Quasi Maximum Likelihood Analysis of High Dimensional Constrained Factor Models*

Kunpeng Li[†], Qi Li[‡] and Lina Lu[§]

August 11, 2016

Abstract

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 by 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. This paper investigates the estimation and inferential theory of constrained factor models under large-N and large-T setup, where N denotes the number of cross sectional units and T the time periods. We propose using the quasi maximum likelihood method to estimate the model and investigate the asymptotic properties of the quasi maximum likelihood estimators, including consistency, rates of convergence and limiting distributions. 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 our theoretical results and show that the quasi maximum likelihood estimators and the proposed new statistic perform well in finite samples. We also consider the extension of an approximate constrained factor model.

Key Words: Constrained factor models, Maximum likelihood estimation, High dimensionality, Inferential theory.

JEL #: C13, C38.

*We would like to thank Jushan Bai for the stimulating comments that greatly improve the paper. We also thank the guest editors and two anonymous referees for their constructive comments. Kunpeng Li gratefully acknowledges the financial support from China NSF No.71571122 and No.71201031. Qi Li's research is partially supported by China NSF project No.71133001.

[†]International School of Economics and Management, Capital University of Economics and Business, Fengtai District, Beijing, 100070, China. Email: kunpenglithu@126.com.

[‡]Department of Economics, Texas A&M University, College Station, TX 77843, USA; International School of Economics and Management, Capital University of Economics and Business, Beijing, China. Corresponding author: Qi Li, Email: qi-li@tamu.edu.

[§]Department of Economics, Columbia University. Email: ll2582@columbia.edu

1 Introduction

With the rapid development of data collecting, 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, Conner and Korajczyk (1986, 1988) and Fan, Liao and Shi (2014) 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, Stixrud and Urzua (2006) use factor models to capture unobservable personal abilities. In international economics, Kose, Otrok and Whiteman (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, Liao and Mincheva (2011) and Fan, Liao and Mincheva (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 (2002)). 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 (2002) 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 (2002) and gives the asymptotic representations of the PC estimates of loadings, factors and common components. Doz, Giannone and Reichlin (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 the limiting distributions of the MLE. Fan, Liao and Wang (2014) 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, 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 is of our 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 paper differs from Tsai and Tsay (2010) in several dimensions. First, although Tsai and Tsay 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 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 given sample size. When k is large, the benefit of constrained factor models against standard factor models becomes weak which makes constrained factor 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 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 paper, we propose new statistics for testing model specifications that are applicable to the large-N and large-T setups.

The rest of the paper is organized as follows. Section 2 provides more empirical examples of the constrained factor model. Section 3 introduces the model and lists the assumptions needed for the subsequent analysis. Section 4 delivers the consistency and limiting distribution results of the MLE. Section 5 considers testing issues within constrained factor models. Section 6 considers a partially constrained factor model and presents the asymptotic properties of the MLE for this model. Section 7 presents the Expectation Maximization (EM) algorithm for computation of the MLE. Section 8 conducts Monte Carlo simulations to investigate the finite sample performance of the MLE and to study the size and power of

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

our proposed statistic on the model specification. In Section 9, we extend Assumption B to a more general weak dependence structure and study the MLE in this extension. Section 10 concludes the paper. All technical contents are delegated to several appendices.

2 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), Conner and Korajczyk (1988) among others. This motivates using

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

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. The Rosenberg's specification is very close to the one studied in this paper. With a light modification, the analysis in this paper can easily be extended to cover the Resenberg'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, Conner and Linton (2007) and Conner, Hagmann and Linton (2012) consider the following nonparametric specification

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

where $g_j(\cdot)$ is an unknown smooth function. Conner, Hagmann 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.,

$$g_j(x) \approx \lambda_{j0} + \lambda_{j1}x + \cdots + \lambda_{jq}x^q. \tag{2.1}$$

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 Conner and Linton (2007) and Conner, Hagmann and Linton (2012) but the computational burden has been much reduced. A primary issue related with our method is whether the approximation (2.1) is good enough. This work can be partially addressed by the W statistic proposed in Section 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.

3 Constrained Factor Models

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

$$z_t = M\Lambda f_t + e_t, \tag{3.1}$$

where $z_t = (z_{1t}, z_{2t}, \ldots, z_{Nt})'$ 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; $f_t = (f_{1t}, f_{2t}, \ldots, f_{rt})'$ 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 $k \geq r$. If $k < r$, the expression Λf_t achieves no dimension reduction and we can simply consider the linear regression $z_t = Mf_t^* + e_t$ with $f_t^* = \Lambda f_t$.

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 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 $\overline{M}_{ff} = \lim\limits_{T\to\infty} M_{ff}$.

Assumption B: The idiosyncratic error term e_{it} is independent across the i index and the t index with $E(e_t) = 0$, $E(e_t e_t') = \Sigma_{ee} = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2)$ 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 $P_\infty = \lim\limits_{N\to\infty} P$ and $R_\infty = \lim\limits_{N\to\infty} R$ exist. In addition, $\lim\limits_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \sigma_i^{-4}(m_i \otimes m_i)(m_i' \otimes m_i') = V_\infty$ 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 cross-sectional and temporal heteroskedasticities and correlations, see Bai and Li (2016) for a related development in this direction. 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 $\Lambda'(\frac{1}{N} M' \Sigma_{ee}^{-1} M)\Lambda = P$, where P is a diagonal matrix whose diagonal elements are distinct and arranged in 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 estimator 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

$$\mathcal{L}(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}\mathrm{tr}[M_{zz}\Sigma_{zz}^{-1}], \tag{3.2}$$

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

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

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:

$$\hat{\Lambda}'M'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}M = 0; \tag{3.3}$$

$$\mathrm{diag}(\hat{\Sigma}_{zz}^{-1}) = \mathrm{diag}(\hat{\Sigma}_{zz}^{-1}M_{zz}\hat{\Sigma}_{zz}^{-1}), \tag{3.4}$$

where $\hat{\Lambda}$, and $\hat{\Sigma}_{ee}$ denote MLE of Λ and Σ_{ee} 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 Expectation Maximization (EM) algorithm to compute the MLE. Details are given in Section 7.

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 4.1 (Consistency) *Let $\hat{\theta} = (\hat{\Lambda}, \hat{\Sigma}_{ee})$ be the MLE that maximizes (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 (2002), Bai (2003), Doz, Giannone and Reichlin (2012) and Bai and Li (2012, 2015). 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 4.1 (Convergence rates) *Under the assumptions of Proposition 4.1, we have*

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

According to Theorem 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.

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

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

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

Theorem 4.2 (Asymptotic representation) *Under assumptions of Theorem 4.1, we have*

$$\begin{aligned} \mathrm{vec}(\hat{\Lambda}' - \Lambda') = \mathbb{B}_1 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (m_i \otimes f_t) e_{it} - \mathbb{B}_2 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^4} (m_i \otimes m_i)(e_{it}^2 - \sigma_i^2) \\ + \frac{1}{T} \Delta + O_p\Big(\frac{1}{N\sqrt{T}}\Big) + O_p\Big(\frac{1}{\sqrt{NT}}\Big) + O_p\Big(\frac{1}{T^{3/2}}\Big), \end{aligned} \tag{4.1}$$

where the symbols $\mathbb{B}_1$, $\mathbb{B}_2$ and Δ are defined above Theorem 4.2.

The first two terms on the right hand side of (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 4.2 reaffirms the convergence rates asserted in Theorem 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 4.2, invoking a Central Limit Theorem, we have the following theorem.

Theorem 4.3 (Limiting distribution) *Under assumptions of Theorem 4.1, as $N, T \to \infty, N/T^2 \to 0$, we have*

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

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 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 4.3, we have the following corollary.

Corollary 4.1 *Under assumptions of Theorem 4.3, with normality of e_{it}, we have*

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

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

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

Remark 4.1 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

$$\begin{aligned}\hat{\mathbb{B}}_1&=K_{kr}[(\hat{P}^{-1}\hat{\Lambda}')\otimes\hat{\Lambda}]+\hat{R}^{-1}\otimes I_r-K_{kr}(I_r\otimes\hat{\Lambda})\hat{\mathbb{D}}_1^{-1}\mathbb{D}_2[(\hat{P}^{-1}\hat{\Lambda}')\otimes I_r],\\ \hat{\mathbb{B}}_2&=K_{kr}(I_r\otimes\hat{\Lambda})\hat{\mathbb{D}}^{-1}\mathbb{D}_3(\hat{\Lambda}\otimes\hat{\Lambda})'.\end{aligned}$$

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 4.2 Theorem 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 (A.16) in Appendix A, together with Lemma B.3 (e), (f) and Lemma B.5 (a), we have

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

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

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

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

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

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

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

Theorem 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 4.5 *Under assumptions of Theorem 4.1, we have*

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

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.

5 Testing

Corollary 4.1 in the previous section gives the limiting distribution of the MLE, which allows one to test whether the loading matrix Λ is equal to some known matrix. First, we consider the following hypothesis:

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

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

We consider a wald statistic

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

where the symbols $\hat{\Delta}$ and $\hat{\Omega}$ are given in Remark 4.1. The following theorem, which is a direct result of Theorem 4.3, gives the limiting distribution of W_Λ.

Theorem 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 degree of freedom kr.

Next, we consider the problem of testing whether specification (3.1) is appropriate in a general factor model. That is, the correctness of the decomposition of loadings matrix $L = M\Lambda$. For a given M, the null and alternative hypotheses are

$$\begin{aligned} H_0: &\quad L = M\Lambda \quad \text{for} \quad \text{some} \quad \Lambda, \\ H_1: &\quad L \neq M\Lambda \quad \text{for} \quad \text{all} \quad \Lambda. \end{aligned}$$

In traditional (low-dimensional) factor analysis, testing restrictions can be conducted by using the likelihood ratio principle. Because the number of parameters is finite, the number of restrictions imposed on these parameters is therefore finite. Consequently, under the null hypothesis, the likelihood ratio has an asymptotic χ^2 distribution with a finite number of the degrees of freedom. 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. If the traditional likelihood ratio test is used, the limiting distribution of the statistic would depend on N, an undesirable feature which can make a test unstable when N is large. 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

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

where $\widetilde{\Sigma}_{ee}$ is an estimator of Σ_{ee} under the alternative hypothesis.

[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 $N/T^2 \to 0$.

Theorem 5.2 *Under the same assumptions of Proposition 4.1 and $N/T^2 \to 0$, under H_0, we have*

$$W \xrightarrow{d} N(0, 2r).$$

Remark 5.1 As pointed out in Section 2, the identification condition IC in this paper 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 such a 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 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.

6 Partially Constrained Factor Models

In this section, we consider the following partially constrained factor model

$$z_t = M\Lambda f_t + \Gamma g_t + e_t \triangleq \Phi h_t + e_t, \tag{6.1}$$

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 (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

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

Here Θ is the parameter space specified by Assumption D and the identification condition IC$'$. In appendix D, we show that the first order condition for Λ can be written as

$$\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M = 0. \tag{6.2}$$

The first order condition for Γ can be written as

$$\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz}) = 0. \tag{6.3}$$

The first order condition for Σ_{ee} can be written as

$$\mathrm{diag}\Big[(M_{zz}-\hat{\Sigma}_{zz})-M\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz}-\hat{\Sigma}_{zz})-(M_{zz}-\hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\Big]=0. \quad (6.4)$$

Before we present the asymptotic results for the MLE, we first introduce some notations:

$$\begin{aligned}
&\mathbb{B}_1^\star=R^{-1}\otimes I_{r_1}+K_{kr_1}[(P^{-1}\Lambda')\otimes\Lambda]-K_{kr_1}(E_1'\otimes\Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}^{-1}E_1\Lambda')\otimes E_1],\\
&\mathbb{B}_2^\star=K_{kr_1}[P^{-1}\otimes\psi]-K_{kr_1}(E_1'\otimes\Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}^{-1}E_1)\otimes E_2],\\
&\mathbb{B}_3^\star=-K_{kr_1}(E_1'\otimes\Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}^{-1}E_2)\otimes E_1],\\
&\mathbb{B}_4^\star=-K_{kr_1}(E_1'\otimes\Psi)\mathbb{D}_1^{-1}\mathbb{D}_2[(\mathcal{H}^{-1}E_2)\otimes E_2],\qquad \mathbb{B}_5^\star=-K_{kr_1}(E_1'\otimes\Psi)\mathbb{D}_1^{-1}\mathbb{D}_3,\\
&\Delta^\star=K_{kr_1}(E_1'\otimes\Psi)\mathbb{D}_1^{-1}\mathbb{D}_3\Big[\frac{1}{N}\sum_{i=1}^N\frac{1}{\sigma_i^6}(\phi_i\otimes\phi_i)(\kappa_{i,4}-\sigma_i^4)+\mathrm{vec}(r_1\mathcal{H}-E_2E_2')\Big],
\end{aligned}$$

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 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 6.1 *Under Assumptions A$'$, B, C$'$ and D, when $N,T\to\infty$, we have, for all i,*

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

In addition, if IC$'$ is imposed, we have, for all i,

$$\hat{\gamma}_i-\gamma_i=\frac{1}{T}\sum_{t=1}^T g_te_{it}+O_p(\frac{1}{T})$$

and

$$\begin{aligned}
\mathrm{vec}(\hat{\Lambda}'-\Lambda')=\;&\mathbb{B}_1^\star\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it}+\mathbb{B}_2^\star\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\sigma_i^2}(\Lambda'm_i\otimes g_t)e_{it}\\
&+\mathbb{B}_3^\star\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\sigma_i^2}(\gamma_i\otimes f_t)e_{it}+\mathbb{B}_4^\star\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\sigma_i^2}(\gamma_i\otimes g_t)e_{it}\\
&+\mathbb{B}_5^\star\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\frac{1}{\sigma_i^4}(\phi_i\otimes\phi_i)(e_{it}^2-\sigma_i^2)+\frac{1}{T}\Delta^\star\\
&+O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T^{3/2}}),
\end{aligned}$$

where $\mathbb{B}_1^\star,\dots,\mathbb{B}_5^\star$ and $\Delta^\star$ are defined above this theorem.

Given the above theorem, we have the following distribution results for the MLE.

Corollary 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

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

where $\Omega^\star = \lim\limits_{N\to\infty} \Omega_N^\star$ with

$$\begin{aligned}\Omega_N^\star = {} & \mathbb{B}_1^\star(R \otimes I_{r_1})\mathbb{B}_1^{\star\prime} + \mathbb{B}_2^\star(P \otimes I_{r_1})\mathbb{B}_2^{\star\prime} + \mathbb{B}_3^\star(Q \otimes I_{r_1})\mathbb{B}_3^{\star\prime} + \mathbb{B}_4^\star(Q \otimes I_{r_2})\mathbb{B}_4^{\star\prime} \\ & + \mathbb{B}_1^\star(S \otimes I_{r_1})\mathbb{B}_3^{\star\prime} + \mathbb{B}_3^\star(S' \otimes I_{r_1})\mathbb{B}_1^{\star\prime} + \mathbb{B}_5^\star\Big[\frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^8}(\phi_i\phi_i') \otimes (\phi_i\phi_i')(\kappa_{i,4} - \sigma_i^4)\Big]\mathbb{B}_5^{\star\prime},\end{aligned}$$

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 method in the proof of Theorem 4.5, we have the following asymptotic representation and limiting distribution results on $\hat{h}_t$.

Theorem 6.2 *Under Assumptions A', B, C′ and D, together with IC′, we have, for all t,*

$$\hat{h}_t - h_t = \mathcal{H}^{-1}\frac{1}{N}\Phi'\Sigma_{ee}^{-1}e_t + O_p\Big(\frac{1}{\sqrt{NT}}\Big) + O_p\Big(\frac{1}{T}\Big),$$

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 $\bar{\mathcal{H}} = \lim\limits_{N\to\infty} \mathcal{H}$ is defined in Assumption C′.

7 EM algorithm

The ML estimation can be implemented via an Expectation-Maximization (EM) algorithm. The EM algorithm is an iteration approach. In this section, we present iterating formulas of the EM algorithm for both the pure constrained factor model case considered in Section 3 and the partially constrained factor model case considered in Section 6.

7.1 EM algorithm for the pure constrained factor model

In this subsection, we provide the iterating formulas of the 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

$$\Lambda^{(k+1)} = (M'\Sigma_{ee}^{(k)-1}M)^{-1}\left[M'\Sigma_{ee}^{(k)-1}\frac{1}{T}\sum_{t=1}^{T}E(z_t f_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(f_t f_t'|Z,\theta^{(k)})\right]^{-1},$$

$$\begin{aligned}\operatorname{diag}(\Sigma_{ee}^{(k+1)}) = \operatorname{diag}\Big\{ & M_{zz} - \frac{2}{T}\sum_{t=1}^{T}E(z_t f_t'|Z,\theta^{(k)})\Lambda^{(k+1)\prime}M' \\ & + M\Lambda^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(f_t f_t'|Z,\theta^{(k)})\Lambda^{(k+1)\prime}M'\Big\},\end{aligned}$$

where $\Sigma_{zz}^{(k)} = M\Lambda^{(k)}\Lambda^{(k)\prime}M' + \Sigma_{ee}^{(k)}$ and

$$\frac{1}{T}\sum_{t=1}^{T}E(f_t f_t'|Z,\theta^{(k)}) = \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M_{zz}(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)} + I_r - \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)},$$

$$\frac{1}{T}\sum_{t=1}^{T}E(z_t f_t'|Z,\theta^{(k)}) = M_{zz}(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)}.$$

The above iteration continues until $\|\theta^{(k+1)} - \theta^{(k)}\|$ is smaller than a preset tolerance. For the initial value of the iteration, we use the PC estimates proposed in Tsai and Tsay (2010) for the constrained factor model. One thing to mention is that once we get the estimates of the final round of iteration, denoted as $(\Lambda^\dagger, \Sigma_{ee}^\dagger)$, normalization is needed to transfer them to satisfy the identification conditions imposed in our paper, i.e. IC in Section 3. Such normalization is defined in a similar way as in Bai and Li (2012), details are following. Let $V^\dagger$ be the orthogonal matrix consisting of the eigenvectors of the matrix $\frac{1}{N}\Lambda^{\dagger\prime}M'(\Sigma_{ee}^\dagger)^{-1}M\Lambda^\dagger$ associated to its eigenvalues arranged in a descending order. Calculate $\hat{\Lambda} = \Lambda^\dagger V^\dagger$ and simply let $\hat{\Sigma}_{ee} = \Sigma_{ee}^\dagger$. Then $\hat{\theta} = (\hat{\Lambda}, \hat{\Sigma}_{ee})$ satisfies IC.

We can show that the limit of the iterated EM solutions satisfy the first order conditions in (3.3) and (3.4), and are stationary points of our objective function. The proof would be similar to Section E in the Supplement of Bai and Li (2012) with slight modification from unconstrained factor models to constrained factor models, and hence omitted in this paper.

7.2 EM algorithm for the partially constrained factor model

The iterating formulas of the EM algorithm for the partially constrained factor model are given in this subsection. 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

$$\Lambda^{(k+1)} = (M'\Sigma_{ee}^{(k)-1}M)^{-1}\left[M'\Sigma_{ee}^{(k)-1}\frac{1}{T}\sum_{t=1}^{T}E(z_t f_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(f_t f_t'|Z,\theta^{(k)})\right]^{-1}$$

$$
- (M'\Sigma_{ee}^{(k)-1}M)^{-1}\left[M'\Sigma_{ee}^{(k)-1}\Gamma^{(k)}\frac{1}{T}\sum_{t=1}^{T}E(g_tf_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)})\right]^{-1},
$$

$$
\Gamma^{(k+1)} = \left[\frac{1}{T}\sum_{t=1}^{T}E(z_tg_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(g_tg_t'|Z,\theta^{(k)})\right]^{-1}
$$

$$
- \left[M\Lambda^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(f_tg_t'|Z,\theta^{(k)})\right]\left[\frac{1}{T}\sum_{t=1}^{T}E(g_tg_t'|Z,\theta^{(k)})\right]^{-1},
$$

$$
\begin{aligned}
\operatorname{diag}(\Sigma_{ee}^{(k+1)}) = \operatorname{diag}\bigg\{ & M_{zz} - \frac{2}{T}\sum_{t=1}^{T}E(z_tf_t'|Z,\theta^{(k)})\Lambda^{(k+1)\prime}M' - \frac{2}{T}\sum_{t=1}^{T}E(z_tg_t'|Z,\theta^{(k)})\Gamma^{(k+1)\prime} \\
& + M\Lambda^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)})\Lambda^{(k+1)\prime}M' + \Gamma^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(g_tg_t'|Z,\theta^{(k)})\Gamma^{(k+1)\prime} \\
& + 2M\Lambda^{(k+1)}\frac{1}{T}\sum_{t=1}^{T}E(f_tg_t'|Z,\theta^{(k)})\Gamma^{(k+1)\prime}\bigg\},
\end{aligned}
$$

where $\Sigma_{zz}^{(k)} = M\Lambda^{(k)}\Lambda^{(k)\prime}M' + \Gamma^{(k)}\Gamma^{(k)\prime} + \Sigma_{ee}^{(k)}$ and

$$
\frac{1}{T}\sum_{t=1}^{T}E(f_tf_t'|Z,\theta^{(k)}) = \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M_{zz}(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)} + I_{r_1} - \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)},
$$

$$
\frac{1}{T}\sum_{t=1}^{T}E(f_tg_t'|Z,\theta^{(k)}) = \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}M_{zz}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)} - \Lambda^{(k)\prime}M'(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)},
$$

$$
\frac{1}{T}\sum_{t=1}^{T}E(g_tg_t'|Z,\theta^{(k)}) = \Gamma^{(k)\prime}(\Sigma_{zz}^{(k)})^{-1}M_{zz}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)} + I_{r_2} - \Gamma^{(k)\prime}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)},
$$

$$
\frac{1}{T}\sum_{t=1}^{T}E(z_tf_t'|Z,\theta^{(k)}) = M_{zz}(\Sigma_{zz}^{(k)})^{-1}M\Lambda^{(k)},
$$

$$
\frac{1}{T}\sum_{t=1}^{T}E(z_tg_t'|Z,\theta^{(k)}) = M_{zz}(\Sigma_{zz}^{(k)})^{-1}\Gamma^{(k)}.
$$

Similar to the procedure in Section 7.1, we use the PC estimates as the starting value, and iterate the above formulas until $\|\theta^{(k+1)} - \theta^{(k)}\|$ is smaller than a preset tolerance, and denote the estimates from the final round of iteration as $\theta^{\diamond} = (\Lambda^{\diamond}, \Gamma^{\diamond}, \Sigma_{ee}^{\diamond})$. Finally we need transfer $\theta^{\diamond}$ to satisfy the IC$''$ by the following normalization. Let $V^{\diamond}$ be the orthogonal matrix consisting of the eigenvectors of the matrix $\frac{1}{N}\Phi^{\diamond\prime}(\Sigma_{ee}^{\diamond})^{-1}\Phi^{\diamond}$ associated to its eigenvalues arranged in a descending order, where $\Phi^{\diamond} = (M\Lambda^{\diamond}, \Gamma^{\diamond})$. Compute $\Phi^{\diamond}V^{\diamond}$, and denote it as $\Phi^{\triangle} = (\Phi_1^{\triangle}, \Phi_2^{\triangle})$ with $\Phi_1^{\triangle}$ being the left $N \times r_1$ subblock and $\Phi_2^{\triangle}$ being the right $N \times r_2$ subblock. 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 $\hat{\theta} = (\hat{\Lambda}, \hat{\Gamma}, \hat{\Sigma}_{ee})$ satisfies IC$''$.

Again, we can show that the limit of the iterated EM solutions satisfy the first order conditions in (6.2), (6.3) and (6.4). The proof is similar to the pure constrained factor model case and therefore skipped here.

8 Simulation results

In this section, we run simulations to investigate the finite sample performance of the MLE, the empirical size and power of the W test.

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 (2002), Doz, Giannone and Reichlin (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 4.1. The calculation formulas based on S simulations are as follows

1. MAD $= \frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{kr}\sum_{p=1}^{k}\sum_{i=1}^{r}|\hat{\Lambda}_{pi}^{s} - \Lambda_{pi}^{s}|\right)$,
2. RMSE $= \sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\frac{1}{kr}\sum_{p=1}^{k}\sum_{i=1}^{r}(\hat{\Lambda}_{pi}^{s} - \Lambda_{pi}^{s})^2\right)}$,
3. RAvar $= \sqrt{NT} \times$ RMSE.

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 with mean zero and variance one. For the values of k and r, we consider two cases: $(k,r) = (3,1)$ and $(k,r) = (8,3)$.

Throughout the whole 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 method above to estimate it. Tables 1 and 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 errors and chi-square errors are almost the same as those for normal errors and are given in Table E1-E4 in Appendix E for space sake. All these results are obtained based on 1000 repetitions.

From Tables 1 and 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 (the root asymptotic variance), we see that the MLE has constant RAvar 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 4. In addition, it suggests that the PC estimate also has $\sqrt{NT}$ convergence rate. Finally, we note that the MLE's RAvar is smaller than PC's RAvar, indicating that the MLE is more efficient than the PC estimate.

Table 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	2.1467	0.0943	0.1386	4.1580
50	30	0.0349	0.0540	2.0908	0.0654	0.0934	3.6155
100	30	0.0262	0.0417	2.2833	0.0474	0.0677	3.7090
150	30	0.0216	0.0340	2.2792	0.0410	0.0582	3.9035
30	50	0.0333	0.0533	2.0629	0.0787	0.1145	4.4330
50	50	0.0237	0.0368	1.8408	0.0546	0.0800	4.0018
100	50	0.0190	0.0306	2.1663	0.0375	0.0541	3.8273
150	50	0.0159	0.0255	2.2092	0.0293	0.0417	3.6084
30	100	0.0232	0.0374	2.0492	0.0674	0.0964	5.2793
50	100	0.0172	0.0263	1.8626	0.0443	0.0611	4.3191
100	100	0.0105	0.0168	1.6771	0.0253	0.0358	3.5843
150	100	0.0102	0.0165	2.0226	0.0200	0.0288	3.5242

Table 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	15.0188	0.5655	0.8071	24.2137
50	30	0.2307	0.3310	12.8213	0.3744	0.5363	20.7700
100	30	0.1537	0.2307	12.6352	0.2224	0.3131	17.1484
150	30	0.1245	0.1881	12.6169	0.1735	0.2452	16.4517
30	50	0.2637	0.3744	14.4999	0.5130	0.7521	29.1290
50	50	0.1794	0.2689	13.4453	0.3184	0.4679	23.3929
100	50	0.1082	0.1578	11.1594	0.1763	0.2545	17.9928
150	50	0.0860	0.1291	11.1797	0.1382	0.2091	18.1078
30	100	0.1846	0.2698	14.7778	0.4570	0.6882	37.6916
50	100	0.1213	0.1937	13.6960	0.2622	0.4064	28.7400
100	100	0.0774	0.1258	12.5832	0.1440	0.2157	21.5706
150	100	0.0620	0.1021	12.5041	0.1033	0.1633	19.9975

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, which is obtained from 1000 repetitions.

From the results in Table 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 closely under three different error distributions. This is consistent with the theoretical result in Section 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 N, we see that the W statistic performs well even when $N = 30$. The reason we conjecture is that when T is small, the variance σ_i^2 would be estimated poorly, which leads to a bad performance of W. In addition, we also consider the case $(k, r) = (8, 3)$. Overall, the performance of the W statistic deteriorates to some extent in this case but is still satisfactory. The results are available upon request.

Tsai and Tsay (2010) propose the traditional likelihood ratio (LR) statistic to perform the model specification testing. In factor model literature, the LR test is usually considered under the fixed-N, large-T setup, see Lawley and Maxwell (1971). As mentioned in the introduction part, 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 under the current data generating setup. We find that the performance of the W statistic dominates that of the LR one. Details are given in Appendix F in the supplementary material of this paper.

Table 3: The empirical size of the test statistic W for the case $(k, r) = (3, 1)$

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%

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 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 the 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 in the hypothesis. Directly deriving the local power region for

W is challenging. We conjecture that this local power region is of $O(N^{-1/4}T^{-1/2})$. The 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 4: The empirical power of the W test for the case $(k,r) = (3,1)$

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%

Table 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 power when α is smaller. This is not surprising. 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 in a diminishing region of order $N^{-1/4}T^{-1/2}$, we conclude that the W has good performance in terms of empirical power. We also make a comparison of the W statistic and the LR one on the empirical power and find that the overall performance of the W test is also better than that of the LR one. Details are given in Appendix F in the supplement.

9 Extension

In this section, we consider the same constrained factor model (3.1) but extending Assumption B to a more general weak dependence structure of the idiosyncratic errors following Bai and Ng (2002), Bai (2003) and Bai and Li (2016), which leads to the approximate factor structure of Chamberlain and Rothschild (1983). We introduce the following new assumptions. There exists a large enough constant C such that

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 $\mathbb{W} = \text{diag}(\mathbb{O})$ being the diagonal matrix that sets the off-diagonal elements of $\mathbb{O}$ to zero. 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.

Assumption B$''$ allows for heteroskedasticity and weak correlations in both cross section and time dimensions, which is more general than the strict factor structure under Assumption B considered in Section 3. Assumption B$''$.3 imposes the boundness of the time average variance w_i^2. Assumption B$''$.4 is used to control the magnitude of the cross-sectional correlation of e_{it}, while Assumption B$''$.5 and B$''$.6 are for the serial correlation.

As pointed out in Bai and Li (2016), under the approximate factor structure, there are a lot of free parameters included in $\mathbb{O}_t$, which are as many as the elements contained in the sample variance of the observations. So directly estimating $\mathbb{O}_t$ (together with the loadings and factors) is difficult, due to the incidental problem. Therefore, using the similar approach as in Bai and Li (2016), we estimate $\mathbb{W}$ which is the time average of $\mathbb{O}_t$, instead of $\mathbb{O}_t$ itself, to avoid the incidental problem.

To facilitate the theoretical analysis, we make more assumptions as following.

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 $[-C, C]$.

Assumption E$''$:

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 C$''$ and D$''$ are similar to Assumption C and D respectively in Section 3, with slight modification. Assumption E$''$ will be useful in deriving the limiting distribution of the MLE.

To remove the rotational indeterminacy in estimation, we impose the following identification conditions, similar to the IC in Section 3.

Identification condition, IC$''$:

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

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

In this extension, we are interested in estimating Λ and $\mathbb{W}$. Let $\theta = (\Lambda, \mathbb{W})$ and $\Sigma_{zz} = M\Lambda\Lambda' M + \mathbb{W}$, we consider the similar objective function as (3.2),

$$\mathcal{L}^\dagger(\theta) = -\frac{1}{2N}\ln|\Sigma_{zz}| - \frac{1}{2N}tr[M_{zz}\Sigma_{zz}^{-1}], \tag{9.1}$$

where $M_{zz} = T^{-1}\sum_{t=1}^T z_t z_t'$. Notice that now under the general weak dependence structure in Assumption B$''$, the above discrepancy function is no longer the likelihood function even when f_t are independently and normally distributed with mean zero and variance I_r, due to the cross-sectional and serial correlations involved in the errors e_{it}. We define the quasi-MLE (or just call it MLE for simplicity) as

$$\hat{\theta} = (\hat{\Lambda}, \hat{\mathbb{W}}) = \operatorname*{argmax}_{\theta\in\Theta} \mathcal{L}^\dagger(\theta),$$

where Θ is the parameter space specified by Assumption D$''$ and IC$''$.

Taking derivatives of (9.1) with respect to Λ and $\mathbb{W}$, we get the following first order conditions, which are similar to (3.3) and (3.4),

$$\hat{\Lambda}'M'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}M = 0; \tag{9.2}$$

$$\text{diag}(\hat{\Sigma}_{zz}^{-1}) = \text{diag}(\hat{\Sigma}_{zz}^{-1}M_{zz}\hat{\Sigma}_{zz}^{-1}), \tag{9.3}$$

where $\hat{\Sigma}_{zz} = M\hat{\Lambda}\hat{\Lambda}'M' + \hat{\mathbb{W}}$. Similar to the constrained factor model under Assumption B, the above two first order conditions are useful in deriving the asymptotic properties of the MLE, but will not be used in the computation of the MLE. Instead, we can use the EM algorithm to compute the MLE.

The following theorem presents the convergence rates of the MLE. The consistency is implied by the theorem.

Theorem 9.1 (Convergence rates) *Under Assumptions A,B$''$, C$''$ and D$''$, together with IC$''$, when $N, T \to \infty$, we have*

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

The above theorem implies the consistency of the MLE still holds under the weak cross-sectional and serial correlations imposed in Assumption B$''$. However, the limiting distributions of the MLE change, as shown in following theorems.

Compared to the results in Theorem 4.1 in Section 4, now under the weak dependence structure, there exits an extra term $O_p(\frac{1}{N})$ in $(\hat{\Lambda} - \Lambda)$ and another extra term $O_p(\frac{1}{N^2})$ in $\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2$. This finding implies that the MLE of loadings in the approximate constrained factor models will not be consistent under fixed N, but will become consistent under large N. This result is consistent with that in an approximate unconstrained factor model in Bai and Li (2016).

Before we state the asymptotic result of $\hat{\Lambda}$, we first introduce some symbols as below.

$$
\begin{aligned}
&\mathbb{D}_1^{\dagger} = \begin{bmatrix} 2D_r^{+} \\ \mathcal{D}[(\mathbb{P} \otimes I_r) + (I_r \otimes \mathbb{P})K_r] \end{bmatrix}, \\
&\mathbb{B}_1^{\dagger} = K_{kr}[(\mathbb{P}^{-1}\Lambda') \otimes \Lambda] + \mathbb{R}^{-1} \otimes I_r - K_{kr}(I_r \otimes \Lambda)(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_2[(\mathbb{P}^{-1}\Lambda') \otimes I_r], \\
&\mathbb{B}_2^{\dagger} = K_{kr}(I_r \otimes \Lambda)(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)', \quad \mathbb{B}_3^{\dagger} = K_{kr}(I_r \otimes \Lambda)(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_3, \\
&\mathbb{B}_4^{\dagger} = \Big((\mathbb{R}^{-1}) \otimes (\mathbb{P}^{-1}\Lambda')\Big) - \frac{1}{2}K_{kr}(I_r \otimes \Lambda)(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_2\Big((\mathbb{P}^{-1}\Lambda') \otimes (\mathbb{P}^{-1}\Lambda')\Big), \\
&\Delta^{\dagger} = \mathbb{B}_2^{\dagger}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\varpi_i^2}{w_i^6}(m_i \otimes m_i), \\
&\Pi^{\dagger} = \mathbb{B}_4^{\dagger}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N}\frac{\mathbb{O}_{ij}}{w_i^2 w_j^2}(m_j \otimes m_i) - \mathbb{B}_3^{\dagger}\frac{1}{N}\sum_{i=1}^{N}\frac{\varsigma_i}{w_i^4}(m_i \otimes m_i).
\end{aligned}
$$

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 $(\mathbb{O} - \mathbb{W})$; $\varpi_i^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[(e_{it}^2 - w_i^2)(e_{is}^2 - w_i^2)]$ 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 9.2 (Asymptotic representation for $\hat{\Lambda}$) *Under assumptions of Theorem 9.1,*

$$
\begin{aligned}
\mathrm{vec}(\hat{\Lambda}' - \Lambda') = \mathbb{B}_1^{\dagger}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i \otimes f_t)e_{it} - \mathbb{B}_2^{\dagger}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(m_i \otimes m_i)(e_{it}^2 - w_i^2) \\
+ \frac{1}{T}\Delta^{\dagger} + \frac{1}{N}\Pi^{\dagger} + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}), \quad (9.4)
\end{aligned}
$$

where the symbols $\mathbb{B}_1^{\dagger}, \mathbb{B}_2^{\dagger}, \Delta^{\dagger}$ and $\Pi^{\dagger}$ are defined in the preceding paragraph.

Compared to the result in Theorem 4.2, there exists an extra bias term $\frac{1}{N}\Pi^{\dagger}$ in the asymptotic representation of $\hat{\Lambda}$ under Assumption B$''$, which is of order $O_p(\frac{1}{N})$ and comes from the weak dependence structure imposed on the errors. The two leading terms (i.e. the first two terms) on the right hand side of (9.4) are similar as these in (4.1) in the strict factor structure case. As we can see from Theorem 4.3 and 4.4, under Assumption B, these two leading terms are asymptotically independent with each other, and hence converge to a normal distribution with a simple expression of its limiting variance. However, it becomes more complicated in the general weak dependence structure case, since these two leading terms are no longer asymptotically independent, and therefore converge to a normal distribution with a more complex limiting variance, as shown in the following theorem.

Theorem 9.3 (Limiting distribution for $\hat{\Lambda}$) *Under assumptions of Theorem 9.1 and Assumption E$''$, as $N, T \to \infty, N/T^2 \to 0$ and $T/N^3 \to 0$, we have*

$$\sqrt{NT}\Big[\mathrm{vec}(\hat{\Lambda}' - \Lambda') - \frac{1}{T}\Delta^\dagger - \frac{1}{N}\Pi^\dagger\Big] \xrightarrow{d} N(0, \Xi),$$

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 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$''$.

Notice that the limiting variance Ξ_{NT} now is much more complicated than Ω_N as defined in Theorem 4.2, due to the weak dependence structure on the errors.

Theorem 9.4 (Asymptotic properties for $\hat{w}_i^2$) *Under assumptions of Theorem 9.1,*

$$\hat{w}_i^2 - w_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}) + O_p(\frac{1}{N}).$$

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.

There are two things worth to address from the above theorem, comparing to the results in the strict factor structure case in Section 4. First, similar to Theorem 4.4, the limiting result for $\hat{w}_i^2$ in the above theorem is still independent with the identification condition IC$''$. Second, different from Theorem 4.4, now there exists an extra term $O_p(\frac{1}{N})$ in the asymptotic representation of $\hat{w}_i^2$, due to the weak dependence structure of the error, resulting in an extra rate condition $T/N^2 \to 0$ in order to derive the limiting distribution of it. In addition, the above limiting result is the same as that in an approximate unconstrained factor model in Bai and Li (2016).

10 Conclusion

This paper 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 converge much faster than the 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 new 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. The asymptotic theories of the corresponding MLE are provided. Monte carlo simulations show that our proposed MLE has better finite sample performances than that of PC estimates. We also run simulations to study the size and power of our proposed statistic, which imply our statistic works well in different cases and a variety of sample sizes. Simulation results are consistent with our theoretical analysis. In addition, we extend Assumption B to a more general weak dependence structure in Section 9 and study the MLE in this extension.

Appendix: Proofs of the theoretical results in Section 4

The following notations will be used in the following appendices.

$$\hat{P} = \frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}; \qquad \hat{R} = \frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M; \qquad \hat{G} = (I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1};$$
$$\hat{P}_N = N\cdot\hat{P} = \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}; \quad \hat{R}_N = N\cdot\hat{R} = M'\hat{\Sigma}_{ee}^{-1}M, \quad \hat{G}_N = N\cdot\hat{G}.$$

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

$$\Sigma_{zz}^{-1} = \Sigma_{ee}^{-1} - \Sigma_{ee}^{-1}M\Lambda(I_r + \Lambda'M'\Sigma_{ee}^{-1}M\Lambda)^{-1}\Lambda'M'\Sigma_{ee}^{-1}. \tag{S.1}$$

It follows that

$$\hat{\Lambda}'M'\hat{\Sigma}_{zz}^{-1} = \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1} - \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1} = \hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}. \tag{S.2}$$

Appendix A: Proof for Proposition 4.1 (consistency)

In this section, 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 a parameter set such that Λ take values in a compact set and $C^{-2} \le \sigma_i^2 \le C^2$ 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 A.1 *Under assumptions of A-D, we have*

$$(a)\quad \sup_{\theta\in\Theta}\frac{1}{NT}\Big|\mathrm{tr}\Big[\Lambda^{*\prime}M'\Sigma_{zz}^{-1}\sum_{t=1}^{T}e_t f_t^{*\prime}\Big]\Big| \xrightarrow{p} 0;$$

$$(b)\quad \sup_{\theta\in\Theta}\frac{1}{NT}\Big|\mathrm{tr}\Big[\sum_{t=1}^{T}(e_t e_t' - \Sigma_{ee}^{*})\Sigma_{zz}^{-1}\Big]\Big| \xrightarrow{p} 0;$$

where $\theta^* = (\Lambda^*, \Sigma_{ee}^*)$ denotes the true parameters and $\Sigma_{zz} = M\Lambda\Lambda' M' + \Sigma_{ee}$.

Proof of Lemma A.1. First, we consider (a). Let m_{ip} be the (i,p)th element of M for $i = 1,\dots,N, p = 1,\dots,k$ and $\Lambda = [\lambda_1, \lambda_2, \dots, \lambda_k]'$. By equation (S.1), we have

$$\begin{aligned}\frac{1}{NT}\Lambda'^{*}M'\Sigma_{zz}^{-1}\sum_{t=1}^{T}e_t f_t^{*\prime} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big(\sum_{p=1}^{k}\lambda_p^* m_{ip}\Big)\frac{1}{\sigma_i^2}e_{it}f_t^{*\prime} \qquad (A.1)\\ -\Lambda^{*\prime}M'\Sigma_{ee}^{-1}M\Lambda(I_r + \Lambda' M'\Sigma_{ee}^{-1}M\Lambda)^{-1}&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Big(\sum_{p=1}^{k}\lambda_p m_{ip}\Big)\frac{1}{\sigma_i^2}e_{it}f_t^{*\prime}.\end{aligned}$$

By the Cauchy-Schwartz inequality, the first term on the right side of (A.1) 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 (A.1) is $o_p(1)$ uniformly on θ. For the second part, we rewrite it in terms of P_N as

$$\Lambda^{*\prime}M'\Sigma_{ee}^{-1}M\Lambda P_N^{-1/2}(P_N^{-1} + I_r)^{-1}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}P_N^{-1/2}\Big(\sum_{p=1}^{k}\lambda_p m_{ip}\Big)\frac{1}{\sigma_i^2}e_{it}f_t^{*\prime}. \qquad (A.2)$$

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

$$\begin{aligned}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\Big\|P_N^{-1/2}\sum_{p=1}^{k}\lambda_p m_{ip}\Big\|^2 &= \sum_{i=1}^{N}\frac{1}{\sigma_i^2}\Big(\sum_{p=1}^{k}\lambda_p' m_{ip}P_N^{-1}\sum_{q=1}^{k}\lambda_q m_{iq}\Big)\\ &= \mathrm{tr}\Big[P_N^{-1}\Lambda' M'\Sigma_{ee}^{-1}M\Lambda\Big] = \mathrm{tr}[P_N^{-1}P_N] = r.\end{aligned} \qquad (A.3)$$

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 (A.3). In addition, term $(P_N^{-1}+I_r)^{-1}=O_p(1)$ uniformly on Θ. So the expression in (A.2) is $O_p(T^{-1/2})$ uniformly on θ. Then result (a) follows.

Next, we consider (b). By equation (S.1), we have

$$\begin{aligned}
&\operatorname{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\Sigma_{zz}^{-1}\Big]\\
&=\operatorname{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)(\Sigma_{ee}^{-1}-\Sigma_{ee}^{-1}M\Lambda(I_r+\Lambda'M'\Sigma_{ee}^{-1}M\Lambda)^{-1}\Lambda'M'\Sigma_{ee}^{-1})\Big]\\
&=\operatorname{tr}\big[\frac{1}{NT}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}^*)\Sigma_{ee}^{-1}\big]\\
&\quad-\operatorname{tr}\Big[\frac{1}{NT}\sum_{t=1}^{T}(\Lambda'M'\Sigma_{ee}^{-1}(e_te_t'-\Sigma_{ee}^*)\Sigma_{ee}^{-1}M\Lambda)(I_r+\Lambda'M'\Sigma_{ee}^{-1}M\Lambda)^{-1}\Big].
\end{aligned}$$

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

$$C^2\Big[\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\|P_N^{-1/2}\sum_{p=1}^{k}\lambda_pm_{ip}\|^2\Big]\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)^2\Big]^{1/2}$$

which is $O_p(T^{-1/2})$ uniformly on θ by (A.3). 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). $\square$

Lemma A.2 *Under Assumptions A-D, we show*

(*a*) $\Big\|\frac{1}{N}\Lambda^{*\prime}M'(\hat\Sigma_{ee}^{-1}-\Sigma_{ee}^{*-1})M\Lambda^*\Big\|=O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat\sigma_i^2-\sigma_i^{*2})^2\Big]^{\frac{1}{2}}\Big);$

(b) $\left\|\frac{1}{N}M'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{*-1})M\right\| = O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^{*2})^2\Big]^{\frac{1}{2}}\Big).$

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) $\hat{R}_N = O_p(N), \quad \hat{R} = \frac{1}{N}\hat{R}_N = O_p(1);$

(d) $\|\hat{R}^{-1/2}\| = O_p(1).$

where $\hat{R}$ *and* $\hat{R}_N$ *are defined above appendix A.*

Proof of Lemma A.2. We first consider (a). The left hand side of (a) can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{p=1}^{k}\lambda_p^* m_{ip}\Big)\Big(\sum_{q=1}^{k}m_{qi}\lambda_q^{*\prime}\Big)\frac{\hat{\sigma}_i^2-\sigma_i^{*2}}{\hat{\sigma}_i^2\sigma_i^2},$$

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 $\|\hat{R}^{-1/2}\|^2 = \mathrm{tr}(\hat{R}^{-1}) = O_p(1)$ by $\hat{R}\xrightarrow{p}R>0$ from result (c).

This completes the proof of Lemma A.2. □

Lemma A.3 *Under Assumptions A-D, we have*

(a) $\frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}^{-1} = \|\hat{P}^{-1/2}\|^2\cdot O_p(T^{-1/2});$

(b) $\frac{1}{N}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t' = \|\hat{P}^{-1/2}\|\cdot O_p(T^{-1/2});$

(c) $\frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}^{-1} = \|\hat{P}_N^{-1}\|\cdot O_p(1);$

(d) $\frac{1}{NT}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}^{-1} = O_p(T^{-1/2});$

$$(e)\quad \frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_t'-\Sigma_{ee}]\hat{\Sigma}_{ee}^{-1}M\hat{R}^{-1}=\|\hat{P}^{-1/2}\|\cdot O_p(T^{-1/2});$$

$$(f)\quad \frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}^{-1}=\|\hat{P}^{-1/2}\|\cdot O_p\Big(\Big[\frac{1}{N^3}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{\frac{1}{2}}\Big).$$

Proof of Lemma A.3. We first consider (a). The left hand side can be rewritten as

$$\frac{1}{N^2}\hat{P}^{-1/2}\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{P}^{-1/2}\Big(\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\Big)\frac{1}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big(\sum_{q=1}^{k}m_{jq}\hat{\lambda}_q'\Big)\hat{P}^{-1/2}\Big]\hat{P}^{-1/2},$$

which is bounded in norm by

$$C^2\|\hat{P}^{-1/2}\|^2\Big[\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\Big\|^2\Big]\Big[\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big|^2\Big]^{1/2},$$

which is $\|\hat{P}^{-1/2}\|^2\cdot O_p(T^{-1/2})$ by (A.3). Thus, (a) follows.

Next, we consider (b). The left hand side can be rewritten as

$$\frac{1}{\sqrt{N}}\hat{P}^{-1/2}\sum_{i=1}^{N}\hat{P}_N^{-1/2}\frac{1}{\hat{\sigma}_i^2}\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\frac{1}{T}\sum_{t=1}^{T}e_{it}f_t',$$

which is bounded in norm by

$$C\|\hat{P}^{-1/2}\|\Big(\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\Big\|^2\Big)^{1/2}\Big(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}e_{it}f_t'\Big\|^2\Big)^{1/2},$$

which is $\|\hat{P}^{-1/2}\|\cdot O_p(T^{-1/2})$ by (A.3). This proves result (b).

To prove result (c), notice that $\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})$ is bounded by $2C^4I_N$ by $C^{-2}\le\hat{\sigma}_i^2\le C^2$ and $C^{-2}\le\sigma_i^2\le C^2$. So the left hand side is bounded in norm by

$$\Big\|\hat{P}_N^{-1}\hat{\Lambda}'M'\Big(2C^4I_N\Big)\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}\Big\|=2C^4\|\hat{P}_N^{-1}\|.$$

Result (c) then follows.

We now consider (d). The left hand side is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_te_{it}m_i'\hat{R},$$

which is bounded in norm by

$$C\|\hat{R}\|\cdot\Big[\frac{1}{N}\sum_{i=1}^{N}\|m_i\|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big\|^2\Big]^{1/2},$$

which is $O_p(T^{-1/2})$ by Lemma A.2(c) and Assumption C. Hence, result (d) follows.

For result (e), the left hand side is equal to

$$\frac{1}{N^{3/2}}\hat{P}^{-1/2}\Big[\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{P}_N^{-1/2}\Big(\sum_{p=1}^{k}\hat{\lambda}_pm_{ip}\Big)\frac{1}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]m_j'\Big]\hat{R}^{-1},$$

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 (A.3) and Lemma A.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 (A.3) and Lemma A.2(c), gives result (f). $\square$

Proof of Proposition 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

$$R(\theta) = -2\frac{1}{NT}\mathrm{tr}\Big[M\Lambda^*\sum_{t=1}^{T}f_t^*e_t'\Sigma_{zz}^{-1}\Big] - \frac{1}{NT}\mathrm{tr}\Big[\sum_{t=1}^{T}(e_te_t' - \Sigma_{ee}^*)\Sigma_{zz}^{-1}\Big].$$

By Lemma A.1, we have $\sup_\theta|R(\theta)| = o_p(1)$. Since $\hat{\theta}$ maximizes $L(\theta)$, it follows $\overline{L}(\hat{\theta}) + R(\hat{\theta})) \ge \overline{L}(\theta^*) + R(\theta^*)$. This implies that $\overline{L}(\hat{\theta}) \ge \overline{L}(\theta^*) + R(\theta^*) - R(\hat{\theta}) \ge \overline{L}(\theta^*) - 2\sup_{\theta\in\Theta}|R(\theta)| = -|o_p(1)|$, where $\overline{L}(\theta^*)$ is normalized to be zero.

Now consider $\overline{L}(\hat{\theta})$ which is equivalent to

$$\overline{L}(\hat{\theta}) = -\frac{1}{N}\ln|\hat{\Sigma}_{zz}| - \frac{1}{N}\mathrm{tr}(\Sigma^*_{zz}\hat{\Sigma}^{-1}_{zz}) + 1 + \frac{1}{N}\ln|\Sigma^*_{zz}|. \tag{A.4}$$

By $\Sigma_{zz} = M\Lambda\Lambda'M' + \Sigma_{ee}$, we have $|\Sigma_{zz}| = |\Sigma_{ee}| \cdot |I_r + \Lambda'M'\Sigma^{-1}_{ee}M\Lambda|$. Similarly, $|\Sigma^*_{zz}| = |\Sigma^*_{ee}| \cdot |I_r + \Lambda^{*'}M'\Sigma^{*-1}_{ee}M\Lambda^*|$. Then equation (A.4) can be written as

$$\begin{aligned}
\overline{L}(\hat{\theta}) = & -\frac{1}{N}\ln|\hat{\Sigma}_{ee}| - \frac{1}{N}\ln|I_r + \hat{\Lambda}'M'\Sigma^{-1}_{ee}M\hat{\Lambda}| - \frac{1}{N}\mathrm{tr}[M\Lambda^*\Lambda^{*\prime}M'\hat{\Sigma}^{-1}_{zz}] \\
& - \frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{zz}] + \frac{1}{N}\ln|\Sigma^*_{ee}| + \frac{1}{N}\ln|I_r + \Lambda^{*\prime}M'\Sigma^{*-1}_{ee}M\Lambda^*| + 1 \\
= & \left\{-\frac{1}{N}\ln|\hat{\Sigma}_{ee}| + \frac{1}{N}\ln|\Sigma^*_{ee}| - \frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{zz}] + 1\right\} \\
& + \left\{-\frac{1}{N}\mathrm{tr}[M\Lambda^*\Lambda^{*\prime}M'\hat{\Sigma}^{-1}_{zz}]\right\} + \left\{-\frac{1}{N}\ln|I_r + \hat{\Lambda}'M'\hat{\Sigma}^{-1}_{ee}M\hat{\Lambda}|\right\} \\
& + \left\{\frac{1}{N}\ln|I_r + \Lambda^{*\prime}M'\Sigma^{*-1}_{ee}M\Lambda^*|\right\}.
\end{aligned}$$

Notice that

$$\frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{zz}] = \frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{ee}] - \frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{ee}M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}^{-1}_{ee}] = \frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{ee}] + o_p(1)$$

by

$$0 < \frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{ee}M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}^{-1}_{ee}] \leq C\frac{1}{N}\mathrm{tr}[\hat{\Lambda}'M'\hat{\Sigma}^{-1}_{ee}M\hat{\Lambda}\hat{G}] \leq C\frac{r}{N},$$

where we use the fact that there exists a constant C such that $\Sigma^*_{ee}\ hat\Sigma^{-1}_{ee} \leq C \cdot I_N$ due to the boundedness of $\hat{\sigma}^2_i$ and σ^{*2}_i.

Given the above result, together with $\frac{1}{N}\ln|I_r + \Lambda^{*\prime}M'\Sigma^{*-1}_{ee}M\Lambda^*| = O(\ln N/N)$, we can further write $\overline{L}(\hat{\theta})$ as

$$\begin{aligned}
\overline{L}(\hat{\theta}) = & -\left\{\frac{1}{N}\ln|\hat{\Sigma}_{ee}| - \frac{1}{N}\ln|\Sigma^*_{ee}| + \frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{ee}] - 1\right\} \\
& - \left\{\frac{1}{N}\mathrm{tr}[M\Lambda^*\Lambda^{*\prime}M'\hat{\Sigma}^{-1}_{zz}]\right\} - \left\{\frac{1}{N}\ln|I_r + \hat{\Lambda}'M'\hat{\Sigma}^{-1}_{ee}M\hat{\Lambda}|\right\} + o_p(1).
\end{aligned}$$

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,

$$\frac{1}{N}\ln|\hat{\Sigma}_{ee}| - \frac{1}{N}\ln|\Sigma^*_{ee}| + \frac{1}{N}\mathrm{tr}[\Sigma^*_{ee}\hat{\Sigma}^{-1}_{ee}] - 1 \xrightarrow{p} 0, \tag{A.5}$$

$$\frac{1}{N}\mathrm{tr}[M\Lambda^*\Lambda^{*\prime}M'\hat{\Sigma}^{-1}_{zz}] \xrightarrow{p} 0. \tag{A.6}$$

Equation (A.5) is equivalent to

$$\frac{1}{N}\sum_{i=1}^{N}(\ln\hat{\sigma}^2_i - \ln\sigma^{*2}_i + \frac{\sigma^{*2}_i}{\hat{\sigma}^2_i} - 1) \xrightarrow{p} 0.$$

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

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^{*2})^2 \xrightarrow{p} 0. \tag{A.7}$$

This proves the first result of Proposition 4.1.

Next, we consider (A.6), which is equivalent to

$$\frac{1}{N}\mathrm{tr}(M\Lambda^*\Lambda'^*M'\hat{\Sigma}_{zz}^{-1}) = \frac{1}{N}\mathrm{tr}\Big[\Lambda'^*M'(\hat{\Sigma}_{ee}^{-1} - \hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1})M\Lambda^*\Big].$$

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

$$\begin{aligned}
&\frac{1}{N}\mathrm{tr}(M\Lambda^*\Lambda'^*M'\hat{\Sigma}_{zz}^{-1}) \\
=&\frac{1}{N}\mathrm{tr}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^* - \Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big] \\
&+ \frac{1}{N}\mathrm{tr}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}(I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big]
\end{aligned}$$

Both terms on the right hand side are non-negative. By (A.6), it follows that

$$\frac{1}{N}\mathrm{tr}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^* - \Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big] \xrightarrow{p} 0, \tag{A.8}$$

$$\frac{1}{N}\mathrm{tr}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}(I_r + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big] \xrightarrow{p} 0. \tag{A.9}$$

By (A.7) and Lemma A.2(a), we know $\frac{1}{N}\mathrm{tr}(\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*)$ converges to a positive constant. Then (A.8) implies that $\frac{1}{N}\mathrm{tr}(\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*)$ converges to the same positive constant. Together with (A.9), 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

$$\hat{G} = o_p(1); \qquad \hat{P}_N^{-1} = o_p(1). \tag{A.10}$$

Consider (A.8) 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

$$\frac{1}{N}\Big[\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\Lambda^* - \Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Big] \xrightarrow{p} 0. \tag{A.11}$$

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 (A.11) 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 A.2(a) and (A.7), 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,

$$\frac{1}{N}\Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^* - (I_r - A)\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(I_r - A)' \xrightarrow{p} 0. \tag{A.12}$$

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 (A.11) can be rewritten as

$$\frac{1}{N}(\hat{\Lambda} - \Lambda^*)'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda} - \Lambda^*) - A\Big(\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\Big)A' \xrightarrow{p} 0. \tag{A.13}$$

We now make use of the first-order conditions to proceed the proof. The first-order condition (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 (S.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

$$\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} = -\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}$$

$$+\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda^*\frac{1}{T}\sum_{t=1}^{T} f_t^* e_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}$$

$$+\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}\Lambda^{*\prime}M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} + \hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}.$$

With notations of $\hat{P}$ and A, we have

$$\begin{aligned} I_r &= (I_r - A)'(I_r - A) + \frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}^{-1} \\ &+(I_r - A)'\frac{1}{NT}\sum_{t=1}^{T} f_t^* e_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}^{-1} + \frac{1}{N}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}(I_r - A) \\ &-\frac{1}{N^2}\hat{P}^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}^{-1} = i_1 + i_2 + \cdots + i_5, \quad \text{say} \end{aligned} \tag{A.14}$$

Term i_2 is $\|\hat{P}^{-1/2}\|^2 \cdot O_p(T^{-1/2})$ by Lemma A.3(a). Term i_3 is $\|I - A\| \cdot \|\hat{P}^{-1/2}\| \cdot O_p(T^{-1/2})$ by Lemma A.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 A.3(c) and (A.10). Given these results, we have

$$I_r = (I - A)'(I - A) + \|\hat{P}^{-1/2}\|^2 O_p(T^{-1/2}) + \|I - A\| \cdot \|\hat{P}^{-1/2}\| \cdot O_p(T^{-1/2}) + o_p(1). \tag{A.15}$$

However, by the definition of $\hat{P}$, equation (A.12) 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 (A.15), we obtain $A = O_p(1)$. To see this, notice that the left hand side of equation (A.15) is bounded. Hence, if $A \neq O_p(1)$, then A is stochastically unbounded, the right hand side of (A.15) 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 (A.15), $\|\hat{P}^{-1/2}\| = O_p(1)$ and $A = O_p(1)$, we have

$$(I_r - A)'(I_r - A) - I_r \xrightarrow{p} 0.$$

And from (A.12),

$$\frac{1}{N}\Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^* - (I_r - A)\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}(I_r - A)' = o_p(1).$$

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)). By (S.2), we can simplify equation (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

$$\hat{\Lambda}' - \Lambda^{*\prime} = -A'\Lambda^{*\prime} + (I - A)'\frac{1}{T}\sum_{t=1}^{T} f_t^* e_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} + \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}\Lambda^{*\prime} \quad \text{(A.16)}$$
$$+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \Sigma_{ee}^*]\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} - \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee}^*)\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}.$$

By $A = o_p(1)$ and Lemma A.3 (d), we have that the first two terms are $o_p(1)$. By $\|\hat{P}^{-1}\| = O_p(1)$ and Lemma A.3 (b), the third term is $o_p(1)$. By $\|\hat{P}^{-1}\| = O_p(1)$ and Lemma A.3 (e), the fourth term is $o_p(1)$. By $\|\hat{P}^{-1}\| = O_p(1)$ and Lemma A.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 4.1. □

Corollary A.1 *Under Assumptions A-D,*

(*a*) $\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} - \frac{1}{N}\Lambda^{*\prime}M'\Sigma_{ee}^{*-1}M\Lambda^* = o_p(1)$;

(*b*) $\hat{P}_N = O_p(N)$, $\hat{P} = O_p(1)$, $\hat{G} = O_p(N^{-1})$, $\hat{G}_N = O_p(1)$;

(*c*) $\frac{1}{N}(\hat{\Lambda} - \Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda} = o_p(1)$.

Proof of Corollary A.1. Result (a) follows from equation (A.12) 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 by $\hat{P} = O_p(N)$ and $A = o_p(1)$. □

Appendix B: Proofs of Theorems 4.1, 4.2 and 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 4.1 and 4.2.

Lemma B.1 *Under Assumptions A-D,*

$$(a) \quad \hat{P}_N^{-1} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} \frac{1}{T} \sum_{t=1}^{T} (e_t e_t' - \Sigma_{ee}) \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} \hat{P}_N^{-1} = O_p(T^{-1/2});$$

$$(b) \quad \hat{P}_N^{-1} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} \frac{1}{T} \sum_{t=1}^{T} e_t f_t' = O_p(T^{-1/2});$$

$$(c) \quad \hat{P}_N^{-1} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} (\hat{\Sigma}_{ee} - \Sigma_{ee}) \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} \hat{P}_N^{-1} = \frac{1}{\sqrt{N}} O_p\Big(\Big[\frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_i^2 - \sigma_i^2)^2 \Big]^{\frac{1}{2}} \Big);$$

$$(d) \quad \frac{1}{T} \sum_{t=1}^{T} f_t e_t' \hat{\Sigma}_{ee}^{-1} M \hat{R}_N^{-1} = O_p(T^{-1/2});$$

$$(e) \quad \hat{P}_N^{-1} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} \frac{1}{T} \sum_{t=1}^{T} [e_t e_t' - \Sigma_{ee}] \hat{\Sigma}_{ee}^{-1} M \hat{R}_N^{-1} = O_p(T^{-1/2});$$

$$(f) \quad \hat{P}_N^{-1} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} (\hat{\Sigma}_{ee} - \Sigma_{ee}) \hat{\Sigma}_{ee}^{-1} M \hat{R}_N^{-1} = \frac{1}{\sqrt{N}} O_p\Big(\Big[\frac{1}{N} \sum_{i=1}^{N} (\hat{\sigma}_i^2 - \sigma_i^2)^2 \Big]^{\frac{1}{2}} \Big).$$

Proof of Lemma B.1. First, we consider (a). The left hand side is equal to

$$\hat{P}^{-1} \frac{1}{N^2} \Big[\sum_{i=1}^{N} \sum_{j=1}^{N} \Big(\sum_{p=1}^{k} \hat{\lambda}_p m_{ip} \Big) \frac{1}{\hat{\sigma}_i^2 \hat{\sigma}_j^2} \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - E(e_{it} e_{jt})] \Big(\sum_{q=1}^{k} m_{jq} \hat{\lambda}_q' \Big) \Big] \hat{P}^{-1},$$

which is bounded in norm by

$$C^2 \|\hat{P}^{-1}\|^2 \Big[\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2} \Big\| \sum_{p=1}^{k} \hat{\lambda}_p m_{ip} \Big\|^2 \Big] \Big[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \Big| \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - E(e_{it} e_{jt})] \Big|^2 \Big]^{1/2}.$$

However, by Corollary A.1(a), we have

$$\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\hat{\sigma}_i^2} \Big\| \sum_{p=1}^{k} \hat{\lambda}_p m_{ip} \Big\|^2 = \mathrm{tr}[\frac{1}{N} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda}] \xrightarrow{p} \mathrm{tr}[\frac{1}{N} \Lambda' M' \Sigma_{ee}^{-1} M \Lambda] = \mathrm{tr}(P). \quad \text{(B.1)}$$

By

$$E \Big[\frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \Big| \frac{1}{T} \sum_{t=1}^{T} [e_{it} e_{jt} - E(e_{it} e_{jt})] \Big|^2 \Big] = O(T^{-1}),$$

together with Corollary A.1(b) and (B.1), we have (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 (B.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

$$C^2\|\hat{P}_N^{-1/2}\|^2\cdot\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\Big\|\hat{P}_N^{-1/2}\Big(\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big)\Big\|^2(\hat{\sigma}_i^2-\sigma_i^2). \tag{B.2}$$

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 (A.3), this gives

$$\frac{1}{\hat{\sigma}_i}\Big\|\hat{P}_N^{-1/2}\sum_{p=1}^{k}\hat{\lambda}_p m_{ip}\Big\|\le\sqrt{r}.$$

Hence, expression in (B.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 B.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 B.2. Consider equation (A.14). In the proof of Proposition 4.1, we had shown $A = o_p(1)$. So term AA' is of a smaller order and hence negligible. With Lemma B.2 (a), (b) and (c), equation (A.14) can be simplified as

$$A + A' = O_p(T^{-1/2}) + o_p\Big(\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2\Big]^{\frac{1}{2}}\Big). \tag{B.3}$$

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

$$\text{Ndg}\Big\{\Lambda'(\frac{1}{N}M'\Sigma_{ee}^{-1}M)\Lambda - \hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda}\Big\} = 0,$$

where Ndg denotes the operator which sets the diagonal elements of its input to zeros. By adding and subtracting terms,

$$\text{Ndg}\Big\{(\hat{\Lambda} - \Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda} + \hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda} - \Lambda) \tag{B.4}$$

$$-(\hat{\Lambda} - \Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda} - \Lambda) + \Lambda'\Big[\frac{1}{N}M'(\hat{\Sigma}_{ee}^{-1} - \Sigma_{ee}^{-1})M\Big]\Lambda\Big\} = 0.$$

By Lemma A.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 A.2(a), we have

$$\text{Ndg}\Big\{(\hat{\Lambda} - \Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda} + \hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda} - \Lambda)\Big\} \tag{B.5}$$

$$= O_p(\|\hat{\Lambda} - \Lambda\|^2) + O_p([\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2]^{1/2}).$$

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 A.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 (B.5) as

$$\text{Ndg}\{AP + PA'\} = O_p(\|\hat{\Lambda} - \Lambda\|^2) + O_p([\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2]^{1/2}), \tag{B.6}$$

where $o_p(A)$ is discarded since it is of smaller order term.

Now equation (B.3) has $\frac{1}{2}r(r+1)$ restrictions and equation (B.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 4.1. We first consider the first order condition (3.4), which can be written as

$$\operatorname{diag}\Big\{(M_{zz}-\hat{\Sigma}_{zz})-(M_{zz}-\hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'-M\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(M_{zz}-\hat{\Sigma}_{zz})\Big\}=0,$$

where "diag" denotes the diagonal operator and $\hat{G}=(I_r+\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda})^{-1}$. By

$$M_{zz}=M\Lambda\Lambda'M'+\Sigma_{ee}+M\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_t'+\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'M'+\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee}),$$

with some algebra manipulations, we can further write the preceding equation as

$$\begin{aligned}
\hat{\sigma}_i^2-\sigma_i^2=&\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2+2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_{it}-2m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\\
&-2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i-2m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})]\\
&+m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i-2m_i'(\hat{\Lambda}-\Lambda)\hat{\Lambda}'m_i+2m_i'(\hat{\Lambda}-\Lambda)\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i\\
&+2m_i'\Lambda(\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i+2\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i.
\end{aligned}\tag{B.7}$$

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 facttor 2) is equal to

$$m_i'\hat{\Lambda}(I_N-\hat{G})\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_{it}=m_i'\hat{\Lambda}(I_N-\hat{G})(I-A)'\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\tag{B.8}$$

and the sum of the seventh and eighth terms is equal to $-2m_i'(\hat{\Lambda}-\Lambda)\hat{G}\hat{\Lambda}'m_i$. Define

$$\psi=\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1};\qquad \phi=\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}.$$

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 ψ, this term is equal to

$$\begin{aligned}
-2m_i'\Lambda&\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i+2m_i'\Lambda(\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i\\
&=-2m_i'\Lambda\psi(I_N-\hat{G})\hat{\Lambda}'m_i+2m_i'\Lambda A(I_N-\hat{G})\hat{\Lambda}'m_i\\
&=2m_i'\Lambda\psi\hat{G}\hat{\Lambda}'m_i-2m_i'\Lambda A\hat{G}\hat{\Lambda}'m_i-2m_i'\Lambda\psi(\hat{\Lambda}-\Lambda)'m_i+2m_i'\Lambda A(\hat{\Lambda}-\Lambda)'m_i\\
&\quad+m_i'\Lambda(A+A'-\psi-\psi')\Lambda'm_i.
\end{aligned}$$

Also, by (A.14), we have

$$A'+A=A'A+\phi+(I_r-A)'\psi+\psi'(I_r-A)-\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1},$$

or equivalently

$$A'+A-\psi-\psi'=A'A+\phi-A'\psi-\psi'A-\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}.$$

Thus, it follows that

$$-2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i+2m_i'\Lambda(\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i \tag{B.9}$$
$$=2m_i'\Lambda\psi\hat{G}\hat{\Lambda}'m_i-2m_i'\Lambda A\hat{G}\hat{\Lambda}'m_i-2m_i'\Lambda\psi(\hat{\Lambda}-\Lambda)'m_i+2m_i'\Lambda A(\hat{\Lambda}-\Lambda)'m_i-m_i'\Lambda A'A\Lambda'm_i$$
$$-m_i'\Lambda\phi\Lambda'm_i+2m_i'\Lambda A'\psi\Lambda'm_i+m_i'\Lambda\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}\Lambda'm_i.$$

Using (B.8) and (B.9), we can rewrite (B.7) as

$$\begin{aligned}\hat{\sigma}_i^2-\sigma_i^2=&\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)-2m_i'(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_te_{it}+2m_i'\hat{\Lambda}\hat{G}\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\\
&+2m_i'\hat{\Lambda}A'\frac{1}{T}\sum_{t=1}^{T}f_te_{it}-2m_i'\hat{\Lambda}\hat{G}A'\frac{1}{T}\sum_{t=1}^{T}f_te_{it}+2m_i'\Lambda\psi\hat{G}\hat{\Lambda}'m_i\\
&-2m_i'\Lambda A\hat{G}\hat{\Lambda}'m_i-2m_i'\Lambda\psi(\hat{\Lambda}-\Lambda)'m_i+2m_i'\Lambda A(\hat{\Lambda}-\Lambda)'m_i\\
&+m_i'\Lambda A'A\Lambda'm_i-2m_i'\Lambda A'\psi\Lambda'm_i-2m_i'(\hat{\Lambda}-\Lambda)\hat{G}\hat{\Lambda}'m_i+2\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i\\
&+m_i'\Lambda\phi\Lambda'm_i-m_i'\Lambda\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}\Lambda'm_i\\
&-2m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})]+m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i\\
=&\,a_{i,1}+a_{i,2}+\cdots+a_{i,17},\quad\text{say.}\end{aligned}\tag{B.10}$$

By the Cauchy-Schwartz inequality, we have

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\leq 17\frac{1}{N}\sum_{i=1}^{N}(\|a_{i,1}\|^2+\cdots+\|a_{i,17}\|^2).$$

The first term $N^{-1}\sum_{i=1}^{N}\|a_{1i}\|^2=O_p(T^{-1})$ by

$$E\Big[\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big|^2\Big]=O(T^{-1}).$$

The second term is bounded in norm by

$$4C^2\|\hat{\Lambda}-\Lambda\|^2\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big\|^2=o_p(T^{-1})$$

by $\hat{\Lambda}-\Lambda=o_p(1)$ and

$$E\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big\|^2\Big]=O(T^{-1}).$$

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 B.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 B.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

$$\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2 - \sigma_i^2)^2 = O_p(T^{-1}) + o_p(\|\hat{\Lambda} - \Lambda\|^2). \tag{B.11}$$

Next, we derive bounds for $\|\hat{\Lambda} - \Lambda\|^2$. By equation (A.16), together with Lemma B.1(b), (d), (e) and (f) and Lemma B.2, we have

$$\hat{\Lambda} - \Lambda = O_p(T^{-1/2}) + O_p([\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2 - \sigma_i^2)^2]^{1/2}). \tag{B.12}$$

Substituting equation (B.12) into (B.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 4.1. $\square$

To prove the first result of Theorem 4.1, we need the following lemmas.

Lemma B.3 *Under Assumptions A-D, we show*

$$\begin{aligned}
&(a)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T(e_te_t' - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1} \\
&\qquad = O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}); \\
&(b)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T e_tf_t' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}); \\
&(c)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1} = O_p(N^{-1}T^{-1/2}); \\
&(d)\quad \frac{1}{T}\sum_{t=1}^T f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}); \\
&(e)\quad \hat{P}_N^{-1}\hat{\Lambda}'\Big(M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T[e_te_t' - \Sigma_{ee}]\hat{\Sigma}_{ee}^{-1}M\Big)\hat{R}_N^{-1} \\
&\qquad = O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}); \\
&(f)\quad \hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee} - \Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} = O_p(N^{-1}T^{-1/2}).
\end{aligned}$$

Proof of Lemma B.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

$$\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j' \frac{1}{\hat{\sigma}_i^2\hat{\sigma}_j^2}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]$$

$$= \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j' \Big(\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\Big)\Big(\frac{1}{\hat{\sigma}_j^2} - \frac{1}{\sigma_j^2}\Big)\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j' \frac{1}{\sigma_i^2}\Big(\frac{1}{\hat{\sigma}_j^2} - \frac{1}{\sigma_j^2}\Big)\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j' \frac{1}{\sigma_j^2}\Big(\frac{1}{\hat{\sigma}_i^2} - \frac{1}{\sigma_i^2}\Big)\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} m_i m_j' \frac{1}{\sigma_i^2\sigma_j^2}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})].$$

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 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 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 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 B.4 *Under Assumptions A-D,*

$$A\equiv(\hat{\Lambda}-\Lambda)'M'\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})+O_p(\|\hat{\Lambda}-\Lambda\|^2).$$

Proof of Lemma B.4. Consider equation (A.14). Using the results in Lemma B.3 and the fact that $A'A$ is of smaller order term than A and therefore negligible, we have

$$A+A'=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T}).\tag{B.13}$$

Now consider the term $\frac{1}{N}\Lambda'M'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})M\Lambda$, which can be written as

$$\begin{aligned}\frac{1}{N}\Lambda'M'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})M\Lambda&=-\Lambda'\Big[\frac{1}{N}\sum_{i=1}^{N}m_im_i'\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2\sigma_i^2}\Big]\Lambda\\&=-\Lambda'\Big[\frac{1}{N}\sum_{i=1}^{N}m_im_i'\frac{\hat{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}\Big]\Lambda+\Lambda'\Big[\frac{1}{N}\sum_{i=1}^{N}m_im_i'\frac{(\hat{\sigma}_i^2-\sigma_i^2)^2}{\hat{\sigma}_i^2\sigma_i^4}\Big]\Lambda.\end{aligned}\tag{B.14}$$

The norm of the second expression on the right hand side of (B.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 (B.10) into the first expression on the right hand side of (B.14) and using the arguments before (B.11), one can show that the first expression is $O_p(\frac{1}{\sqrt{NT}})+o_p(\frac{1}{T})$. Hence, we have

$$\frac{1}{N}\Lambda'M'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})M\Lambda=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T}).\tag{B.15}$$

Now consider (B.4). Using the same arguments as in the derivation of (B.6) except that the result for $\frac{1}{N}\Lambda' M'(\hat{\Sigma}_{ee}^{-1} - \Sigma_{ee}^{-1})M\Lambda$ is given by (B.15) instead of $o_p([\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2]^{1/2})$, we have

$$\text{Ndg}\{AP + PA'\} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}) + O_p(\|\hat{\Lambda} - \Lambda\|^2). \tag{B.16}$$

Solving the equation system (B.13) and (B.16), we have

$$A = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}) + O_p(\|\hat{\Lambda} - \Lambda\|^2),$$

as asserted in this lemma. This proves Lemma B.4. $\square$

Proof of Theorem 4.1 (continued). Using the results in Lemma B.3 and Lemma B.4 and noticing that $\|\hat{\Lambda} - \Lambda\|^2$ is of smaller order than $\hat{\Lambda} - \Lambda$ and therefore negligible, we have from (A.16)

$$\hat{\Lambda} - \Lambda = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}),$$

as asserted by the first result of Theorem 4.1. This completes the proof of Theorem 4.1.

Corollary B.1 *Under Assumptions A-D,*

$$A \equiv (\hat{\Lambda} - \Lambda)' M' \hat{\Sigma}_{ee}^{-1} M \hat{\Lambda} \hat{P}_N^{-1} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}).$$

Corollary B.1 is a direct result of Lemma B.4 and Theorem 4.1.

Lemma B.5 *Under Assumptions A-D,*

$$(a)\quad \frac{1}{T}\sum_{t=1}^{T} f_t e_t' \hat{\Sigma}_{ee}^{-1} M \hat{R}_N^{-1} = \frac{1}{T}\sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R_N^{-1} + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});$$

$$(b)\quad \hat{P}_N^{-1}\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t f_t' = P_N^{-1}\Lambda' M' \Sigma_{ee}^{-1} \frac{1}{T}\sum_{t=1}^{T} e_t f_t' + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});$$

$$\begin{aligned}(c)\quad \frac{1}{N} M'(\hat{\Sigma}_{ee}^{-1} - \Sigma_{ee}^{-1})M = & -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4} m_i m_i'(e_{it}^2 - \sigma_i^2) + \frac{1}{NT}\sum_{i=1}^{N} m_i m_i' \frac{\kappa_{i,4} - \sigma_i^4}{\sigma_i^4} \\ & + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).\end{aligned}$$

Proof of Lemma B.5. Equation (B.10) can be written as

$$\hat{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - \sigma_i^2) + \mathcal{R}_i, \tag{B.17}$$

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

$$\mathcal{S}_i = -2m_i'(\hat{\Lambda} - \Lambda)\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}\hat{G}\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}A'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{G}A'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}$$

$$+ 2m_i'\Lambda\psi\hat{G}\hat{\Lambda}'m_i - 2m_i'\Lambda A\hat{G}\hat{\Lambda}'m_i - 2m_i'\Lambda\psi(\hat{\Lambda}-\Lambda)'m_i + 2m_i'\Lambda A(\hat{\Lambda}-\Lambda)'m_i$$

$$+ m_i'\Lambda A'A\Lambda'm_i - 2m_i'\Lambda A'\psi\Lambda'm_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{G}\hat{\Lambda}'m_i + 2\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'m_i$$

$$+ m_i'\Lambda\phi\Lambda'm_i - m_i'\Lambda\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{\Lambda}\hat{P}_N^{-1}\Lambda'm_i + m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i.$$

Given that $\psi = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Lemma B.3 (b), $\hat{\Lambda}-\Lambda = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Theorem 4.1, $A = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Corollary B.1, by the same arguments in the derivation of (B.10), we have

$$\frac{1}{N}\sum_{i=1}^N \mathcal{S}_i^2 = O_p(N^{-1}T^{-2}) + O_p(N^{-2}T^{-1}) + O_p(T^{-3}). \tag{B.18}$$

We now consider

$$\frac{1}{N}\sum_{i=1}^N \Big| m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T[e_te_{it} - E(e_te_{it})]\Big|^2,$$

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

$$\frac{1}{N}\sum_{i=1}^N\Big|m_i'\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T[e_te_{it} - E(e_te_{it})]\Big|^2 = O_p(\frac{1}{NT}) + O_p(\frac{1}{T^2}).$$

This results, together with (B.18), gives

$$\frac{1}{N}\sum_{i=1}^N\mathcal{R}_i^2 = O_p(\frac{1}{NT}) + O_p(\frac{1}{T^2}). \tag{B.19}$$

Notice that

$$\frac{1}{NT}\sum_{t=1}^{T} f_t e_t'\hat{\Sigma}_{ee}^{-1}M = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{\sigma}_i^2}f_t e_{it} m_i'$$
$$= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}f_t e_{it} m_i' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2\sigma_i^2}f_t e_{it} m_i'.$$

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

$$C^5\Big[\frac{1}{N}\sum_{i=1}^{N}\|\mathcal{R}_i\|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big\|^2\Big]^{1/2},$$

which is $O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2})$ by (B.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

$$C^5\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_t e_{it}\Big\|^2\cdot\Big\|\frac{1}{T}\sum_{t=1}^{T}e_{is}^2-\sigma_i^2\Big\|^2\Big]^{1/2},$$

which is $O_p(T^{-3/2})$. Given the above results, we have

$$\frac{1}{NT}\sum_{t=1}^{T} f_t e_t'\hat{\Sigma}_{ee}^{-1}M = \frac{1}{NT}\sum_{t=1}^{T} f_t e_t'\Sigma_{ee}^{-1}M + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}). \tag{B.20}$$

Given (B.20), together with $\hat{R} = R + O_p(T^{-1/2})$, we immediately obtain (a). Given (B.20), together with $\hat{P} = P + O_p(T^{-1/2})$ and $\hat{\Lambda} = \Lambda + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$, 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 (B.17) into this term, we obtain

$$i_1 = -\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}m_i m_i' = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}(e_{it}^2-\sigma_i^2)m_i m_i'$$
$$+2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\mathrm{tr}\Big[\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})]m_i'\Big]m_i m_i' - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\mathcal{S}_i m_i m_i'.$$

Consider the second expression. The (v,u) element of this expression $(v,u=1,\dots,k)$ is

$$\operatorname{tr}\Big[\frac{1}{N}\sum_{i=1}^{N}\hat{\Lambda}\hat{G}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})]\frac{1}{\sigma_i^4}m_i'm_{iv}m_{iu}\Big]$$

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 B.3(a). The third term is bounded by

$$C^6\Big[\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_i^2\Big]^{1/2}=O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2})$$

by (B.18). Hence, we have

$$i_1=-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}(e_{it}^2-\sigma_i^2)m_im_i'+O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{\sqrt{N}T})+O_p(\frac{1}{T^{3/2}}).$$

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

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2\sigma_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big]^2m_im_i'+2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2\sigma_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big]\mathcal{R}_im_im_i'+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2\sigma_i^4}\mathcal{R}_i^2m_im_i'.$$

We analyze the three terms at right-hand-side of the above equation one by one. The second term is bounded in norm by

$$2C^8\Big[\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\mathcal{R}_i^2\Big]^{1/2},$$

which is $O_p(N^{-1/2}T^{-1})$ by (B.19). The third term is bounded in norm by

$$C^8\frac{1}{N}\sum_{i=1}^{N}\mathcal{R}_i^2=O_p(\frac{1}{NT})+O_p(\frac{1}{T^2})$$

by (B.19). Finally, the first term can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^6}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big]^2m_im_i'-\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2\sigma_i^6}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big]^2m_im_i'$$

The first term of the above expression is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\frac{\kappa_{i,4}-\sigma_i^4}{\sigma_i^6}m_im_i'+O_p(N^{-1/2}T^{-1}).$$

The second term is bounded in norm by

$$C^{10}\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2-\sigma_i^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big|^4\Big]^{1/2}=O_p(T^{-3/2}).$$

Hence, we have

$$i_2 = \frac{1}{NT}\sum_{i=1}^N \frac{\kappa_{i,4}-\sigma_i^4}{\sigma_i^6} m_i m_i' + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).$$

Summarizing the results on i_1 and i_2, we have (c). $\square$

Proof of Theorem 4.2. We first derive the asymptotic behavior of A. Consider equation (A.14), using Lemma B.3 (a) and (f), Lemma B.5 (b) and Lemma B.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 $\text{vech}(B)$ 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 $\text{vech}(B + B') = 2D_r^+\text{vec}(B)$, for any $r \times r$ matrix B, we have

$$2D_r^+\text{vec}(A) = 2D_r^+\text{vec}(\eta) + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}). \tag{B.21}$$

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 (B.4). By Lemma B.5(c) and $\hat{\Lambda} - \Lambda = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Theorem 4.1, we have

$$\begin{aligned}
&\text{Ndg}\Big\{\hat{\Lambda}'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)(\hat{\Lambda} - \Lambda) + (\hat{\Lambda} - \Lambda)'(\frac{1}{N}M'\hat{\Sigma}_{ee}^{-1}M)\hat{\Lambda}\Big\} \\
&= \text{Ndg}\{\zeta - \mu\} + O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).
\end{aligned}$$

Using the same arguments in the derivation of (B.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 $\text{veck}(B)$ 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

$$\mathcal{D}[(P\otimes I_r)+(I_r\otimes P)K_r]\text{vec}(A) = \mathcal{D}\text{vec}(\zeta-\mu)+O_p(N^{-1}T^{-1/2})+O_p(N^{-1/2}T^{-1})+O_p(T^{-3/2}), \tag{B.22}$$

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 (B.21) and (B.22), we have

$$\begin{aligned}\begin{bmatrix} 2D_r^+ \\ \mathcal{D}[(P \otimes I_r) + (I_r \otimes P)K_r] \end{bmatrix} \text{vec}(A) = \begin{bmatrix} 2D_r^+\text{vec}(\eta) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{D}\text{vec}(\zeta) \end{bmatrix} - \begin{bmatrix} 0 \\ \mathcal{D}\text{vec}(\mu) \end{bmatrix} \\ + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).\end{aligned} \tag{B.23}$$

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

$$\mathbb{D}_1\text{vec}(A) = \mathbb{D}_2\text{vec}(\eta)+\mathbb{D}_3\text{vec}(\zeta)-\mathbb{D}_3\text{vec}(\mu)+O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T^{3/2}}). \tag{B.24}$$

Also, notice that

$$\text{vec}(\eta) = \text{vec}\Big[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2} f_t e_{it} m_i' \Lambda P^{-1}\Big] = (P^{-1}\Lambda' \otimes I_r)\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it}$$

$$\text{vec}(\zeta) = \text{vec}\Big[\Lambda'\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{m_i m_i'}{\sigma_i^4}(e_{it}^2 - \sigma_i^2)\Lambda\Big] = (\Lambda \otimes \Lambda)'\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^4}(m_i \otimes m_i)(e_{it}^2 - \sigma_i^2)$$

and

$$\text{vec}(\mu) = \text{vec}\Big[\Lambda'\frac{1}{NT}\sum_{i=1}^N \frac{\kappa_{i,4} - \sigma_i^4}{\sigma_i^6} m_i m_i' \Lambda\Big] = (\Lambda \otimes \Lambda)'\frac{1}{NT}\sum_{i=1}^N \frac{1}{\sigma_i^6}(m_i \otimes m_i)(\kappa_{i,4} - \sigma_i^4).$$

Given the above three results, we can rewrite (B.24) as

$$\begin{aligned}\text{vec}(A) = {} & \mathbb{D}_1^{-1}\mathbb{D}_2(P^{-1}\Lambda' \otimes I_r)\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}(m_i \otimes f_t)e_{it} \\ & + \mathbb{D}_1^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)'\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^4}(m_i \otimes m_i)(e_{it}^2 - \sigma_i^2) \\ & - \mathbb{D}_1^{-1}\mathbb{D}_3(\Lambda \otimes \Lambda)'\frac{1}{NT}\sum_{i=1}^N \frac{1}{\sigma_i^6}(m_i \otimes m_i)(\kappa_{i,4} - \sigma_i^4) \\ & + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).\end{aligned} \tag{B.25}$$

Consider equation (A.16). Using the results of Lemma B.5 (a) and (b) and Lemma B.3 (e) and (f), we have

$$\begin{aligned}\hat{\Lambda}' - \Lambda' = &-A'\Lambda' + \frac{1}{NT}\sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R^{-1} + P^{-1}\Lambda' \frac{1}{NT} M' \Sigma_{ee}^{-1} \sum_{t=1}^{T} e_t f_t' \Lambda' \\ &+ O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).\end{aligned} \tag{B.26}$$

Notice that

$$\begin{aligned}\text{vec}\Big[\frac{1}{NT}\sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R^{-1}\Big] &= \text{vec}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{\sigma_i^2} f_t e_{it} m_i' R^{-1}\Big] \\ &= (R^{-1} \otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i \otimes f_t) e_{it}\end{aligned}$$

and

$$\begin{aligned}\text{vec}\Big[P^{-1}\Lambda'\frac{1}{NT} M'\Sigma_{ee}^{-1}\sum_{t=1}^{T} e_t f_t'\Lambda'\Big] &= \text{vec}\Big[P^{-1}\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2} m_i e_{it} f_t'\Lambda'\Big] \\ &= K_{kr}\text{vec}\Big[\Lambda\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2} f_t e_{it} m_i' \Lambda P^{-1}\Big] \\ &= K_{kr}[(P^{-1}\Lambda')\otimes\Lambda]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it},\end{aligned}$$

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 (B.26), we have

$$\begin{aligned}\text{vec}(\hat{\Lambda}' - \Lambda') = &\Big[K_{kr}[(P^{-1}\Lambda')\otimes\Lambda] + R^{-1}\otimes I_r\Big]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it} \\ &- K_{kr}(I_r\otimes\Lambda)\text{vec}(A) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).\end{aligned} \tag{B.27}$$

Substituting (B.25) into (B.27),

$$\begin{aligned}\text{vec}(\hat{\Lambda}' - \Lambda') = &\mathbb{B}_1\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it} - \mathbb{B}_2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^4}(m_i\otimes m_i)(e_{it}^2 - \sigma_i^2) \\ &+ \frac{1}{T}\Delta + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}),\end{aligned} \tag{B.28}$$

where

$$\begin{aligned}\mathbb{B}_1 &= K_{kr}[(P^{-1}\Lambda')\otimes\Lambda] + R^{-1}\otimes I_r - K_{kr}(I_r\otimes\Lambda)\mathbb{D}_1^{-1}\mathbb{D}_2[(P^{-1}\Lambda')\otimes I_r], \\ \mathbb{B}_2 &= K_{kr}(I_r\otimes\Lambda)\mathbb{D}_1^{-1}\mathbb{D}_3(\Lambda\otimes\Lambda)', \\ \Delta &= \mathbb{B}_2\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^6}(m_i\otimes m_i)(\kappa_{i,4} - \sigma_i^4).\end{aligned}$$

Given the above results and by a Central Limit Theorem, we obtain as $N, T \to \infty$ and $N/T^2 \to 0$,

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

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 4.2. $\square$

Proof of Theorem 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 A.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 (B.10). Term a_3 can be shown to be $O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T})$ by equation (A.16). Then we have

$$\frac{1}{N}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}e_t = \frac{1}{N}\Lambda'M'\Sigma_{ee}^{-1}e_t + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}).$$

Therefore,

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

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 4.5. $\square$

References

Ahn, S.C. and Horenstein, A.R. (2003). Eigenvalue ratio test for the number of factors. *Econometrica*, 81 (3), 1203-1227.

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 Li, K. (2012). Statistical analysis of factor models of high dimension, *The Annals of Statistics*, 40 (1), 436–465.

Bai, J. and Li, K. (2016). Maximum likelihood estimation and inference of approximate factor mdoels of high dimension, *Review of Economcis and Statistics*, 98(2), 298-309.

Bai, J. and Ng, S. (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.

Carhart, M. M. (1997) On persistence of mutual fund performance. *Journal of Finance*, 52, 57-82.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, vol. 51(5), 1281-1304.

Connor, G. and Korajczyk, R.A. (1986). Performance measurement with the arbitrage pricing theory : A new framework for analysis. *Journal of Financial Economics*, vol. 15(3), 373-394.

Connor, G. and Korajczyk, R.A. (1988). Risk and return in an equilibrium APT: Application of a new test methodology. *Journal of Financial Economics*, vol. 21, 255-289.

Connor, G. and O. Linton (2007). Semiparametric estimation of a characteristic-based factor model of stock returns. *Journal of Empirical Finance*, vol. 14, 694-717.

Connor, G., M. Hagmann, and O. Linton (2012). Efficient semiparametric estimation of the Fama-French Model and extensions. *Econometrica*, vol. 80(2), 713-754.

Connor, G. and Korajczyk, R.A. (1988). Risk and return in an equilibrium APT : Application of a new test methodology. *Journal of Financial Economics*, vol. 21 (2), 255-289.

Doz, C., Giannone D., and Reichlin L. (2012). A quasi-maximum likelihood approach for large approximate dynamic factor models, *The Review of Economics and Statistics*, MIT Press, vol. 94(4), 1014-1024.

Fama, E.F. and French, K.R. (1993). Common risk factors in the returns to stocks and bonds. *Journal of Financial Economics*, vol. 33, 3-56.

Fan, J., Liao, Y., and Mincheva, M. (2011). High dimensional covariance matrix estimation in approximate factor models. *Annals of statistics*, 39(6), 3320.

Fan, J., Liao, Y., and Mincheva, M. (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 Shi X. (2014). Risks of large portfolios, *Journal of Econometrics*, forthcoming.

Fan, J., Liao Y. and Wang W. (2014). Projected principal component analysis in factor models, *manuscript*.

Goyal, A. and Santa-Clara, P. (2003). Idiosyncratic risk matters, *Journal of Finance*, 58, 975-1008.

Geweke, J. (1977). The dynamic factor analysis of economic time series, in: D.J. Aigner and A.S. Goldberger, eds., Latent Variables in Socio-Economic Models, (North-Holland, Amsterdam).

Heckman, J.J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411-482.

Jennrich, R.I. (1969). Asymptotic properties of non-linear Least Squares estimation. *Annals of Mathematical Statistics*, 40, 633-643.

Kose, A., C. Otrok, and Whiteman, C. (2003). International business cycles: world region and country specific factors. *American Economic Review*, 93:4, 1216-1239.

Lawley, D. N., and A. E., Maxwell (1971). *Factor analysis as a statitical method.* New York: Ameriacan Elsevier.

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.

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues, *Review of Economics and Statistics*, 92(4), 1004-1016.

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 thoery of capital asset pricing. *Journal of Economic Theory*, 13, 341-360.

Sargent, T.J., and C.A. Sims (1977). Business cycle modeling without pretending to have too much a-priori economic theory, in: C. Sims et al., eds., New Methods in Business Cycle Research (Federal Reserve Bank of Minneapolis, Minneapolis).

Stock, J.H. and Watson, M.W (2002). Forecasting using principal components from a large number of predictors, *Journal of the American Statistical Association*, 97, 1167-1179.

Tsai, H. and Tsay, R. (2010). Constrained factor models, *Journal of the American Statistical Association*, vol. 105(492), 1593-1605.

Tsai, H., Tsay R.S, Lin M.H and Cheng, C.W. (2016). Doubly constrained factor models with applications, *Statistica Sinica*, forthcoming.

Zhou, Z. and Matteson, D.S. (2015). Temporal and spatio-temporal models for ambulance demand, book chapter in H.Yang and E.K.Lee, eds, *Healthcare Data Analytics*, Wiley Series in Operations Research and Management Science.

SUPPLEMENTARY MATERIALS

This supplement includes Appendices C-G, where we provide detailed proofs for the theorems in Sections 5,6 and 9, and more simulation results in addition to Section 8.

Appendix C: Proof of Theorem 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

$$
\begin{aligned}
\frac{1}{N}(M\hat{\Lambda} - \hat{L})'\widetilde{\Sigma}_{ee}^{-1}(M\hat{\Lambda} - \hat{L}) &= \frac{1}{N}\Big[M(\hat{\Lambda} - \Lambda) - (\hat{L} - L)\Big]'\widetilde{\Sigma}_{ee}^{-1}\Big[M(\hat{\Lambda} - \Lambda) - (\hat{L} - L)\Big] \\
&= (\hat{\Lambda} - \Lambda)'\Big[\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}M\Big](\hat{\Lambda} - \Lambda) - (\hat{\Lambda} - \Lambda)'\Big[\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}(\hat{L} - L)\Big] \\
&-\Big[\frac{1}{N}(\hat{L} - L)'\widetilde{\Sigma}_{ee}^{-1}M\Big](\hat{\Lambda} - \Lambda) + \frac{1}{N}(\hat{L} - L)'\widetilde{\Sigma}_{ee}^{-1}(\hat{L} - L) = I_a - I_b - I_c + I_d, \quad \text{say}
\end{aligned}
$$

Consider the first term I_a. Notice that

$$
\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}M - \frac{1}{N}M'\Sigma_{ee}^{-1}M = o_p(1) \tag{C.1}
$$

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 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

$$
\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}(\hat{L} - L) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_i(\hat{l}_i - l_i)'. \tag{C.2}
$$

According to (A.14) in the supplement of Bai and Li (2012), we know that

$$
\begin{aligned}
\hat{l}_i - l_i &= (\hat{L} - L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L} - L)(\hat{L} - L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i \\
&-\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}L\Big(\frac{1}{T}\sum_{t=1}^{T}f_te_t'\Big)\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Big)L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i \\
&-\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}\hat{l}_i\hat{l}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt} - E(e_{it}e_{jt})]\Big)\hat{H}l_i + \hat{H}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^4}\hat{l}_i\hat{l}_i'(\tilde{\sigma}_i^2 - \sigma_i^2)\hat{H}l_i \\
&+\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Big)l_i + \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}L\Big(\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big) \\
&+\hat{H}\Big(\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}\hat{l}_j\frac{1}{T}\sum_{t=1}^{T}[e_{jt}e_{it} - E(e_{jt}e_{it})]\Big) - \hat{H}l_i\frac{1}{\tilde{\sigma}_i^2}(\tilde{\sigma}_i^2 - \sigma_i^2).
\end{aligned} \tag{C.3}
$$

Substituting (C.3) into the right hand side of (C.2),

$$
\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}(\hat{L} - L) = \Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'\Big)\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L} - L) \tag{C.4}
$$

$$-\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'\Big)\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)(\hat{L}-L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}+\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'\Big)\hat{H}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^4}\hat{l}_i\hat{l}_i'(\tilde{\sigma}_i^2-\sigma_i^2)\hat{H}$$

$$-\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'\Big)\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Big)L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}+\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'\Big)\Big(\frac{1}{T}\sum_{t=1}^{T}f_te_t'\Big)\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}$$

$$-\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'\Big)\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}L\Big(\frac{1}{T}\sum_{t=1}^{T}f_te_t'\Big)\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}+\Big(\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_ie_{it}f_t'\Big)L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}$$

$$-\Big(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'\Big)\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}\hat{l}_i\hat{l}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)\hat{H}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}m_i\hat{l}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\hat{H}-\frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\tilde{\sigma}_i^4}m_il_i'\hat{H}.$$

Similar to (C.1), we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}m_il_i'=o_p(1), \tag{C.5}$$

which implies that $\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_il_i'=O_p(1)$. Now we analyze the terms on the right hand side of (C.4) one by one. The first term is $O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$ due to (C.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.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 $\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\frac{1}{T}\sum_{t=1}^{T}e_tf_t')=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$ 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 $\frac{1}{NT}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}m_ie_{it}f_t'=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T})$, 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

$$\frac{1}{N}M'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)=O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T}).$$

This result, together with Theorem 4.1, shows that

$$I_b=O_p(\frac{1}{NT})+O_p(\frac{1}{T^2}).$$

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) as

$$\hat{l}_i-l_i=\frac{1}{T}\sum_{t=1}^{T}f_te_{it}+\mathcal{T}_i, \tag{C.6}$$

where

$$\mathcal{T}_i = (\hat{L}-L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)(\hat{L}-L)'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i$$
$$-\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}L\Big(\frac{1}{T}\sum_{t=1}^{T}f_te_t'\Big)\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Big)L'\widetilde{\Sigma}_{ee}^{-1}\hat{L}\hat{H}l_i$$
$$-\hat{H}\Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2}\hat{l}_i\hat{l}_j'\frac{1}{T}\sum_{t=1}^{T}[e_{it}e_{jt}-E(e_{it}e_{jt})]\Big)\hat{H}l_i + \hat{H}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^4}\hat{l}_i\hat{l}_i'(\tilde{\sigma}_i^2-\sigma_i^2)\hat{H}l_i$$
$$+\hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}\Big(\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Big)l_i - \hat{H}\hat{L}'\widetilde{\Sigma}_{ee}^{-1}(\hat{L}-L)\Big(\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big)$$
$$+\hat{H}\Big(\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}\hat{l}_j\frac{1}{T}\sum_{t=1}^{T}[e_{jt}e_{it}-E(e_{jt}e_{it})]\Big) - \hat{H}l_i\frac{1}{\tilde{\sigma}_i^2}(\tilde{\sigma}_i^2-\sigma_i^2).$$

Now term I_d can be written as

$$I_d = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}(\hat{l}_i-l_i)(\hat{l}_i-l_i)' = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}+\mathcal{T}_i\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}+\mathcal{T}_i\Big]'$$
$$= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]' + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\mathcal{T}_i'$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\mathcal{T}_i\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]' + \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\mathcal{T}_i\mathcal{T}_i' = II_a + II_b + II_c + II_d.$$

First consider II_a, which can be written as

$$II_a = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]' - \frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\tilde{\sigma}_i^2\sigma_i^2}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'. \quad (C.7)$$

The first expression of (C.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.7) can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]' - \frac{1}{N}\sum_{i=1}^{N}\frac{(\tilde{\sigma}_i^2-\sigma_i^2)^2}{\tilde{\sigma}_i^2\sigma_i^4}\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'. \quad (C.8)$$

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.8), which can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\sigma_i^4}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]+\frac{1}{NT}\sum_{i=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\sigma_i^2}I_r. \tag{C.9}$$

The first term of the preceding equation can be further written as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathcal{S}_i}{\sigma_i^4}\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]$$

$$+\frac{1}{NT^3}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^4}f_tf_s'[\varepsilon_{i,uts}-E(\varepsilon_{i,uts})]+\frac{1}{NT^3}\sum_{i=1}^{N}\sum_{u=1}^{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^4}f_tf_s'E(\varepsilon_{i,uts}),$$

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

$$C^4\Big[\frac{1}{N}\sum_{i=1}^{N}\mathcal{S}_i^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]\Big\|^2\Big]^{1/2},$$

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.9) is $O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2})$. Consider the second term of (C.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.8) is $O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2})$.

The second term of (C.8) is bounded by

$$C^6\frac{1}{N}\sum_{i=1}^{N}(\tilde{\sigma}_i^2-\sigma_i^2)^2\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]',$$

which is further bounded in norm by

$$2C^6\frac{1}{N}\sum_{i=1}^{N}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)\Big]^2\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'$$

$$+2C^6\frac{1}{N}\sum_{i=1}^{N}\Big[\frac{1}{T}\sum_{t=1}^{T}\mathcal{S}_i\Big]^2\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]\Big[\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big]'.$$

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

$$II_a=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]+\frac{1}{T}I_r+O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2}).$$

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} \\
& - \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]'.
\end{aligned}
$$

Notice that

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\tilde{\sigma}_i^2}l_i f_t' e_{it} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}),$$

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

$$\hat{H}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}\hat{l}_j\Big\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t' e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\} \tag{C.10}$$

$$-\frac{1}{N}\hat{H}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\tilde{\sigma}_i^2-\sigma_i^2}{\tilde{\sigma}_i^2\tilde{\sigma}_j^2\sigma_i^2}\hat{l}_j\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})].$$

The first term of (C.10) can be written as

$$\frac{1}{NT^2}\hat{H}\sum_{j=1}^{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2\sigma_j^2}l_jf_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]$$
$$-\hat{H}\sum_{j=1}^{N}\frac{\tilde{\sigma}_j^2-\sigma_j^2}{\tilde{\sigma}_j^2\sigma_j^2}l_j\Big\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\}$$
$$-\hat{H}\sum_{j=1}^{N}\frac{1}{\tilde{\sigma}_j^2}(\hat{l}_j-l_j)\Big\{\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_t'e_{it}[e_{js}e_{is}-E(e_{js}e_{is})]\Big\}.$$

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.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

$$II_c=O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2}).$$

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

$$I_d=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]+\frac{1}{T}I_r+O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2}).$$

Summarizing the results on $I_a,\dots,I_d$, we have

$$\begin{aligned}&\frac{1}{N}(M\hat{\Lambda}-\hat{L})'\widetilde{\Sigma}_{ee}^{-1}(M\hat{\Lambda}-\hat{L})\\&=\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]+\frac{1}{T}I_r+O_p(\frac{1}{\sqrt{NT^3}})+O_p(\frac{1}{T^2}),\end{aligned}$$

Now consider the term $\frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_tf_s'[e_{it}e_{is}-E(e_{it}e_{is})]$, which we use ω to denote. Then the variance of $\operatorname{tr}(\omega)$ is

$$\operatorname{var}(\operatorname{tr}(\omega))=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\operatorname{var}\Big[\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}f_t'f_se_{it}e_{is}\Big]=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^4}\operatorname{var}\Big[e_i'\frac{FF'}{T}e_i\Big]$$

where $e_i=(e_{i1},e_{i2},\dots,e_{iT})'$. By the well-known result that

$$\operatorname{var}(V'BV)=(\mu_4^v-3\sigma^4)\sum_{t=1}^{T}b_{tt}^2+\sigma^4\Big[\operatorname{tr}(BB')+\operatorname{tr}(B^2)\Big]$$

where $V=(v_1,v_2,\dots,v_T)'$ with each v_t is iid over t with mean zero and variance σ^2 and $\mu_4^v=E(v_t^4)$, 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

$$\operatorname{var}\Big[e_i'\frac{FF'}{T}e_i\Big]=(\mu_4-3\sigma_i^4)\sum_{t=1}^{T}\Big(\frac{f_t'f_t}{T}\Big)^2+\sigma_i^4\Big[\operatorname{tr}\Big(\frac{FF'}{T}\frac{FF'}{T}\Big)+\operatorname{tr}\Big(\frac{FF'}{T}\frac{FF'}{T}\Big)\Big]$$

where $\mu_4 = E(e_{it}^4)$. By the IC that $\frac{F'F}{T} = I_r$, the above equation can be rewritten as

$$\operatorname{var}\Big[e_i'\frac{FF'}{T}e_i\Big] = (\mu_4 - 3\sigma_i^4)\sum_{t=1}^{T}\Big(\frac{f_t'f_t}{T}\Big)^2 + \sigma_i^4 2r$$

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

$$\operatorname{var}\Big[e_i'\frac{FF'}{T}e_i\Big] = \sigma_i^4 2r$$

which implies that $\operatorname{var}(\operatorname{tr}(\omega)) = 2r$. Hence as $N, T \to \infty$ and $N/T^2 \to 0$,

$$\begin{aligned} W &\triangleq \operatorname{tr}\Big[\sqrt{NT^2}\Big(\frac{1}{N}(M\hat{\Lambda} - \hat{L})'\widetilde{\Sigma}_{ee}^{-1}(M\hat{\Lambda} - \hat{L}) - \frac{1}{T}I_r\Big)\Big] \\ &= \frac{1}{\sqrt{NT^2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\sigma_i^2}f_s'f_t[e_{it}e_{is} - E(e_{it}e_{is})] + o_p(1) \xrightarrow{d} N(0, 2r). \end{aligned}$$

This completes the whole proof of Theorem 5.2. $\square$

Appendix D: Partially constrained factor models

We first give detailed derivations of equations (6.2)-(6.4). The first order condition for Λ is

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

The first order condition for Γ is

$$\hat{\Gamma}'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1} = 0. \tag{D.2}$$

The first order condition for Σ_{ee} is

$$\operatorname{diag}[\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}] = 0. \tag{D.3}$$

By (D.1) and (D.2), together with the definition of Φ, we have

$$\hat{\Phi}'\hat{\Sigma}_{zz}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{zz}^{-1}\hat{\Phi} = 0, \tag{D.4}$$

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

$$\hat{\Sigma}_{zz}^{-1} = \hat{\Sigma}_{ee}^{-1} - \hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{G}}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}, \tag{D.5}$$

we have $\hat{\Phi}'\hat{\Sigma}_{zz}^{-1} = \hat{\mathcal{G}}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}$. Given this result, together with (D.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

$$\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\hat{\Sigma}_{ee}^{-1}\hat{\Phi} = 0. \tag{D.6}$$

Now equation (D.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 (D.6), we have

$$[I_{r_1}, 0] \hat{\mathcal{G}} \hat{\Phi}' \hat{\Sigma}_{ee}^{-1} (M_{zz} - \hat{\Sigma}_{zz}) \hat{\Sigma}_{ee}^{-1} M = 0. \tag{D.7}$$

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 (D.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

$$\hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} (M_{zz} - \hat{\Sigma}_{zz}) \hat{\Sigma}_{ee}^{-1} M = 0. \tag{D.8}$$

Proceed to consider (D.2). Post-multiplying $\hat{\Sigma}_{zz}$ on both side of (D.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

$$\hat{\Gamma}' \hat{\Sigma}_{ee}^{-1} (M_{zz} - \hat{\Sigma}_{zz}) = 0. \tag{D.9}$$

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 (D.8) as

$$\hat{\Lambda}' - \Lambda' = -A_{11}' \Lambda' - A_{21}' \Gamma' \hat{\Sigma}_{ee}^{-1} M \hat{R}_N^{-1} - \hat{P}_N^{-1} \hat{\Lambda}' M' \hat{\Sigma}_{ee}^{-1} (\hat{\Sigma}_{ee} - \Sigma_{ee}) \hat{\Sigma}_{ee}^{-1} M \hat{R}_N^{-1}$$

$$+(I-A_{11})'\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}-A_{21}'\frac{1}{T}\sum_{t=1}^{T}g_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'$$
$$+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\Gamma'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}.$$

The above result can be alternatively written as

$$\hat{\Lambda}'-\Lambda'=-A_{11}'\Lambda'-A_{21}'\Gamma'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}+\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1} \tag{D.10}$$
$$+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\Gamma'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}+\mathcal{J}_\Lambda,$$

where

$$\mathcal{J}_\Lambda=-\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}-A_{11}'\frac{1}{T}\sum_{t=1}^{T}f_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}$$
$$-A_{21}'\frac{1}{T}\sum_{t=1}^{T}g_te_t'\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}+\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}M\hat{R}_N^{-1}.$$

By similar arguments as above, the first order condition (D.9) can be written as

$$\hat{\gamma}_i-\gamma_i=\frac{1}{T}\sum_{t=1}^{T}g_te_{it}+\mathcal{J}_{i,\Gamma} \tag{D.11}$$

where

$$\mathcal{J}_{i,\Gamma}=-A_{22}'\gamma_i-A_{12}'\Lambda'm_i-A_{22}'\frac{1}{T}\sum_{t=1}^{T}g_te_{it}+\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i-\hat{Q}_N^{-1}\gamma_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}$$
$$-A_{12}'\frac{1}{T}\sum_{t=1}^{T}f_te_{it}+\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i+\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})].$$

Similarly, we can rewrite the first order condition (D.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

$$\hat{\sigma}_i^2-\hat{\sigma}_i^2=\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-\sigma_i^2)+\mathcal{J}_{i,\sigma^2}, \tag{D.12}$$

where

$$\mathcal{J}_{i,\sigma^2}=-2\gamma_i'\mathcal{J}_{i,\Gamma}-(\hat{\gamma}_i-\gamma_i)'(\hat{\gamma}_i-\gamma_i)-2m_i'(\hat{\Lambda}-\Lambda)\Lambda'm_i$$
$$-m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i-2m_i'(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_te_{it}+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\frac{1}{T}\sum_{t=1}^{T}f_te_{it}$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_te_{it}-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i$$

$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda'm_i+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M\Lambda(\hat{\Lambda}-\Lambda)'m_i$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\Gamma\mathcal{J}_{i,\Gamma}+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)(\hat{\gamma}_i-\gamma_i)$$
$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'m_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T e_tg_t'\gamma_i$$
$$-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T[e_te_{it}-E(e_te_{it})].$$

Equation (D.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

$$A+A'=A'A+(I-A)'\frac{1}{T}\sum_{t=1}^T h_te_t'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}+\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T e_th_t'(I-A) \tag{D.13}$$
$$+\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^T(e_te_t'-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}-\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}.$$

By identification condition IC', we have

$$\mathrm{Ndg}\left\{\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}-\frac{1}{N}\Phi'\Sigma_{ee}^{-1}\Phi\right\}=0.$$

The expression on the left hand side of the preceding equation is equal to

$$\mathrm{Ndg}\left\{\frac{1}{N}(\hat{\Phi}-\Phi)'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}+\frac{1}{N}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(\hat{\Phi}-\Phi)-\frac{1}{N}(\hat{\Phi}-\Phi)'\hat{\Sigma}_{ee}^{-1}(\hat{\Phi}-\Phi)+\frac{1}{N}\Phi'(\hat{\Sigma}_{ee}^{-1}-\Sigma_{ee}^{-1})\Phi\right\}.$$

Given the above result, by the definition of A, we have

$$\mathrm{Ndg}(A\hat{\mathcal{H}}+\hat{\mathcal{H}}A') \tag{D.14}$$
$$=\mathrm{Ndg}\left\{\frac{1}{N}(\hat{\Phi}-\Phi)'\hat{\Sigma}_{ee}^{-1}(\hat{\Phi}-\Phi)-\frac{1}{N}\sum_{i=1}^N\frac{\phi_i\phi_i'}{\hat{\sigma}_i^2\sigma_i^4}(\hat{\sigma}_i^2-\sigma_i^2)^2+\frac{1}{N}\sum_{i=1}^N\frac{\phi_i\phi_i'}{\sigma_i^4}(\hat{\sigma}_i^2-\sigma_i^2)\right\},$$

where $\hat{\mathcal{H}}=\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\hat{\mathcal{H}}/N$. Now we use the above results to prove Theorem 6.1. First we can show that

$$\frac{1}{N}\sum_{i=1}^N\frac{1}{\hat{\sigma}_i^2}\|\hat{\phi}_i-\phi_i\|^2\xrightarrow{p}0 \tag{D.15}$$

and

$$\frac{1}{N}\sum_{i=1}^N(\hat{\sigma}_i^2-\sigma_i^2)^2\xrightarrow{p}0. \tag{D.16}$$

Notice that the present model is a mixture of a standard factor model and a constrained factor model. In Proposition 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 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 (D.9), (D.10), one can further show that

$$\begin{aligned}
&\hat{\Lambda} - \Lambda = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}),\\
&\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\hat{\sigma}_i^2}\|\hat{\gamma}_i - \gamma_i\|^2 = O_p(\frac{1}{T}),\\
&\frac{1}{N}\sum_{i=1}^{N}(\hat{\sigma}_i^2 - \sigma_i^2)^2 = O(\frac{1}{T}).
\end{aligned} \tag{D.17}$$

Equation (D.13) corresponds to equation (A.14) in the pure constrained factor model. Using the arguments as in the derivation of (B.13), one can obtain a similar result

$$A + A' = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}). \tag{D.18}$$

By the consistency results (D.15) and (D.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 (B.16), one can show that

$$\text{Ndg}(A\mathcal{H} + \mathcal{H}A') = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}). \tag{D.19}$$

The equation system (D.18) and (D.19) gives

$$A = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}). \tag{D.20}$$

Using the above result, it can be shown that

$$\mathcal{J}_{i,\sigma^2} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}).$$

The above result, together with (D.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 B.3 and (D.20), we have

$$\mathcal{J}_{i,\Gamma} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}).$$

This result, together with (D.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 B.3 and B.5 continue to hold for a constrained factor model. Given this, we can rewrite (D.10) as

$$\hat{\Lambda}' - \Lambda' = -A_{11}'\Lambda' - A_{21}'\psi' + \frac{1}{T}\sum_{t=1}^{T} f_t e_t' \Sigma_{ee}^{-1} M R_N^{-1} + P_N^{-1}\Lambda' M' \Sigma_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t' \Lambda' \tag{D.21}$$
$$+ P_N^{-1}\Lambda' M'\Sigma_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t g_t'\psi' + O_p\Big(\frac{1}{N\sqrt{T}}\Big) + O_p\Big(\frac{1}{\sqrt{N}T}\Big) + O_p\Big(\frac{1}{T^{3/2}}\Big).$$

We note that

$$\begin{aligned}
\mathrm{vec}\Big(\frac{1}{T}\sum_{t=1}^{T} f_t e_t'\Sigma_{ee}^{-1}MR_N^{-1}\Big) &= \mathrm{vec}\Big(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2} f_t e_{it} m_i' R^{-1}\Big) \\
&= (R^{-1}\otimes I_{r_1})\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it}, \\
\mathrm{vec}\Big(P_N^{-1}\Lambda' M'\Sigma_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t'\Lambda'\Big) &= \mathrm{vec}\Big(P^{-1}\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2} m_i f_t' e_{it}\Lambda'\Big) \\
&= K_{kr_1}\mathrm{vec}\Big(\Lambda\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2} f_t m_i' e_{it}\Lambda P^{-1}\Big) \\
&= K_{kr_1}[(P^{-1}\Lambda')\otimes\Lambda]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it}, \\
\mathrm{vec}\Big(P_N^{-1}\Lambda' M'\Sigma_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t g_t'\psi'\Big) &= \mathrm{vec}\Big(P^{-1}\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2} m_i g_t' e_{it}\psi'\Big) \\
&= K_{kr_1}\mathrm{vec}\Big(\psi\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2} g_t m_i' e_{it}\Lambda P^{-1}\Big) \\
&= K_{kr_1}[(P^{-1}\Lambda')\otimes\psi]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes g_t)e_{it}.
\end{aligned}$$

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

$$\mathrm{vec}\Big(A_{11}'\Lambda' + A_{21}'\psi'\Big) = \mathrm{vec}(E_1'A'\Psi') = K_{kr_1}\mathrm{vec}(\Psi A E_1) = K_{kr_1}(E_1'\otimes\Psi)\mathrm{vec}(A).$$

Taking the vectorization operation on both sides of (D.21), we get

$$\mathrm{vec}(\hat{\Lambda}' - \Lambda') = \Big[(R^{-1}\otimes I_{r_1}) + K_{kr_1}[(P^{-1}\Lambda')\otimes\Lambda]\Big]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it} \tag{D.22}$$
$$+ K_{kr_1}[(P^{-1}\Lambda')\otimes\psi]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes g_t)e_{it} - K_{kr_1}(E_1'\otimes\Psi)\mathrm{vec}(A)$$

$$+ O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{N}T}) + O_p(\frac{1}{T^{3/2}}).$$

Now consider (D.13) and (D.14). Again, using similar arguments as in the derivation of (B.21), one can show by (D.13) that

$$2D_r^+\text{vec}(A) = 2D_r^+\text{vec}(\eta^\star) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{N}T}) + O_p(\frac{1}{T^{3/2}}), \tag{D.23}$$

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} \\
&\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
\end{aligned}
$$

$$+ 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{11}'\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{21}'\gamma_i$$

$$+ 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i$$

$$+ 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i$$

$$= \phi_i'A'A\phi_i - 2\phi_i'A'\frac{1}{T}\sum_{t=1}^{T}h_te_t'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i - \phi_i'\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i + 2\gamma_i'A_{22}'\frac{1}{T}\sum_{t=1}^{T}g_te_{it}$$

$$+ \phi_i'\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t' - \Sigma_{ee}^{-1})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i + 2\gamma_i'A_{12}'\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2\gamma_i'\hat{Q}_N^{-1}\gamma_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}$$

$$- 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i + 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i$$

$$+ 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{11}'\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{21}'\gamma_i$$

$$+ 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i$$

$$+ 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i - 2\gamma_i'\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it} - E(e_te_{it})].$$

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

$$\mathcal{J}_{i,\sigma^2}^* = m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i - 2m_i'(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\frac{1}{T}\sum_{t=1}^{T}f_te_{it}$$

$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i$$

$$-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\mathcal{J}_{i,\Gamma} + 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)(\hat{\gamma}_i-\gamma_i)$$

$$+2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'m_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2} - 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it} - E(e_te_{it})]$$

$$-2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1(\hat{\Lambda}-\Lambda)'m_i - 2m_i'\hat{\Lambda}\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i$$

$$+\phi_i'A'A\phi_i - 2\phi_i'A'\frac{1}{T}\sum_{t=1}^{T}h_te_t'\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i - \phi_i'\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'(\hat{\Sigma}_{ee}-\Sigma_{ee})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1} + 2\gamma_i'A_{22}'\frac{1}{T}\sum_{t=1}^{T}g_te_{it}$$

$$+\phi_i'\hat{\mathcal{H}}_N^{-1}\hat{\Phi}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t' - \Sigma_{ee}^{-1})\hat{\Sigma}_{ee}^{-1}\hat{\Phi}\hat{\mathcal{H}}_N^{-1}\phi_i + 2\gamma_i'A_{12}'\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2\gamma_i'\hat{Q}_N^{-1}\gamma_i\frac{\hat{\sigma}_i^2-\sigma_i^2}{\hat{\sigma}_i^2}$$

$$-2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i + 2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t'\Lambda'm_i$$

$$+2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}M(\hat{\Lambda}-\Lambda)\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{11}'\Lambda'm_i - 2m_i'\Lambda\hat{\mathcal{G}}_1A_{21}'\gamma_i$$

$$+2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}(\hat{\Gamma}-\Gamma)\gamma_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathcal{G}}_1\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i$$

$$+2m_i'\Lambda\hat{\mathcal{G}}_1\hat{P}_N^{-1}\hat{\Lambda}'M'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tg_t'\gamma_i - 2\gamma_i'\hat{Q}_N^{-1}\hat{\Gamma}'\hat{\Sigma}_{ee}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it}-E(e_te_{it})].$$

Given the expression of $\mathcal{J}_{i,\sigma^2}^*$, one can show that

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\phi_i\phi_i'}{\sigma_i^4}\mathcal{J}_{i,\sigma_i^2}^* = O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).$$

Given this result, we have

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\phi_i\phi_i'}{\sigma_i^4}(\hat{\sigma}_i^2-\sigma_i^2) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\phi_i\phi_i'}{\sigma_i^4}(e_{it}^2-\sigma_i^2) - \frac{1}{T}r_1\mathcal{H} + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).$$

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 (B.22), one can show that

$$\mathcal{D}[(\mathcal{H}_N\otimes I_r)+(I_r\otimes K_r)K_r]\text{vec}(A) = \mathcal{D}\text{vec}(\zeta^\star-\mu^\star) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).$$

Let $\mathbb{D}_1, \mathbb{D}_2$ and $\mathbb{D}_3$ be defined the same as in the main text. Similar to (B.24), we have

$$\mathbb{D}_1\text{vec}(A) = \mathbb{D}_2\text{vec}(\eta^\star) + \mathbb{D}_3\text{vec}(\zeta^\star) - \mathbb{D}_3\text{vec}(\mu^\star) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).$$

Also notice that

$$\begin{aligned}
\text{vec}(\eta^\star) &= \text{vec}\Big[\frac{1}{T}\sum_{t=1}^{T}h_te_t'\Sigma_{ee}^{-1}\Phi\mathcal{H}_N^{-1}\Big] = \text{vec}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}h_t\phi_i'e_{it}\mathcal{H}^{-1}\Big],\\
&= (\mathcal{H}^{-1}\otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\phi_i\otimes h_t)e_{it}\\
&= (\mathcal{H}^{-1}\otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(E_1\Lambda'm_i+E_2\gamma_i)\otimes(E_1f_t+E_2g_t)e_{it}\\
&= [(\mathcal{H}^{-1}E_1\Lambda')\otimes E_1]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes f_t)e_{it}\\
&\quad + [(\mathcal{H}^{-1}E_1\Lambda')\otimes E_2]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(m_i\otimes g_t)e_{it}\\
&\quad + [(\mathcal{H}^{-1}E_2)\otimes E_1]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\sigma_i^2}(\gamma_i\otimes f_t)e_{it}
\end{aligned}$$

$$+ [(\mathcal{H}^{-1}E_2) \otimes E_2] \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (\gamma_i \otimes g_t) e_{it},$$

$$\text{vec}(\zeta^\star) = \text{vec}\Big[\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{\phi_i \phi_i'}{\sigma_i^4} (e_{it}^2 - \sigma_i^2)\Big] = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^4} (\phi_i \otimes \phi_i)(e_{it}^2 - \sigma_i^2),$$

$$\text{vec}(\mu^\star) = \text{vec}\Big[\frac{1}{T} r_1 \mathcal{H} + \frac{1}{NT} \sum_{i=1}^{N} \frac{\phi_i \phi_i'}{\sigma_i^6} (\kappa_{i,4} - \sigma_i^4) - \frac{1}{T} E_2 E_2'\Big]$$

$$= \frac{1}{NT} \sum_{i=1}^{N} \frac{1}{\sigma_i^6} (\phi_i \otimes \phi_i)(\kappa_{i,4} - \sigma_i^2) + \frac{1}{T} \text{vec}\Big[r_1 \mathcal{H} - E_2 E_2'\Big].$$

Given the above result, we have

$$\begin{aligned}
\text{vec}(A) = {} & \mathbb{D}_1^{-1} \mathbb{D}_2 [(\mathcal{H}_N^{-1} E_1 \Lambda') \otimes E_1] \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (m_i \otimes f_t) e_{it} \\
& + \mathbb{D}_1^{-1} \mathbb{D}_2 [(\mathcal{H}_N^{-1} E_1 \Lambda') \otimes E_2] \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (m_i \otimes g_t) e_{it} \\
& + \mathbb{D}_1^{-1} \mathbb{D}_2 [(\mathcal{H}_N^{-1} E_2) \otimes E_1] \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (\gamma_i \otimes f_t) e_{it} \\
& + \mathbb{D}_1^{-1} \mathbb{D}_2 [(\mathcal{H}_N^{-1} E_2) \otimes E_2] \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (\gamma_i \otimes g_t) e_{it} \qquad \text{(D.24)} \\
& + \mathbb{D}_1^{-1} \mathbb{D}_3 \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^4} (\phi_i \otimes \phi_i)(e_{it}^2 - \sigma_i^2) \\
& - \mathbb{D}_1 \mathbb{D}_3 \Big\{ \frac{1}{NT} \sum_{i=1}^{N} \frac{1}{\sigma_i^6} (\phi_i \otimes \phi_i)(\kappa_{i,4} - \sigma_i^2) + \frac{1}{T} \text{vec}\Big[r_1 \mathcal{H}_N - E_2 E_2'\Big] \Big\} \\
& + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).
\end{aligned}$$

Now we define

$$\begin{aligned}
\mathbb{B}_1^\star &= R^{-1} \otimes I_{r_1} + K_{kr_1} [(P^{-1} \Lambda') \otimes \Lambda] - K_{kr_1} (E_1' \otimes \Psi) \mathbb{D}_1^{-1} \mathbb{D}_2 [(\mathcal{H}_N^{-1} E_1 \Lambda') \otimes E_1], \\
\mathbb{B}_2^\star &= K_{kr_1} [P^{-1} \otimes \psi] - K_{kr_1} (E_1' \otimes \Psi) \mathbb{D}_1^{-1} \mathbb{D}_2 [(\mathcal{H}_N^{-1} E_1) \otimes E_2], \\
\mathbb{B}_3^\star &= -K_{kr_1} (E_1' \otimes \Psi) \mathbb{D}_1^{-1} \mathbb{D}_2 [(\mathcal{H}_N^{-1} E_2) \otimes E_1], \\
\mathbb{B}_4^\star &= -K_{kr_1} (E_1' \otimes \Psi) \mathbb{D}_1^{-1} \mathbb{D}_2 [(\mathcal{H}_N^{-1} E_2) \otimes E_2], \\
\mathbb{B}_5^\star &= -K_{kr_1} (E_1' \otimes \Psi) \mathbb{D}_1^{-1} \mathbb{D}_3, \\
\Delta^\star &= K_{kr_1} (E_1' \otimes \Psi) \mathbb{D}_1^{-1} \mathbb{D}_3 \Big[\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^6} (\phi_i \otimes \phi_i)(\kappa_{i,4} - \sigma_i^4) + \text{vec}(r_1 \mathcal{H}_N - E_2 E_2')\Big].
\end{aligned}$$

Substituting (D.24) into (D.22), we can rewrite (D.22) in terms of $\mathbb{B}_i^\star$ as

$$\text{vec}(\hat{\Lambda}' - \Lambda') = \mathbb{B}_1^\star \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (m_i \otimes f_t) e_{it} + \mathbb{B}_2^\star \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{1}{\sigma_i^2} (\Lambda' m_i \otimes g_t) e_{it}$$

$$+\mathbb{B}_3^\star \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}(\gamma_i\otimes f_t)e_{it} + \mathbb{B}_4^\star \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^2}(\gamma_i\otimes g_t)e_{it}$$
$$+\mathbb{B}_5^\star \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \frac{1}{\sigma_i^4}(\phi_i\otimes\phi_i)(e_{it}^2-\sigma_i^2) + \frac{1}{T}\Delta^\star$$
$$+O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).$$

Given the above result, by a Central Limit Theorem, 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

$$\Omega_N^\star = \mathbb{B}_1^\star(R\otimes I_{r_1})\mathbb{B}_1^{\star\prime} + \mathbb{B}_2^\star(P\otimes I_{r_1})\mathbb{B}_2^{\star\prime} + \mathbb{B}_3^\star(Q\otimes I_{r_1})\mathbb{B}_3^{\star\prime} + \mathbb{B}_4^\star(Q\otimes I_{r_2})\mathbb{B}_4^{\star\prime}$$
$$+\mathbb{B}_1^\star(S\otimes I_{r_1})\mathbb{B}_3^{\star\prime} + \mathbb{B}_3^\star(S'\otimes I_{r_1})\mathbb{B}_1^{\star\prime} + \mathbb{B}_5^\star\Big[\frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^8}(\phi_i\phi_i')\otimes(\phi_i\phi_i')(\kappa_{i,4}-\sigma_i^4)\Big]\mathbb{B}_5^{\star\prime}.$$

Appendix E: More simulation results

In Section 7.1, we present the comparison results of MLE and PC estimates when errors follow normal distribution. In this appendix, we provide addition comparison results when errors follow t-distribution and χ^2-distribution, in the following Table E1-E4.

Table E1: $k=3$, $r=1$, 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	2.1513	0.1016	0.1499	4.4964
50	30	0.0328	0.0523	2.0249	0.0682	0.0997	3.8633
100	30	0.0229	0.0346	1.8956	0.0465	0.0676	3.7001
150	30	0.0198	0.0293	1.9675	0.0384	0.0547	3.6675
30	50	0.0319	0.0495	1.9184	0.0781	0.1114	4.3136
50	50	0.0227	0.0365	1.8257	0.0558	0.0804	4.0183
100	50	0.0166	0.0262	1.8536	0.0367	0.0522	3.6946
150	50	0.0142	0.0220	1.9064	0.0302	0.0426	3.6906
30	100	0.0227	0.0371	2.0298	0.0679	0.0965	5.2859
50	100	0.0154	0.0251	1.7734	0.0448	0.0642	4.5430
100	100	0.0111	0.0179	1.7883	0.0280	0.0394	3.9425
150	100	0.0094	0.0151	1.8436	0.0221	0.0313	3.8328

Table E2: $k=8$, $r=3$, and $\epsilon_{it}\sim t_5$.

$\Lambda_{3\times1}$		MLE			PC		
N	T	MAD	RMSE	RAvar	MAD	RMSE	RAvar
30	30	0.3478	0.4961	14.8830	0.5800	0.8257	24.7707
50	30	0.2379	0.3498	13.5469	0.3959	0.5677	21.9853
100	30	0.1461	0.2217	12.1448	0.2236	0.3244	17.7679
150	30	0.1156	0.1751	11.7477	0.1661	0.2415	16.2004
30	50	0.2584	0.3742	14.4935	0.5165	0.7541	29.2059
50	50	0.1727	0.2530	12.6510	0.3226	0.4753	23.7656
100	50	0.1154	0.1826	12.9116	0.1816	0.2686	18.9961
150	50	0.0930	0.1429	12.3777	0.1402	0.2069	17.9150
30	100	0.1880	0.2761	15.1209	0.4626	0.7075	38.7519
50	100	0.1249	0.1928	13.6358	0.2734	0.4208	29.7560
100	100	0.0812	0.1321	13.2061	0.1410	0.2144	21.4392
150	100	0.0639	0.1025	12.5593	0.1065	0.1592	19.5004

Table E3: $k = 3$, $r = 1$, and $\epsilon_{it} \sim \chi^2(2)$.

$\Lambda_{3\times1}$		MLE			PC		
N	T	MAD	RMSE	RAvar	MAD	RMSE	RAvar
30	30	0.0409	0.0649	1.9461	0.0941	0.1394	4.1808
50	30	0.0319	0.0497	1.9248	0.0707	0.1011	3.9147
100	30	0.0225	0.0351	1.9249	0.0459	0.0654	3.5813
150	30	0.0207	0.0320	2.1499	0.0388	0.0553	3.7093
30	50	0.0335	0.0541	2.0942	0.0841	0.1216	4.7105
50	50	0.0229	0.0362	1.8116	0.0569	0.0826	4.1296
100	50	0.0172	0.0281	1.9877	0.0371	0.0526	3.7186
150	50	0.0135	0.0208	1.7992	0.0285	0.0401	3.4761
30	100	0.0220	0.0362	1.9845	0.0673	0.0959	5.2502
50	100	0.0165	0.0274	1.9405	0.0456	0.0647	4.5738
100	100	0.0109	0.0175	1.7453	0.0281	0.0397	3.9739
150	100	0.0088	0.0141	1.7298	0.0219	0.0311	3.8117

Table E4: $k = 8$, $r = 3$, 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	14.7279	0.5657	0.8061	24.1843
50	30	0.2353	0.3481	13.4800	0.3746	0.5424	21.0072
100	30	0.1547	0.2475	13.5542	0.2242	0.3258	17.8441
150	30	0.1203	0.1893	12.6995	0.1752	0.2559	17.1631
30	50	0.2632	0.3831	14.8379	0.5189	0.7618	29.5025
50	50	0.1795	0.2697	13.4851	0.3214	0.4769	23.8459
100	50	0.1160	0.1803	12.7504	0.1813	0.2632	18.6091
150	50	0.0959	0.1656	14.3393	0.1417	0.2096	18.1551
30	100	0.1839	0.2687	14.7185	0.4666	0.7114	38.9666
50	100	0.1271	0.1945	13.7531	0.2718	0.4124	29.1630
100	100	0.0854	0.1452	14.5163	0.1439	0.2214	22.1428
150	100	0.0676	0.1151	14.1010	0.1045	0.1617	19.8014

Appendix F: More comparison of W and LR

In this appendix, we compare the empirical size and power of our W test with the LR test proposed in Tsai and Tsay (2010). Their LR is a modified likelihood ratio test statistic with the Bartlett's correction factor,[①] which is defined as

$$LR = \left(T - \frac{2N+11}{6} - \frac{2r}{3}\right)\left(\ln|\hat{\Sigma}_c| - \ln|\hat{\Sigma}_u|\right)$$

where $\hat{\Sigma}_c = M\hat{\Lambda}\hat{\Lambda}'M + \hat{\Sigma}_{ee}$ and $\hat{\Sigma}_u = \hat{L}\hat{L}' + \tilde{\Sigma}_{ee}$, with $\hat{\Lambda}, \hat{\Sigma}_{ee}$ being the MLE from the constrained factor model and $\hat{L}, \tilde{\Sigma}_{ee}$ being the MLE from the unconstrained factor model. We run simulations based on the same data generating processes as in Tables 3 and 4, with the empirical size and power results of LR provided in the following Tables F1 and F2 respectively. Comparison can be made based on these tables as below.

From the empirical size results in Table F1, we have following observations. First, LR is not working when $N \geq T$, as the empirical sizes of LR are close to zero (or exactly equal zero) when $N > T$ and suffer severe size distortion when $N = T$. This is due to the definition of LR, as the correction factor $\left(T - \frac{2N+11}{6} - \frac{2r}{3}\right)$ might be too small or even negative when $N \geq T$ and then LR won't be able to reject H_0.[②] This finding might explain why Tsai and Tsay (2010) only considered the small $N(= 30)$ and large $T(\geq 100)$ cases in their size analysis. Second, when N is too big, LR is also not working even if $T > N$, as the empirical sizes of LR are close to one when $N = 100, 200, 300$ and $T > N$. With comparison, as shown in Table 3, W test statistic works well in all combinations of (N, T) except some small size distortion under small $T(= 30)$. Therefore, in terms of size, W performs better than LR.

From the empirical power results in Table F2, we can see that LR has very low power when $N > T$, due to the same reason as in the size analysis. Although LR has higher power than W in some cases when T is much bigger than N (like $(N, T) = (30, 150)$ or $(100, 500)$), such difference of power between LR and W decreases as α increases and gets close to zero when $\alpha = 2$ and 5. With comparison, as shown in Table 4, the power performance of W is more consistent as it works well in all combinations of (N, T) and very close to one when $\alpha = 2$ or 5. So overall, W also performs better than LR in terms of power.

In conclusion, the overall performance of W test statistic dominates that of the LR one.[③]

[①] Tsai and Tsay (2010) adopted the suggestion of Bartlett (1950) and Anderson (2003, p. 581).

[②] We also check the empirical size for the likelihood ratio test without the correction factor, denoted as $LR_2 = T\left(\ln|\hat{\Sigma}_c| - \ln|\hat{\Sigma}_u|\right)$. In the case $N \geq T$, the empirical size of LR_2 is close to one, which might because without correction factor, LR_2 is too large due to the T part and LR_2 rejects H_0 most of the time.

[③] We also run simulations when errors follow student's or chi-squared distribution, and similar comparison results can be concluded.

Table F1: The empirical size of the LR test for the case $(k, r) = (3, 1)$ 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%

Table F2: The empirical power of the LR test for the case $(k, r) = (3, 1)$ 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%

Appendix G: Proofs of the theoretical results in Section 9

In this appendix, we define the following notations:

$$\hat{\mathbb{P}} = \frac{1}{N}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}; \qquad \hat{\mathbb{R}} = \frac{1}{N}M'\hat{\mathbb{W}}^{-1}M; \qquad \hat{\mathbb{G}} = (I_r + \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda})^{-1};$$
$$\hat{\mathbb{P}}_N = N \cdot \hat{\mathbb{P}} = \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}; \quad \hat{\mathbb{R}}_N = N \cdot \hat{\mathbb{R}} = M'\hat{\mathbb{W}}^{-1}M, \quad \hat{\mathbb{G}}_N = N \cdot \hat{\mathbb{G}}.$$

Then we have $\hat{\mathbb{P}}_N^{-1} = \hat{\mathbb{G}}(I - \hat{\mathbb{G}})^{-1}$ and

$$\Sigma_{zz}^{-1} = \mathbb{W}^{-1} - \mathbb{W}^{-1}M\Lambda(I_r + \Lambda'M'\mathbb{W}^{-1}M\Lambda)^{-1}\Lambda'M'\mathbb{W}^{-1}, \tag{G.1}$$

and

$$\hat{\Lambda}'M'\hat{\Sigma}_{zz}^{-1} = \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1} - \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}(I_r + \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda})^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1} = \hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}. \tag{G.2}$$

Before starting, we first introduce the following lemma, which are useful throughout the proofs in this appendix. C is a large enough constant.

Lemma G.1 *From assumptions of A and B″, we have*

$$(a)\quad E\bigg(\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_te_{it}\Big\|^2\bigg)\le C,\quad \textit{for all } i;$$

$$(b)\quad E\bigg(\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}f_te_{it}\Big\|^2\bigg)\le C;$$

$$(c)\quad E\bigg(\Big|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(e_{it}^2-w_i^2)\Big|^2\bigg)\le C.$$

Further, we have

$$(d)\quad \frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\Big\|^2=O_p(T^{-1});$$

$$(e)\quad \frac{1}{N}\sum_{i=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-w_i^2)\Big)^2=O_p(T^{-1});$$

$$(f)\quad \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\Big(\frac{1}{T}\sum_{t=1}^{T}\big[e_{it}e_{jt}-E(e_{it}e_{jt})\big]\Big)^2=O_p(T^{-1});$$

Proof of Lemma G.1 follows directly from Assumption A and B″, so omitted here.

Appendix G1: Proof of the consistency of the MLE in Section 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 $C^{-2}\le w_i^2\le C^2$ 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 G1.1, and Proposition G1.1 will be used in the proofs in the following Appendix G2.

Lemma G1.1 *Under assumptions of A, B″, C″ and D″, we have*

$$(a)\quad \sup_{\theta\in\Theta}\frac{1}{NT}\Big|\mathrm{tr}\Big[\Lambda^{*\prime}M'\Sigma_{zz}^{-1}\sum_{t=1}^{T}e_tf_t^{*\prime}\Big]\Big|\xrightarrow{p}0;$$

$$(b)\quad \sup_{\theta\in\Theta}\frac{1}{NT}\Big|\mathrm{tr}\Big[\sum_{t=1}^{T}(e_te_t'-\mathbb{O}^*)\Sigma_{zz}^{-1}\Big]\Big|\xrightarrow{p}0;$$

$$(c)\quad \sup_{\theta\in\Theta}\frac{1}{N}\Big|\mathrm{tr}\Big[(\mathbb{O}^*-\mathbb{W}^*)\Sigma_{zz}^{-1}\Big]\Big|\xrightarrow{p}0;$$

where $\theta^*=(\Lambda^*,\mathbb{W}^*)$ *denotes the true parameters and* $\Sigma_{zz}=M\Lambda\Lambda'M'+\mathbb{W}$.

Proof of Lemma G1.1 (a)(b) is similar to that of Lemma A.1, and proof of G1.1(c) is similar to that of Lemma S.3(b) in Bai and Li (2016), so omitted here.

Lemma G1.2 *Under assumptions of A, B″, C′ and D″, we have*

$$(a)\quad \Big\|\frac{1}{N}\Lambda^{*\prime}M'(\hat{\mathbb{W}}^{-1}-\mathbb{W}^{*-1})M\Lambda^*\Big\| = O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^N(\hat{w}_i^2-w_i^{*2})^2\Big]^{\frac{1}{2}}\Big);$$

$$(b)\quad \Big\|\frac{1}{N}M'(\hat{\mathbb{W}}^{-1}-\mathbb{W}^{*-1})M\Big\| = O_p\Big(\Big[\frac{1}{N}\sum_{i=1}^N(\hat{w}_i^2-w_i^{*2})^2\Big]^{\frac{1}{2}}\Big).$$

Given the above results, if $N^{-1}\sum_{i=1}^N(\hat{w}_i^2-w_i^{*2})^2 = o_p(1)$, *we have*

$$(c)\quad \hat{\mathbb{R}}_N = O_p(N), \quad \hat{\mathbb{R}} = \frac{1}{N}\hat{\mathbb{R}}_N = O_p(1);$$

$$(d)\quad \|\hat{\mathbb{R}}^{-1/2}\| = O_p(1).$$

where $\hat{\mathbb{R}}$ *and* $\hat{\mathbb{R}}_N$ *are defined in the beginning of Appendix G.*

Proof of the above lemma is similar to that of Lemma A.2 and hence omitted here.

Lemma G1.3 *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 G1.3. Proofs for (a)-(c) and (e)-(g) are similar to those for Lemma A.3, so we only include the proofs for (d) and (h) which are different from Lemma A.3. Consider (d). The left hand side can be rewritten as

$$\frac{1}{N}\hat{\mathbb{P}}^{-1/2}\Big[\sum_{i=1}^N\sum_{j=1}^N\hat{\mathbb{P}}_N^{-1/2}\frac{1}{\hat{w}_i^2}\sum_{p=1}^k\hat{\lambda}_pm_{ip}\Big[\mathbb{O}_{ij}-1(i=j)w_i^2\Big]\frac{1}{\hat{w}_j^2}\sum_{l=1}^k\hat{\lambda}_l'm_{jl}\hat{\mathbb{P}}_N^{-1/2}\Big]\hat{\mathbb{P}}^{-1/2},$$

where $1(i=j)$ is the indicator function, equals 1 if $i=j$ and 0 otherwise. The above expression is bounded in norm by

$$C\frac{1}{\sqrt{N}}\|\hat{\mathbb{P}}^{-1/2}\|^2\Big(\sum_{i=1}^N\frac{1}{\hat{w}_i^2}\Big\|\hat{\mathbb{P}}_N^{-1/2}\sum_{p=1}^k\hat{\lambda}_pm_{ip}\Big\|^2\Big)\Big(\frac{1}{N}\sum_{i=1}^N\sum_{j=1,j\neq i}^N(\mathbb{O}_{ij})^2\Big)^{1/2},$$

which is $\|\hat{\mathbb{P}}^{-1/2}\|^2 \cdot O_p(N^{-1/2})$ by the fact that $\Big(\sum_{i=1}^N \frac{1}{\hat{w}_i^2}\Big\|\hat{\mathbb{P}}_N^{-1/2}\sum_{p=1}^k \hat{\lambda}_p m_{ip}\Big\|^2\Big) = r$ and $\Big(\frac{1}{N}\sum_{i=1}^N \sum_{j=1,j\neq i}^N (\mathbb{O}_{ij})^2\Big)$ 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 G1.2(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 G1.1 (Consistency) *Let $\hat{\theta} = (\hat{\Lambda}, \hat{\mathbb{W}})$ be the MLE that maximizes (9.1). 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 G1.1. Similar to the proof of Proposition 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 G1.1, we have $\sup_\theta |R^\dagger(\theta)| = o_p(1)$. Then using the same approach as in the proof of Proposition 4.1, we get $\overline{L}^\dagger(\hat{\theta}) \geq -2|o_p(1)|$, which implies

$$\frac{1}{N}\ln|\hat{\mathbb{W}}| - \frac{1}{N}\ln|\mathbb{W}^*| + \frac{1}{N}\mathrm{tr}[\mathbb{W}^*\hat{\mathbb{W}}^{-1}] - 1 \xrightarrow{p} 0, \tag{G.3}$$

$$\frac{1}{N}\mathrm{tr}[M\Lambda^*\Lambda^{*\prime}M'\hat{\Sigma}_{zz}^{-1}] \xrightarrow{p} 0. \tag{G.4}$$

The above arguments further imply

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^{*2})^2 \xrightarrow{p} 0. \tag{G.5}$$

which is the second result of Proposition G1.1, and other results as following:

$$\hat{\mathbb{G}} = o_p(1); \qquad \hat{\mathbb{P}}_N^{-1} = o_p(1); \tag{G.6}$$

$$\frac{1}{N}\Lambda^{*\prime}M'\mathbb{W}^{*-1}M\Lambda^* - (I_r - \mathbb{A})\frac{1}{N}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}(I_r - \mathbb{A})' \xrightarrow{p} 0, \tag{G.7}$$

$$\frac{1}{N}(\hat{\Lambda} - \Lambda^*)'M'\hat{\mathbb{W}}^{-1}M(\hat{\Lambda} - \Lambda^*) - \mathbb{A}\Big(\frac{1}{N}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\Big)\mathbb{A}' \xrightarrow{p} 0. \tag{G.8}$$

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 (9.2) 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 (G.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

$$\begin{aligned}
\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} &= -\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda} \\
&+\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\Lambda^*\Lambda^{*\prime}M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} + \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}M\Lambda^*\frac{1}{T}\sum_{t=1}^{T}f_t^*e_t'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} \\
&+\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t^{*\prime}\Lambda^{*\prime}M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda} + \hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t' - \mathbb{O}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda} \\
&+\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}^* - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}.
\end{aligned}$$

By the definitions of $\hat{\mathbb{P}}$ and $\mathbb{A}$, we have

$$\begin{aligned}
I_r &= (I_r - \mathbb{A})'(I_r - \mathbb{A}) + \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_te_t' - \mathbb{O}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} \\
&+(I_r - \mathbb{A})'\frac{1}{NT}\sum_{t=1}^{T}f_t^*e_t'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} + \frac{1}{N}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t^{*\prime}(I_r - \mathbb{A}) \\
&-\frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} + \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}^* - \mathbb{W}^*)\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} \\
&= i_1 + i_2 + \cdots + i_6, \quad \text{say}
\end{aligned} \tag{G.9}$$

Compared to (A.14), there exists an extra term i_6 in the above equation, due to the weak dependence structure of the error. Based on (G.9) and (G.8), together with Lemma G1.3, 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 (9.2), which can be simplified as (by (G.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

$$\hat{\Lambda}' - \Lambda^{*\prime} = -\mathbb{A}'\Lambda^{*\prime} + (I - \mathbb{A})'\frac{1}{T}\sum_{t=1}^{T} f_t^* e_t' \hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}_N^{-1} + \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T} e_t f_t^{*\prime}\Lambda^{*\prime} \quad \text{(G.10)}$$
$$+\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \mathbb{O}^*]\hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}_N^{-1} - \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W}^*)\hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}_N^{-1}$$
$$+\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O}^* - \mathbb{W}^*)\hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}_N^{-1}$$

We want 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 G1.3 (e), we know the first two terms are $o_p(1)$. From $\|\hat{\mathbb{P}}^{-1}\| = O_p(1)$ and Lemma G1.3 (b)(f)(g)(h), we get that the rest 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 G1.1. $\square$

Corollary G1.1 *Under Assumptions A, B$''$, C$''$ and D$''$,*

(*a*) $\dfrac{1}{N}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda} - \dfrac{1}{N}\Lambda^{*\prime} M' \mathbb{W}^{*-1} M \Lambda^* = o_p(1)$;

(*b*) $\hat{\mathbb{P}}_N = O_p(N)$, $\hat{\mathbb{P}} = O_p(1)$, $\hat{\mathbb{G}} = O_p(N^{-1})$, $\hat{\mathbb{G}}_N = O_p(1)$;

(*c*) $\dfrac{1}{N}(\hat{\Lambda} - \Lambda)' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda} = o_p(1)$.

Proof of Corollary A.1. Proof for the above Corollary G1.1 is similar to Lemma A.1, and therefore omitted here.

Appendix G2: Proofs of Theorem 9.1, 9.2 and 9.3

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 9.1 and 9.2.

Lemma G2.1 *Under Assumptions A, B$''$, C$''$ and D$''$, we have*

(*a*) $\dfrac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}\dfrac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \mathbb{O})\hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{P}}^{-1} = O_p(T^{-1/2})$;

(*b*) $\dfrac{1}{N}\hat{\mathbb{P}}^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}\dfrac{1}{T}\sum_{t=1}^{T} e_t f_t' = O_p(T^{-1/2})$;

(*c*) $\dfrac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W})\hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{P}}^{-1} = \dfrac{1}{\sqrt{N}} O_p\Big(\Big[\dfrac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\Big]^{\frac{1}{2}}\Big)$;

(*d*) $\dfrac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{P}}^{-1} = O_p(N^{-1/2})$;

(e) $\dfrac{1}{NT}\sum_{t=1}^{T} f_t e_t' \hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}^{-1} = O_p(T^{-1/2})$;

(f) $\dfrac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}\dfrac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \mathbb{O}]\hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}^{-1} = O_p(T^{-1/2})$;

(g) $\dfrac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}} - \mathbb{W})\hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}^{-1} = \dfrac{1}{\sqrt{N}} O_p\Big(\Big[\dfrac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\Big]^{\frac{1}{2}}\Big)$;

(h) $\dfrac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1} M \hat{\mathbb{R}}^{-1} = O_p(N^{-1})$.

The above lemma is strengthened from Lemma G1.3, with its proof similar to Lemma B.1 and hence omitted here.

Based on (G.9) and IC2$''$, together with Lemma G2.1, we have the following Lemma G2.2, which corresponds to Lemma B.2 with modification.

Lemma G2.2 *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 G2.2 is similar to Lemma B.2 and hence omitted here.

Proof of Theorem 4.1. We can rewrite the first order condition (9.3) as

$$\operatorname{diag}\Big\{(M_{zz} - \hat{\Sigma}_{zz}) - (M_{zz} - \hat{\Sigma}_{zz})\hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' - M\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(M_{zz} - \hat{\Sigma}_{zz})\Big\} = 0.$$

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 first order condition (9.3) as

$$\begin{aligned}
\hat{w}_i^2 - w_i^2 = {} & \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2) + 2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} M\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} \\
& -2m_i'\Lambda\frac{1}{T}\sum_{t=1}^{T} f_t e_t' \hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' m_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] \qquad \text{(G.11)} \\
& +m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)' m_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\Lambda}' m_i + 2m_i'(\hat{\Lambda}-\Lambda)\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' m_i \\
& +2m_i'\Lambda(\hat{\Lambda}-\Lambda)' M' \hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' m_i + 2\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2} m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' m_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})_i.
\end{aligned}$$

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 (B.10), we can rewrite (G.11) as

$$\begin{aligned}
\hat{w}_i^2 - w_i^2 = {} & \frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2) - 2m_i'(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} \\
& + 2m_i'\hat{\Lambda}\mathbb{A}'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\mathbb{A}'\frac{1}{T}\sum_{t=1}^{T} f_t e_{it} + 2m_i'\Lambda\psi_1\hat{\mathbb{G}}\hat{\Lambda}'m_i \\
& - 2m_i'\Lambda\mathbb{A}\hat{\mathbb{G}}\hat{\Lambda}'m_i - 2m_i'\Lambda\psi_1(\hat{\Lambda}-\Lambda)'m_i + 2m_i'\Lambda\mathbb{A}(\hat{\Lambda}-\Lambda)'m_i \\
& + m_i'\Lambda\mathbb{A}'\mathbb{A}\Lambda'm_i - 2m_i'\Lambda\mathbb{A}'\psi_1\Lambda'm_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathbb{G}}\hat{\Lambda}'m_i + 2\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2}m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'m_i \\
& + m_i'\Lambda\varphi_1\Lambda'm_i - m_i'\Lambda\varphi_2\Lambda'm_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] \\
& + m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i + m_i'\Lambda\varphi_3\Lambda'm_i - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})_i \\
= {} & a_{i,1} + a_{i,2} + \cdots + a_{i,19}, \quad \text{say.}
\end{aligned} \tag{G.12}$$

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 (B.11)), and notice that the last two terms are $O_p(N^{-2})$, we have

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2 = O_p(T^{-1}) + O_p(N^{-2}) + o_p(\|\hat{\Lambda}-\Lambda\|^2). \tag{G.13}$$

Next, we consider the term $\|\hat{\Lambda}-\Lambda\|$. Using Lemma G2.1(b), (e)-(h) and Lemma G2.2, together with equation (G.10), we have

$$\hat{\Lambda} - \Lambda = O_p(T^{-1/2}) + O_p(N^{-1}) + O_p([\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2]^{1/2}). \tag{G.14}$$

Substituting equation (G.14) into (G.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 9.1. The proof for the first result of Theorem 9.1 is provided after Lemma G2.4. $\square$

The following two lemmas will be useful in proving the first result of Theorem 9.1.

Lemma G2.3 *Under Assumptions A, B$''$, C$''$ and D$''$, we have*

$$\begin{aligned}
(a)\quad & \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}(e_t e_t' - \mathbb{O})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} \\
& = O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});
\end{aligned}$$

$$(b)\quad \frac{1}{N}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_t f_t' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1});$$

$$(c)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = O_p(N^{-1}T^{-1/2}) + O_p(N^{-2});$$

$$(d)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}^{-1} = O_p(N^{-1});$$

$$(e)\quad \frac{1}{NT}\sum_{t=1}^{T}f_t e_t'\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1});$$

$$(f)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_t e_t' - \mathbb{O}]\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1}$$
$$= O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2});$$

$$(g)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\hat{\mathbb{W}}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(N^{-1}T^{-1/2}) + O_p(N^{-2});$$

$$(h)\quad \frac{1}{N^2}\hat{\mathbb{P}}^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}^{-1} = O_p(N^{-1}).$$

The above lemma is strengthened from Lemma G2.1, with its proof similar to Lemma B.3 and hence omitted here.

Lemma G2.4 *Under Assumptions A, B″, C″ and D″, we have*

$$A \equiv (\hat{\Lambda}-\Lambda)'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}) + O_p(\frac{1}{N}) + O_p(\|\hat{\Lambda}-\Lambda\|^2).$$

Proof of the above lemma is similar to that of Lemma B.4 with Lemma G2.3 (a)-(d) and the second result of Theorem 9.1, and therefore omitted here.

Proof of Theorem 4.1 (continued). Now we prove the first result of Theorem 9.1. Notice that the term $\|\hat{\Lambda}-\Lambda\|^2$ is of smaller order than $\hat{\Lambda}-\Lambda$ and hence negligible. Then from (G.10), together with Lemma G2.3 and Lemma G2.4, we have

$$\hat{\Lambda}-\Lambda = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}) + O_p(\frac{1}{N}).$$

This completes the proof of Theorem 9.1. □

From Lemma G2.4 and Theorem 9.1, we have the following corollary directly.

Corollary G2.1 *Under Assumptions A, B″, C″ and D″, we have*

$$A \equiv (\hat{\Lambda}-\Lambda)'M'\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1} = O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T}) + O_p(\frac{1}{N}).$$

The following lemma will be useful in proving Theorem 9.2.

Lemma G2.5 *Under Assumptions A, B″, C″ and D″, we have*

$$(a)\quad \frac{1}{T}\sum_{t=1}^{T}f_t e_t'\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}_N^{-1} = \frac{1}{T}\sum_{t=1}^{T}f_t e_t'\mathbb{W}^{-1}M\mathbb{R}_N^{-1} + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{T^{3/2}});$$

$$(b)\quad \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t' = \mathbb{P}_N^{-1}\Lambda'M'\mathbb{W}^{-1}\frac{1}{T}\sum_{t=1}^{T}e_tf_t' + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{T^{3/2}});$$

$$(c)\quad \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\mathbb{R}}_N^{-1} = \mathbb{P}_N^{-1}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\mathbb{R}_N^{-1} + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{N^2});$$

$$(d)\quad \hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1} = \mathbb{P}_N^{-1}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}_N^{-1} + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{N^2});$$

$$\begin{aligned}(e)\quad \frac{1}{N}M'(\hat{\mathbb{W}}^{-1}-\mathbb{W}^{-1})M = &-\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}m_im_i'(e_{it}^2-w_i^2) + \frac{1}{NT}\sum_{i=1}^{N}m_im_i'\frac{\varpi_i^2}{w_i^4}\\ &-\frac{1}{N}\sum_{i=1}^{N}\frac{m_im_i'}{w_i^4}m_i'\Lambda\mathbb{P}_N^{-1}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}_N^{-1}\Lambda'm_i\\ &+\frac{1}{N}\sum_{i=1}^{N}\frac{m_im_i'}{w_i^4}2m_i'\Lambda\mathbb{G}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})_i\\ &+O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}).\end{aligned}$$

where $\varpi_i^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E[(e_{it}^2-w_i^2)(e_{is}^2-w_i^2)]$.

PROOF OF LEMMA G2.5. First we reconsider the equation (G.12), which can be written as

$$\begin{aligned}\hat{w}_i^2 - w_i^2 = &\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2-w_i^2) + m_i'\Lambda\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})\hat{\mathbb{W}}^{-1}M\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\Lambda'm_i\\ &-2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}(\mathbb{O}-\mathbb{W})_i + \tilde{\mathcal{R}}_i,\end{aligned}\tag{G.15}$$

where

$$\tilde{\mathcal{R}}_i = -2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'M'\hat{\mathbb{W}}^{-1}\frac{1}{T}\sum_{t=1}^{T}[e_te_{it} - E(e_te_{it})] + \tilde{\mathcal{S}}_i$$

with Using the argument deriving (B.10), we can rewrite (G.11) as

$$\begin{aligned}\tilde{\mathcal{S}}_i = &-2m_i'(\hat{\Lambda}-\Lambda)\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\frac{1}{T}\sum_{t=1}^{T}f_te_{it}\\ &+2m_i'\hat{\Lambda}\mathbb{A}'\frac{1}{T}\sum_{t=1}^{T}f_te_{it} - 2m_i'\hat{\Lambda}\hat{\mathbb{G}}\mathbb{A}'\frac{1}{T}\sum_{t=1}^{T}f_te_{it} + 2m_i'\Lambda\psi_1\hat{\mathbb{G}}\hat{\Lambda}'m_i\\ &-2m_i'\Lambda\mathbb{A}\hat{\mathbb{G}}\hat{\Lambda}'m_i - 2m_i'\Lambda\psi_1(\hat{\Lambda}-\Lambda)'m_i + 2m_i'\Lambda\mathbb{A}(\hat{\Lambda}-\Lambda)'m_i\\ &+m_i'\Lambda\mathbb{A}'\mathbb{A}\Lambda'm_i - 2m_i'\Lambda\mathbb{A}'\psi_1\Lambda'm_i - 2m_i'(\hat{\Lambda}-\Lambda)\hat{\mathbb{G}}\hat{\Lambda}'m_i + 2\frac{\hat{w}_i^2-w_i^2}{\hat{w}_i^2}m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}'m_i\\ &+m_i'\Lambda\varphi_1\Lambda'm_i - m_i'\Lambda\varphi_2\Lambda'm_i + m_i'(\hat{\Lambda}-\Lambda)(\hat{\Lambda}-\Lambda)'m_i.\end{aligned}\tag{G.16}$$

By the same arguments in the derivation of (B.18) and (B.19), we have

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{S}}_i^2 = O_p(N^{-1}T^{-2}) + O_p(N^{-2}T^{-1}) + O_p(T^{-3}).\tag{G.17}$$

and further

$$\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{R}}_i^2 = O_p(\frac{1}{NT}) + O_p(\frac{1}{T^2}). \tag{G.18}$$

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

$$\begin{aligned}
j_2 = &\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{\hat{w}_i^2 w_i^2} f_t e_{it}(e_{is}^2 - w_i^2) m_i' - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{w}_i^2 w_i^2}\Big[2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})_i\Big] f_t e_{it} m_i' \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{w}_i^2 w_i^2}\Big[m_i'\Lambda\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\Lambda' m_i\Big] f_t e_{it} m_i' \\
&+ \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{\hat{w}_i^2 w_i^2}\tilde{\mathcal{R}}_i f_t e_{it} m_i' = j_{21} + j_{22} + j_{23} + j_{24}, \quad \text{say.}
\end{aligned}$$

The term j_{24} is bounded in norm by

$$C^5\Big[\frac{1}{N}\sum_{i=1}^{N}\|\tilde{\mathcal{R}}_i\|^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}\Big\|^2\Big]^{1/2},$$

which is $O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2})$ by (G.18). Similarly by

$$\frac{1}{N}\sum_{i=1}^{N}\Big\|2m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})_i\Big\|^2 = O_p(N^{-2}), \tag{G.19}$$

and

$$\frac{1}{N}\sum_{i=1}^{N}\Big\|m_i'\Lambda\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\Lambda' m_i\Big\|^2 = O_p(N^{-2}), \tag{G.20}$$

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

$$\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{1}{w_i^4} f_t e_{it}(e_{is}^2 - w_i^2) m_i' - \frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2 w_i^4} f_t e_{it}(e_{is}^2 - w_i^2) m_i'.$$

The first term of the above expression is $O_p(N^{-1/2}T^{-1})$. The second term is bounded in norm by

$$C^5\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big\|\frac{1}{T}\sum_{t=1}^{T} f_t e_{it}\Big\|^2 \cdot \Big\|\frac{1}{T}\sum_{t=1}^{T} e_{is}^2 - w_i^2\Big\|^2\Big]^{1/2},$$

which is $O_p(T^{-3/2})$. By the preceding results, we have

$$\frac{1}{NT}\sum_{t=1}^{T} f_t e_t' \hat{\mathbb{W}}^{-1} M = \frac{1}{NT}\sum_{t=1}^{T} f_t e_t' \mathbb{W}^{-1} M + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{T^{3/2}}). \quad \text{(G.21)}$$

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 G2.3 (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 (G.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 (G.15), l_1 can be rewritten as

$$\begin{aligned} l_1 = -\frac{1}{N}\sum_{i=1}^{N}\frac{\hat{w}_i^2 - w_i^2}{w_i^4} m_i m_i' &= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(e_{it}^2 - w_i^2) m_i m_i' \\ -\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}&\Big[m_i'\hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})\hat{\mathbb{W}}^{-1} M \hat{\Lambda}\hat{\mathbb{P}}_N^{-1}\hat{\Lambda}' m_i\Big] \\ &+\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}\Big[2 m_i'\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}(\mathbb{O} - \mathbb{W})_i\Big] \\ +2\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_i^4}\mathrm{tr}\Big[\hat{\Lambda}\hat{\mathbb{G}}\hat{\Lambda}' M' \hat{\mathbb{W}}^{-1}&\frac{1}{T}\sum_{t=1}^{T}[e_t e_{it} - E(e_t e_{it})] m_i'\Big] m_i m_i' - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_i^4}\tilde{\mathcal{S}}_i m_i m_i' \\ &= l_{11} + \cdots + l_{15}, \quad \text{say.} \end{aligned}$$

First consider l_{12}. Using the argument to prove (c), we have

$$l_{12} = -\frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4} m_i'\Lambda\mathbb{P}_N^{-1}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O} - \mathbb{W})\mathbb{W}^{-1} M \Lambda\mathbb{P}_N^{-1}\Lambda' m_i + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{N^2}).$$

Similarly, by the fact that $[m_i'\Lambda\mathbb{G}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O} - \mathbb{W})_i] = O_p(N^{-1})$, we have

$$l_{13} = \frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4} 2 m_i'\Lambda\mathbb{G}\Lambda' M' \mathbb{W}^{-1}(\mathbb{O} - \mathbb{W})_i + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{N^2}).$$

Then consider l_{14}, whose (v, u) element $(v, u = 1, \dots, k)$ 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 G2.3(a). The last term l_{15} is bounded by (using (G.17))

$$C^6\Big[\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{S}}_i^2\Big]^{1/2} = O_p(N^{-1}T^{-1/2}) + O_p(N^{-1/2}T^{-1}) + O_p(T^{-3/2}).$$

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 (G.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 (G.18). The term l_{23} is bounded in norm by

$$C^8\frac{1}{N}\sum_{i=1}^{N}\tilde{\mathcal{R}}_i^2 = O_p(\frac{1}{NT}) + O_p(\frac{1}{T^2}).$$

Similarly, by (G.19) and (G.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

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{w_i^6}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2)\Big]^2 m_i m_i' - \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{w}_i^2 - w_i^2}{\hat{w}_i^2 w_i^6}\Big[\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2)\Big]^2 m_i m_i'$$

The first term of the above expression is equal to

$$\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i' + O_p(N^{-1/2}T^{-1}).$$

where ϖ_i^2 is defined in Lemma G2.5. The second term is bounded in norm by

$$C^{10}\Big[\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i^2 - w_i^2)^2\Big]^{1/2}\Big[\frac{1}{N}\sum_{i=1}^{N}\Big|\frac{1}{T}\sum_{t=1}^{T}(e_{it}^2 - w_i^2)\Big|^4\Big]^{1/2} = O_p(T^{-3/2}).$$

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

$$l_2 = \frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i' + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}).$$

Combining the preceding results on l_1 and l_2, we have result (e). $\square$

Proof of Theorem 9.2. In order to derive the asymptotic representation of $\hat{\Lambda}$, we need to derive the asymptotic behavior of $\mathbb{A}$ first. By equation (G.9), together with Lemma G2.3 (a)(c)(d), Lemma G2.4 and Lemma G2.5 (d), we have

$$\mathbb{A} + \mathbb{A}' = \eta_1 + \eta_1' + \xi_1 + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}),$$

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 of the above equation, we get

$$\text{vech}(\mathbb{A} + \mathbb{A}') = \text{vech}(\eta_1 + \eta_1') + \text{vech}(\xi_1) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}).$$

further implying

$$2D_r^+\text{vec}(\mathbb{A}) = 2D_r^+\text{vec}(\eta_1) + D_r^+\text{vec}(\xi_1) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}). \tag{G.22}$$

where D_r^+ is defined the same as in Theorem 4.2. The above equation has $\frac{r(r+1)}{2}$ restrictions. Then by the identification condition IC$''$1, 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

$$\text{Ndg}\Big\{\Lambda'(\frac{1}{N}M'\mathbb{W}^{-1}M)\Lambda - \hat{\Lambda}'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)\hat{\Lambda}\Big\} = 0,$$

further implying (by adding and subtracting terms)

$$\text{Ndg}\Big\{(\hat{\Lambda} - \Lambda)'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)\hat{\Lambda} + \hat{\Lambda}'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)(\hat{\Lambda} - \Lambda) \tag{G.23}$$

$$-(\hat{\Lambda} - \Lambda)'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)(\hat{\Lambda} - \Lambda) + \Lambda'\Big[\frac{1}{N}M'(\hat{\mathbb{W}}^{-1} - \mathbb{W}^{-1})M\Big]\Lambda\Big\} = 0.$$

Using Lemma G2.5(e) and $\hat{\Lambda} - \Lambda = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) + O_p(N^{-1})$ from Theorem 9.1, we have

$$\text{Ndg}\Big\{\hat{\Lambda}'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)(\hat{\Lambda} - \Lambda) + (\hat{\Lambda} - \Lambda)'(\frac{1}{N}M'\hat{\mathbb{W}}^{-1}M)\hat{\Lambda}\Big\}$$

$$= \mathrm{Ndg}\{\zeta_1 - \mu_1 + \xi_2\} + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}).$$

where

$$\zeta_1 = \Lambda'\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{m_i m_i'}{w_i^4}(e_{it}^2 - w_i^2)\Big]\Lambda, \quad \mu_1 = \Lambda'\Big[\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i'\Big]\Lambda,$$

and

$$\xi_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}m_i'\Lambda\mathbb{P}_N^{-1}\Lambda' M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}_N^{-1}\Lambda' m_i - \frac{1}{N}\sum_{i=1}^{N}\frac{m_i m_i'}{w_i^4}2m_i'\Lambda\mathbb{G}\Lambda' M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})_i,$$

with $\varpi_i^2 = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}E\big[(e_{it}^2 - w_i^2)(e_{is}^2 - w_i^2)\big]$. With the same definition of $\mathcal{D}$ as given in Theorem 4.2, together with the definition of $\mathbb{P}$, the preceding equation can be rewritten as

$$\mathrm{veck}(\mathbb{A}\mathbb{P} + \mathbb{P}\mathbb{A}') = \mathrm{veck}(\zeta_1 - \mu_1 + \xi_2) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}),$$

or equivalently

$$\mathcal{D}\mathrm{vec}(\mathbb{A}\mathbb{P} + \mathbb{P}\mathbb{A}') = \mathcal{D}\mathrm{vec}(\zeta_1 - \mu_1 + \xi_2) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}).$$

Furthermore, we can rewrite the above equation as

$$\mathcal{D}[(\mathbb{P}\otimes I_r) + (I_r\otimes\mathbb{P})K_r]\mathrm{vec}(\mathbb{A}) = \mathcal{D}\mathrm{vec}(\zeta_1 - \mu_1 + \xi_2) + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}), \tag{G.24}$$

where K_r is defined the same as in Theorem 4.2. The above equation has $\frac{r(r-1)}{2}$ restrictions. Then combining (G.22) and (G.24), we have

$$\begin{aligned}\begin{bmatrix} 2D_r^+ \\ \mathcal{D}[(\mathbb{P}\otimes I_r) + (I_r\otimes\mathbb{P})K_r] \end{bmatrix}\mathrm{vec}(\mathbb{A}) &= \begin{bmatrix} 2D_r^+\mathrm{vec}(\eta_1) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{D}\mathrm{vec}(\zeta_1) \end{bmatrix} - \begin{bmatrix} 0 \\ \mathcal{D}\mathrm{vec}(\mu_1) \end{bmatrix} \\ &+ \begin{bmatrix} D_r^+\mathrm{vec}(\xi_1) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathcal{D}\mathrm{vec}(\xi_2) \end{bmatrix} \\ &+ O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}).\end{aligned} \tag{G.25}$$

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 4.2, the above equation can be rewritten as

$$\begin{aligned}\mathbb{D}_1^\dagger\mathrm{vec}(\mathbb{A}) &= \mathbb{D}_2\mathrm{vec}(\eta_1) + \mathbb{D}_3\mathrm{vec}(\zeta_1) - \mathbb{D}_3\mathrm{vec}(\mu_1) + \frac{1}{2}\mathbb{D}_2\mathrm{vec}(\xi_1) + \mathbb{D}_3\mathrm{vec}(\xi_2) \\ &+ O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{NT}}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}).\end{aligned} \tag{G.26}$$

Noticing that

$$\text{vec}(\eta_1) = \text{vec}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}f_t e_{it} m_i'\Lambda\mathbb{P}^{-1}\Big] = (\mathbb{P}^{-1}\Lambda'\otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it}$$

$$\text{vec}(\zeta_1) = \text{vec}\Big[\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{m_i m_i'}{w_i^4}(e_{it}^2 - w_i^2)\Lambda\Big] = (\Lambda\otimes\Lambda)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(m_i\otimes m_i)(e_{it}^2 - w_i^2)$$

$$\text{vec}(\mu_1) = \text{vec}\Big[\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}m_i m_i'\Lambda\Big] = (\Lambda\otimes\Lambda)'\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}(m_i\otimes m_i)$$

$$\text{vec}(\xi_1) = \frac{1}{N}\Big((\mathbb{P}^{-1}\Lambda')\otimes(\mathbb{P}^{-1}\Lambda')\Big)\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\mathbb{O}_{ij}}{w_i^2 w_j^2}(m_j\otimes m_i)$$

and

$$\text{vec}(\xi_2) = \frac{1}{N^2}\sum_{i=1}^{N}\frac{\varsigma_i}{w_i^4}(m_i\otimes m_i)$$

where

$$\varsigma_i = \frac{1}{N}m_i'\Lambda\mathbb{P}^{-1}\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})\mathbb{W}^{-1}M\Lambda\mathbb{P}^{-1}\Lambda'm_i - 2m_i'\Lambda\mathbb{G}_N\Lambda'M'\mathbb{W}^{-1}(\mathbb{O}-\mathbb{W})_i,$$

we can further rewrite (G.26) to get the asymptotic behavior of $\mathbb{A}$ as following

$$\begin{aligned}\text{vec}(\mathbb{A}) = {}& (\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_2(\mathbb{P}^{-1}\Lambda'\otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it} \\ &+ (\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_3(\Lambda\otimes\Lambda)'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(m_i\otimes m_i)(e_{it}^2 - w_i^2) \\ &- (\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_3(\Lambda\otimes\Lambda)'\frac{1}{NT}\sum_{i=1}^{N}\frac{\varpi_i^2}{w_i^6}(m_i\otimes m_i) \\ &+ \frac{1}{2}(\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_2\frac{1}{N}\Big((\mathbb{P}^{-1}\Lambda')\otimes(\mathbb{P}^{-1}\Lambda')\Big)\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\mathbb{O}_{ij}}{w_i^2 w_j^2}(m_j\otimes m_i) \\ &+ (\mathbb{D}_1^{\dagger})^{-1}\mathbb{D}_3\frac{1}{N^2}\sum_{i=1}^{N}\frac{\varsigma_i}{w_i^4}(m_i\otimes m_i) \\ &+ O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{N}T}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}).\end{aligned} \tag{G.27}$$

Next consider equation (G.10), which is derived from the first order condition of $\hat{\Lambda}$. By Lemma G2.3 (f)(g) and Lemma G2.5 (a)(b)(c), we have

$$\begin{aligned}\hat{\Lambda}' - \Lambda' = {}& -\mathbb{A}'\Lambda' + \frac{1}{NT}\sum_{t=1}^{T}f_t e_t'\mathbb{W}^{-1}M\mathbb{R}^{-1} + \mathbb{P}^{-1}\Lambda'\frac{1}{NT}M'\mathbb{W}^{-1}\sum_{t=1}^{T}e_t f_t'\Lambda' \\ &+ \xi_3 + O_p(\frac{1}{N\sqrt{T}}) + O_p(\frac{1}{\sqrt{N}T}) + O_p(\frac{1}{T^{3/2}}) + O_p(\frac{1}{N^2}).\end{aligned} \tag{G.28}$$

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 (G.28) and noticing that
$$\begin{aligned}\operatorname{vec}\Big[\frac{1}{NT}\sum_{t=1}^{T} f_t e_t'\mathbb{W}^{-1}M\mathbb{R}^{-1}\Big] &= \operatorname{vec}\Big[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2} f_t e_{it} m_i'\mathbb{R}^{-1}\Big]\\ &= (\mathbb{R}^{-1}\otimes I_r)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it},\end{aligned}$$

$$\begin{aligned}\operatorname{vec}\Big[\mathbb{P}^{-1}\Lambda'\frac{1}{NT}M'\mathbb{W}^{-1}\sum_{t=1}^{T} e_t f_t'\Lambda'\Big] &= \operatorname{vec}\Big[\mathbb{P}^{-1}\Lambda'\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2} m_i e_{it} f_t'\Lambda'\Big]\\ &= K_{kr}\operatorname{vec}\Big[\Lambda\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2} f_t e_{it} m_i'\Lambda\mathbb{P}^{-1}\Big]\\ &= K_{kr}[(\mathbb{P}^{-1}\Lambda')\otimes\Lambda]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it},\end{aligned}$$

and
$$\operatorname{vec}(\xi_3) = \frac{1}{N}\Big((\mathbb{R}^{-1})\otimes(\mathbb{P}^{-1}\Lambda')\Big)\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\frac{\mathbb{O}_{ij}}{w_i^2 w_j^2}(m_j\otimes m_i),$$
where K_{kr} is defined the same as in Theorem 4.2, we have
$$\begin{aligned}\operatorname{vec}(\hat\Lambda'-\Lambda') = {}& \Big[K_{kr}[(\mathbb{P}^{-1}\Lambda')\otimes\Lambda]+\mathbb{R}^{-1}\otimes I_r\Big]\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it} \qquad \text{(G.29)}\\ &- K_{kr}(I_r\otimes\Lambda)\operatorname{vec}(\mathbb{A})+\operatorname{vec}(\xi_3)+O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T^{3/2}})+O_p(\frac{1}{N^2}).\end{aligned}$$
Plug (G.27) into (G.29), then we have
$$\begin{aligned}\operatorname{vec}(\hat\Lambda'-\Lambda') = {}& \mathbb{B}_1^\dagger\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^2}(m_i\otimes f_t)e_{it} - \mathbb{B}_2^\dagger\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{w_i^4}(m_i\otimes m_i)(e_{it}^2-w_i^2)\\ &+\frac{1}{T}\Delta^\dagger+\frac{1}{N}\Pi^\dagger+O_p(\frac{1}{N\sqrt{T}})+O_p(\frac{1}{\sqrt{NT}})+O_p(\frac{1}{T^{3/2}})+O_p(\frac{1}{N^2}), \qquad \text{(G.30)}\end{aligned}$$
where $\mathbb{B}_1^\dagger, \mathbb{B}_2^\dagger, \Delta^\dagger$ and $\Pi^\dagger$ are defined in the paragraph before Theorem 9.2. This completes the proof of Theorem 9.2. □

Proof of Theorem 9.3. Given the results as in Theorem 9.2 and let $N, T\to\infty$ and $N/T^2\to 0$ and $T/N^3\to 0$, we can derive the following limiting distribution
$$\sqrt{NT}\Big[\operatorname{vec}(\hat\Lambda'-\Lambda')-\frac{1}{T}\Delta^\dagger-\frac{1}{N}\Pi^\dagger\Big]\xrightarrow{d} N(0,\Xi),$$
where $\Xi = \lim_{N\to\infty}\Xi_{NT}$, and Ξ_{NT} is defined in Theorem 9.3. This completes the proof of Theorem 9.3. □

Proof of Theorem 9.4. From equation (G.15) and the analysis in the proof of Lemma G2.5(e), we know both the second and third terms on the right hand side of (G.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 9.4. Hence we prove Theorem 9.4. □