sigma_j^2)^2\Big]\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]\Big|^2\Big] = O_p(T^{-2}).$$

Given the above result, we have

$$\frac{1}{N}\sum_{i=1}^{N}\Big|m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it} - E(e_te_{it})]\Big|^2 = O_p\left(\frac{1}{NT}\right) + O_p\left(\frac{1}{T^2}\right).$$

This result, together with (C.2.18), gives

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{R}_i^2 = O_p\left(\frac{1}{NT}\right) + O_p\left(\frac{1}{T^2}\right). \tag{C.2.19}$$

Notice that

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T} f_t e_t' \hat{\Sigma}_{ee}^{-1} M &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2} f_t e_{it} m_i' \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2} f_t e_{it} m_i' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2\sigma_i^2} f_t e_{it} m_i'.\end{aligned}$$

The second term can be written as

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\hat{\sigma}_i^2\sigma_i^2} f_t e_{it}(e_{is}^2-\sigma_i^2)m_i' + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2\sigma_i^2}\mathcal{R}_i f_t e_{it} m_i'$$

The second term of the above equation is bounded in norm by

$$C^5\left[\frac{1}{N}\sum_{i=1}^{N}\|\mathcal{R}_i\|^2\right]^{1/2}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}\right\|^2\right]^{1/2},$$

which is $O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2})$ by (C.2.19). The first term can be written as

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^4} f_t e_{it}(e_{is}^2-\sigma_i^2)m_i' - \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2\sigma_i^4} f_t e_{it}(e_{is}^2-\sigma_i^2)m_i'.$$

The first term of the above expression is $O_p(N^{-1/2}T^{-1})$. The second term is bounded in norm by

$$C^5 \left[\frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_i^2 - \sigma_i^2)^2 \right]^{1/2} \left[\frac{1}{N} \sum_{i=1}^{N} \Big\| \frac{1}{T} \sum_{t=1}^{T} f_t e_{it} \Big\|^2 \cdot \Big\| \frac{1}{T} \sum_{t=1}^{T} e_{is}^2 - \sigma_i^2 \Big\|^2 \right]^{1/2},$$

which is $O_p(T^{-3/2})$. Given the above results, we have

$$\frac{1}{NT} \sum_{t=1}^{T} f_t e_t' \hat{\Sigma}_{ee}^{-1} M = \frac{1}{NT} \sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right). \qquad \text{(C.2.20)}$$

Given (C.2.20), together with $\hat{R} = R + O_p(T^{-1/2})$, we immediately obtain (a). Given (C.2.20), together with $\hat{P} = P + O_p(T^{-1/2})$ and $\hat{\Lambda} = \Lambda + O_p(\frac{1}{\sqrt{N}T}) + O_p(\frac{1}{T})$, we also have (b).

We now consider (c). The left hand side of (c) is equal to

$$-\frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\sigma}_i^2 - \sigma_i^2}{\hat{\sigma}_i^2 \sigma_i^2} m_i m_i' = -\frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\sigma}_i^2 - \sigma_i^2}{\sigma_i^4} m_i m_i' + \frac{1}{N} \sum_{i=1}^{N} \frac{(\hat{\sigma}_i^2 - \sigma_i^2)^2}{\hat{\sigma}_i^2 \sigma_i^4} m_i m_i'.$$

We use i_1 and i_2 to denote the two expressions on the right hand side of the above equation. We first consider i_1. Substituting (C.2.17) into this term, we obtain

$$i_1 = -\frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\sigma}_i^2 - \sigma_i^2}{\sigma_i^4} m_i m_i' = -\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^4} (e_{it}^2 - \sigma_i^2) m_i m_i'$$
$$+2 \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^4} \operatorname{tr}\Big[\hat{\Lambda} \hat{G} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} \frac{1}{T} \sum_{t=1}^{T} [e_t e_{it} - E(e_t e_{it})] m_i' \Big] m_i m_i' - \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^4} \mathcal{S}_i m_i m_i'.$$

Consider the second expression. The (v, u) element of this expression $(v, u = 1, \ldots, k)$ is

$$\operatorname{tr}\Big[\frac{1}{N} \sum_{i=1}^{N} \hat{\Lambda} \hat{G} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} \frac{1}{T} \sum_{t=1}^{T} [e_t e_{it} - E(e_t e_{it})] \frac{1}{\sigma_i^4} m_i' m_{iv} m_{iu} \Big]$$

which can be proved to be $O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2})$ similarly as Lemma C.2.3(a). The third term is bounded by

$$C^6 \left[\frac{1}{N} \sum_{i=1}^{N} \mathcal{S}_i^2 \right]^{1/2} = O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2})$$

by (C.2.18). Hence, we have

$$i_1 = -\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^4}(e_{it}^2 - \sigma_i^2) m_i m_i' + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right).$$

Proceed to consider i_2. By

$$\hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2) + \mathcal{R}_i,$$

we can write i_2 as

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2 \sigma_i^4} \left[\frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2) \right]^2 m_i m_i' + 2\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2 \sigma_i^4} \left[\frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2) \right] \mathcal{R}_i m_i m_i' + \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2 \sigma_i^4} \mathcal{R}_i^2 m_i m_i'.$$

We analyze the three terms at right-hand-side of the above equation one by one. The second term is bounded in norm by

$$2C^8 \left[\frac{1}{N} \sum_{i=1}^{N} \left| \frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2) \right|^2 \right]^{1/2} \left[\frac{1}{N} \sum_{i=1}^{N} \mathcal{R}_i^2 \right]^{1/2},$$

which is $O_p(N^{-1/2}T^{-1})$ by (C.2.19). The third term is bounded in norm by

$$C^8 \frac{1}{N} \sum_{i=1}^{N} \mathcal{R}_i^2 = O_p\left(\frac{1}{NT}\right) + O_p\left(\frac{1}{T^2}\right)$$

by (C.2.19). Finally, the first term can be written as

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^6} \Big[\frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2)\Big]^2 m_i m_i' - \frac{1}{N} \sum_{i=1}^{N} \frac{\hat{\sigma}_i^2 - \sigma_i^2}{\hat{\sigma}_i^2 \sigma_i^6} \Big[\frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2)\Big]^2 m_i m_i'$$

The first term of the above expression is equal to

$$\frac{1}{NT} \sum_{i=1}^{N} \frac{\kappa_{i,4} - \sigma_i^4}{\sigma_i^6} m_i m_i' + O_p(N^{-1/2}T^{-1}).$$

The second term is bounded in norm by

$$C^{10} \left[\frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_i^2 - \sigma_i^2)^2\right]^{1/2} \left[\frac{1}{N} \sum_{i=1}^{N} \Big|\frac{1}{T} \sum_{t=1}^{T} (e_{it}^2 - \sigma_i^2)\Big|^4\right]^{1/2} = O_p(T^{-3/2}).$$

Hence, we have

$$i_2 = \frac{1}{NT} \sum_{i=1}^{N} \frac{\kappa_{i,4} - \sigma_i^4}{\sigma_i^6} m_i m_i' + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right).$$

Summarizing the results on i_1 and i_2, we have (c). $\square$

PROOF OF THEOREM 3.4.2. We first derive the asymptotic behavior of A. Consider equation (C.1.16), using Lemma C.2.3 (a) and (f), Lemma C.2.5 (b) and Lemma C.2.4, we have

$$A + A' = \eta + \eta' + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}),$$

where

$$\eta = \frac{1}{NT}\sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M \Lambda P^{-1}.$$

Let vech(B) be the operation which stacks the elements on and below the diagonal of matrix B into a vector, for any square matrix B. Taking vech operation on both sides, we get

$$\text{vech}(A + A') = \text{vech}(\eta + \eta') + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).$$

Let D_r be the r-dimensional duplication matrix and D_r^+ be its Moore-Penrose inverse. By the basic fact that vech$(B + B') = 2D_r^+\text{vec}(B)$, for any $r \times r$ matrix B, we have

$$2D_r^+\text{vec}(A) = 2D_r^+\text{vec}(\eta) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}). \tag{C.2.21}$$

Furthermore, define

$$\zeta = \Lambda'\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{m_i m_i'}{\sigma_i^4}(e_{it}^2 - \sigma_i^2)\Big]\Lambda, \quad \mu = \Lambda'\Big[\frac{1}{NT}\sum_{i=1}^{N}\frac{\kappa_{i,4} - \sigma_i^4}{\sigma_i^6} m_i m_i'\Big]\Lambda.$$

Proceed to consider equation (C.2.4). By Lemma C.2.5(c) and $\hat{\Lambda} - \Lambda = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Theorem 3.4.1, we have

$$\text{Ndg}\Big\{\hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda} - \Lambda) + (\hat{\Lambda} - \Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda}\Big\}$$

$$= \text{Ndg}\{\zeta - \mu\} + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).$$

Using the same arguments in the derivation of (C.2.16), we have

$$\text{Ndg}(AP + PA') = \text{Ndg}(\zeta - \mu) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).$$

Let veck(B) be the operation which stacks the elements below the diagonal of matrix B into a vector, for any square matrix B. Let $\mathcal{D}$ be the matrix such that $\text{veck}(B) = \mathcal{D}\text{vec}(B)$ for any $r \times r$ matrix B. By the preceding equation,

$$\text{veck}(AP + PA') = \text{veck}(\zeta - \mu) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}),$$

or equivalently

$$\mathcal{D}\text{vec}(AP + PA') = \mathcal{D}\text{vec}(\zeta - \mu) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).$$

Using $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, we can rewrite the preceding equation as

$$\mathcal{D}[(P \otimes I_r) + (I_r \otimes P)K_r]\text{vec}(A) = \mathcal{D}\text{vec}(\zeta - \mu) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}), \tag{C.2.22}$$

where K_r is the r-dimensional communication matrix such that $K_r\text{vec}(B') = \text{vec}(B)$ for any $r \times r$ matrix B. By (C.2.21) and (C.2.22), we have

$$\begin{bmatrix} 2D_r^+ \\ \mathcal{D}[(P \otimes I_r) + (I_r \otimes P)K_r] \end{bmatrix} \text{vec}(A) = \begin{bmatrix} 2D_r^+\text{vec}(\eta) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{D}\text{vec}(\zeta) \end{bmatrix} - \begin{bmatrix} 0 \\ \mathcal{D}\text{vec}(\mu) \end{bmatrix} \tag{C.2.23}$$

$$+ O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).$$

Define

$$\mathbb{D}_1 = \begin{bmatrix} 2D_r^+ \\ \mathcal{D}[(P \otimes I_r) + (I_r \otimes P)K_r] \end{bmatrix}, \quad \mathbb{D}_2 = \begin{bmatrix} 2D_r^+ \\ 0_{\frac{1}{2}r(r-1)\times r^2} \end{bmatrix}, \quad \mathbb{D}_3 = \begin{bmatrix} 0_{\frac{1}{2}r(r+1)\times r^2} \\ \mathcal{D} \end{bmatrix}.$$

The above result can be rewritten as

$$\mathbb{D}_1\text{vec}(A) = \mathbb{D}_2\text{vec}(\eta) + \mathbb{D}_3\text{vec}(\zeta) - \mathbb{D}_3\text{vec}(\mu) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}). \quad \text{(C.2.24)}$$

Also, notice that

$$\text{vec}(\eta) = \text{vec}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}f_t e_{it} m_i' \Lambda P^{-1}\right] = (P^{-1}\Lambda' \otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it},$$

$$\text{vec}(\zeta) = \text{vec}\left[\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{m_i m_i'}{\sigma_i^4}(e_{it}^2 - \sigma_i^2)\Lambda\right] = (\Lambda \otimes \Lambda)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}(m_i \otimes m_i)(e_{it}^2 - \sigma_i^2)$$

and

$$\text{vec}(\mu) = \text{vec}\left[\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\frac{\kappa_{i,4} - \sigma_i^4}{\sigma_i^6}m_i m_i' \Lambda\right] = (\Lambda \otimes \Lambda)'\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\sigma_i^6}(m_i \otimes m_i)(\kappa_{i,4} - \sigma_i^4).$$

Given the above three results, we can rewrite (C.2.24) as

$$\text{vec}(A) = \mathbb{D}_1^{-1}\mathbb{D}_2(P^{-1}\Lambda' \otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it} \quad \text{(C.2.25)}$$

$$+ \mathbb{D}_1^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)' \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^4}(m_i \otimes m_i)(e_{it}^2 - \sigma_i^2)$$

$$- \mathbb{D}_1^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)' \frac{1}{NT} \sum_{i=1}^{N} \frac{1}{\sigma_i^6}(m_i \otimes m_i)(\kappa_{i,4} - \sigma_i^4)$$

$$+ O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{N}T}) + O_p(\frac{1}{T^{3/2}}).$$

Consider equation (C.1.18). Using the results of Lemma C.2.5 (a) and (b) and Lemma C.2.3 (e) and (f), we have

$$\hat{\Lambda}' - \Lambda' = -A'\Lambda' + \frac{1}{NT} \sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R^{-1} + P^{-1}\Lambda' \frac{1}{NT} M' \Sigma_{ee}^{-1} \sum_{t=1}^{T} e_t f_t' \Lambda'$$
$$+ O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right). \tag{C.2.26}$$

Notice that

$$\text{vec}\left[\frac{1}{NT} \sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R^{-1}\right] = \text{vec}\left[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} f_t e_{it} m_i' R^{-1}\right]$$
$$= (R^{-1} \otimes I_r) \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2}(m_i \otimes f_t) e_{it}$$

and

$$\text{vec}\left[P^{-1}\Lambda' \frac{1}{NT} M' \Sigma_{ee}^{-1} \sum_{t=1}^{T} e_t f_t' \Lambda'\right] = \text{vec}\left[P^{-1}\Lambda' \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} m_i e_{it} f_t' \Lambda'\right]$$
$$= K_{kr} \text{vec}\left[\Lambda \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} f_t e_{it} m_i' \Lambda P^{-1}\right]$$
$$= K_{kr}[(P^{-1}\Lambda') \otimes \Lambda] \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2}(m_i \otimes f_t) e_{it},$$

where K_{mn} is the commutation matrix such that $K_{mn}\text{vec}(B) = \text{vec}(B')$ for any $m \times n$ matrix B.

Taking vectorization operation on the both sides of (C.2.26), we have

$$\begin{aligned}\text{vec}(\hat{\Lambda}' - \Lambda') = \Big[K_{kr}[(P^{-1}\Lambda') \otimes \Lambda] + R^{-1} \otimes I_r\Big] \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it} \\ - K_{kr}(I_r \otimes \Lambda)\text{vec}(A) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right).\end{aligned} \tag{C.2.27}$$

Substituting (C.2.25) into (C.2.27),

$$\begin{aligned}\text{vec}(\hat{\Lambda}' - \Lambda') = \mathbb{B}_1 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it} - \mathbb{B}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^4}(m_i \otimes m_i)(e_{it}^2 - \sigma_i^2) \\ + \frac{1}{T}\Delta + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right),\end{aligned} \tag{C.2.28}$$

where

$$\mathbb{B}_1 = K_{kr}[(P^{-1}\Lambda') \otimes \Lambda] + R^{-1} \otimes I_r - K_{kr}(I_r \otimes \Lambda)\mathbb{D}_1^{-1}\mathbb{D}_2[(P^{-1}\Lambda') \otimes I_r],$$

$$\mathbb{B}_2 = K_{kr}(I_r \otimes \Lambda)\mathbb{D}_1^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)',$$

$$\Delta = \mathbb{B}_2 \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^6}(m_i \otimes m_i)(\kappa_{i,4} - \sigma_i^4).$$

Given the above results and by a Central Limit Theorem, we obtain as $N, T \to \infty$ and $N/T^2 \to 0$,

$$sqrtNT\Big[\text{vec}(\hat{\Lambda}' - \Lambda') - \frac{1}{T}\Delta\Big] \xrightarrow{d} N(0, \Omega),$$

where $\Omega = \lim_{N\to\infty} \Omega_N$ with

$$\Omega_N = \mathbb{B}_1(R \otimes I_r)\mathbb{B}_1' + \mathbb{B}_2\Big[\frac{1}{N}\sum_{i=1}^{N}\frac{\kappa_{i,4}-\sigma_i^4}{\sigma_i^8}(m_i m_i') \otimes (m_i m_i')\Big]\mathbb{B}_2'.$$

This completes the proof of Theorem 3.4.2. □

Proof of Theorem 3.4.5. By the definition of $\hat{f}_t = (\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda})^{-1} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} z_t$ and A, we have

$$\hat{f}_t - f_t = -A' f_t + \hat{P}^{-1}\frac{1}{N}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} e_t.$$

From Corollary B.1, we know $A = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$, then the first term of the above equation is $O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$. From Corollary C.1.1 (a)(b), we know $\hat{P} = P + o_p(1)$ and $\hat{P} = O_p(1)$, and from Assumption C.3, we know $P_\infty = \lim_{N\to\infty} P$ where P_∞ is positive definite matrix. Consider the part $\frac{1}{N}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} e_t$, which can be rewritten as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\hat{\Lambda}' m_i e_{it} = \frac{1}{N}\Lambda' M' \Sigma_{ee}^{-1} e_t, -\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2 - \sigma_i^2}{\hat{\sigma}_i^2 \sigma_i^2}\Lambda' m_i e_{it} + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}(\hat{\Lambda} - \Lambda)' m_i e_{it},$$

where m_i is the transpose of the ith row of M. Use a_1, a_2, a_3 to denote the three terms on the right hand side of the above equation. Term a_2 can be shown to be $O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}})$ by the equation (C.2.10). Term a_3 can be shown to be $O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$ by equation (C.1.18). Then we have

$$\frac{1}{N}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} e_t = \frac{1}{N}\Lambda' M' \Sigma_{ee}^{-1} e_t + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right).$$

Therefore,

$$\hat{f}_t - f_t = P^{-1}\frac{1}{N}\Lambda' M'\Sigma_{ee}^{-1}e_t + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right)$$

Based on the above result, by a Central Limit Theorem, we obtain as $N, T \to \infty$ and $N/T^2 \to 0$,

$$\sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} N(0, P_\infty^{-1}).$$

This completes the proof of Theorem 3.4.5. $\square$

C.3 Proof of Theorem 3.5.2

We only derive the asymptotic result under $H_0 : L = M\Lambda$. The consistency of the test can be easily verified. In addition, we note that since $\hat{\Lambda}^\dagger - \Lambda = O_p(\frac{1}{\sqrt{NT}}) + o_p(\frac{1}{T})$, the proof for the statistic calculated by $\hat{\Lambda}^\dagger$ is almost the same as the statistic calculated by $\hat{\Lambda}$. Hence, we will only consider the statistic calculated by $\hat{\Lambda}$ in the proofs below. We first consider the term

$$\frac{1}{N}(M\hat{\Lambda} - \hat{L})'\widetilde{\Sigma}_{ee}^{-1}(M\hat{\Lambda} - \hat{L}) = \frac{1}{N}\Big[M(\hat{\Lambda} - \Lambda) - (\hat{L} - L)\Big]'\widetilde{\Sigma}_{ee}^{-1}\Big[M(\hat{\Lambda} - \Lambda) - (\hat{L} - L)\Big]$$
$$= (\hat{\Lambda} - \Lambda)'\Big[\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}M\Big](\hat{\Lambda} - \Lambda) - (\hat{\Lambda} - \Lambda)'\Big[\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}(\hat{L} - L)\Big]$$
$$-\Big[\frac{1}{N}(\hat{L} - L)'\widetilde{\Sigma}_{ee}^{-1}M\Big](\hat{\Lambda} - \Lambda) + \frac{1}{N}(\hat{L} - L)'\widetilde{\Sigma}_{ee}^{-1}(\hat{L} - L) = I_a - I_b - I_c + I_d, \quad \text{say}$$

Consider the first term I_a. Notice that

$$\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}M - \frac{1}{N}M'\Sigma_{ee}^{-1}M = o_p(1) \tag{C.3.1}$$

by Lemma A.4 in the supplement of Bai and Li (2012). This result, together with $\hat{\Lambda} - \Lambda = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$ by Theorem 3.4.1, gives $I_a = O_p(\frac{1}{NT}) + O_p(\frac{1}{T^2})$.

For the second term I_b, the term inside the squared parenthesis is

$$\frac{1}{N} M' \widetilde{\Sigma}_{ee}^{-1} (\hat{L} - L) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\tilde{\sigma}_i^2} m_i (\hat{l}_i - l_i)'. \tag{C.3.2}$$

According to (A.14) in the supplement of Bai and Li (2012), we know that

$$\begin{aligned}
\hat{l}_i - l_i &= (\hat{L} - L)' \widetilde{\Sigma}_{ee}^{-1} \hat{L} \hat{H} l_i - \hat{H} \hat{L}' \widetilde{\Sigma}_{ee}^{-1} (\hat{L} - L)(\hat{L} - L)' \widetilde{\Sigma}_{ee}^{-1} \hat{L} \hat{H} l_i \\
&- \hat{H} \hat{L}' \widetilde{\Sigma}_{ee}^{-1} L \Big(\frac{1}{T} \sum_{t=1}^{T} f_t e_t' \Big) \widetilde{\Sigma}_{ee}^{-1} \hat{L} \hat{H} l_i - \hat{H} \hat{L}' \widetilde{\Sigma}_{ee}^{-1} \Big(\frac{1}{T} \sum_{t=1}^{T} e_t f_t' \Big) L' \widetilde{\Sigma}_{ee}^{-1} \hat{L} \hat{H} l_i \\
&- \hat{H} \Big(\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\tilde{\sigma}_i^2 \tilde{\sigma}_j^2} \hat{l}_i \hat{l}_j' \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - E(e_{it} e_{jt})] \Big) \hat{H} l_i + \hat{H} \sum_{i=1}^{N} \frac{1}{\tilde{\sigma}_i^4} \hat{l}_i \hat{l}_i' (\tilde{\sigma}_i^2 - \sigma_i^2) \hat{H} l_i \\
&+ \hat{H} \hat{L}' \widetilde{\Sigma}_{ee}^{-1} \Big(\frac{1}{T} \sum_{t=1}^{T} e_t f_t' \Big) l_i + \hat{H} \hat{L}' \widetilde{\Sigma}_{ee}^{-1} L \Big(\frac{1}{T} \sum_{t=1}^{T} f_t e_{it} \Big) \\
&+ \hat{H} \Big(\sum_{j=1}^{N} \frac{1}{\tilde{\sigma}_j^2} \hat{l}_j \frac{1}{T} \sum_{t=1}^{T} [e_{jt} e_{it} - E(e_{jt} e_{it})] \Big) - \hat{H} l_i \frac{1}{\tilde{\sigma}_i^2} (\tilde{\sigma}_i^2 - \sigma_i^2).
\end{aligned} \tag{C.3.3}$$

Substituting (C.3.3) into the right hand side of (C.3.2),

$$\frac{1}{N} M' \widetilde{\Sigma}_{ee}^{-1} (\hat{L} - L) = \Big(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\tilde{\sigma}_i^2} m_i l_i' \Big) \hat{H} \hat{L}' \widetilde{\Sigma}_{ee}^{-1} (\hat{L} - L) \tag{C.3.4}$$

$$- \Big(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\tilde{\sigma}_i^2} m_i l_i' \Big) \hat{H} \hat{L}' \widetilde{\Sigma}_{ee}^{-1} (\hat{L} - L)(\hat{L} - L)' \widetilde{\Sigma}_{ee}^{-1} \hat{L} \hat{H} + \Big(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\tilde{\sigma}_i^2} m_i l_i' \Big) \hat{H} \sum_{i=1}^{N} \frac{1}{\tilde{\sigma}_i^4} \hat{l}_i \hat{l}_i' (\tilde{\sigma}_i^2 - \sigma_i^2) \hat{H}$$

$$- \Big(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\tilde{\sigma}_i^2} m_i l_i' \Big) \hat{H} \hat{L}' \widetilde{\Sigma}_{ee}^{-1} \Big(\frac{1}{T} \sum_{t=1}^{T} e_t f_t' \Big) L' \widetilde{\Sigma}_{ee}^{-1} \hat{L} \hat{H} + \Big(\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\tilde{\sigma}_i^2} m_i l_i' \Big) \Big(\frac{1}{T} \sum_{t=1}^{T} f_t e_t' \Big) \widetilde{\Sigma}_{ee}^{-1} \hat{L} \hat{H}$$

$$-\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_i l_i'\Big)\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}L\Big(\frac{1}{T}\sum_{t=1}^{T}f_t e_t'\Big)\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}+\Big(\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_i e_{it}f_t'\Big)L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}$$

$$-\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_i l_i'\Big)\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}\hat{l}_i\hat{l}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)\hat{H}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}m_i\hat{l}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\hat{H}-\frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\tilde{\sigma}_i^4}m_i l_i'\hat{H}.$$

Similar to (C.3.1), we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_i l_i'-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}m_i l_i'=o_p(1), \tag{C.3.5}$$

which implies that $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_i l_i'=O_p(1)$. Now we analyze the terms on the right hand side of (C.3.4) one by one. The first term is $O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$ due to (C.3.5) and $\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$ by (C.10) in the supplement of Bai and Li (2012). The second term is $O_p(\frac{1}{NT})+O_p(\frac{1}{T^2})$ by the same argument. The third term is $O_p(\frac{1}{N\sqrt{T}})$ by (C.3.5) and Lemma C.1 (f) of Bai and Li (2012). The fourth, fifth and sixth terms are all $O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$ because $L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}=O_p(1)$ by Lemma C.1 (a) and $\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\frac{1}{T}\sum_{t=1}^{T}e_t f_t')=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$ by Lemma C.1 (e) of Bai and Li (2012). The seventh term is also $O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$ since $L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}=O_p(1)$ and $\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_i e_{it}f_t'=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$, where the proof of the second result is implicitly contained in the one of Lemma C.1 (e) of Bai and Li (2012). The eighth and ninth terms are both $O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{T})$ by Lemma C.1 (c) of Bai and Li (2012). The last term is $O_p(\frac{1}{N\sqrt{T}})$ by the same arguments as the third term. Summarizing all the above results, we have

$$\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)=O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T}\right).$$

This result, together with Theorem 3.4.1, shows that

$$I_b = O_p\left(\frac{1}{NT}\right) + O_p\left(\frac{1}{T^2}\right).$$

Term I_c is also $O_p(\frac{1}{NT}) + O_p(\frac{1}{T^2})$ since it is the transpose of I_b.

We now consider the last term I_d. We first rewrite equation (C.3.3) as

$$\hat{l}_i - l_i = \frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + \mathcal{T}_i, \tag{C.3.6}$$

where

$$\begin{aligned}
\mathcal{T}_i = (\hat{L}-L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)(\hat{L}-L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i \\
-\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}L\Big(\frac{1}{T}\sum_{t=1}^{T} f_t e_t'\Big)\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Big)L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i \\
-\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}\hat{l}_i\hat{l}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]\Big)\hat{H}l_i + \hat{H}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^4}\hat{l}_i\hat{l}_i'(\tilde{\sigma}_i^2 - \sigma_i^2)\hat{H}l_i \\
+\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Big)l_i - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)\Big(\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}\Big) \\
+\hat{H}\Big(\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}\hat{l}_j\frac{1}{T}\sum_{t=1}^{T}[e_{jt}e_{it} - E(e_{jt}e_{it})]\Big) - \hat{H}l_i\frac{1}{\tilde{\sigma}_i^2}(\tilde{\sigma}_i^2 - \sigma_i^2).
\end{aligned}$$

Now term I_d can be written as

$$\begin{aligned}
I_d &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}(\hat{l}_i - l_i)(\hat{l}_i - l_i)' = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + \mathcal{T}_i\Big]\Big[\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + \mathcal{T}_i\Big]' \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}\Big]' + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}\Big]\mathcal{T}_i'
\end{aligned}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\mathcal{T}_i\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\mathcal{T}_i\mathcal{T}_i'=II_a+II_b+II_c+II_d.$$

First consider II_a, which can be written as

$$II_a=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'-\frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\tilde{\sigma}_i^2\sigma_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'. \tag{C.3.7}$$

The first expression of (C.3.7) is equal to

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]+\frac{1}{T}I_r.$$

The second expression of (C.3.7) can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'-\frac{1}{N}\sum_{i=1}^{N}\frac{(\tilde{\sigma}_i^2-\sigma_i^2)^2}{\tilde{\sigma}_i^2\sigma_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'. \tag{C.3.8}$$

Equation (B.9) in the supplement of Bai and Li (2012) implies that

$$\tilde{\sigma}_i^2-\sigma_i^2=\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)+\mathcal{S}_i$$

with

$$\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_i^2=O_p(\frac{1}{NT})+O_p(\frac{1}{T^2}).$$

Consider the first term of (C.3.8), which can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]+\frac{1}{NT}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\sigma_i^2}I_r. \tag{C.3.9}$$

The first term of the preceding equation can be further written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathcal{S}_i}{\sigma_i^4}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t f_s'[e_{it}e_{is}-E(e_{it}e_{is})]$$

$$+\frac{1}{NT^3}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^4}f_t f_s'[\varepsilon_{i,uts}-E(\varepsilon_{i,uts})]+\frac{1}{NT^3}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^4}f_t f_s' E(\varepsilon_{i,uts}),$$

where $\varepsilon_{i,uts}=(e_{iu}^2-\sigma_i^2)[e_{it}e_{is}-E(e_{it}e_{is})]$. The first term of the above equation is bounded in norm by

$$C^4\left[\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_i^2\right]^{1/2}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t f_s'[e_{it}e_{is}-E(e_{it}e_{is})]\right\|^2\right]^{1/2},$$

which is $O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2})$. The second term is $O_p(\frac{1}{\sqrt{NT^3}})$. The third term is $O(\frac{1}{T^2})$. Given the above analysis, we have that the first expression of (C.3.9) is $O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2})$. Consider the second term of (C.3.9). Ignoring I_r, this term is equal to

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(e_{it}^2-\sigma_i^2)+\frac{1}{NT}\sum_{i=1}^{N}\frac{\mathcal{S}_i}{\sigma_i^2}.$$

The first term is $O_p(\frac{1}{\sqrt{NT^3}})$. The second term is bounded in norm by $C^2\frac{1}{T}(\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_i^2)^{1/2}$, which is $O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2})$. Summarizing all the results, we have shown that the first term of (C.3.8) is $O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2})$.

The second term of (C.3.8) is bounded by

$$C^6\frac{1}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i^2-\sigma_i^2)^2\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]',$$

which is further bounded in norm by

$$2C^6 \frac{1}{N}\sum_{i=1}^{N}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big]^2\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]'$$

$$+2C^6 \frac{1}{N}\sum_{i=1}^{N}\Big[\frac{1}{T}\sum_{t=1}^{T}\mathcal{S}_i\Big]^2\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]'.$$

The first term is $O_p(\frac{1}{T^2})$ and the second term is $O_p(\frac{1}{T^3}) + O_p(\frac{1}{NT^2})$. Given these results, we have

$$II_a = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t f_s'[e_{it}e_{is} - E(e_{it}e_{is})] + \frac{1}{T}I_r + O_p\left(\frac{1}{\sqrt{NT^3}}\right) + O_p\left(\frac{1}{T^2}\right).$$

The derivations of II_b and II_c are similar. So we only consider II_c. Substituting the expression of $\mathcal{T}_i$ into II_c, we have

$$\begin{aligned}
II_c = {} & (\hat{L}-L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it} \\
& - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)(\hat{L}-L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it} \\
& - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}L\Big(\frac{1}{T}\sum_{t=1}^{T}f_t e_t'\Big)\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it} \\
& - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Big)L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it} \\
& - \hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}\hat{l}_i\hat{l}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]\Big)\hat{H}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it} \\
& + \hat{H}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^4}\hat{l}_i\hat{l}_i'(\tilde{\sigma}_i^2-\sigma_i^2)\hat{H}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it} \\
& + \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Big)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it}
\end{aligned}$$

$$- \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]'$$
$$+ \hat{H}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2\tilde{\sigma}_i^2}\hat{l}_j\frac{1}{T}\sum_{t=1}^{T}[e_{jt}e_{it}-E(e_{jt}e_{it})]\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]'$$
$$- \hat{H}\frac{1}{N}\sum_{i=1}^{N}l_i\frac{1}{\tilde{\sigma}_i^4}(\tilde{\sigma}_i^2-\sigma_i^2)\Big[\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big]'.$$

Notice that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it} = O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T}\right),$$

which is shown in Lemma C.1 (e) of Bai and Li (2012). Given the above result, together with $(\hat{L}-L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H} = O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$ by (C.10) in the supplement of Bai and Li (2012), we have that the first term is $O_p(\frac{1}{NT})+O_p(\frac{1}{T^2})$. By similar arguments, one can show that the second term is $O_p(\frac{1}{\sqrt{N^3T^3}})+O_p(\frac{1}{T^3})$, the third and the fourth terms are both $O_p(\frac{1}{NT})+O_p(\frac{1}{T^2})$. The fifth term is $O_p(\frac{1}{\sqrt{N^3T^2}})+O_p(\frac{1}{T^2})$. The sixth term is $O_p(\frac{1}{\sqrt{N^3T^2}})$. The seventh term is $O_p(\frac{1}{NT})+O_p(\frac{1}{T^2})$. The eighth term is bounded in norm by

$$C\Big\|\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)\Big\|\cdot\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big\|^2,$$

which is $O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2})$. The ninth term can be written as

$$\hat{H}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}\hat{l}_j\Big\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\} \tag{C.3.10}$$
$$-\frac{1}{N}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2\sigma_i^2}\hat{l}_j\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})].$$

The first term of (C.3.10) can be written as

$$\frac{1}{NT^2}\hat{H}\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2\sigma_j^2}l_j f_t' e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]$$
$$-\hat{H}\sum_{j=1}^{N}\frac{\tilde{\sigma}_j^2-\sigma_j^2}{\tilde{\sigma}_j^2\sigma_j^2}l_j\Big\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t' e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\}$$
$$-\hat{H}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}(\hat{l}_j-l_j)\Big\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t' e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\}.$$

The first term is $O_p(\frac{1}{NT})$ since its variance is $O(\frac{1}{N^2T^2})$. The second term is bounded in norm by

$$C\cdot\|N\hat{H}\|\cdot\Big[\frac{1}{N}\sum_{j=1}^{N}(\tilde{\sigma}_j^2-\sigma_j^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t' e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\|^2\Big]^{1/2},$$

which is $O_p(\frac{1}{\sqrt{NT^3}})$ by Theorem 5.1 of Bai and Li (2012). The third term is bounded in norm by

$$C\cdot\|N\hat{H}\|\cdot\Big[\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}\|\hat{l}_j-l_j\|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{j=1}^{N}\Big\|\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t' e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\|^2\Big]^{1/2},$$

which is also $O_p(\frac{1}{\sqrt{NT^3}})$ by Theorem 5.1 of Bai and Li (2012). The second term of (C.3.10) can be written as

$$-\frac{1}{N}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(\tilde{\sigma}_i^2-\sigma_i^2)(\tilde{\sigma}_j^2-\sigma_j^2)}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2\sigma_i^2\sigma_j^2}l_j\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t' e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]$$

$$+\frac{1}{N}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2\sigma_i^2}(\hat{l}_j-l_j)\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t' e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]$$

$$+\frac{1}{N}\hat{H}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\tilde{\sigma}_i^2\sigma_i^2}\frac{1}{NT^2}\sum_{j=1}^{N}\frac{1}{\sigma_j^2}l_j\sum_{t=1}^{T}\sum_{s=1}^{T}f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})].$$

The first term is bounded in norm by

$$C\cdot\|N\hat{H}\|\cdot\Big[\frac{1}{N}\sum_{j=1}^{N}(\tilde{\sigma}_j^2-\sigma_j^2)^2\Big]\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\|^2\Big]^{1/2},$$

which is $O_p(\frac{1}{T^2})$ by Theorem 5.1 of Bai and Li (2012). The second term is bounded in norm by

$$C\cdot\|N\hat{H}\|\Big[\frac{1}{N}\sum_{j=1}^{N}(\tilde{\sigma}_j^2-\sigma_j^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}\|\hat{l}_j-l_j\|^2\Big]^{1/2}$$

$$\times\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big\|f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\|^2\Big]^{1/2},$$

which is also $O_p(\frac{1}{T^2})$ by Theorem 5.1 of Bai and Li (2012). The third term is bounded in norm by

$$C\cdot\|N\hat{H}\|\Big[\frac{1}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i^2-\sigma_i^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{NT^2}\sum_{j=1}^{N}\frac{1}{\sigma_j^2}l_j\sum_{t=1}^{T}\sum_{s=1}^{T}f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\|^2\Big]^{1/2},$$

which is $O_p(\frac{1}{\sqrt{NT^3}})$ by Theorem 5.1 of Bai and Li (2012). Summarizing all the results, we have that that the ninth term is $O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2})$. The last term is bounded in norm by

$$C\|\hat{H}\|\Big[\frac{1}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i^2-\sigma_i^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big\|^2\Big]^{1/2},$$

which is $O_p(\frac{1}{NT})$. Given the above analysis, we have

$$II_c=O_p\left(\frac{1}{\sqrt{NT^3}}\right)+O_p\left(\frac{1}{T^2}\right).$$

Term II_d is bounded in norm by $C\frac{1}{N}\sum_{i=1}^{N}\|\mathcal{T}_i\|^2$. Using the argument to prove II_c, we can show that it is bounded in norm by $O_p(\frac{1}{\sqrt{NT^3}}) + O_p(\frac{1}{T^2})$.

Given the above analysis, we have

$$I_d = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t f_s'[e_{it}e_{is} - E(e_{it}e_{is})] + \frac{1}{T}I_r + O_p\left(\frac{1}{\sqrt{NT^3}}\right) + O_p\left(\frac{1}{T^2}\right).$$

Summarizing the results on $I_a, \dots, I_d$, we have

$$\begin{aligned}
&\frac{1}{N}(M\hat{\Lambda} - \hat{L})'\widetilde{\Sigma}_{ee}^{-1}(M\hat{\Lambda} - \hat{L}) \\
&\quad = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t f_s'[e_{it}e_{is} - E(e_{it}e_{is})] + \frac{1}{T}I_r + O_p\left(\frac{1}{\sqrt{NT^3}}\right) + O_p\left(\frac{1}{T^2}\right),
\end{aligned}$$

Now consider the term $\frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t f_s'[e_{it}e_{is} - E(e_{it}e_{is})]$, which we use ω to denote. Then the variance of $\mathrm{tr}(\omega)$ is

$$\mathrm{var}(\mathrm{tr}(\omega)) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\mathrm{var}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t' f_s e_{it}e_{is}\Big] = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\mathrm{var}\Big[e_i'\frac{FF'}{T}e_i\Big]$$

where $e_i = (e_{i1}, e_{i2}, \dots, e_{iT})'$. By the well-known result that

$$\mathrm{var}(V'BV) = (\mu_4^v - 3\sigma^4)\sum_{t=1}^{T}b_{tt}^2 + \sigma^4\Big[\mathrm{tr}(BB') + \mathrm{tr}(B^2)\Big]$$

where $V = (v_1, v_2, \dots, v_T)'$ with each v_t is iid over t with mean zero and variance σ^2, $\mu_4^v = E(v_t^4)$, and B is a $T \times T$ matrix with its tth diagonal element denoted as b_{tt}, together with the fact that e_{it} is iid over t with mean zero and variance σ_i^2, then we have

$$\mathrm{var}\Big[e_i'\frac{FF'}{T}e_i\Big] = (\mu_4 - 3\sigma_i^4)\sum_{t=1}^{T}\Big(\frac{f_t'f_t}{T}\Big)^2 + \sigma_i^4\Big[\mathrm{tr}\Big(\frac{FF'}{T}\frac{FF'}{T}\Big) + \mathrm{tr}\Big(\frac{FF'}{T}\frac{FF'}{T}\Big)\Big],$$

where $\mu_4 = E(e_{it}^4)$. By the identification condition that $F'F/T = I_r$, the above equation can be rewritten as

$$\mathrm{var}\Big[e_i'\frac{FF'}{T}e_i\Big] = (\mu_4 - 3\sigma_i^4)\sum_{t=1}^{T}\Big(\frac{f_t'f_t}{T}\Big)^2 + \sigma_i^4 2r.$$

Notice that $\sum_{t=1}^{T}\Big(\frac{f_t'f_t}{T}\Big)^2 = \frac{1}{T}\frac{1}{T}\sum_{t=1}^{T}(f_t'f_t)^2$ is $O_p(\frac{1}{T})$, since $\frac{1}{T}\sum_{t=1}^{T}(f_t'f_t)^2$ is $O_p(1)$ from Assumption A. Meanwhile from Assumption B, we know both σ_i^2 and μ_4 are bounded. Therefore as $T \to \infty$, the first term on the right hand side of the above equation goes to zero, hence

$$\mathrm{var}\Big[e_i'\frac{FF'}{T}e_i\Big] = \sigma_i^4 2r,$$

which implies that $\mathrm{var}(\mathrm{tr}(\omega)) = 2r$. Hence as $N, T \to \infty$ and $N/T^2 \to 0$,

$$\begin{aligned} W &\triangleq \mathrm{tr}\Big[\sqrt{NT^2}\Big(\frac{1}{N}(M\hat{\Lambda} - \hat{L})'\widetilde{\Sigma}_{ee}^{-1}(M\hat{\Lambda} - \hat{L}) - \frac{1}{T}I_r\Big)\Big] \\ &= \frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_s'f_t[e_{it}e_{is} - E(e_{it}e_{is})] + o_p(1) \xrightarrow{d} N(0, 2r). \end{aligned}$$

This completes the whole proof of Theorem 3.5.2. □

C.4 Partially constrained factor models

We first give detailed derivations of equations (3.6.2)-(3.6.4). The first order condition for Λ is

$$\hat{\Lambda}' M' \hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1} M = 0. \tag{C.4.1}$$

The first order condition for Γ is

$$\hat{\Gamma}' \hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1} = 0. \tag{C.4.2}$$

The first order condition for Σ_{ee} is

$$\text{diag}[\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}] = 0. \tag{C.4.3}$$

By (C.4.1) and (C.4.2), together with the definition of Φ, we have

$$\hat{\Phi}' \hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1} \hat{\Phi} = 0, \tag{C.4.4}$$

where $\hat{\Phi} = [M\hat{\Lambda}, \hat{\Gamma}]$. Let $\hat{\mathcal{G}} = (I_r + \hat{\Phi}' \hat{\Sigma}_{ee}^{-1} \hat{\Phi})^{-1}$. By the Woodbury formula

$$\hat{\Sigma}_{zz}^{-1} = \hat{\Sigma}_{ee}^{-1} - \hat{\Sigma}_{ee}^{-1} \hat{\Phi} \hat{\mathcal{G}} \hat{\Phi}' \hat{\Sigma}_{ee}^{-1}, \tag{C.4.5}$$

we have $\hat{\Phi}' \hat{\Sigma}_{zz}^{-1} = \hat{\mathcal{G}} \hat{\Phi}' \hat{\Sigma}_{ee}^{-1}$. Given this result, together with (C.4.4), we have

$$\hat{\mathcal{G}} \hat{\Phi}' \hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1} \hat{\Phi} \hat{\mathcal{G}} = 0,$$

or equivalently

$$\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}\hat{\Phi} = 0. \tag{C.4.6}$$

Now equation (C.4.1) can be written as

$$0 = [I_{r_1}, 0] \begin{bmatrix} \hat{\Lambda}'M' \\ \hat{\Gamma}' \end{bmatrix} \hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}M = [I_{r_1}, 0]\hat{\Phi}'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}M$$

$$= [I_{r_1}, 0]\hat{\mathcal{G}}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}M = [I_{r_1}, 0]\hat{\mathcal{G}}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})(\hat{\Sigma}_{ee}^{-1} - \hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{G}}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1})M.$$

Using (C.4.6), we have

$$[I_{r_1}, 0]\hat{\mathcal{G}}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M = 0. \tag{C.4.7}$$

By identification condition IC′, we see that $\hat{\mathcal{G}}$ is a diagonal matrix, which we partition into

$$\hat{\mathcal{G}} = \begin{bmatrix} \hat{\mathcal{G}}_1 & 0 \\ 0 & \hat{\mathcal{G}}_2 \end{bmatrix}.$$

So we can rewrite (C.4.7) as

$$\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M = 0,$$

or equivalently

$$\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M = 0. \tag{C.4.8}$$

Proceed to consider (C.4.2). Post-multiplying $\hat{\Sigma}_{zz}$ on both side of (C.4.2) gives,

$$0 = \hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = [0, I_{r_2}] \begin{bmatrix} \hat{\Lambda}'M' \\ \hat{\Gamma}' \end{bmatrix} \hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})$$

$$= [0, I_{r_2}]\hat{\Phi}'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = [0, I_{r_2}]\hat{\mathcal{G}}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = \hat{\mathcal{G}}_2\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz}),$$

which implies that

$$\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = 0. \tag{C.4.9}$$

For ease of exposition, we introduce a matrix A in a partial constrained factor model, which is defined as

$$A \triangleq (\hat{\Phi} - \Phi)'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}(\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\hat{\Phi})^{-1} = (\hat{\Phi} - \Phi)'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1},$$

where $\hat{\mathcal{H}}_N = \hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}$. We partition matrix A as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

By definition, we have

$$A_{11} = (\hat{\Lambda} - \Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}, \qquad A_{12} = (\hat{\Lambda} - \Lambda)'M'\hat{\Sigma}_{ee}^{-1}\hat{\Gamma}\hat{Q}_N^{-1},$$

$$A_{21} = (\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}, \qquad A_{22} = (\hat{\Gamma} - \Gamma)'\hat{\Sigma}_{ee}^{-1}\hat{\Gamma}\hat{Q}_N^{-1},$$

where $\hat{P}_N = \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}$ and $\hat{Q}_N = \hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\hat{\Gamma}$. With some algebra manipulations, together with $\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\hat{\Gamma} = 0$ by the identification condition, we can rewrite the first order condition

(C.4.8) as

$$\hat{\Lambda}' - \Lambda' = -A_{11}'\Lambda' - A_{21}'\Gamma'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} - \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}$$
$$+(I - A_{11})'\frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} - A_{21}'\frac{1}{T}\sum_{t=1}^{T} g_t e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} + \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Lambda'$$
$$+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t g_t'\Gamma'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} + \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}.$$

The above result can be alternatively written as

$$\hat{\Lambda}' - \Lambda' = -A_{11}'\Lambda' - A_{21}'\Gamma'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} + \frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} \tag{C.4.10}$$
$$+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Lambda' + \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t g_t'\Gamma'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} + \mathcal{J}_\Lambda,$$

where

$$\mathcal{J}_\Lambda = -\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} - A_{11}'\frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}$$
$$-A_{21}'\frac{1}{T}\sum_{t=1}^{T} g_t e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} + \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}.$$

By similar arguments as above, the first order condition (C.4.9) can be written as

$$\hat{\gamma}_i - \gamma_i = \frac{1}{T}\sum_{t=1}^{T} g_t e_{it} + \mathcal{J}_{i,\Gamma}, \tag{C.4.11}$$

where

$$\mathcal{J}_{i,\Gamma} = -A'_{22}\gamma_i - A'_{12}\Lambda' m_i - A'_{22}\frac{1}{T}\sum_{t=1}^{T} g_t e_{it} + \hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t g_t'\gamma_i - \hat{Q}_N^{-1}\gamma_i\frac{\hat{\sigma}_i^2 - \sigma_i^2}{\hat{\sigma}_i^2}$$
$$-A'_{12}\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + \hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Lambda' m_i + \hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})].$$

Similarly, we can rewrite the first order condition (C.4.3) as

$$\operatorname{diag}\Big((M_{zz} - \hat{\Sigma}_{zz}) - M\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) - (M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\Big) = 0.$$

Given the above result, with some algebra computation, we have

$$\hat{\sigma}_i^2 - \hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2) + \mathcal{J}_{i,\sigma^2}, \tag{C.4.12}$$

where

$$\mathcal{J}_{i,\sigma^2} = -2\gamma_i'\mathcal{J}_{i,\Gamma} - (\hat{\gamma}_i - \gamma_i)'(\hat{\gamma}_i - \gamma_i) - 2m_i'(\hat{\Lambda} - \Lambda)\Lambda' m_i$$
$$-m_i'(\hat{\Lambda} - \Lambda)(\hat{\Lambda} - \Lambda)'m_i - 2m_i'(\hat{\Lambda} - \Lambda)\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda} - \Lambda)\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Lambda' m_i$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda} - \Lambda)\Lambda' m_i + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda(\hat{\Lambda} - \Lambda)'m_i$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda} - \Lambda)(\hat{\Lambda} - \Lambda)'m_i + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma} - \Gamma)\gamma_i$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\Gamma\mathcal{J}_{i,\Gamma} + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma} - \Gamma)(\hat{\gamma}_i - \gamma_i)$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'m_i\frac{\hat{\sigma}_i^2 - \sigma_i^2}{\hat{\sigma}_i^2} - 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t g_t'\gamma_i$$

$$-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})].$$

Equation (C.4.6) is equal to

$$\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\Big[\Phi\Phi'+\Sigma_{ee}-\hat{\Phi}\hat{\Phi}'-\hat{\Sigma}_{ee}+\Phi\frac{1}{T}\sum_{t=1}^{T}h_te_t'+\frac{1}{T}\sum_{t=1}^{T}e_th_t'\Phi'+\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\Big]\hat{\Sigma}_{ee}^{-1}\hat{\Phi}=0.$$

The above equation can be written as

$$\begin{aligned}A+A'=A'A+(I-A)'\frac{1}{T}\sum_{t=1}^{T}h_te_t'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}+\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_th_t'(I-A)\qquad\text{(C.4.13)}\\+\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}-\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}.\end{aligned}$$

By identification condition IC', we have

$$\text{Ndg}\left\{\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}-\frac{1}{N}\Phi'\Sigma_{ee}^{-1}\Phi\right\}=0.$$

The expression on the left hand side of the preceding equation is equal to

$$\text{Ndg}\left\{\frac{1}{N}(\hat{\Phi}-\Phi)'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}+\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(\hat{\Phi}-\Phi)-\frac{1}{N}(\hat{\Phi}-\Phi)'\hat{\Sigma}_{ee}^{-1}(\hat{\Phi}-\Phi)+\frac{1}{N}\Phi'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})\Phi\right\}.$$

Given the above result, by the definition of A, we have

$$\begin{aligned}&\text{Ndg}(A\hat{\mathcal{H}}+\hat{\mathcal{H}}A')\qquad\text{(C.4.14)}\\&=\text{Ndg}\left\{\frac{1}{N}(\hat{\Phi}-\Phi)'\hat{\Sigma}_{ee}^{-1}(\hat{\Phi}-\Phi)-\frac{1}{N}\sum_{i=1}^{N}\frac{\phi_i\phi_i'}{\hat{\sigma}_i^2\sigma_i^4}(\hat{\sigma}_i^2-\sigma_i^2)^2+\frac{1}{N}\sum_{i=1}^{N}\frac{\phi_i\phi_i'}{\sigma_i^4}(\hat{\sigma}_i^2-\sigma_i^2)\right\},\end{aligned}$$

where $\hat{\mathcal{H}} = \hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\hat{\mathcal{H}}/N$. Now we use the above results to prove Theorem 3.6.1. First we can show that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\|\hat{\phi}_i - \phi_i\|^2 \xrightarrow{p} 0 \tag{C.4.15}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2 \xrightarrow{p} 0. \tag{C.4.16}$$

Notice that the present model is a mixture of a standard factor model and a constrained factor model. In Proposition 3.4.1, we have shown the consistency of the MLE for a constrained factor model. In Proposition 5.1 of Bai and Li (2012), the consistency of the MLE for a standard factor model is shown. By combining the arguments in the proofs of Proposition 3.4.1 and Proposition 5.1 of Bai and Li (2012), one can prove the above two results.

Along with the argument of consistency, using (C.4.9), (C.4.10), one can further show that

$$\begin{aligned}
\hat{\Lambda} - \Lambda &= O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right), \\
\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\|\hat{\gamma}_i - \gamma_i\|^2 &= O_p\left(\frac{1}{T}\right), \\
\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2 &= O\left(\frac{1}{T}\right).
\end{aligned} \tag{C.4.17}$$

Equation (C.4.13) corresponds to equation (C.1.16) in the pure constrained factor model. Using the arguments as in the derivation of (C.2.13), one can obtain a similar result

$$A + A' = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right). \tag{C.4.18}$$

By the consistency results (C.4.15) and (C.4.16), one can show that $\hat{\mathcal{H}} = \mathcal{H} + o_p(1)$. So $A(\hat{\mathcal{H}} - \mathcal{H})$ is of smaller order term than A and therefore negligible. Similar to the derivation of (C.2.16), one can show that

$$\text{Ndg}(A\mathcal{H} + \mathcal{H}A') = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right). \tag{C.4.19}$$

The equation system (C.4.18) and (C.4.19) gives

$$A = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right). \tag{C.4.20}$$

Using the above result, it can be shown that

$$\mathcal{J}_{i,\sigma^2} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right).$$

The above result, together with (C.4.9), gives

$$\sqrt{T}(\hat{\sigma}_i^2 - \sigma_i^2) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2) + o_p(1).$$

Similarly, using the results in Lemma C.2.3 and (C.4.20), we have

$$\mathcal{J}_{i,\Gamma} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right).$$

This result, together with (C.4.10), gives

$$\sqrt{T}(\hat{\gamma}_i - \gamma_i) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} g_t e_{it} + o_p(1).$$

Let $\psi = (M'\Sigma_{ee}^{-1}M)^{-1}M'\Sigma_{ee}^{-1}\Gamma$. It can be shown that Lemmas C.2.3 and C.2.5 continue to hold for a constrained factor model. Given this, we can rewrite (C.4.10) as

$$\hat{\Lambda}' - \Lambda' = -A'_{11}\Lambda' - A'_{21}\psi' + \frac{1}{T}\sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R_N^{-1} + P_N^{-1}\Lambda' M' \Sigma_{ee}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t f_t' \Lambda' \quad \text{(C.4.21)}$$
$$+ P_N^{-1}\Lambda' M' \Sigma_{ee}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t g_t' \psi' + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right).$$

We note that

$$\begin{aligned}
\text{vec}\left(\frac{1}{T}\sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R_N^{-1}\right) &= \text{vec}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2} f_t e_{it} m_i' R^{-1}\right) \\
&= (R^{-1} \otimes I_{r_1}) \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2}(m_i \otimes f_t) e_{it}, \\
\text{vec}\left(P_N^{-1}\Lambda' M' \Sigma_{ee}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t f_t' \Lambda'\right) &= \text{vec}\left(P^{-1}\Lambda' \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2} m_i f_t' e_{it} \Lambda'\right) \\
&= K_{kr_1}\text{vec}\left(\Lambda \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2} f_t m_i' e_{it} \Lambda P^{-1}\right) \\
&= K_{kr_1}[(P^{-1}\Lambda') \otimes \Lambda] \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2}(m_i \otimes f_t) e_{it}, \\
\text{vec}\left(P_N^{-1}\Lambda' M' \Sigma_{ee}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t g_t' \psi'\right) &= \text{vec}\left(P^{-1}\Lambda' \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2} m_i g_t' e_{it} \psi'\right) \\
&= K_{kr_1}\text{vec}\left(\psi \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2} g_t m_i' e_{it} \Lambda P^{-1}\right) \\
&= K_{kr_1}[(P^{-1}\Lambda') \otimes \psi] \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2}(m_i \otimes g_t) e_{it}.
\end{aligned}$$

In addition

$$-A'_{11}\Lambda' - A'_{21}\psi' = -[I_{r_1}, 0_{r_1\times r_2}]\begin{bmatrix} A'_{11} & A'_{21} \\ A'_{12} & A'_{22}\end{bmatrix}\begin{bmatrix}\Lambda' \\ \psi'\end{bmatrix} = -E'_1 A'\Psi',$$

where $\Psi = [\Lambda, \psi]$, $E_1 = \begin{bmatrix} I_{r_1} \\ 0_{r_2\times r_1}\end{bmatrix}$ and $E_2 = \begin{bmatrix} 0_{r_1\times r_2} \\ I_{r_2}\end{bmatrix}$. Given the above result, we have

$$\text{vec}\Big(A'_{11}\Lambda' + A'_{21}\psi'\Big) = \text{vec}(E'_1 A'\Psi') = K_{kr_1}\text{vec}(\Psi A E_1) = K_{kr_1}(E'_1 \otimes \Psi)\text{vec}(A).$$

Taking the vectorization operation on both sides of (C.4.21), we get

$$\begin{aligned}\text{vec}(\hat{\Lambda}' - \Lambda') = \Big[(R^{-1}\otimes I_{r_1}) + K_{kr_1}[(P^{-1}\Lambda')\otimes\Lambda]\Big]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it} \qquad \text{(C.4.22)}\\ + K_{kr_1}[(P^{-1}\Lambda')\otimes\psi]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes g_t)e_{it} - K_{kr_1}(E'_1\otimes\Psi)\text{vec}(A)\\ + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right).\end{aligned}$$

Now consider (C.4.13) and (C.4.14). Again, using similar arguments as in the derivation of (C.2.21), one can show by (C.4.13) that

$$2D_r^+\text{vec}(A) = 2D_r^+\text{vec}(\eta^\star) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right), \qquad \text{(C.4.23)}$$

where $\eta^\star = \frac{1}{T}\sum_{t=1}^{T} h_t e'_t \Sigma_{ee}^{-1}\Phi\mathcal{H}_N^{-1}$ with $\mathcal{H}_N = \Phi'\Sigma_{ee}^{-1}\Phi$. To proceed the analysis, we first consider the expression $\mathcal{J}_{i,\sigma^2}$. The sum of the 3rd term and the 10th term is equal to

$$-2m_i'(\hat{\Lambda}-\Lambda)\Lambda' m_i + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda(\hat{\Lambda}-\Lambda)'m_i$$

$$= 2m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i - 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1(\hat{\Lambda}-\Lambda)'m_i - 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i.$$

By $\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\hat{\Gamma}=0$, we can rewrite the 13th term as $-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\mathcal{J}_{i,\Gamma}$. Further consider the sum of the 1st, 8th, 9th, 12th and 16th terms, which is equal to

$$\begin{aligned}
&-2\gamma_i'\mathcal{J}_{i,\Gamma} - 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Lambda' m_i + 2m_i'\hat{\Lambda}'\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda' m_i \\
&\quad + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i - 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t g_t'\gamma_i \\
&= 2\gamma_i'A_{22}'\gamma_i + 2\gamma_i'A_{12}'\Lambda' m_i + 2\gamma_i'A_{22}'\frac{1}{T}\sum_{t=1}^{T}g_t e_{it} - 2\gamma_i'\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t g_t'\gamma_i + 2\gamma_i'A_{12}'\frac{1}{T}\sum_{t=1}^{T}f_t e_{it} \\
&\quad - 2\gamma_i'\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Lambda' m_i - 2\gamma_i'\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] + 2\gamma_i'\hat{Q}_N^{-1}\gamma_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2} \\
&\quad - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Lambda' m_i + 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Lambda' m_i \\
&\quad - 2m_i'\Lambda\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t'\Lambda' m_i + 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda' m_i \\
&\quad - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{11}'\Lambda' m_i + 2m_i'\Lambda A_{11}'\Lambda' m_i + 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{21}'\gamma_i \\
&\quad + 2m_i'\Lambda A_{21}'\gamma_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t g_t'\gamma_i + 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t g_t'\gamma_i \\
&\quad - 2m_i'\Lambda\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t g_t'\gamma_i \\
&= \phi_i'\Big[A + A' - \hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t h_t' - \frac{1}{T}\sum_{t=1}^{T}h_t e_t'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\Big]\phi_i + 2\gamma_i'A_{22}'\frac{1}{T}\sum_{t=1}^{T}g_t e_{it} \\
&\quad + 2\gamma_i'A_{12}'\frac{1}{T}\sum_{t=1}^{T}f_t e_{it} - 2\gamma_i'\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] + 2\gamma_i'\hat{Q}_N^{-1}\gamma_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}
\end{aligned}$$

$$
\begin{aligned}
&- 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i + 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i \\
&+ 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{11}'\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{21}'\gamma_i \\
&+ 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i \\
&+ 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i \\
= {}& \phi_i'A'A\phi_i - 2\phi_i'A'\frac{1}{T}\sum_{t=1}^{T}h_te_t'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i - \phi_i'\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i + 2\gamma_i'A_{22}'\frac{1}{T}\sum_{t=1}^{T}g_te_{it} \\
&+ \phi_i'\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^{-1})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i + 2\gamma_i'A_{12}'\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2\gamma_i'\hat{Q}_N^{-1}\gamma_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2} \\
&- 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i + 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i \\
&+ 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{11}'\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{21}'\gamma_i \\
&+ 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i \\
&+ 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i - 2\gamma_i'\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})].
\end{aligned}
$$

Given the above result, we can rewrite $\hat{\sigma}_i^2-\sigma_i^2$ as

$$
\hat{\sigma}_i^2-\sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2) - (\hat{\gamma}_i-\gamma_i)'(\hat{\gamma}_i-\gamma_i) + \mathcal{J}_{i,\sigma^2}^*,
$$

where

$$
\begin{aligned}
\mathcal{J}_{i,\sigma^2}^* &= m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i - 2m_i'(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\frac{1}{T}\sum_{t=1}^{T}f_te_{it} \\
&+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i
\end{aligned}
$$

$$
-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\mathcal{J}_{i,\Gamma}+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)(\hat{\gamma}_i-\gamma_i)
$$

$$
+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'m_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})]
$$

$$
-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1(\hat{\Lambda}-\Lambda)'m_i-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i
$$

$$
+\phi_i'A'A\phi_i-2\phi_i'A'\frac{1}{T}\sum_{t=1}^{T}h_te_t'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i-\phi_i'\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}+2\gamma_i'A_{22}'\frac{1}{T}\sum_{t=1}^{T}g_te_{it}
$$

$$
+\phi_i'\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^{-1})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i+2\gamma_i'A_{12}'\frac{1}{T}\sum_{t=1}^{T}f_te_{it}+2\gamma_i'\hat{Q}_N^{-1}\gamma_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}
$$

$$
-2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i+2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i
$$

$$
+2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda'm_i-2m_i'\Lambda\hat{\mathcal{G}}_1A_{11}'\Lambda'm_i-2m_i'\Lambda\hat{\mathcal{G}}_1A_{21}'\gamma_i
$$

$$
+2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i-2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i
$$

$$
+2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i-2\gamma_i'\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})].
$$

Given the expression of $\mathcal{J}_{i,\sigma^2}^*$, one can show that

$$
\frac{1}{N}\sum_{i=1}^{N}\frac{\phi_i\phi_i'}{\sigma_i^4}\mathcal{J}_{i,\sigma_i^2}^*=O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}T}\right)+O_p\left(\frac{1}{T^{3/2}}\right).
$$

Given this result, we have

$$
\frac{1}{N}\sum_{i=1}^{N}\frac{\phi_i\phi_i'}{\sigma_i^4}(\hat{\sigma}_i^2-\sigma_i^2)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\phi_i\phi_i'}{\sigma_i^4}(e_{it}^2-\sigma_i^2)-\frac{1}{T}r_1\mathcal{H}+O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}T}\right)+O_p\left(\frac{1}{T^{3/2}}\right).
$$

Let $E_2 = [0_{r_2 \times r_1}, I_{r_2}]'$. We introduce the following notation for ease of exposition:

$$\zeta^\star = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\phi_i\phi_i'}{\sigma_i^4}(e_{it}^2 - \sigma_i^2),$$
$$\mu^\star = \frac{1}{T}r_1\mathcal{H} + \frac{1}{NT}\sum_{i=1}^{N}\frac{\phi_i\phi_i'}{\sigma_i^6}(\kappa_{i,4} - \sigma_i^4) - \frac{1}{T}E_2E_2'.$$

Using similar arguments as in the derivation of (C.2.22), one can show that

$$\mathcal{D}[(\mathcal{H}_N \otimes I_r) + (I_r \otimes K_r)K_r]\text{vec}(A) = \mathcal{D}\text{vec}(\zeta^\star - \mu^\star) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right).$$

Let $\mathbb{D}_1, \mathbb{D}_2$ and $\mathbb{D}_3$ be defined the same as in the main text. Similar to (C.2.24), we have

$$\mathbb{D}_1\text{vec}(A) = \mathbb{D}_2\text{vec}(\eta^\star) + \mathbb{D}_3\text{vec}(\zeta^\star) - \mathbb{D}_3\text{vec}(\mu^\star) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right).$$

Also notice that

$$\begin{aligned}
\text{vec}(\eta^\star) &= \text{vec}\Big[\frac{1}{T}\sum_{t=1}^{T}h_t e_t'\Sigma_{ee}^{-1}\Phi\mathcal{H}_N^{-1}\Big] = \text{vec}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}h_t\phi_i' e_{it}\mathcal{H}^{-1}\Big], \\
&= (\mathcal{H}^{-1} \otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\phi_i \otimes h_t)e_{it} \\
&= (\mathcal{H}^{-1} \otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(E_1\Lambda' m_i + E_2\gamma_i) \otimes (E_1 f_t + E_2 g_t)e_{it} \\
&= [(\mathcal{H}^{-1}E_1\Lambda') \otimes E_1]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it} \\
&\quad + [(\mathcal{H}^{-1}E_1\Lambda') \otimes E_2]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes g_t)e_{it} \\
&\quad + [(\mathcal{H}^{-1}E_2) \otimes E_1]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i \otimes f_t)e_{it}
\end{aligned}$$

$$+ [(\mathcal{H}^{-1}E_2) \otimes E_2]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i \otimes g_t)e_{it},$$

$$\text{vec}(\zeta^\star) = \text{vec}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\phi_i\phi_i'}{\sigma_i^4}(e_{it}^2 - \sigma_i^2)\Big] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}(\phi_i \otimes \phi_i)(e_{it}^2 - \sigma_i^2),$$

$$\begin{aligned}\text{vec}(\mu^\star) &= \text{vec}\Big[\frac{1}{T}r_1\mathcal{H} + \frac{1}{NT}\sum_{i=1}^{N}\frac{\phi_i\phi_i'}{\sigma_i^6}(\kappa_{i,4} - \sigma_i^4) - \frac{1}{T}E_2E_2'\Big] \\ &= \frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\sigma_i^6}(\phi_i \otimes \phi_i)(\kappa_{i,4} - \sigma_i^2) + \frac{1}{T}\text{vec}\Big[r_1\mathcal{H} - E_2E_2'\Big].\end{aligned}$$

Given the above result, we have

$$\begin{aligned}\text{vec}(A) = \mathbb{D}_1^{-1}\mathbb{D}_2&[(\mathcal{H}_N^{-1}E_1\Lambda') \otimes E_1]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it} \\ &+ \mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}_N^{-1}E_1\Lambda') \otimes E_2]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes g_t)e_{it} \\ &+ \mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}_N^{-1}E_2) \otimes E_1]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i \otimes f_t)e_{it} \\ &+ \mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}_N^{-1}E_2) \otimes E_2]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i \otimes g_t)e_{it} \\ &+ \mathbb{D}_1^{-1}\mathbb{D}_3\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}(\phi_i \otimes \phi_i)(e_{it}^2 - \sigma_i^2) \\ &- \mathbb{D}_1\mathbb{D}_3\Big\{\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\sigma_i^6}(\phi_i \otimes \phi_i)(\kappa_{i,4} - \sigma_i^2) + \frac{1}{T}\text{vec}\Big[r_1\mathcal{H}_N - E_2E_2'\Big]\Big\} \\ &+ O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right).\end{aligned} \tag{C.4.24}$$

Now we define

$$\mathbb{B}_1^\star = R^{-1} \otimes I_{r_1} + K_{kr_1}[(P^{-1}\Lambda') \otimes \Lambda] - K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}_N^{-1}E_1\Lambda') \otimes E_1],$$

$$\mathbb{B}_2^\star = K_{kr_1}[P^{-1} \otimes \psi] - K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}_N^{-1}E_1) \otimes E_2],$$

$$\mathbb{B}_3^\star = -K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}_N^{-1}E_2) \otimes E_1],$$

$$\mathbb{B}_4^\star = -K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}_N^{-1}E_2) \otimes E_2],$$

$$\mathbb{B}_5^\star = -K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_3,$$

$$\Delta^\star = K_{kr_1}(E_1' \otimes \Psi)\mathbb{D}_1^{-1}\mathbb{D}_3\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^6}(\phi_i \otimes \phi_i)(\kappa_{i,4} - \sigma_i^4) + \text{vec}(r_1\mathcal{H}_N - E_2E_2')\right].$$

Substituting (C.4.24) into (C.4.22), we can rewrite (C.4.22) in terms of $\mathbb{B}_i^\star$ as

$$\begin{aligned}
\text{vec}(\hat{\Lambda}' - \Lambda') = {} & \mathbb{B}_1^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it} + \mathbb{B}_2^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\Lambda' m_i \otimes g_t)e_{it} \\
& + \mathbb{B}_3^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i \otimes f_t)e_{it} + \mathbb{B}_4^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i \otimes g_t)e_{it} \\
& + \mathbb{B}_5^\star \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}(\phi_i \otimes \phi_i)(e_{it}^2 - \sigma_i^2) + \frac{1}{T}\Delta^\star \\
& + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right).
\end{aligned}$$

Given the above result, by a Central Limit Theorem, we have

$$\sqrt{NT}\left[\text{vec}(\hat{\Lambda}' - \Lambda') - \frac{1}{T}\Delta^\star\right] \xrightarrow{d} N(0, \Omega^\star),$$

where $\Omega^\star = \lim_{N\to\infty}\Omega_N^\star$ with

$$\begin{aligned}
\Omega_N^\star = {} & \mathbb{B}_1^\star(R \otimes I_{r_1})\mathbb{B}_1^{\star\prime} + \mathbb{B}_2^\star(P \otimes I_{r_1})\mathbb{B}_2^{\star\prime} + \mathbb{B}_3^\star(Q \otimes I_{r_1})\mathbb{B}_3^{\star\prime} + \mathbb{B}_4^\star(Q \otimes I_{r_2})\mathbb{B}_4^{\star\prime} \\
& + \mathbb{B}_1^\star(S \otimes I_{r_1})\mathbb{B}_3^{\star\prime} + \mathbb{B}_3^\star(S' \otimes I_{r_1})\mathbb{B}_1^{\star\prime} + \mathbb{B}_5^\star\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^8}(\phi_i\phi_i') \otimes (\phi_i\phi_i')(\kappa_{i,4} - \sigma_i^4)\right]\mathbb{B}_5^{\star\prime}.
\end{aligned}$$

Table C.51: Simulation results under $k = 3$, $r = 1$, and $\epsilon_{it} \sim t_5$

$\Lambda_{3\times1}$		MLE			PC		
N	T	MAD	RMSE	RAvar	MAD	RMSE	RAvar
30	30	0.0451	0.0717	2.2151	0.1016	0.1499	N/A
50	30	0.0328	0.0523	2.1456	0.0682	0.0997	N/A
100	30	0.0229	0.0346	1.8912	0.0465	0.0676	N/A
150	30	0.0198	0.0293	2.0935	0.0384	0.0547	N/A
30	50	0.0319	0.0495	1.9587	0.0781	0.1114	N/A
50	50	0.0227	0.0365	2.0295	0.0558	0.0804	N/A
100	50	0.0166	0.0262	1.8357	0.0367	0.0522	N/A
150	50	0.0142	0.0220	1.9402	0.0302	0.0426	N/A
30	100	0.0227	0.0371	1.8139	0.0679	0.0965	N/A
50	100	0.0154	0.0251	1.9126	0.0448	0.0642	N/A
100	100	0.0111	0.0179	1.7941	0.0280	0.0394	N/A
150	100	0.0094	0.0151	1.7799	0.0221	0.0313	N/A

C.5 More simulation results

In this appendix, we provide additional simulation results when errors have t-distribution and χ^2-distribution. The results are given in Tables C.51-C.54.

C.6 More comparison of W and LR tests

In this appendix, we make a comparison on the proposed W test and the traditional LR test. The LR test is advocated in Tsai and Tsay (2010). Following Bartlett (1950) and Anderson (2003), Tsai and Tsay consider a modified version of the LR statistic to improve the finite sample performance. The modified LR statistic is defined as

$$LR = \Big(T - \frac{2N+11}{6} - \frac{2r}{3}\Big)\Big(\ln|\hat{\Sigma}_c| - \ln|\hat{\Sigma}_u|\Big),$$

Table C.52: Simulation results under $k = 8$, $r = 3$, and $\epsilon_{it} \sim t_5$

$\Lambda_{3\times 1}$		MLE			PC		
N	T	MAD	RMSE	RAvar	MAD	RMSE	RAvar
30	30	0.3478	0.4961	15.1723	0.5800	0.8257	N/A
50	30	0.2379	0.3498	13.1208	0.3959	0.5677	N/A
100	30	0.1461	0.2217	12.3297	0.2236	0.3244	N/A
150	30	0.1156	0.1751	11.8396	0.1661	0.2415	N/A
30	50	0.2584	0.3742	14.6463	0.5165	0.7541	N/A
50	50	0.1727	0.2530	13.2355	0.3226	0.4753	N/A
100	50	0.1154	0.1826	13.1610	0.1816	0.2686	N/A
150	50	0.0930	0.1429	11.5573	0.1402	0.2069	N/A
30	100	0.1880	0.2761	15.5842	0.4626	0.7075	N/A
50	100	0.1249	0.1928	12.8791	0.2734	0.4208	N/A
100	100	0.0812	0.1321	12.3295	0.1410	0.2144	N/A
150	100	0.0639	0.1025	14.4627	0.1065	0.1592	N/A

Table C.53: Simulation results under $k = 3$, $r = 1$, and $\epsilon_{it} \sim \chi^2(2)$

$\Lambda_{3\times 1}$		MLE			PC		
N	T	MAD	RMSE	RAvar	MAD	RMSE	RAvar
30	30	0.0409	0.0649	2.0501	0.0941	0.1394	N/A
50	30	0.0319	0.0497	1.9461	0.0707	0.1011	N/A
100	30	0.0225	0.0351	1.9543	0.0459	0.0654	N/A
150	30	0.0207	0.0320	2.1578	0.0388	0.0553	N/A
30	50	0.0335	0.0541	1.8213	0.0841	0.1216	N/A
50	50	0.0229	0.0362	1.8956	0.0569	0.0826	N/A
100	50	0.0172	0.0281	1.9791	0.0371	0.0526	N/A
150	50	0.0135	0.0208	1.9470	0.0285	0.0401	N/A
30	100	0.0220	0.0362	1.9443	0.0673	0.0959	N/A
50	100	0.0165	0.0274	1.8368	0.0456	0.0647	N/A
100	100	0.0109	0.0175	1.7312	0.0281	0.0397	N/A
150	100	0.0088	0.0141	1.7539	0.0219	0.0311	N/A

Table C.54: Simulation results under $k = 8$, $r = 3$, and $\epsilon_{it} \sim \chi^2(2)$

$\Lambda_{3\times1}$		MLE			PC		
N	T	MAD	RMSE	RAvar	MAD	RMSE	RAvar
30	30	0.3446	0.4909	15.2244	0.5657	0.8061	N/A
50	30	0.2353	0.3481	13.6764	0.3746	0.5424	N/A
100	30	0.1547	0.2475	12.9084	0.2242	0.3258	N/A
150	30	0.1203	0.1893	13.3989	0.1752	0.2559	N/A
30	50	0.2632	0.3831	15.0428	0.5189	0.7618	N/A
50	50	0.1795	0.2697	13.7256	0.3214	0.4769	N/A
100	50	0.1160	0.1803	12.4406	0.1813	0.2632	N/A
150	50	0.0959	0.1656	13.1984	0.1417	0.2096	N/A
30	100	0.1839	0.2687	14.8799	0.4666	0.7114	N/A
50	100	0.1271	0.1945	15.0769	0.2718	0.4124	N/A
100	100	0.0854	0.1452	13.9679	0.1439	0.2214	N/A
150	100	0.0676	0.1151	14.4559	0.1045	0.1617	N/A

where $\hat{\Sigma}_c = M\hat{\Lambda}\hat{\Lambda}'M + \hat{\Sigma}_{ee}$ is the estimated variance for the constrained model and $\hat{\Sigma}_u = \hat{L}\hat{L}' + \tilde{\Sigma}_{ee}$ the estimated variance for the unconstrained one. Here $\hat{\Lambda}$ and $\hat{\Sigma}_{ee}$ are the MLEs for the constrained model and $\hat{L}$ and $\tilde{\Sigma}_{ee}$ the MLEs for the unconstrained one. We run simulations based on the same data generating processes as in Section 3.8.2. The empirical sizes and powers of the modified LR statistic are given in Tables and below.

Table presents the empirical sizes in all combinations of N and T. We are surprised to find that the modified LR statistic has severe size distortions in all the sample sizes. In some cases, the LR test over-accepts the null hypothesis with empirical sizes deceasing to zero. In other cases, the LR test over-rejects the null hypothesis with empirical sizes larger than 50%. As far as we see, the poor performance of the LR test is not related with the adjusted factor $T - (2N + 11)/6 - 2r/3$ since we also consider the unmodified LR statistic and the results are not good either.

Table presents the empirical powers of the modified LR test. We see that the LR test

Table C.65: The empirical size of the LR test with $(k, r) = (3, 1)$ under normal errors

Empirical size of LR										
$\epsilon_{it} \sim$		$N(0,1)$			t_5			$\chi^2(2)$		
N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%
30	30	0.3%	10.5%	27.4%	1.3%	11.0%	28.6%	0.9%	10.0%	26.7%
50	30	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
100	30	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
150	30	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
30	50	23.7%	72.4%	90.6%	25.0%	70.3%	88.4%	25.0%	72.4%	90.0%
50	50	5.0%	27.8%	55.1%	4.3%	29.3%	55.8%	4.5%	30.8%	56.7%
100	50	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.1%
150	50	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
30	100	64.4%	95.3%	99.6%	67.7%	96.1%	99.8%	69.2%	96.7%	99.6%
50	100	77.3%	98.4%	99.7%	78.7%	98.5%	99.9%	80.4%	98.2%	99.6%
100	100	29.4%	74.4%	91.1%	27.6%	77.9%	92.7%	28.5%	75.0%	91.0%
150	100	0.1%	0.1%	0.3%	0.0%	0.0%	0.3%	0.1%	0.1%	0.1%
30	150	79.3%	98.2%	99.9%	79.3%	98.7%	99.8%	78.5%	98.5%	100.0%
50	150	95.7%	99.9%	100.0%	95.0%	99.7%	100.0%	93.8%	99.6%	100.0%
100	150	96.3%	100.0%	100.0%	95.8%	100.0%	100.0%	96.5%	100.0%	100.0%
150	150	65.1%	95.2%	98.5%	65.2%	93.6%	98.3%	65.2%	95.0%	98.9%
100	100	29.4%	74.4%	91.1%	27.6%	77.9%	92.7%	28.5%	75.0%	91.0%
200	100	0.1%	0.1%	0.1%	0.1%	0.1%	0.1%	0.2%	0.2%	0.2%
300	100	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
100	200	100.0%	100.0%	100.0%	99.6%	99.9%	99.9%	99.8%	100.0%	100.0%
200	200	81.5%	93.4%	93.5%	82.7%	94.2%	94.8%	83.2%	94.3%	94.7%
300	200	0.3%	0.3%	0.4%	0.1%	0.2%	0.5%	0.3%	0.3%	0.4%
100	300	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
200	300	94.7%	94.7%	94.7%	94.3%	94.3%	94.3%	95.0%	95.0%	95.0%
300	300	74.0%	74.8%	74.8%	76.6%	76.8%	76.9%	74.0%	74.3%	74.4%
100	500	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
200	500	93.4%	93.4%	93.4%	94.7%	94.7%	94.7%	93.8%	93.8%	93.8%
300	500	77.4%	77.4%	77.4%	75.0%	75.0%	75.0%	77.0%	77.0%	77.0%

does not have stable powers. If N is comparable to or smaller than T, the LR test would have good powers. However, if $N \gg T$, say $N = 150, T = 30$, the power decreases to zero. This is in contrast with the proposed W test, which has stable powers in all combinations of N and T.

From Tables and , we conclude that the proposed W test dominates the LR test in terms of empirical size and power. Moreover, we also consider the normalized LR test which is W_{LRN} as mentioned in Remark 3.5.3. The performance of W_{LRN} is similar to W_{LR} and hence not the simulation results are not presented here. In conclusion, our proposed W test outperforms both the W_{LR} and W_{LRN}.

C.7 Proofs of the theoretical results in Section 3.9

In this appendix, we define the following notation:

$$\hat{\mathbb{P}} = \frac{1}{N}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda}; \qquad \hat{\mathbb{R}} = \frac{1}{N} M' \hat{\mathbb{W}}^{-1} M; \qquad \hat{\mathbb{G}} = (I_r + \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda})^{-1};$$
$$\hat{\mathbb{P}}_N = N \cdot \hat{\mathbb{P}} = \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda}; \quad \hat{\mathbb{R}}_N = N \cdot \hat{\mathbb{R}} = M' \hat{\mathbb{W}}^{-1} M, \quad \hat{\mathbb{G}}_N = N \cdot \hat{\mathbb{G}}.$$

Then we have $\hat{\mathbb{P}}_N^{-1} = \hat{\mathbb{G}}(I - \hat{\mathbb{G}})^{-1}$ and

$$\Sigma_{zz}^{-1} = \mathbb{W}^{-1} - \mathbb{W}^{-1} M \Lambda (I_r + \Lambda' M' \mathbb{W}^{-1} M \Lambda)^{-1} \Lambda' M' \mathbb{W}^{-1}, \tag{C.7.1}$$

Table C.66: The empirical power of the LR test with $(k, r) = (3, 1)$ under normal errors

Empirical power of LR													
α		0.2			0.5			2			5		
N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
30	30	16.9%	35.4%	54.0%	44.4%	60.8%	73.5%	89.0%	93.6%	96.6%	99.6%	100.0%	100.0%
50	30	6.0%	9.5%	11.2%	25.3%	31.4%	34.9%	71.9%	76.2%	78.6%	97.5%	98.5%	98.7%
100	30	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
150	30	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
30	50	54.8%	84.8%	95.9%	72.6%	91.3%	97.3%	96.2%	99.5%	99.9%	99.9%	100.0%	100.0%
50	50	33.3%	60.0%	77.7%	61.5%	78.2%	87.5%	95.6%	98.4%	99.4%	99.9%	100.0%	100.0%
100	50	6.4%	7.4%	8.3%	26.3%	31.6%	33.9%	68.2%	70.5%	72.7%	94.3%	95.3%	96.1%
150	50	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
30	100	79.3%	97.4%	99.6%	90.9%	99.4%	99.7%	99.2%	100.0%	100.0%	100.0%	100.0%	100.0%
50	100	91.0%	99.2%	99.9%	95.6%	99.8%	100.0%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%
100	100	66.4%	92.2%	98.1%	83.0%	95.8%	99.1%	99.0%	99.9%	99.9%	100.0%	100.0%	100.0%
150	100	28.9%	36.1%	41.1%	57.1%	61.4%	63.5%	85.6%	89.1%	92.4%	99.8%	99.9%	100.0%
30	150	88.4%	99.5%	100.0%	94.9%	99.8%	100.0%	99.7%	100.0%	100.0%	100.0%	100.0%	100.0%
50	150	97.7%	99.8%	100.0%	99.2%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
100	150	99.0%	100.0%	100.0%	99.3%	99.9%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
150	150	85.7%	97.9%	99.0%	92.1%	98.3%	98.8%	99.1%	99.3%	99.3%	100.0%	100.0%	100.0%
100	100	69.3%	90.4%	97.6%	84.2%	96.0%	98.9%	98.2%	99.9%	100.0%	100.0%	100.0%	100.0%
200	100	8.2%	10.6%	11.4%	34.6%	38.0%	40.1%	70.9%	72.8%	73.5%	93.9%	95.0%	95.2%
300	100	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
100	200	99.9%	100.0%	100.0%	99.9%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
200	200	90.2%	93.9%	94.1%	92.9%	94.3%	94.3%	95.8%	95.9%	95.9%	98.2%	98.2%	98.2%
300	200	19.5%	23.8%	26.6%	37.0%	39.9%	42.5%	66.7%	70.6%	72.4%	82.0%	82.2%	82.2%
100	300	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
200	300	93.6%	93.6%	93.6%	93.8%	93.8%	93.8%	95.1%	95.1%	95.1%	97.4%	97.4%	97.4%
300	300	75.7%	75.8%	75.8%	76.0%	76.1%	76.1%	77.3%	77.3%	77.3%	85.3%	85.3%	85.3%
100	500	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
200	500	93.1%	93.1%	93.1%	94.9%	94.9%	94.9%	94.8%	94.8%	94.8%	96.8%	96.8%	96.8%
300	500	79.7%	79.7%	79.7%	75.6%	75.6%	75.6%	80.9%	80.9%	80.9%	79.9%	79.9%	79.9%

and

$$\hat{\Lambda}' M' \hat{\Sigma}_{zz}^{-1} = \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} - \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda} (I_r + \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda})^{-1} \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} = \hat{\mathbb{G}} \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}. \tag{C.7.2}$$

The following lemma is a direct result of Assumptions A and B″, which will be used throughout the whole proof.

Lemma C.7.1. *From assumptions of A and B″, we have*

$$(a) \quad E\left(\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t e_{it}\Big\|^2\right) \leq C, \quad \textit{for all } i;$$

$$(b) \quad E\left(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T} f_t e_{it}\Big\|^2\right) \leq C;$$

$$(c) \quad E\left(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(e_{it}^2 - w_i^2)\Big|^2\right) \leq C.$$

Further, we have

$$(d) \quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}\Big\|^2 = O_p(T^{-1});$$

$$(e) \quad \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2)\Big)^2 = O_p(T^{-1});$$

$$(f) \quad \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}\big[e_{it}e_{jt} - E(e_{it}e_{jt})\big]\Big)^2 = O_p(T^{-1});$$

C.7.1 Proof of the consistency of the MLE in Section 3.9

Similar to Appendix A, we use symbols with superscript "*" to denote the true parameters and variables without superscript "*" denote the arguments of the likelihood function in this section. Let $\theta = (\Lambda, w_1^2, \cdots, w_N^2)$ and let Θ be a parameter set such that Λ take values in a compact set and $C^{-2} \leq w_i^2 \leq C^2$ for all $i = 1, ..., N$. We assume $\theta^* = (\Lambda^*, w_1^{*2}, \cdots, w_N^{*2})$ is an interior point of Θ. For simplicity, we write $\theta = (\Lambda, \mathbb{W})$ and $\theta^* = (\Lambda^*, \mathbb{W}^*)$.

The following lemmas are useful to prove the following Proposition C.7.1, and Proposition C.7.1 will be used in the proofs in the following Appendix C.7.2.

Lemma C.7.2. *Under assumptions of A, B″, C″ and D″, we have*

$$(a) \quad \sup_{\theta\in\Theta} \frac{1}{NT} \left| \text{tr}\Big[\Lambda^{*\prime} M' \Sigma_{zz}^{-1} \sum_{t=1}^{T} e_t f_t^{*\prime}\Big] \right| \xrightarrow{p} 0;$$

$$(b) \quad \sup_{\theta\in\Theta} \frac{1}{NT} \left| \text{tr}\Big[\sum_{t=1}^{T} (e_t e_t' - \mathbb{O}^*) \Sigma_{zz}^{-1}\Big] \right| \xrightarrow{p} 0;$$

$$(c) \quad \sup_{\theta\in\Theta} \frac{1}{N} \left| \text{tr}\Big[(\mathbb{O}^* - \mathbb{W}^*) \Sigma_{zz}^{-1}\Big] \right| \xrightarrow{p} 0;$$

where $\theta^* = (\Lambda^*, \mathbb{W}^*)$ denotes the true parameters and $\Sigma_{zz} = M\Lambda\Lambda' M' + \mathbb{W}$.

Results (a) and (b) in Lemma C.7.2 can be proved in the same way as in Lemma C.1.1, and proof of C.7.2(c) is similar to that of Lemma S.3(b) in Bai and Li (2016). Details are therefore omitted.

Lemma C.7.3. *Under assumptions of A, B″, C″ and D″, we have*

$$(a) \quad \left\| \frac{1}{N} \Lambda^{*\prime} M' (\hat{\mathbb{W}}^{-1} - \mathbb{W}^{*-1}) M \Lambda^* \right\| = O_p\bigg(\Big[\frac{1}{N}\sum_{i=1}^{N} (\hat{w}_i^2 - w_i^{*2})^2\Big]^{\frac{1}{2}}\bigg);$$

$$(b) \quad \left\| \frac{1}{N} M' (\hat{\mathbb{W}}^{-1} - \mathbb{W}^{*-1}) M \right\| = O_p\bigg(\Big[\frac{1}{N}\sum_{i=1}^{N} (\hat{w}_i^2 - w_i^{*2})^2\Big]^{\frac{1}{2}}\bigg).$$

Given the above results, if $N^{-1}\sum_{i=1}^{N} (\hat{w}_i^2 - w_i^{*2})^2 = o_p(1)$, we have

$$(c) \quad \hat{\mathbb{R}}_N = O_p(N), \quad \hat{\mathbb{R}} = \frac{1}{N}\hat{\mathbb{R}}_N = O_p(1);$$

(d) $\|\hat{\mathbb{R}}^{-1/2}\| = O_p(1)$.

where $\hat{\mathbb{R}}$ and $\hat{\mathbb{R}}_N$ are defined in the beginning of Appendix G.

The proof of this lemma is similar to that of Lemma C.1.2 and hence omitted here.

Lemma C.7.4. *Under assumptions of A, B$''$, C$''$ and D$''$, we have*

$$(a)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t' - \mathbb{O})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = \|\hat{\mathbb{P}}^{-1/2}\|^2 \cdot O_p(T^{-1/2});$$

$$(b)\quad \frac{1}{N}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t' = \|\hat{\mathbb{P}}^{-1/2}\| \cdot O_p(T^{-1/2});$$

$$(c)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = \|\hat{\mathbb{P}}_N^{-1}\| \cdot O_p(1);$$

$$(d)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = \|\hat{\mathbb{P}}^{-1/2}\|^2 \cdot O_p(N^{-1/2});$$

$$(e)\quad \frac{1}{NT}\sum_{t=1}^{T}f_te_t'\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(T^{-1/2});$$

$$(f)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_t' - \mathbb{O}]\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = \|\hat{\mathbb{P}}^{-1/2}\| \cdot O_p(T^{-1/2});$$

$$(g)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = \|\hat{\mathbb{P}}^{-1/2}\| \cdot O_p\Big(\Big[\frac{1}{N^3}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\Big]^{\frac{1}{2}}\Big);$$

$$(h)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = \|\hat{\mathbb{P}}^{-1/2}\| \cdot O_p(N^{-1}).$$

Proof of Lemma C.7.4. Proofs for (a)-(c) and (e)-(g) are similar to those for Lemma C.1.3, so we only include the proofs for (d) and (h) which are different from Lemma C.1.3. Consider (d). The left hand side can be rewritten as

$$\frac{1}{N}\mathbb{P}^{-1/2}\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathbb{P}}_N^{-1/2}\frac{1}{\hat{w}_i^2}\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\Big[\mathbb{O}_{ij} - 1(i=j)w_i^2\Big]\frac{1}{\hat{w}_j^2}\sum_{l=1}^{k}\hat{\lambda}_l'm_{jl}\hat{\mathbb{P}}_N^{-1/2}\Big]\hat{\mathbb{P}}^{-1/2},$$

where $1(i = j)$ is the indicator function, equals 1 if $i = j$ and 0 otherwise. The above expression is bounded in norm by

$$C\frac{1}{\sqrt{N}}\|\hat{\mathbb{P}}^{-1/2}\|^2\Big(\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2}\Big\|\hat{\mathbb{P}}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}(\mathbb{O}_{ij})^2\Big)^{1/2},$$

which is $\|\hat{\mathbb{P}}^{-1/2}\|^2 \cdot O_p(N^{-1/2})$ by the fact that $\left(\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2}\Big\|\hat{\mathbb{P}}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|^2\right) = r$ and $\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}(\mathbb{O}_{ij})^2\right)$ is $O_p(1)$ from Assumption B$''$. So result (d) follows.

Next consider (h). Similarly, the left hand side can be rewritten as

$$\frac{1}{N^{3/2}}\hat{\mathbb{P}}^{-1/2}\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\mathbb{P}}_N^{-1/2}\frac{1}{\hat{w}_i^2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big[\mathbb{O}_{ij} - 1(i=j)w_i^2\Big]\frac{1}{\hat{w}_j^2}m_j'\Big]\hat{\mathbb{R}}^{-1},$$

which is bounded in norm by

$$C\frac{1}{N}\|\hat{\mathbb{P}}^{-1/2}\|\|\hat{\mathbb{R}}^{-1}\|\Big(\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2}\Big\|\hat{\mathbb{P}}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\sum_{j=1,j\neq i}^{N}\mathbb{O}_{ij}m_j\Big\|^2\Big)^{1/2},$$

which is $\|\hat{\mathbb{P}}^{-1/2}\| \cdot O_p(N^{-1})$ by $\hat{\mathbb{R}}^{-1} = O_p(1)$ from Lemma C.7.3(c) and $\Big\|\sum_{j=1,j\neq i}^{N}\mathbb{O}_{ij}m_j\Big\| = O_p(1)$ from Assumption B$''$. Hence we have result (h). $\square$

Proposition C.7.1 (Consistency)**.** *Let $\hat{\theta} = (\hat{\Lambda}, \hat{\mathbb{W}})$ be the MLE that maximizes (3.3.2). Then under Assumptions A,B$''$, C$''$ and D$''$, together with IC$''$, when $N, T \to \infty$, we have*

$$\hat{\Lambda} - \Lambda \xrightarrow{p} 0; \qquad \frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2 \xrightarrow{p} 0.$$

Proof of Proposition C.7.1. Similar to the proof of Proposition 3.4.1, we consider the following centered objective function

$$L^{\dagger}(\theta) = \overline{L}^{\dagger}(\theta) + R^{\dagger}(\theta),$$

where

$$\overline{L}^{\dagger}(\theta) = -\frac{1}{N}\ln|\Sigma_{zz}| - \frac{1}{N}\mathrm{tr}\Big(\Sigma_{zz}^{*}\Sigma_{zz}^{-1}\Big) + 1 + \frac{1}{N}\ln|\Sigma_{zz}^{*}|$$

and

$$R^{\dagger}(\theta) = -\frac{1}{N}\mathrm{tr}\Big[(M_{zz} - \Sigma_{zz}^{*})\Sigma_{zz}^{-1}\Big],$$

where $\Sigma_{zz} = M\Lambda\Lambda' M' + \mathbb{W}$ and $\Sigma_{zz}^{*} = M\Lambda^{*}\Lambda^{*\prime}M' + \mathbb{W}^{*}$. By the definition of M_{zz}, we have

$$R^{\dagger}(\theta) = -2\frac{1}{NT}\mathrm{tr}\Big[M\Lambda^{*}\sum_{t=1}^{T} f_t^{*}e_t'\Sigma_{zz}^{-1}\Big] - \frac{1}{NT}\mathrm{tr}\Big[\sum_{t=1}^{T}(e_te_t' - \mathbb{O}^{*})\Sigma_{zz}^{-1}\Big] - \frac{1}{N}\mathrm{tr}\Big[(\mathbb{O}^{*} - \mathbb{W}^{*})\Sigma_{zz}^{-1}\Big].$$

By Lemma C.7.2, we have $\sup_{\theta}|R^{\dagger}(\theta)| = o_p(1)$. Then using the same approach as in the proof of Proposition 3.4.1, we get $\overline{L}^{\dagger}(\hat{\theta}) \geq -2|o_p(1)|$, which implies

$$\frac{1}{N}\ln|\hat{\mathbb{W}}| - \frac{1}{N}\ln|\mathbb{W}^{*}| + \frac{1}{N}\mathrm{tr}[\mathbb{W}^{*}\hat{\mathbb{W}}^{-1}] - 1 \xrightarrow{p} 0, \tag{C.7.3}$$

$$\frac{1}{N}\mathrm{tr}[M\Lambda^{*}\Lambda^{*\prime}M'\hat{\Sigma}_{zz}^{-1}] \xrightarrow{p} 0. \tag{C.7.4}$$

The above arguments further imply

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^{*2})^2 \xrightarrow{p} 0. \tag{C.7.5}$$

which is the second result of Proposition C.7.1, and other results as following:

$$\hat{\mathbb{G}} = o_p(1); \qquad \hat{\mathbb{P}}_N^{-1} = o_p(1); \tag{C.7.6}$$

$$\frac{1}{N}\Lambda^{*\prime}M'\mathbb{W}^{*-1}M\Lambda^* - (I_r - \mathbb{A})\frac{1}{N}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}(I_r - \mathbb{A})' \xrightarrow{p} 0, \tag{C.7.7}$$

$$\frac{1}{N}(\hat{\Lambda} - \Lambda^*)'M'\hat{\mathbb{W}}^{-1}M(\hat{\Lambda} - \Lambda^*) - \mathbb{A}\Big(\frac{1}{N}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\Big)\mathbb{A}' \xrightarrow{p} 0. \tag{C.7.8}$$

where $\mathbb{A} \equiv (\hat{\Lambda} - \Lambda^*)'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}$.

We now consider the first-order condition for $\hat{\Lambda}$. Post multiplying (3.3.3) by $\hat{\Lambda}$ implies

$$\hat{\Lambda}'M'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}M\hat{\Lambda} = 0.$$

By (C.7.2), we can simplify the above equation as

$$\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\mathbb{W}}^{-1}M\hat{\Lambda} = 0,$$

which can be further rewritten as

$$\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} = -\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}$$

$$+\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\Lambda^*\Lambda^{*\prime}M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} + \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\Lambda^*\frac{1}{T}\sum_{t=1}^{T} f_t^* e_t'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}$$

$$+\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}\Lambda^{*\prime}M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} + \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \mathbb{O}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}$$

$$+\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}^* - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}.$$

By the definitions of $\hat{\mathbb{P}}$ and $\mathbb{A}$, we have

$$
\begin{aligned}
I_r &= (I_r - \mathbb{A})'(I_r - \mathbb{A}) + \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \mathbb{O}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} \\
&+ (I_r - \mathbb{A})'\frac{1}{NT}\sum_{t=1}^{T} f_t^* e_t'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} + \frac{1}{N}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}(I_r - \mathbb{A}) \qquad \text{(C.7.9)} \\
&- \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} + \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}^* - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} \\
&= i_1 + i_2 + \cdots + i_6, \quad \text{say}
\end{aligned}
$$

Compared to (C.1.16), there exists an extra term i_6 in the above equation, due to the weak dependence structure of the error. Based on (C.7.9) and (C.7.8), together with Lemma C.7.4, we can show that $\mathbb{A} = O_p(1)$ and $\|\hat{\mathbb{P}}^{-1}\| = O_p(1)$. Furthermore, applying Lemma A.1 of the supplement of Bai and Li (2012) and using the identification condition IC2$''$, we can prove that $\mathbb{A} = o_p(1)$.

Again, we consider the first-order condition (3.3.3), which can be simplified as (by (C.7.2))

$$
\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\mathbb{W}}^{-1}M = 0.
$$

By the definition of M_{zz}, the above equation can be rewritten as

$$
\begin{aligned}
\hat{\Lambda}' - \Lambda^{*\prime} &= -\mathbb{A}'\Lambda^{*\prime} + (I - \mathbb{A})'\frac{1}{T}\sum_{t=1}^{T} f_t^* e_t'\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}_N^{-1} + \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}\Lambda^{*\prime} \qquad \text{(C.7.10)} \\
&+\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \mathbb{O}^*]\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}_N^{-1} - \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}_N^{-1} \\
&+\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}^* - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}_N^{-1}
\end{aligned}
$$

We need to show all the six terms on the right hand side of the above equation are $o_p(1)$. From the preceding results that $\mathbb{A} = o_p(1)$ and Lemma C.7.4(e), we know the first two terms are $o_p(1)$. From $\|\hat{\mathbb{P}}^{-1}\| = O_p(1)$ and the results in Lemma C.7.4, we see that the remaining four terms are also $o_p(1)$. Therefore we have $\hat{\Lambda}' - \Lambda^{*\prime} = o_p(1)$, which implies that $\hat{\Lambda} \xrightarrow{p} \Lambda^{*\prime}$. This completes the proof of Proposition C.7.1. □

Corollary C.7.1. *Under Assumptions A, B″, C″ and D″,*

$$(a) \quad \frac{1}{N}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} - \frac{1}{N}\Lambda^{*\prime}M'\mathbb{W}^{*-1}M\Lambda^{*} = o_p(1);$$

$$(b) \quad \hat{\mathbb{P}}_N = O_p(N),\ \hat{\mathbb{P}} = O_p(1),\ \hat{\mathbb{G}} = O_p(N^{-1}),\ \hat{\mathbb{G}}_N = O_p(1);$$

$$(c) \quad \frac{1}{N}(\hat{\Lambda} - \Lambda)'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} = o_p(1).$$

Proof of Corollary C.1.1. Proof for the above corollary is similar to Corollary C.1.1, and therefore omitted here.

C.7.2 Proofs of Theorem 3.9.1, 3.9.2 and 3.9.1

In this appendix, we drop "*" from the symbols of underlying true values for notational simplicity. The following lemmas will be useful in the proofs of Theorems 3.9.1 and 3.9.2.

Lemma C.7.5. *Under Assumptions A, B″, C′ and D′, we have*

$$(a) \quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t' - \mathbb{O})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = O_p(T^{-1/2});$$

$$(b) \quad \frac{1}{N}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t' = O_p(T^{-1/2});$$

$$(c) \quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = \frac{1}{\sqrt{N}}O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\Big]^{\frac{1}{2}}\Big);$$

(d) $\frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = O_p(N^{-1/2})$;

(e) $\frac{1}{NT}\sum_{t=1}^{T} f_t e_t'\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(T^{-1/2})$;

(f) $\frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \mathbb{O}]\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(T^{-1/2})$;

(g) $\frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = \frac{1}{\sqrt{N}}O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\Big]^{\frac{1}{2}}\Big)$;

(h) $\frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(N^{-1})$.

The above lemma is strengthened from Lemma C.7.4, with its proof similar to Lemma C.2.1 and hence omitted here.

Based on (C.7.9) and IC2$''$, together with Lemma C.7.5, we have the following Lemma C.7.6, which corresponds to Lemma C.2.2 with modification.

Lemma C.7.6. *Under Assumptions A, B$''$, C$'$ and D$'$, we have*

$$A \equiv (\hat{\Lambda}-\Lambda)'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1} = O_p(\frac{1}{\sqrt{T}}) + O_p(\frac{1}{N}) + O_p(\|\hat{\Lambda}-\Lambda\|^2) + O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\Big]^{\frac{1}{2}}\Big).$$

Proof of Lemma C.7.6 is similar to Lemma C.2.2 and hence omitted here.

Proof of Theorem 3.4.1. We can rewrite the first order condition of $\hat{\mathbb{W}}$ as

$$\mathrm{diag}\Big\{(M_{zz}-\hat{\Sigma}_{zz}) - (M_{zz}-\hat{\Sigma}_{zz})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M' - M\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(M_{zz}-\hat{\Sigma}_{zz})\Big\} = 0.$$

With

$$M_{zz} = M\Lambda\Lambda' M' + \mathbb{W} + M\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_t' + \frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Lambda' M' + \frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \mathbb{O}) + (\mathbb{O} - \mathbb{W}),$$

we can further rewrite the above first order condition as

$$\begin{aligned}
\hat{w}_i^2 - w_i^2 &= \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2) + 2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\mathbb{G}\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1} M\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} \\
&-2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\mathbb{W}}^{-1} M\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' m_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] \qquad \text{(C.7.11)} \\
&+m_i'(\hat{\Lambda} - \Lambda)(\hat{\Lambda} - \Lambda)' m_i - 2m_i'(\hat{\Lambda} - \Lambda)\hat{\Lambda}' m_i + 2m_i'(\hat{\Lambda} - \Lambda)\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1} M\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' m_i \\
&+2m_i'\Lambda(\hat{\Lambda} - \Lambda)' M'\hat{\mathbb{W}}^{-1} M\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' m_i + 2\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2} m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' m_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})_i.
\end{aligned}$$

where $(\mathbb{O} - \mathbb{W})_i$ denotes the ith column of the $N \times N$ matrix $(\mathbb{O} - \mathbb{W})$. Define

$$\psi_1 = \frac{1}{T}\sum_{t=1}^{T} f_t e_t'\hat{\mathbb{W}}^{-1} M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}; \qquad \varphi_1 = \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \mathbb{O})\hat{\mathbb{W}}^{-1} M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1};$$

$$\varphi_2 = \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W})\hat{\mathbb{W}}^{-1} M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1};$$

$$\varphi_3 = \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1} M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}.$$

Using the argument deriving (C.2.10), we can rewrite (C.7.11) as

$$\begin{aligned}
\hat{w}_i^2 - w_i^2 &= \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2) - 2m_i'(\hat{\Lambda} - \Lambda)\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} \qquad \text{(C.7.12)} \\
&+ 2m_i'\hat{\Lambda}\mathbb{A}'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\mathbb{A}'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\Lambda\psi_1\hat{\mathbb{G}}\hat{\Lambda}' m_i
\end{aligned}$$

$$
\begin{aligned}
& - 2m_i'\Lambda\mathbb{A}\hat{\mathbb{G}}\hat{\Lambda}'m_i - 2m_i'\Lambda\psi_1(\hat{\Lambda}-\Lambda)'m_i + 2m_i'\Lambda\mathbb{A}(\hat{\Lambda}-\Lambda)'m_i \\
& + m_i'\Lambda\mathbb{A}'\mathbb{A}\Lambda'm_i - 2m_i'\Lambda\mathbb{A}'\psi_1\Lambda'm_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathbb{G}}\hat{\Lambda}'m_i + 2\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2}m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'m_i \\
& + m_i'\Lambda\varphi_1\Lambda'm_i - m_i'\Lambda\varphi_2\Lambda'm_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it} - E(e_te_{it})] \\
& + m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i + m_i'\Lambda\varphi_3\Lambda'm_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})_i \\
= {} & a_{i,1} + a_{i,2} + \cdots + a_{i,19}, \quad \text{say.}
\end{aligned}
$$

Using the Cauchy-Schwartz inequality, we have

$$
\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2 \leq 19\frac{1}{N}\sum_{i=1}^{N}(\|a_{i,1}\|^2 + \cdots + \|a_{i,19}\|^2).
$$

Analyzing term by term of the first 17 terms on the left hand side of the above inequality (similar to the derivation of (C.2.11)), and notice that the last two terms are $O_p(N^{-2})$, we have

$$
\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2 = O_p(T^{-1}) + O_p(N^{-2}) + o_p(\|\hat{\Lambda}-\Lambda\|^2). \tag{C.7.13}
$$

Next, we consider the term $\|\hat{\Lambda}-\Lambda\|$. Using Lemma C.7.5(b), (e)-(h) and Lemma C.7.6, together with equation (C.7.10), we have

$$
\hat{\Lambda} - \Lambda = O_p(T^{-1/2}) + O_p(N^{-1}) + O_p([\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2]^{1/2}). \tag{C.7.14}
$$

Substituting equation (C.7.14) into (C.7.13), we get $\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2 = O_p(T^{-1}) + O_p(N^{-2})$, which is the second result of Theorem 3.9.1. The proof for the first result of Theorem 3.9.1 is provided after Lemma C.7.8. □

The following two lemmas will be useful in proving the first result of Theorem 3.9.1.

Lemma C.7.7. *Under Assumptions A, B″, C″, D″ and F″, we have*

$$(a)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \mathbb{O})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1}$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});$$

$$(b)\quad \frac{1}{N}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1});$$

$$(c)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = O_p(N^{-1}T^{-1/2}) + O_p(N^{-2});$$

$$(d)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = O_p(N^{-1});$$

$$(e)\quad \frac{1}{NT}\sum_{t=1}^{T}f_t e_t'\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1});$$

$$(f)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \mathbb{O}]\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1}$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});$$

$$(g)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(N^{-1}T^{-1/2}) + O_p(N^{-2});$$

$$(h)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(N^{-1}).$$

The proof of the above lemma is similar to that of Lemma B.3 and the details are therefore omitted.

Lemma C.7.8. *Under Assumptions A, B″, C″, D″ and F″, we have*

$$A \equiv (\hat{\Lambda} - \Lambda)'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}) + O_p(\frac{1}{N}) + O_p(\|\hat{\Lambda} - \Lambda\|^2).$$

Proof of the above lemma is similar to that of Lemma C.2.4 with a slight modification to account for the weak dependence in errors. The results (a)-(d) in Lemma C.7.7 and the second part of Theorem 3.9.1 are used to control the magnitude. Details are omitted.

Proof of Theorem 3.4.1 (continued). Now we prove the first result of Theorem 3.9.1. Notice that the term $\|\hat{\Lambda} - \Lambda\|^2$ is of smaller order than $\hat{\Lambda} - \Lambda$ and hence negligible. Then from (C.7.10), together with Lemma C.7.7 and Lemma C.7.8, we have

$$\hat{\Lambda} - \Lambda = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right).$$

This completes the proof of Theorem 3.9.1. □

From Lemma C.7.8 and Theorem 3.9.1, we have the following corollary directly.

Corollary C.7.2. *Under Assumptions A, B'', C'', D'' and F'', we have*

$$A \equiv (\hat{\Lambda} - \Lambda)' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda} \hat{\mathbb{P}}_N^{-1} = O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{N}\right).$$

The following lemma will be useful in proving Theorem 3.9.2.

Lemma C.7.9. *Under Assumptions A, B'', C'', D'' and F'', we have*

(a) $$\frac{1}{T}\sum_{t=1}^{T} f_t e_t' \hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}_N^{-1} = \frac{1}{T}\sum_{t=1}^{T} f_t e_t' \mathbb{W}^{-1} M \mathbb{R}_N^{-1} + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{T^{3/2}}\right);$$

(b) $$\hat{\mathbb{P}}_N^{-1} \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t f_t'$$
$$= \mathbb{P}_N^{-1} \Lambda' M' \mathbb{W}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t f_t' + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{T^{3/2}}\right);$$

(c) $$\hat{\mathbb{P}}_N^{-1} \hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} (\mathbb{O} - \mathbb{W}) \hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}_N^{-1}$$

$$= \mathbb{P}_N^{-1}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\mathbb{R}_N^{-1} + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{N^2}\right);$$

(d) $\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}$

$$= \mathbb{P}_N^{-1}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}_N^{-1} + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{N^2}\right);$$

(e)
$$\frac{1}{N}M'(\hat{\mathbb{W}}^{-1}-\mathbb{W}^{-1})M = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}m_i m_i'(e_{it}^2-w_i^2) + \frac{1}{NT}\sum_{i=1}^{N} m_i m_i' \frac{\varpi_i^2}{w_i^4}$$
$$-\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4} m_i'\Lambda\mathbb{P}_N^{-1}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}_N^{-1}\Lambda' m_i$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4} 2m_i'\Lambda\mathbb{G}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})_i$$
$$+O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{T^{3/2}}\right) + O_p\left(\frac{1}{N^2}\right).$$

where $\varpi_i^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E\big[(e_{it}^2-w_i^2)(e_{is}^2-w_i^2)\big]$.

Proof of Lemma C.7.9. First we reconsider the equation (C.7.12), which can be written as

$$\hat{w}_i^2 - w_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-w_i^2) + m_i'\Lambda\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\Lambda' m_i \tag{C.7.15}$$
$$-2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})_i + \tilde{\mathcal{R}}_i,$$

where

$$\tilde{\mathcal{R}}_i = -2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] + \tilde{\mathcal{S}}_i$$

with Using the argument deriving (C.2.10), we can rewrite (C.7.11) as

$$\tilde{\mathcal{S}}_i = -2m_i'(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} \tag{C.7.16}$$

$$+ 2m_i'\hat{\Lambda}\mathbb{A}'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\mathbb{A}'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\Lambda\psi_1\hat{\mathbb{G}}\hat{\Lambda}'m_i$$

$$- 2m_i'\Lambda\mathbb{A}\hat{\mathbb{G}}\hat{\Lambda}'m_i - 2m_i'\Lambda\psi_1(\hat{\Lambda}-\Lambda)'m_i + 2m_i'\Lambda\mathbb{A}(\hat{\Lambda}-\Lambda)'m_i$$

$$+ m_i'\Lambda\mathbb{A}'\mathbb{A}\Lambda'm_i - 2m_i'\Lambda\mathbb{A}'\psi_1\Lambda'm_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathbb{G}}\hat{\Lambda}'m_i + 2\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2}m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'m_i$$

$$+ m_i'\Lambda\varphi_1\Lambda'm_i - m_i'\Lambda\varphi_2\Lambda'm_i + m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i.$$

By the same arguments in the derivation of (C.2.18) and (C.2.19), we have

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{S}}_i^2 = O_p\left(N^{-1}T^{-2}\right) + O_p\left(N^{-2}T^{-1}\right) + O_p(T^{-3}). \tag{C.7.17}$$

and further

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{R}}_i^2 = O_p\left(\frac{1}{NT}\right) + O_p\left(\frac{1}{T^2}\right). \tag{C.7.18}$$

Now consider (a). Notice that

$$\begin{aligned}\frac{1}{NT}\sum_{t=1}^{T} f_t e_t'\hat{\mathbb{W}}^{-1}M &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{w}_i^2}f_t e_{it} m_i' \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}f_t e_{it} m_i' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2 w_i^2}f_t e_{it} m_i' = j_1 + j_2, \quad \text{say.}\end{aligned}$$

The term j_2 can be written as

$$j_2 = \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\hat{w}_i^2 w_i^2}f_t e_{it}(e_{is}^2 - w_i^2)m_i' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{w}_i^2 w_i^2}\Big[2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})_i\Big]f_t e_{it} m_i'$$

$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{w}_i^2 w_i^2}\Big[m_i'\Lambda\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\Lambda'm_i\Big]f_te_{it}m_i'$$

$$+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{w}_i^2 w_i^2}\tilde{\mathcal{R}}_i f_t e_{it} m_i' = j_{21}+j_{22}+j_{23}+j_{24}, \quad \text{say.}$$

The term j_{24} is bounded in norm by

$$C^5\left[\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\mathcal{R}}_i\|^2\right]^{1/2}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\right\|^2\right]^{1/2},$$

which is $O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$ by (C.7.18). Similarly by

$$\frac{1}{N}\sum_{i=1}^{N}\left\|2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})_i\right\|^2 = O_p(N^{-2}), \tag{C.7.19}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\left\|m_i'\Lambda\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\Lambda'm_i\right\|^2 = O_p(N^{-2}), \tag{C.7.20}$$

we can show that $j_{22}=O_p(N^{-1}T^{-1/2})$ and $j_{23}=O_p(N^{-1}T^{-1/2})$. Then consider the term j_{21}, which can be rewritten as

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{w_i^4}f_te_{it}(e_{is}^2-w_i^2)m_i' - \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{\hat{w}_i^2-w_i^2}{\hat{w}_i^2 w_i^4}f_te_{it}(e_{is}^2-w_i^2)m_i'.$$

The first term of the above expression is $O_p(N^{-1/2}T^{-1})$ due to Assumption F″.6 in Section

9. The second term is bounded in norm by

$$C^5\left[\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\right]^{1/2}\left[\frac{1}{N}\sum_{i=1}^{N}\left\|\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\right\|^2\cdot\left\|\frac{1}{T}\sum_{t=1}^{T}e_{is}^2 - w_i^2\right\|^2\right]^{1/2},$$

which is $O_p(T^{-3/2})$. By the preceding results, we have

$$\frac{1}{NT}\sum_{t=1}^{T}f_t e_t'\hat{\mathbb{W}}^{-1}M = \frac{1}{NT}\sum_{t=1}^{T}f_t e_t'\mathbb{W}^{-1}M + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{T^{3/2}}\right). \tag{C.7.21}$$

Combining the above result and $\hat{\mathbb{R}} = \mathbb{R} + O_p(T^{-1/2})$, we have (a). Combining the above result and $\hat{\mathbb{P}} = \mathbb{P} + O_p(T^{-1/2})$ and $\hat{\Lambda} = \Lambda + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}) + O_p(\frac{1}{N})$, we have (b).

Next we consider (c). Notice the expression of the left hand side is $O_p(N^{-1})$ from Lemma C.7.7 (h). Then by $\hat{\mathbb{R}} = \mathbb{R}+O_p(T^{-1/2})$, $\hat{\mathbb{P}} = \mathbb{P}+O_p(T^{-1/2})$, $\hat{\Lambda} = \Lambda+O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})+O_p(\frac{1}{N})$ and $\hat{w}_i^2 - w_i^2 = O_p(T^{-1/2}) + O_p(N^{-1}) + O_p(N^{-1/2}T^{-1/2})$ from (C.7.15), we have result (c). Result (d) can be proved similarly.

Finally we consider (e). The left hand side of (e) equals

$$-\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2 w_i^2}m_i m_i' = -\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{w}_i^2 - w_i^2}{w_i^4}m_i m_i' + \frac{1}{N}\sum_{i=1}^{N}\frac{(\hat{w}_i^2 - w_i^2)^2}{\hat{w}_i^2 w_i^4}m_i m_i' = l_1 + l_2, \text{ say.}$$

We first consider l_1. By (C.7.15), l_1 can be rewritten as

$$l_1 = -\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{w}_i^2 - w_i^2}{w_i^4}m_i m_i' = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(e_{it}^2 - w_i^2)m_i m_i'$$
$$-\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}\Big[m_i'\Lambda\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\Lambda' m_i\Big]$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}\Big[2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})_i\Big]$$

$$+2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_i^4}\mathrm{tr}\Big[\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it}-E(e_t e_{it})]m_i'\Big]m_i m_i' - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_i^4}\tilde{\mathcal{S}}_i m_i m_i'$$

$$= l_{11}+\cdots+l_{15}, \quad \text{say.}$$

First consider l_{12}. Using the argument to prove (c), we have

$$l_{12} = -\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}m_i'\Lambda\mathbb{P}_N^{-1}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}_N^{-1}\Lambda'm_i + O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{N^2}\right).$$

Similarly, by the fact that $\left[m_i'\Lambda\mathbb{G}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})_i\right] = O_p(N^{-1})$, we have

$$l_{13} = \frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}2m_i'\Lambda\mathbb{G}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})_i + O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{N^2}\right).$$

Then consider l_{14}, whose (v,u) element $(v,u=1,\ldots,k)$ equals

$$\mathrm{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it}-E(e_t e_{it})]\frac{1}{w_i^4}m_i' m_{iv} m_{iu}\Big]$$

which can be proved to be $O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$ similarly as Lemma C.7.7(a). The last term l_{15} is bounded by (using (C.7.17))

$$C^6\left[\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{S}}_i^2\right]^{1/2} = O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2}).$$

Hence, we have

$$l_1 = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(e_{it}^2-w_i^2)m_i m_i'$$

$$-\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}m_i'\Lambda\mathbb{P}_N^{-1}\Lambda' M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}_N^{-1}\Lambda' m_i$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}2m_i'\Lambda\mathbb{G}\Lambda' M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})_i$$

$$+O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{\sqrt{N}T})+O_p(\frac{1}{T^{3/2}})+O_p(\frac{1}{N^2}).$$

Then consider l_2, which can be rewritten as (by (C.7.15))

$$l_2=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2 w_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-w_i^2)\Big]^2 m_i m_i' + 2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2 w_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-w_i^2)\Big]\tilde{\mathcal{R}}_i m_i m_i'$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2 w_i^4}\tilde{\mathcal{R}}_i^2 m_i m_i' + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2 w_i^4}(d_i)^2 m_i m_i' + 2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2 w_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-w_i^2)\Big]d_i m_i m_i'$$

$$+2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{w}_i^2 w_i^4}d_i\tilde{\mathcal{R}}_i m_i m_i' = l_{21}+\cdots+l_{26}, \quad \text{say.}$$

where $d_i = m_i'\Lambda\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\Lambda' m_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})_i$. We analyze the six terms on the right hand side of the above equation one by one. The term l_{22} is bounded in norm by

$$2C^8\Big[\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-w_i^2)\Big|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{R}}_i^2\Big]^{1/2},$$

which is $O_p(N^{-1/2}T^{-1})$ by (C.7.18). The term l_{23} is bounded in norm by

$$C^8\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{R}}_i^2 = O_p\left(\frac{1}{NT}\right)+O_p\left(\frac{1}{T^2}\right).$$

Similarly, by (C.7.19) and (C.7.20), we can show $l_{24}=O_p(N^{-2})$, $l_{25}=O_p(N^{-1}T^{-1/2})$ and

$l_{26} = O_p(N^{-3/2}T^{-1/2}) + O_p(N^{-1}T^{-1})$. Finally, the term l_{21} can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_i^6}\left[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2)\right]^2 m_i m_i' - \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2 w_i^6}\left[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2)\right]^2 m_i m_i'$$

The first term of the above expression is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i' + O_p\left(N^{-1/2}T^{-1}\right),$$

where ϖ_i^2 is defined in Lemma C.7.9. The second term is bounded in norm by

$$C^{10}\left[\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\right]^{1/2}\left[\frac{1}{N}\sum_{i=1}^{N}\left|\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2)\right|^4\right]^{1/2} = O_p\left(T^{-3/2}\right).$$

So

$$l_{21} = \frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i' + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).$$

Hence we have

$$l_2 = \frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i' + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right).$$

Combining the preceding results on l_1 and l_2, we have result (e). $\square$

Proof of Theorem 3.9.2. To derive the asymptotic representation of $\hat{\Lambda}$, we first study the asymptotic behavior of $\mathbb{A}$. By equation (C.7.9), together with Lemma C.7.7(a), (c) and (d), Lemma C.7.8 as well as Lemma C.7.9(d),

$$\mathbb{A} + \mathbb{A}' = \eta_1 + \eta_1' + \xi_1 + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right) + O_p\left(\frac{1}{N^2}\right),$$

where

$$\eta_1 = \frac{1}{NT}\sum_{t=1}^{T} f_t e_t' \mathbb{W}^{-1} M \Lambda \mathbb{P}^{-1}, \quad \xi_1 = \frac{1}{N^2}\mathbb{P}^{-1}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1} M \Lambda \mathbb{P}^{-1}.$$

Taking vech operation on both sides,

$$\text{vech}(\mathbb{A}+\mathbb{A}') = \text{vech}(\eta_1+\eta_1')+\text{vech}(\xi_1)+O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{N}T}\right)+O_p\left(\frac{1}{T^{3/2}}\right)+O_p\left(\frac{1}{N^2}\right),$$

implying

$$2D_r^+\text{vec}(\mathbb{A}) = 2D_r^+\text{vec}(\eta_1)+D_r^+\text{vec}(\xi_1)+O_p(\frac{1}{N\sqrt{T}})+O_p\left(\frac{1}{\sqrt{N}T}\right)+O_p\left(\frac{1}{T^{3/2}}\right)+O_p\left(\frac{1}{N^2}\right), \tag{C.7.22}$$

where D_r^+ is defined the same as in Theorem 3.4.2. By the identification condition, we know both $\Lambda'(\frac{1}{N}M'\mathbb{W}^{-1}M)\Lambda$ and $\hat{\Lambda}'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)\hat{\Lambda}$ are diagonal matrices, which implies

$$\text{Ndg}\left\{\Lambda'(\frac{1}{N}M'\mathbb{W}^{-1}M)\Lambda - \hat{\Lambda}'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)\hat{\Lambda}\right\} = 0,$$

where Ndg$(\cdot)$ denote the non-diagonal elements of its argument. By adding and subtracting terms,

$$\text{Ndg}\left\{(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)\hat{\Lambda} + \hat{\Lambda}'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)(\hat{\Lambda}-\Lambda)\right. \tag{C.7.23}$$

$$\left. -(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)(\hat{\Lambda}-\Lambda) + \Lambda'\left[\frac{1}{N}M'(\hat{\mathbb{W}}^{-1}-\mathbb{W}^{-1})M\right]\Lambda\right\} = 0.$$

Using Lemma C.7.9(e) and $\hat{\Lambda}-\Lambda = O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})+O_p(\frac{1}{N})$ from Theorem 3.9.1, we have

$$\text{Ndg}\Big\{\hat{\Lambda}'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)(\hat{\Lambda}-\Lambda)+(\hat{\Lambda}-\Lambda)'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)\hat{\Lambda}\Big\}$$

$$= \text{Ndg}\{\zeta_1-\mu_1+\xi_2\}+O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T^{3/2}}\right)+O_p\left(\frac{1}{N^2}\right),$$

where

$$\begin{aligned}
\zeta_1 &= \Lambda'\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{m_i m_i'}{w_i^4}(e_{it}^2-w_i^2)\right]\Lambda,\\
\mu_1 &= \Lambda'\left[\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i'\right]\Lambda,\\
\xi_2 &= \Lambda'\left[\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}m_i'\Lambda\mathbb{P}_N^{-1}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}_N^{-1}\Lambda'm_i\right.\\
&\quad\left.-\frac{2}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}m_i'\Lambda\mathbb{G}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})_i\right]\Lambda\\
&= \frac{1}{N}\Lambda'\left[\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}\varsigma_i\right]\Lambda
\end{aligned}$$

where ς_i is a scalar defined in the paragraph before Theorem 3.9.2 and $\varpi_i^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} E\big[(e_{it}^2-w_i^2)(e_{is}^2-w_i^2)\big]$. With the same definition of $\mathcal{D}$ given in Theorem 3.4.2, together with the definition of $\mathbb{P}$, the preceding equation can be rewritten as

$$\text{veck}(\mathbb{A}\mathbb{P}+\mathbb{P}\mathbb{A}') = \text{veck}(\zeta_1-\mu_1+\xi_2)+O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T^{3/2}}\right)+O_p\left(\frac{1}{N^2}\right),$$

or equivalently

$$\mathcal{D}\text{vec}(\mathbb{A}\mathbb{P}+\mathbb{P}\mathbb{A}') = \mathcal{D}\text{vec}(\zeta_1-\mu_1+\xi_2)+O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T^{3/2}}\right)+O_p\left(\frac{1}{N^2}\right).$$

Furthermore, we can rewrite the above equation as

$$\mathcal{D}[(\mathbb{P}\otimes I_r)+(I_r\otimes\mathbb{P})K_r]\text{vec}(\mathbb{A}) = \mathcal{D}\text{vec}(\zeta_1-\mu_1+\xi_2)+O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T^{3/2}}\right)+O_p\left(\frac{1}{N^2}\right), \tag{C.7.24}$$

where K_r is defined the same as in Theorem 3.4.2. The above equation has $\frac{r(r-1)}{2}$ restrictions.

Then combining (C.7.22) and (C.7.24), we have

$$\begin{bmatrix} 2D_r^+ \\ \mathcal{D}[(\mathbb{P}\otimes I_r)+(I_r\otimes\mathbb{P})K_r] \end{bmatrix}\text{vec}(\mathbb{A}) = \begin{bmatrix} 2D_r^+\text{vec}(\eta_1) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{D}\text{vec}(\zeta_1) \end{bmatrix} - \begin{bmatrix} 0 \\ \mathcal{D}\text{vec}(\mu_1) \end{bmatrix} \tag{C.7.25}$$

$$+\begin{bmatrix} D_r^+\text{vec}(\xi_1) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{D}\text{vec}(\xi_2) \end{bmatrix} + O_p\left(\frac{1}{N\sqrt{T}}\right)+O_p\left(\frac{1}{\sqrt{NT}}\right)+O_p\left(\frac{1}{T^{3/2}}\right)+O_p\left(\frac{1}{N^2}\right).$$

Let

$$\mathbb{D}_1^\dagger = \begin{bmatrix} 2D_r^+ \\ \mathcal{D}[(\mathbb{P}\otimes I_r)+(I_r\otimes\mathbb{P})K_r] \end{bmatrix},$$

together with the same definitions of $\mathbb{D}_2$ and $\mathbb{D}_3$ given in Theorem 3.4.2, the above equation can be rewritten as

$$\mathbb{D}_1^{\dagger}\text{vec}(\mathbb{A}) = \mathbb{D}_2\text{vec}(\eta_1) + \mathbb{D}_3\text{vec}(\zeta_1) - \mathbb{D}_3\text{vec}(\mu_1) + \frac{1}{2}\mathbb{D}_2\text{vec}(\xi_1) + \mathbb{D}_3\text{vec}(\xi_2) \tag{C.7.26}$$

$$+O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{NT}}\right) + O_p\left(\frac{1}{T^{3/2}}\right) + O_p\left(\frac{1}{N^2}\right).$$

Noticing that

$$\text{vec}(\eta_1) = \text{vec}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}f_t e_{it} m_i'\Lambda\mathbb{P}^{-1}\right] = (\mathbb{P}^{-1}\Lambda' \otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i \otimes f_t)e_{it},$$

$$\text{vec}(\zeta_1) = \text{vec}\left[\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{m_i m_i'}{w_i^4}(e_{it}^2 - w_i^2)\Lambda\right] = (\Lambda \otimes \Lambda)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(m_i \otimes m_i)(e_{it}^2 - w_i^2),$$

$$\text{vec}(\mu_1) = \text{vec}\left[\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i'\Lambda\right] = (\Lambda \otimes \Lambda)'\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}(m_i \otimes m_i),$$

$$\text{vec}(\xi_1) = \frac{1}{N}\Big((\mathbb{P}^{-1}\Lambda') \otimes (\mathbb{P}^{-1}\Lambda')\Big)\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}\frac{\mathbb{O}_{ij}}{w_i^2 w_j^2}(m_j \otimes m_i),$$

$$\begin{aligned}\text{vec}(\xi_2) = {} & \frac{1}{N}(\Lambda \otimes \Lambda)'\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_i^4}(m_i \otimes m_i)(m_i \otimes m_i)'\right] \\ & \times \left[(\Lambda\mathbb{P}^{-1}\Lambda') \otimes (\Lambda\mathbb{P}^{-1}\Lambda')\right]\text{vec}\left[\frac{1}{N}M'\mathbb{W}^{-1}(\mathbb{O} - \mathbb{W})\mathbb{W}^{-1}M\right] \\ & - 2\frac{1}{N}(\Lambda \otimes \Lambda)'\left[\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_i^4}(m_i\iota_i') \otimes (m_i m_i')\right]\text{vec}\left[\Lambda\mathbb{G}^{-1}\Lambda' M'\mathbb{W}^{-1}(\mathbb{O} - \mathbb{W})\right] \\ = {} & \frac{1}{N}(\Lambda \otimes \Lambda)'\frac{1}{N}\sum_{i=1}^{N}\frac{\varsigma_i}{w_i^4}(m_i \otimes m_i).\end{aligned}$$

Now we can rewrite the asymptotic expression of $\mathbb{A}$ as

$$\text{vec}(\mathbb{A}) = (\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_2(\mathbb{P}^{-1}\Lambda' \otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i \otimes f_t)e_{it} \tag{C.7.27}$$

$$+ (\mathbb{D}_1^\dagger)^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)' \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \frac{1}{w_i^4}(m_i \otimes m_i)(e_{it}^2 - w_i^2)$$

$$- (\mathbb{D}_1^\dagger)^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)' \frac{1}{NT}\sum_{i=1}^N \frac{\varpi_i^2}{w_i^6}(m_i \otimes m_i)$$

$$+ \frac{1}{2}(\mathbb{D}_1^\dagger)^{-1}\mathbb{D}_2 \frac{1}{N}\Big((\mathbb{P}^{-1}\Lambda') \otimes (\mathbb{P}^{-1}\Lambda')\Big)\frac{1}{N}\sum_{i=1}^N \sum_{j=1, j\neq i}^N \frac{\mathbb{O}_{ij}}{w_i^2 w_j^2}(m_j \otimes m_i)$$

$$+ (\mathbb{D}_1^\dagger)^{-1}\mathbb{D}_3 \frac{1}{N}(\Lambda \otimes \Lambda)' \frac{1}{N}\sum_{i=1}^N \frac{\varsigma_i}{w_i^4}(m_i \otimes m_i)$$

$$+ O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right) + O_p\left(\frac{1}{N^2}\right).$$

Next consider equation (C.7.10), which is derived from the first order condition of $\hat{\Lambda}$. By Lemma C.7.7 (f)(g) and Lemma C.7.9 (a)(b)(c), we have

$$\hat{\Lambda}' - \Lambda' = -\mathbb{A}'\Lambda' + \frac{1}{NT}\sum_{t=1}^T f_t e_t' \mathbb{W}^{-1} M \mathbb{R}^{-1} + \mathbb{P}^{-1}\Lambda' \frac{1}{NT} M' \mathbb{W}^{-1} \sum_{t=1}^T e_t f_t' \Lambda' \tag{C.7.28}$$
$$+ \xi_3 + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right) + O_p\left(\frac{1}{N^2}\right),$$

where

$$\xi_3 = \mathbb{P}_N^{-1}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O} - \mathbb{W})\mathbb{W}^{-1} M \mathbb{R}_N^{-1}.$$

Taking vec operation on both sides of the above equation (C.7.28) and noticing that

$$\text{vec}\left[\frac{1}{NT}\sum_{t=1}^T f_t e_t' \mathbb{W}^{-1} M \mathbb{R}^{-1}\right] = \text{vec}\left[\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \frac{1}{w_i^2} f_t e_{it} m_i' \mathbb{R}^{-1}\right]$$
$$= (\mathbb{R}^{-1} \otimes I_r)\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \frac{1}{w_i^2}(m_i \otimes f_t)e_{it},$$

$$\text{vec}\left[\mathbb{P}^{-1}\Lambda'\frac{1}{NT}M'\mathbb{W}^{-1}\sum_{t=1}^{T}e_t f_t'\Lambda'\right] = \text{vec}\left[\mathbb{P}^{-1}\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}m_i e_{it} f_t'\Lambda'\right]$$
$$= K_{kr}\text{vec}\left[\Lambda\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}f_t e_{it} m_i'\Lambda\mathbb{P}^{-1}\right]$$
$$= K_{kr}[(\mathbb{P}^{-1}\Lambda')\otimes\Lambda]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it},$$

and

$$\text{vec}(\xi_3) = \frac{1}{N}\Big((\mathbb{R}^{-1})\otimes(\mathbb{P}^{-1}\Lambda')\Big)\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\mathbb{O}_{ij}}{w_i^2 w_j^2}(m_j\otimes m_i),$$

where K_{kr} is defined the same as in Theorem 3.4.2, we have

$$\text{vec}(\hat{\Lambda}'-\Lambda') = \Big[K_{kr}[(\mathbb{P}^{-1}\Lambda')\otimes\Lambda] + \mathbb{R}^{-1}\otimes I_r\Big]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it} + \text{vec}(\xi_3) \tag{C.7.29}$$
$$- K_{kr}(I_r\otimes\Lambda)\text{vec}(\mathbb{A}) + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right) + O_p\left(\frac{1}{N^2}\right).$$

Plug (C.7.27) into (C.7.29), then we have

$$\text{vec}(\hat{\Lambda}'-\Lambda') = \mathbb{B}_1^{\dagger}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it} - \mathbb{B}_2^{\dagger}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(m_i\otimes m_i)(e_{it}^2 - w_i^2) + \frac{1}{T}\Delta^{\dagger}$$
$$+ \frac{1}{N}\Pi^{\dagger} + O_p\left(\frac{1}{N\sqrt{T}}\right) + O_p\left(\frac{1}{\sqrt{N}T}\right) + O_p\left(\frac{1}{T^{3/2}}\right) + O_p\left(\frac{1}{N^2}\right), \tag{C.7.30}$$

where $\mathbb{B}_1^{\dagger}, \mathbb{B}_2^{\dagger}, \Delta^{\dagger}$ and $\Pi^{\dagger}$ are defined in the paragraph before Theorem 3.9.2. This completes the proof of Theorem 3.9.2. □

Proof of Theorem 3.9.1. Given the results in Theorem 3.9.2, letting $N, T \to \infty$ and $N/T^2 \to 0$ and $T/N^3 \to 0$, by the Central Limit Theorem, we have the following limiting distribution

$$\sqrt{NT}\Big[\text{vec}(\hat{\Lambda}' - \Lambda') - \frac{1}{T}\Delta^{\dagger} - \frac{1}{N}\Pi^{\dagger}\Big] \xrightarrow{d} N(0, \Xi),$$

where $\Xi = \lim_{N\to\infty} \Xi_{NT}$ with Ξ_{NT} defined in Theorem 3.9.1. This completes the proof. □

Proof of Theorem 3.9.3. From equation (C.7.15) and the analysis in the proof of Lemma C.7.9(e), we know both the second and third terms on the right hand side of (C.7.15) are $O_p(N^{-1})$, and the last term $\tilde{\mathcal{R}}_i$ is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$, which directly implies the asymptotic representation of $\hat{w}_i^2$ as in Theorem 3.9.3. Hence we prove Theorem 3.9.3. □