Journal of Business & Economic Statistics

ISSN: 0735-0015 (Print) 1537-2707 (Online) Journal homepage: http://www.tandfonline.com/loi/ubes20

Estimation and Inference of FAVAR Models

Jushan Bai, Kunpeng Li & Lina Lu

Supplementary materials for this article are available online. Please go to http://tandfonline.com/r/JBES

Estimation and Inference of FAVAR Models

Jushan Bai
Department of Economics, Columbia University, New York, NY 10027, and School of Finance, Nankai University, Tianjin, China (jushan.bai@columbia.edu)

Kunpeng Li
International School of Economics and Management, Capital University of Economics and Business, Beijing, China (kunpenglithu@126.com)

Lina Lu
Department of Economics, Columbia University, New York, NY 10027 (lina.lu@columbia.edu)

The factor-augmented vector autoregressive (FAVAR) model is now widely used in macroeconomics and finance. In this model, observable and unobservable factors jointly follow a vector autoregressive process, which further drives the comovement of a large number of observable variables. We study the identification restrictions for FAVAR models, and propose a likelihood-based two-step method to estimate the model. The estimation explicitly accounts for factors being partially observed. We then provide an inferential theory for the estimated factors, factor loadings, and the dynamic parameters in the VAR process. We show how and why the limiting distributions are different from the existing results. Supplementary materials for this article are available online.

KEY WORDS: High-dimensional analysis; Identification restrictions; Impulse response; Inferential theory; Likelihood-based analysis; VAR.

1. INTRODUCTION

Since the seminal work of Sims (1980), vector autoregressive (VAR) models have played an important role in macroeconomic analysis. Because the number of parameters in a VAR system increases rapidly with the number of variables, there is a degree-of-freedom problem when too many variables are included in the system. On the other hand, too few variables may not fully capture the dimension of the structural shocks. These problems may explain some puzzling empirical results in the body of VAR research. For example, various studies commonly find that a contractionary monetary policy often leads to an increase of the price level, rather than a decrease as the standard economic theory alleges (see Sims 1992; Christiano, Eichenbaum and Evans 1999). Sims (1992) proposed a plausible interpretation of this puzzle, suggesting that it results from the VAR analysis not fully capturing the information. Including more series in a VAR model is limited because of the loss of degrees of freedom. (The Bayes method is alternatively considered (Doan, Litterman, and Sims 1984; Litterman 1986; Sims 1993), and by imposing some prior restrictions, the usual VAR model can accommodate more variables (e.g., Leeper, Sims, and Zha 1996).) Furthermore, as Stock and Watson (2005) pointed out, it is doubtful that the larger VAR models with some potentially incredible restrictions would be superior to the smaller ones.

Bernanke, Boivin, and Eliasz (2005) proposed a factor-augmented vector autoregressive (FAVAR) model to address the dilemma arising from the information deficiency and the degree-of-freedom problem in traditional VAR models. In contrast with such models, the FAVAR model includes unobserved low-dimensional factors in the autoregression. These factors, which may not be captured by some specific macroeconomic aggregates, are thought to contain the bulk of information about an economy. With inclusion of these unobserved factors, the FAVAR model is of rich information, but remains tractable in terms of the number of parameters, owing to the low dimension of the factors. Such approach of using a small number of factors to summarize useful information in a large number of indicators have been used in many papers, for instance, Bernanke and Boivin (2003) and Stock and Watson (2002). The FAVAR model is now widely used in economic applications. (For example, Boivin, Giannoni, and Mihov (2009), Bianchi, Mumtaz, and Surico (2009), Forni and Gambetti (2010), Moench (2008), and Ludvigson and Ng (2009), to name a few. Large dimensional factor models are also increasingly used outside macroeconomics and finance, for example, Fan, Liao, and Mincheva (2011), Fan, Liao and Mincheva (2013), and Tsai and Tsay (2010).) Despite its wide applicability, important issues remain to be addressed.

We first derive the number of restrictions needed in the presence of observable factors, and then consider how to impose these restrictions. Two types of restrictions may be considered. One type involves restrictions on the sample moments of factor process, the other involves restrictions on the population moments of the factor process. The first type is more appropriate for factors being a sequence of fixed constants, for example, Bai and Li (2012). The second type is more appropriate for factors being a random sequence. Similar issue was discussed by Anderson (2003, p. 571). In FAVAR models, since the factors are stochastic processes, restrictions on population variance are more reasonable than on sample variance. An important result of this article is that the two types of restrictions, although asymptotically equivalent, lead to different limiting distributions for the estimated factors and factor loadings, as well as different

Journal of Business & Economic Statistics
October 2016, Vol. 34, No. 4
DOI: 10.1080/07350015.2015.1111222

limiting distributions for the estimated parameters in the VAR process.

The second issue is estimation and the related inferential theory. In the FAVAR literature, Bernanke, Boivin, and Eliasz (2005) and Boivin, Giannoni, and Mihov (2009) suggested a two-step method to estimate an FAVAR model, in which the factors are extracted first and their dynamics are estimated next. There are no studies on the inferential theory of the FAVAR model. The deficiency in this respect makes it difficult to construct the confidence intervals for the impulse response function and to interpret the subsequent economic analysis. Possibly for this reason, Bernanke, Boivin, and Eliasz (2005) also considered a Bayesian method to estimate the model. However, the burdensome computation procedure of the Markov chain Monte Carlo (MCMC) method in this context is formidable for many researchers.

In this article, we consider the identification, estimation, and inferential theory of the FAVAR models. We contribute to the FAVAR literature in several ways. First, we investigate the identification problem of the FAVAR model. Due to the presence of partially observable factors, the identification problem here differs from those in standard factor models. We consider three sets of identification conditions. Unlike the usual identification conditions that are imposed on the sample variance of factors, we put the conditions on the variance of innovations to factors. These conditions are similar to those in the standard structural VAR literature. Second, we propose a likelihood-based two-step method to estimate the FAVAR model, which explicitly takes into account of partial factors being observed. Using maximum likelihood (ML) method instead of principal components (PC) method in the first step gives a better estimation of unobserved factors. (See Bai and Li (2016) for a comparison of finite sample performance of the ML and PC methods.) In addition, we find that the iterative estimation procedure advocated by Boivin, Giannoni, and Mihov (2009) can be avoided.

Third, we establish the statistical theory of the two-step estimators including consistency, convergence rates, and the asymptotic representations. We also give an inferential theory for the impulse response functions. Based on this theory, the confidence intervals of the impulse response function can be easily constructed.

There are several studies related to our work. Stock and Watson (2005) considered the identification and estimation issues in the dynamic factor models. Their identification strategies share with ours the same feature that partial conditions are imposed on the variance of innovations. But the remaining conditions are different: their conditions are imposed on the vector moving average representation and ours are imposed on the original factor representation. Which identification strategy is preferred depends on specific applications. Bernanke, Boivin, and Eliasz (2005) suggested a timing-exclusion strategy for identification. Their strategy may lead to over-identification. Han (2015) proposed a statistic to test the over-identification restrictions. There are additional studies considering the bootstrap method to construct confidence intervals for factor-augmented models, such as Goncalves and Perron (2014), Shintani and Guo (2011), and Yamamoto (2011). Our theoretical results also pave ways for future studies in this direction.

The rest of the article is arranged as follows. Section 2 introduces the FAVAR model with its identification problem, and examines three sets of identification restrictions; and presents some regularity conditions. Section 3 states our two-step estimation procedures. Section 4 presents all the asymptotic properties of our estimators. Section 5 focuses the impulse response function and its confidence intervals. Section 6 investigates the finite sample properties of our estimators. Section 7 concludes. Technical proofs are delivered in the online appendix. Throughout the article, the norm of a vector or matrix is that of Frobenius, that is, $\|A\| = \sqrt{[\text{tr}(A'A)]}$ for vector or matrix A.

2. THE FAVAR MODELS

Let g_t be a vector of observable factors, and f_t be a vector of latent factors, both of low dimension. The FAVAR model assumes that g_t and f_t jointly follow a VAR process. That is, let $h_t = (f_t', g_t')'$, then h_t is characterized by a VAR(K) process for some K,

$$h_t = \Phi_1 h_{t-1} + \Phi_2 h_{t-2} + \cdots + \Phi_K h_{t-K} + u_t, \tag{2.1}$$

where $\Phi_1, \Phi_2, \ldots, \Phi_K$ are matrices of coefficients. In general, neither f_t nor g_t alone is a finite-order VAR process. The FAVAR model further assumes that a large number of observable variables $z_t = (z_{1t}, z_{2t}, \ldots, z_{Nt})'$, dimension of $N \times 1$, is affected by h_t through a factor model

$$z_t = [\Lambda \quad \Gamma]\begin{bmatrix} f_t \\ g_t \end{bmatrix} + e_t, \tag{2.2}$$

where Λ and Γ are the factor loadings with $\Lambda = (\lambda_1, \ldots, \lambda_N)'$ and $\Gamma = (\gamma_1, \ldots, \gamma_N)'$, and $e_t = (e_{1t}, e_{2t}, \ldots, e_{Nt})'$ is the idiosyncratic error, where λ_i is of dimension $r_1 \times 1$ and γ_i of $r_2 \times 1$, for all $i = 1, 2, \ldots, N$. Throughout, we assume f_t is of dimension $r_1 \times 1$, g_t of $r_2 \times 1$, and h_t of $r = r_1 + r_2$. We consider estimating the factors (f_t) and factor loadings, the variance of the idiosyncratic errors e_{it}, and the dynamic parameters in the h_t process, and derive their limiting distributions under various identification restrictions.

Model (2.1)–(2.2) is the FAVAR model proposed by Bernanke, Boivin, and Eliasz (2005). Equation (2.1) is a standard specification of VAR(K) model, except that the variables f_t are unobservable. The inclusion of unobservable factors is crucial to the FAVAR model. These unobservable factors usually capture the information of some structural shocks that are important to the economy but cannot be well represented by specific macroeconomic aggregates. As mentioned before, omitting unknown structure shocks may be a primary reason for the failure of the traditional VAR model in some empirical applications. Equation (2.2) specifies that the common factors h_t are related to the observable data z_t by a factor model. This approach is a plausible way to model the relation between observable variables z_t and the latent variable f_t, given the diffusion nature of common shocks in h_t. The FAVAR model can be considered as a special case of Forni et al. (2000), but with more structures.

2.1 The Number of Identification Restrictions Needed

Model (2.1)–(2.2) cannot be fully identified without additional restrictions. To see this, for any invertible $r_1 \times r_1$ matrix

M_{11} and $r_1 \times r_2$ matrix M_{12}, the model can be written as

$$z_t = \Lambda f_t + \Gamma g_t + e_t = \underbrace{(\Lambda M_{11})}_{\Lambda^*} \underbrace{(M_{11}^{-1} f_t - M_{11}^{-1} M_{12} g_t)}_{f_t^*} + \underbrace{(\Gamma + \Lambda M_{12})}_{\Gamma^*} g_t + e_t. \tag{2.3}$$

Then we obtain two observably equivalent models. Since the total number of free parameters of M_{11} and M_{12} is $r_1^2 + r_1 r_2$, we need at least $r_1^2 + r_1 r_2$ restrictions to identity parameters. A subsequent question is whether $r_1^2 + r_1 r_2$ restrictions are enough. To answer this question, we first define some notations for ease of exposition. Let

$$F = (f_1, f_2, \ldots, f_T)', \qquad G = (g_1, g_2, \ldots, g_T)',$$
$$H = (h_1, h_2, \ldots, h_T)' = [F, G].$$

The following proposition shows that the preceding question has a definite answer.

Proposition 1. Suppose that H is of full column rank, the number of restrictions needed to fully identify model (2.2)–(2.1) is $(r_1^2 + r_1 r_2)$.

Proof. Let M be any invertible $r \times r$ rotation matrix, partitioned as

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix},$$

where M_{11}, M_{22} are $r_1 \times r_1$ and $r_2 \times r_2$ square matrices, respectively. Then Equation (2.2) can be written as

$$z_t = [\Lambda\ \Gamma] \begin{bmatrix} f_t \\ g_t \end{bmatrix} + e_t = [\Lambda \Gamma] \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}^{-1} \times \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} f_t \\ g_t \end{bmatrix} + e_t.$$

Let $h_t^\dagger = M h_t$. If M is a qualified rotation matrix, the lower r_2 elements of $h_t^\dagger$ should be g_t. This gives

$$\begin{bmatrix} f_t^\dagger \\ g_t \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} f_t \\ g_t \end{bmatrix},$$

implying $g_t = M_{21} f_t + M_{22} g_t$, or equivalently

$$[M_{21}\ (M_{22} - I_{r_2})] \begin{bmatrix} f_t \\ g_t \end{bmatrix} = 0,$$

for $t = 1, 2, \ldots, T$. The above result is equivalent to

$$[M_{21}\ (M_{22} - I_{r_2})] H' = 0.$$

If H is of full column rank, by post-multiplying $H(H'H)^{-1}$, we have $M_{21} = 0$, $M_{22} = I_{r_2}$. This result indicates that, to fully identify the parameters, we only need to uniquely determine the matrix M_{11} and M_{12}, whose number of free parameters is exactly $r_1^2 + r_1 r_2$. This proves the proposition. □

2.2 Identification Restrictions

The identification problem brings advantages and disadvantage to the FAVAR model. On one hand, it causes difficulties in interpreting the model in a universal way; on the other hand, the model has flexibility to fit specific situations through a careful design of the identification strategy. In what follows, we consider three sets of identification restrictions, which we think are of practical relevance. We first introduce the following notations:

$$u_t = \begin{bmatrix} \varepsilon_t \\ \boldsymbol{v}_t \end{bmatrix}; \quad \Omega = E(u_t u_t') = \begin{bmatrix} E(\varepsilon_t \varepsilon_t') & E(\varepsilon_t \boldsymbol{v}_t') \\ E(\boldsymbol{v}_t \varepsilon_t') & E(\boldsymbol{v}_t \boldsymbol{v}_t') \end{bmatrix} = \begin{bmatrix} \Omega_{\varepsilon\varepsilon} & \Omega_{\varepsilon \boldsymbol{v}} \\ \Omega_{\boldsymbol{v}\varepsilon} & \Omega_{\boldsymbol{v}\boldsymbol{v}} \end{bmatrix}$$

$$h_t = \begin{bmatrix} f_t \\ g_t \end{bmatrix}; \quad \boldsymbol{\Delta} = E(h_t h_t') = \begin{bmatrix} E(f_t f_t') & E(f_t g_t') \\ E(g_t f_t') & E(g_t g_t') \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Delta}_{ff} & \boldsymbol{\Delta}_{fg} \\ \boldsymbol{\Delta}_{gf} & \boldsymbol{\Delta}_{gg} \end{bmatrix}, \tag{2.4}$$

where ε_t and $\boldsymbol{v}_t$ are the innovations corresponding to f_t and g_t, respectively. We consider the following three sets of identification restrictions.

- IRa. The underlying parameter values θ satisfy: $\Omega_{\varepsilon\varepsilon} = I_{r_1}$, $\Omega_{\varepsilon \boldsymbol{v}} = 0$, and $\frac{1}{N} \Lambda' \Sigma_{ee}^{-1} \Lambda = Q$, where Q is a diagonal matrix with its diagonal elements being distinct and arranged in descending order.
- IRb. The underlying parameter values θ satisfy: $\Omega_{\varepsilon\varepsilon} = I_{r_1}$, $\Omega_{\varepsilon \boldsymbol{v}} = 0$ and Λ_1 is a lower triangular matrix, where Λ_1 is the upper $r_1 \times r_1$ submatrix of Λ.
- IRc. The underlying parameter values θ satisfy: $\Omega_{\varepsilon \boldsymbol{v}} = 0$ and $\Lambda_1 = I_{r_1}$, where Λ_1 is the upper $r_1 \times r_1$ submatrix of Λ.

Each set of identification restrictions imposes $r_1^2 + r_1 r_2$ restrictions. There are no restrictions on $\Omega_{\boldsymbol{v}\boldsymbol{v}}$ as $\boldsymbol{v}_t$ is the reduced form residual from the observable g_t. In the next subsection, we explain why it is possible to assume $\Omega_{\varepsilon \boldsymbol{v}} = 0$.

Remark 1. In factor analysis, Anderson (2003, p. 571) considers both types of restrictions $E(f_t f_t') = I_{r_1}$ and $\frac{1}{T}\sum_{t=1}^T f_t f_t' = I_{r_1}$. The former restriction is considered population restriction, and the latter is considered sample version restriction. In our case, since we have dynamics in h_t, the errors ε_t correspond to f_t. Because we assume the errors are random, it is reasonable to make populational assumptions rather than sample version restrictions. However, as we will show, though $E(\varepsilon_t \varepsilon_t') = I_{r_1}$ and $\frac{1}{T}\sum_{t=1}^T \varepsilon_t \varepsilon_t' = I_{r_1}$ are asymptotically equivalent in a certain sense, they imply different distributions for the estimated factor loadings and the estimated factors f_t. The population version restriction implies larger variance than the sample version restriction.

2.3 Discussions on the Identification Restrictions

We give some discussions on the preceding identification restrictions, especially the reason that we can impose the restriction $\Omega_{\varepsilon \boldsymbol{v}} = 0$. Suppose the original FAVAR model is

$$z_t = [\Lambda^\dagger \quad \Gamma^\dagger] \begin{bmatrix} f_t^\dagger \\ g_t \end{bmatrix} + e_t,$$
$$h_t^\dagger = \Phi_1^\dagger h_{t-1}^\dagger + \Phi_2^\dagger h_{t-2}^\dagger + \cdots + \Phi_K^\dagger h_{t-K}^\dagger + u_t^\dagger,$$

where $h_t^\dagger = [\begin{smallmatrix} f_t^\dagger \\ g_t \end{smallmatrix}]$ and $u_t^\dagger = [\begin{smallmatrix} \varepsilon_t^\dagger \\ \boldsymbol{v}_t \end{smallmatrix}]$ with the variance matrix $\Omega^\dagger = E(u_t^\dagger u_t^{\dagger\prime}) = [\begin{smallmatrix} \Omega_{\varepsilon\varepsilon}^\dagger & \Omega_{\varepsilon\boldsymbol{v}}^\dagger \\ \Omega_{\boldsymbol{v}\varepsilon}^\dagger & \Omega_{\boldsymbol{v}\boldsymbol{v}} \end{smallmatrix}]$. Note that this original VAR representation is in a reduced form with $\Omega_{\boldsymbol{v}\varepsilon}^\dagger \neq 0$. Let A be a rotation matrix defined as $A = [\begin{smallmatrix} (\Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\dagger)^{-1/2} & -(\Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\dagger)^{-1/2}\Omega_{\varepsilon\boldsymbol{v}}^\dagger\Omega_{\boldsymbol{v}\boldsymbol{v}}^{-1} \\ 0 & I_{r_2} \end{smallmatrix}]$, then the new FAVAR model after rotation is

$$z_t = \underbrace{[\Lambda^\dagger \quad \Gamma^\dagger]A^{-1}}_{[\Lambda \quad \Gamma]} \cdot \underbrace{A\begin{bmatrix} f_t^\dagger \\ g_t \end{bmatrix}}_{\begin{bmatrix} f_t \\ g_t \end{bmatrix} \equiv h_t} + e_t,$$

$$\underbrace{Ah_t^\dagger}_{h_t} = \underbrace{A\Phi_1^\dagger A^{-1}}_{\Phi_1} \cdot \underbrace{Ah_{t-1}^\dagger}_{h_{t-1}} + \underbrace{A\Phi_2^\dagger A^{-1}}_{\Phi_2} \cdot \underbrace{Ah_{t-2}^\dagger}_{h_{t-2}} + \cdots + \underbrace{A\Phi_K^\dagger A^{-1}}_{\Phi_K} \cdot \underbrace{Ah_{t-K}^\dagger}_{h_{t-K}} + \underbrace{Au_t^\dagger}_{u_t},$$

where we use the notation without $\dagger$ to denote the new parameters. Note that the observable factor g_t and the corresponding innovation $\boldsymbol{v}_t$ do not change. Let Ω be the variance matrix of the new innovation $u_t = [\begin{smallmatrix} \varepsilon_t \\ \boldsymbol{v}_t \end{smallmatrix}]$, then $\Omega = A\Omega^\dagger A' = [\begin{smallmatrix} I_{r_1} & 0 \\ 0 & \Omega_{\boldsymbol{v}\boldsymbol{v}} \end{smallmatrix}]$, where the new innovations satisfy $\Omega_{\varepsilon\varepsilon} = I_{r_1}$ and $\Omega_{\varepsilon\boldsymbol{v}} = 0$. Consequently our imposed identification restrictions on the innovations as stated in the previous subsection are reasonable. The new factor $f_t = (\Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\dagger)^{-1/2} f_t^\dagger - (\Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\dagger)^{-1/2}\Omega_{\varepsilon\boldsymbol{v}}^\dagger\Omega_{\boldsymbol{v}\boldsymbol{v}}^{-1} g_t$ is now a linear combination of $f_t^\dagger$ and g_t. With some appropriate restrictions on the new loadings $[\Lambda \quad \Gamma]$, the factor f_t can now have economic meanings with additional identification restrictions.

The three different identification restrictions in the previous subsection can be interpreted as follows.

IRa requires that $\Lambda'\Sigma_{ee}^{-1}\Lambda$ be diagonal, which is often used in the maximum likelihood estimation, see Lawley and Maxwell (1971). This identification condition is important in terms of the statistical analysis, it can also be of economic interest in some specific cases, as pointed out in Bai and Ng (2013). For example, Λ is block diagonal such as $\Lambda = [\pi_1, 0; 0, \pi_2]$, where π_i is a column vector of N_i elements with $N_1 + N_2 = N$. In this case, the first factor only affects the first N_1 variables, and the second factor only affects the next N_2 variables. Each variable is affected by only a single factor, but we do not need to know which variable is affected by which factor; we have $\Lambda'\Sigma_{ee}^{-1}\Lambda$ being diagonal under arbitrary cross-sectional permutation of individuals.

IRb shares the same feature with IRa by imposing the restrictions on the variance of u_t. In addition, it restricts Λ_1 to being a lower triangular matrix. This allows IRb to endow economic implications with the unobserved factors. Under IRb, only the first unobservable factor affects the first variable, the first two unobservable factors affect the second variable, etc. This scheme somewhat resembles the recursive identification in structural VAR analysis. Through careful selection of the first r_1 variables, the unobservable factors are now explainable. For example, Geweke and Zhou (1996), in their study of pricing error in the Arbitrage Pricing Theory, used the recursive identification conditions.

IRc restricts the upper $r_1 \times r_1$ matrix of the factor loadings Λ to being an identity matrix. Since more restrictions are imposed on the factor loadings Λ, IRc relinquishes the requirement that the innovations to the unobservable factors be orthogonal and have unit variance. Under IRc, the first unobservable factor affects only the first series, the second unobservable factor affects only the second series, etc.

Overall, the identification restrictions considered in this article share the feature that they impose restrictions on the loadings Λ and the variance of the innovations to h_t. This is in contrast with the usual identification conditions in factor models, which impose restrictions on the loadings and the sample variance of factors; see Anderson and Rubin (1956) and Bai and Li (2012) for traditional identification conditions. Imposing restrictions on innovations instead on factors themselves is important and reasonable because the components of f_t are correlated while the innovations ε_t can be assumed uncorrelated, similar to structural analysis.

2.4 Assumptions

To analyze model (2.2)–(2.1), we make the following assumptions:

Assumption A. The factor $h_t = (f_t', g_t')'$ admits a VAR representation (2.1), where u_t is an iid process with $u_t = \Omega^{1/2}\varsigma_t$, where $E(\varsigma_t) = 0$, $\text{var}(\varsigma_t) = I_r$, and $\Omega > 0$, $E(\|\varsigma_t\|^4) < \infty$ and the elements of ς_t are mutually independent. In addition, all the roots of the polynomial $\Phi(L) = I_r - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_K L^K = 0$ are outside of the unit circle.

Assumption B. There exists a positive constant C large enough such that

B.1. $\|\lambda_i\| \leq C < \infty$, $\|\gamma_i\| \leq C < \infty$ for all i.
B.2. $C^{-2} \leq \sigma_i^2 \leq C^2$ for all i, where σ_i^2 is defined in Assumption C.
B.3. $\lim_{N\to\infty} \frac{1}{N}\Lambda'\Sigma_{ee}^{-1}\Lambda = \mathbf{Q}$ exists and is a positive-definite matrix, where Σ_{ee} is defined in Assumption C.

Assumption C. $E(e_t) = 0$; $E(e_t e_t') = \Sigma_{ee} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$; $E(e_{it}^4) < \infty$ for all i and t. The e_{it} are independent over i and t. The $N \times 1$ vector e_t is identically distributed over t. Furthermore, e_{it} is independent with u_s for all i, t, s.

Assumption D. Variances σ_i^2 are estimated in the compact set $[C^{-2}, C^2]$.

Assumption A makes the regularity conditions on factors. It requires factor h_t to be stationary over t. It also guarantees that $H = (h_1, h_2, \ldots, h_T)'$ is of full column rank. So under Assumption A, Proposition 1 holds. Assumption B is made on the factor loadings. This assumption is standard. Notice that Assumption B requires the columns of Λ to be linearly independent; otherwise, $\mathbf{Q}$ will be a singular matrix. Assumption C centers on the idiosyncratic errors. Under Assumption C, the correlations over time and cross-section are ruled out. Meanwhile, the heteroscedasticity over time is also precluded. This assumption can be relaxed to a great extent. In fact, the analysis of this article can be extended to the approximate factor models (Chamberlain and Rothschild 1983). Assumption D requires σ_i^2 to be estimated in a compact set. This assumption is due to the high nonlinearity of the likelihood function, and it is common in the literature for nonlinear problems.

3. ESTIMATION

In this section, we propose a two-step method to estimate the underlying structure parameters that satisfy IRa, IRb, or IRc. Some alternative methods can also be used. Bernanke, Boivin, and Eliasz (2005) considered the MCMC method. Boivin, Giannoni, and Mihov (2009) considered the iterated principal component-ordinary least squares (PC-OLS) method. Our method directly takes into account that g_t is observable, no iteration is necessary. Also, the MLE-based method is more efficient than that of PC-based.

To gain insight into our method, write (2.2) into matrix form as

$$Z = \Lambda F' + \Gamma G' + e. \quad (3.1)$$

Post-multiplying $\mathbb{M}_G = I_T - G(G'G)^{-1}G$, we have

$$Z\mathbb{M}_G = \Lambda F'\mathbb{M}_G + e\mathbb{M}_G.$$

Applying the quasi-maximum likelihood (ML) estimation method to the model, we obtain the quasi-ML estimates (QMLE) $\tilde{\Lambda}$, $\tilde{\Sigma}_{ee}$, and $\tilde{F}$. Let $f_t^\star = R_{11}(f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}g_t)$, where R_{11} is a rotation matrix. It can be shown that $\tilde{f}_t$ consistently estimate $f_t^\star$. To recover f_t from $f_t^\star$ and g_t, we only need to determine $\mathbf{\Delta}_{fg}$ and R_{11}, which is achieved by our identification conditions.

The estimation method is formally stated as follows:

1. Apply quasi-ML method with $Y = Z\mathbb{M}_G$ to get QMLE $\tilde{\lambda}_i, \tilde{\sigma}_i^2$; then calculate $\tilde{F} = Y'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}(\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}$ and $\tilde{\Gamma} = (Z - \tilde{\Lambda}\tilde{F}')G(G'G)^{-1}$, where $\tilde{\Sigma}_{ee} = \operatorname{diag}(\tilde{\sigma}_1^2, \ldots, \tilde{\sigma}_N^2)$.
2. Let $\tilde{h}_t = (\tilde{f}_t', g_t')'$ and run the following regression
$$\tilde{h}_t = \Phi_1\tilde{h}_{t-1} + \Phi_2\tilde{h}_{t-2} + \cdots + \Phi_K\tilde{h}_{st-K} + \text{error} \quad (3.2)$$
to get the estimators $\tilde{\Phi}_1, \tilde{\Phi}_2, \ldots, \tilde{\Phi}_K$.
3. Let $\tilde{u}_t$ be the residuals of the regression (3.2). Calculate $\tilde{\Omega} = \frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{u}_t\tilde{u}_t'$, where $\bar{T} = T - K$ and $\bar{K} = K + 1$. Then $\tilde{\Omega}_{\varepsilon\varepsilon}$, $\tilde{\Omega}_{\varepsilon\boldsymbol{v}}$ and $\tilde{\Omega}_{\boldsymbol{vv}}$ are obtained by the definition. Calculate $\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}} = \tilde{\Omega}_{\varepsilon\varepsilon} - \tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon}$.
4. *Estimation under IRa*: Let $\mathcal{V}$ be the eigenvector matrix of $\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{1/2}(\frac{1}{N}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{1/2}$, whose associated eigenvalues are in descending order. Calculate $\hat{\Lambda} = \tilde{\Lambda}\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{1/2}\mathcal{V}$, $\hat{\Gamma} = \tilde{\Gamma} + \tilde{\Lambda}\tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1}$, $\hat{F} = (\tilde{F} - G\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon})\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{-1/2}\mathcal{V}$. Further construct R as
$$R = \begin{bmatrix} \mathcal{V}'\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{-1/2} & -\mathcal{V}'\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{-1/2}\tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1} \\ 0 & I_{r_2} \end{bmatrix}.$$
Then $\hat{\Phi}_p = R\tilde{\Phi}_pR^{-1}$ for $p = 1, 2, \ldots, K$, and $\hat{\Omega}_{\boldsymbol{vv}} = \tilde{\Omega}_{\boldsymbol{vv}}$.
Estimation under IRb: Let $\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{1/2}\tilde{\Lambda}_1' = \mathcal{QR}$ be the QR decomposition of $\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{1/2}\tilde{\Lambda}_1'$ with $\mathcal{Q}$ an orthogonal matrix and $\mathcal{R}$ an upper triangular matrix, where $\tilde{\Lambda}_1$ is the upper $r_1 \times r_1$ submatrix of $\tilde{\Lambda}$. The parameters are estimated by $\hat{\Lambda} = \tilde{\Lambda}\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{1/2}\mathcal{Q}$, $\hat{\Gamma} = \tilde{\Gamma} + \tilde{\Lambda}\tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1}$, $\hat{F} = (\tilde{F} - G\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon})\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{-1/2}\mathcal{Q}$. Let
$$R = \begin{bmatrix} \mathcal{Q}'\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{-1/2} & -\mathcal{Q}'\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^{-1/2}\tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1} \\ 0 & I_{r_2} \end{bmatrix}.$$
Then $\hat{\Phi}_p = R\tilde{\Phi}_pR^{-1}$ for $p = 1, 2, \ldots, K$, and $\hat{\Omega}_{\boldsymbol{vv}} = \tilde{\Omega}_{\boldsymbol{vv}}$.
Estimation under IRc: The parameters are estimated by $\hat{\Lambda} = \tilde{\Lambda}(\tilde{\Lambda}_1)^{-1}$, $\hat{\Gamma} = \tilde{\Gamma} + \tilde{\Lambda}\tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1}$, and $\hat{F} = (\tilde{F} - G\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon})\tilde{\Lambda}_1'$. Let
$$R = \begin{bmatrix} \tilde{\Lambda}_1 & -\tilde{\Lambda}_1\tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1} \\ 0 & I_{r_2} \end{bmatrix}.$$
Then $\hat{\Phi}_p = R\tilde{\Phi}_pR^{-1}$ for $p = 1, 2, \ldots, K$, and $\hat{\Omega}_{\boldsymbol{vv}} = \tilde{\Omega}_{\boldsymbol{vv}}$, $\hat{\Omega}_{\varepsilon\varepsilon} = \tilde{\Lambda}_1\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}\tilde{\Lambda}_1'$.

Remark 2. The innovations $\boldsymbol{v}_t$ do not involve any identification problem and hence are the same under different identification restrictions, due to the factors g_t being observable. As a result, the estimator $\hat{\Omega}_{\boldsymbol{vv}}$ is the same under different identification restrictions. However, for the innovations ε_t, its variance matrix is restricted to being an identity matrix under IRa and IRb, so we only need estimate $\Omega_{\varepsilon\varepsilon}$ under IRc. The estimator $\hat{\Omega}$ would be useful in the construction of the impulse response function in Section 5.

Remark 3. We explain how we recover f_t from $f_t^\star$ (how to obtain $\hat{f}_t$ from $\tilde{f}_t$) using the given formula above. We take IRc as the example to illustrate. By $f_t^\star = R_{11}(f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}g_t)$, we have $F = (F^\star + G\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf}R_{11}')R_{11}^{-1\prime}$. From the estimation procedure, it is seen that $\tilde{\Lambda}_1^{-1}$ corresponds to R_{11}. Also notice that

$$\begin{bmatrix} f_t \\ g_t \end{bmatrix} = \begin{bmatrix} R_{11}^{-1} & \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} f_t^\star \\ g_t \end{bmatrix} \longrightarrow \begin{bmatrix} \varepsilon_t \\ \boldsymbol{v}_t \end{bmatrix} = \begin{bmatrix} R_{11}^{-1} & \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} \varepsilon_t^\star \\ \boldsymbol{v}_t^\star \end{bmatrix}$$

(notice that $\boldsymbol{v}_t^\star = \boldsymbol{v}_t$), which further implies

$$\begin{bmatrix} \Omega_{\varepsilon\varepsilon} & \Omega_{\varepsilon\boldsymbol{v}} \\ \Omega_{\boldsymbol{v}\varepsilon} & \Omega_{\boldsymbol{vv}} \end{bmatrix} = \begin{bmatrix} * & R_{11}^{-1}\Omega_{\varepsilon\boldsymbol{v}}^\star + \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}\Omega_{\boldsymbol{vv}}^\star \\ \Omega_{\boldsymbol{v}\varepsilon}^\star R_{11}^{-1\prime} + \Omega_{\boldsymbol{vv}}^\star\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf} & \Omega_{\boldsymbol{vv}}^\star \end{bmatrix}.$$

By $\Omega_{\boldsymbol{v}\varepsilon} = 0$, we see that $\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^\star = -\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf}R_{11}'$. So the term $-\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon}$ is an estimator of $\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf}R_{11}'$. This justifies the formula $\hat{F} = (\tilde{F} - G\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon})\tilde{\Lambda}_1'$ in IRc.

Remark 4. The parameters Λ, Γ, Σ_{ee}, $\Phi_1, \ldots, \Phi_k$, and Ω can also be estimated by the state space method using the Kalman smoother as in Watson and Engle (1983), Quah and Sargent (1992), and Doz, Giannone, and Reichlin (2012) (though the latter article considers homoscedastic e_{it}, it can be extended to heteroscedastic errors). But the state space method is computationally more demanding than the two-step method here. That is perhaps the reason that Doz, Giannone, and Reichlin (2011) subsequently also considered a two-step method. Furthermore, it can be shown that, due to the static relationship between z_{it} and h_t, there is no asymptotic efficiency gain by using the Kalman smoother (see Bai and Li (2016)). None of these articles study the limiting distributions of the estimators.

Throughout the article, we use the symbols with a hat to denote the final estimators (e.g., $\hat{\lambda}_i$, $\hat{f}_t$, $\hat{\Phi}_k$) and the symbols with a tilde to denote the intermediate estimators (e.g., $\tilde{\lambda}_i$, $\tilde{f}_t$,

$\tilde{\Phi}_k$). Since σ_i^2 does not have the identification problem, the intermediate estimator and the final estimator are the same. For this reason, we use the two symbols interchangeably, that is, $\hat{\sigma}_i^2 = \tilde{\sigma}_i^2$ and $\hat{\Sigma}_{ee} = \tilde{\Sigma}_{ee}$.

4. ASYMPTOTIC PROPERTIES OF THE ESTIMATORS

In this section, we deliver the asymptotic results on the two-step estimators. The following proposition states that the two-step estimators are individually consistent.

Proposition 2. Under Assumptions A–D, when $N, T \to \infty$, with any one of identification conditions (IRa, IRb, or IRc), we have

$$\hat{\lambda}_i - \lambda_i \xrightarrow{p} 0; \quad \hat{\gamma}_i - \gamma_i \xrightarrow{p} 0; \quad \hat{\sigma}_i^2 - \sigma_i^2 \xrightarrow{p} 0;$$
$$\hat{f}_t - f_t \xrightarrow{p} 0; \quad \hat{\Phi}_k - \Phi_k \xrightarrow{p} 0,$$

for each $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T; k = 1, 2, \ldots, K$.

To give the asymptotic representations for the factor loadings, we introduce the following notations. Let V be an $r_1 \times r_1$ matrix, which is defined as follows:

$$\operatorname{vec}(V) = \begin{cases} \mathbb{B}_Q^{-1}\mathbb{P}_1 D_{r_1}^{+}\frac{1}{T}\sum_{t=\bar{K}}^{T}[\varepsilon_t \otimes \varepsilon_t - \operatorname{vec}(I_{r_1})], & \text{under IRa} \\ \mathbb{D}_2\frac{1}{T}\sum_{t=\bar{K}}^{T}[\varepsilon_t \otimes \varepsilon_t - \operatorname{vec}(I_{r_1})] \\ +\mathbb{D}_3(\Lambda_1 \otimes \mathbf{\Delta}_{\phi\phi})^{-1}\frac{1}{T}\sum_{t=1}^{T}(\xi_t \otimes \phi_t), & \text{under IRb} \\ -(I_{r_1} \otimes \mathbf{\Delta}_{\phi\phi}^{-1})\frac{1}{T}\sum_{t=1}^{T}\xi_t \otimes \phi_t, & \text{under IRc} \end{cases}$$

where D_r is the r-dimensional duplication matrix such that $D_r\operatorname{vech}(M) = \operatorname{vec}(M)$ for any $r \times r$ symmetric matrix M and D_r^{+} is its Moore–Penrose inverse; $\mathbb{B}_Q = [2D_{r_1}^{+\prime}, \big(K_{r_1}'(I_{r_1} \otimes Q) + Q \otimes I_{r_1}\big)\mathbb{D}_1']'$, where $Q = (\Lambda'\Sigma_{ee}^{-1}\Lambda)/N$, K_r is the r-dimensional commutation matrix such that $K_r\operatorname{vec}(M) = \operatorname{vec}(M')$ for any $r \times r$ matrix M and $\mathbb{D}_1$ is the matrix such that $\operatorname{veck}(M) = \mathbb{D}_1\operatorname{vec}(M)$ for any symmetric matrix, and $\operatorname{veck}(M)$ is the operator that stacks the elements of M below the diagonal into a vector; $\mathbb{P}_1 = [I_p, 0_{p\times q}]'$ with $p = (r_1+1)r_1/2$ and $q = r_1(r_1-1)/2$; $\mathbb{D}_2 = K_{r_1}D_{r_1}^{*}(D_{r_1}^{*\prime}S_{r_1}'S_{r_1}D_{r_1}^{*})^{-1}D_{r_1}^{*\prime}S_{r_1}'/2$ where D^{*} is the matrix such that $\operatorname{vec}(M) = D_r^{*}\operatorname{vech}(M)$ for any lower triangular $r \times r$ matrix M and S_{r_1} is the symmetrizer matrix such that $S_r = (I_{r^2} + K_r)/2$; $\mathbb{D}_3 = 2\mathbb{D}_2S_{r_1} - I_{r_1^2}$; Λ_1 is the upper $r_1 \times r_1$ submatrix of Λ; $\mathbf{\Delta}_{\phi\phi} = E(\phi_t\phi_t')$ with $\phi_t = f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}g_t$; $\xi_t = (e_1, e_2, \ldots, e_{r_1t})'$.

Given the consistency, we have the following theorem on the asymptotic representation of the estimator for loadings $\hat{\lambda}_i$:

Theorem 1. Under Assumptions A–D, when $N, T \to \infty$ and $\sqrt{T}/N \to 0$, under IRa, IRb, or IRc, we have,

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) = \sqrt{T}V\lambda_i + \mathbf{\Delta}_{\phi\phi}^{-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\phi_t e_{it}\Big) + o_p(1), \tag{4.1}$$

where $\phi_t = f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}g_t$ and $\mathbf{\Delta}_{\phi\phi} = E(\phi_t\phi_t')$, where $\mathbf{\Delta}_{fg}$ and $\mathbf{\Delta}_{gg}$ are defined in (2.4).

Remark 5. Consider the limiting distribution under IRa. The restrictions under IRa are similar to those for the principal components estimator. The limiting distribution here is different from that of the usual PC in several ways. First because of the presence of observable g_t, the "regressors" f_t is projected onto g_t, and the projection error ϕ_t enters into the distribution. Second, there is an extra term V in the limiting distribution. To better understand this term, consider the situation in which g_t is absent, and the dynamics in h_t is also absent so that $h_t = f_t = \varepsilon_t$. The restriction $E(\varepsilon_t\varepsilon_t') = I_r$ becomes $E(f_tf_t') = I_r$. The limiting distribution under IRa becomes

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) = \sqrt{T}V\lambda_i + \Big(\frac{1}{T}\sum_{t=1}^{T}f_tf_t'\Big)^{-1}\frac{1}{\sqrt{T}} \times \sum_{t=1}^{T}f_te_{it} + o_p(1), \tag{4.2}$$

where V depends on $\frac{1}{T}\sum_{t=1}^{T}f_tf_t' - I_r$. If one assumes the sample version restriction $\frac{1}{T}\sum_t f_tf_t' = I_r$, then the first term disappears. This result is consistent with that of Bai and Li (2012), where the sample version restriction is considered. Thus restrictions on sample covariance and restrictions on population covariance lead to different limiting distributions for the estimated factor loadings. Restrictions on population covariance of f_t imply a larger limiting variance for $\hat{\lambda}_i$. Finally, because we allow dynamics in h_t, the first term V involves the innovations of ε_t rather than f_t.

Also in the absence of g_t and without dynamics in h_t so that $h_t = f_t = \varepsilon_t$, Equation (4.2) is also the asymptotic representation of the PC estimator under the population restriction $E(f_tf_t') = I_r$; but in the definition of V, the matrix $Q = (\Lambda'\Sigma_{ee}^{-1}\Lambda)/N$ is replaced by $\Lambda'\Lambda/N$ because the PC estimator treats the error covariance matrix as an identify matrix. This result is also consistent with that of Bai (2003), where he considers a limiting distribution of the form $\sqrt{T}(\hat{\lambda}_i - R\lambda_i)$ for some rotation matrix R. If we let $R = I_r + V$, then the representation will be equivalent to that of Bai (2003).

Under IRb, the population restriction $E(\varepsilon_t\varepsilon_t') = I_{r_1}$ continues to affect the limiting distribution. Now V itself is composed of two expressions. The second expression in V is analogous to a term in Bai and Li (2012) under IC5 .

Under IRc, there are no restrictions on the population variance of ε_t, and instead, the restrictions are imposed on the factor loadings. The limiting distribution is analogous to that of Bai and Li (2012) under IC1.

Remark 6. Theorem 1 shows that the asymptotic representation for $\hat{\lambda}_i$ under different IRs has a similar expression, which justifies our treatment that the asymptotic properties for $\hat{\lambda}_i$ under different IRs are studied in a unified framework. The symbol ϕ_t in the asymptotic representation is the residual from projecting f_t on g_t. Hence, it is orthogonal with g_t. The expression of V is different under different identification restrictions.

To derive the limiting distribution of $\hat{\lambda}_i$, we consider the covariance between the first and second term on the right-hand side of (4.1). Under IRa, V only involves the variance of the VAR innovations $\varepsilon_t\varepsilon_t'$, and the second term only involves ϕ_te_{it}, so the first and second term are asymptotically independent because of the absence of correlation between $\varepsilon_t\varepsilon_t'$ and ϕ_te_{it}. Uncorrelation implies asymptotic independence because each term is asymptotically normal. Under IRc, we only need to estimate λ_i

with $i > r_1$, so the second term only involves e_{it} where $i > r_1$. But V involves $\xi_t\phi_t'$, where $\xi_t = (e_{1t}, e_{2t}, \ldots, e_{r_1t})$, so the first term is asymptotically independent with the second term. Under IRb, for $i > r_1$, these two terms are asymptotically independent for the same reason as under IRa and IRc. But for $i \le r_1$, these two terms are correlated. Based on the preceding analysis and Theorem 1, we have the following corollary.

Corollary 1. Under the assumptions of Theorem 1, we have

Under IRa:

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} \mathcal{N}\left(0, (\lambda_i' \otimes I_{r_1})\mathbb{B}_{\mathbf{Q}}^{-1}\mathbb{P}_1\left[2(D_{r_1}'D_{r_1})^{-1} + \mathcal{G}\right] \times \mathbb{P}_1'\mathbb{B}_{\mathbf{Q}}^{-1\prime}(\lambda_i \otimes I_{r_1}) + \sigma_i^2\mathbf{\Delta}_{\phi\phi}^{-1}\right).$$

Under IRb: for $i > r_1$,

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} \mathcal{N}\left(0, (\lambda_i' \otimes I_{r_1})\left[\mathbb{D}_2(2I_{r_1^2} + \mathcal{H})\mathbb{D}_2' + \mathbb{D}_3[(\Lambda_1'\Sigma_{\xi\xi}^{-1}\Lambda_1) \otimes \mathbf{\Delta}_{\phi\phi}]^{-1}\mathbb{D}_3'\right](\lambda_i \otimes I_{r_1}) + \sigma_i^2\mathbf{\Delta}_{\phi\phi}^{-1}\right),$$

for $1 \le i \le r_1$,

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} \mathcal{N}\left(0, (\lambda_i' \otimes I_{r_1})\mathbb{D}_2\left[2I_{r_1^2} + \mathcal{H} + 4S_{r_1}[(\Lambda_1'\Sigma_{\xi\xi}^{-1}\Lambda_1) \otimes \mathbf{\Delta}_{\phi\phi}]^{-1}S_{r_1}'\right]\mathbb{D}_2'(\lambda_i \otimes I_{r_1})\right),$$

Under IRc: for $i > r_1$,

$$\sqrt{T}(\hat{\lambda}_i - \lambda_i) \xrightarrow{d} \mathcal{N}\left(0, (\lambda_i'\Sigma_{\xi\xi}\lambda_i + \sigma_i^2)\mathbf{\Delta}_{\phi\phi}^{-1}\right),$$

where $\Sigma_{\xi\xi} = \mathrm{var}(\xi_t)$ with $\xi_t = (e_{1t}, e_{2t}, \ldots, e_{r_1t})'$, $\mathcal{G} = D_{r_1}^{+}\mathscr{D}_{r_1}\mathcal{W}_\varepsilon\mathscr{D}_{r_1}'D_{r_1}^{+\prime}$, and $\mathcal{H} = \mathscr{D}_{r_1}\mathcal{W}_\varepsilon\mathscr{D}_{r_1}'$, where $D_{r_1}^{+} = (D_{r_1}'D_{r_1})^{-1}D_{r_1}'$ is the Moore–Penrose inverse of D_{r_1}, $\mathscr{D}_r$ is an $r^2 \times r$ transformation matrix such that $\mathrm{vec}(M) = \mathscr{D}_r\mathrm{diag}(M)$ for any diagonal r-dimensional matrix M and $\mathcal{W}_\varepsilon$ is an r_1-dimensional diagonal matrix with its ith diagonal element $\kappa_i^* - 3$ where $\kappa_i^* = E(\varepsilon_{it}^4)$. The notations $\mathbb{B}_{\mathbf{Q}}, \mathbb{P}_1, D_{r_1}, \mathbb{D}_2, \mathbb{D}_3, \Lambda_1, S_{r_1}$, and $\mathbf{\Delta}_{\phi\phi}$ are defined in the paragraph before Theorem 1.

Remark 7. Term V in the asymptotic representation under IRa and IRb involves $\varepsilon_t \otimes \varepsilon_t - \mathrm{vec}(I_{r_1})$, where ε_t is a component of u_t. The limiting variances therefore depend on the kurtosis of ε_{it}, where ε_{it} is the ith element of ε_t. If normality of u_t is assumed, then terms $\mathcal{G}$ and $\mathcal{H}$ are zeros since $\mathcal{W} = 0$, and the limiting distribution in the preceding theorem is simplified. Similarly, with normality assumption of u_t, the asymptotic variances in Corollaries 3 and 4 and Theorems 7–9 will also be simplified. The simplified limiting distributions can be found in an earlier (online) version of this article.

Now we consider the asymptotic results for $\hat{\gamma}_i - \gamma_i$. We have the following theorem.

Theorem 2. Under Assumptions A–D, when $N, T \to \infty$, and $\sqrt{T}/N \to 0$, under IRa, IRb, or IRc, we have

$$\sqrt{T}(\hat{\gamma}_i - \gamma_i) = \sqrt{T}W\lambda_i + \mathbf{\Delta}_{\eta\eta}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t e_{it}\right) + o_p(1),$$

where $\eta_t = g_t - \mathbf{\Delta}_{gf}\mathbf{\Delta}_{ff}^{-1}f_t$ and $\mathbf{\Delta}_{\eta\eta} = E(\eta_t\eta_t')$ where $\mathbf{\Delta}_{gf}$ and $\mathbf{\Delta}_{ff}$ are defined (2.4). In addition,

$$W = \Omega_{\boldsymbol{v}\boldsymbol{v}}^{-1}\frac{1}{T}\sum_{t=\bar{K}}^{T}\boldsymbol{v}_t\varepsilon_t'.$$

Similar to Theorem 1, the asymptotic representation of $\hat{\gamma}_i$ under different IRs also has a unified expression. Symmetric to the symbol ϕ_t in Theorem 1, the symbol η_t here is the residual of projecting g_t on f_t. The matrix W has a unified expression under different IRs. If the population restriction $E(\boldsymbol{v}_t\varepsilon_t') = 0$ is replaced by the sample version $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{v}_t\varepsilon_t' = 0$, then the first term W disappears.

Note that the two terms in the asymptotic representation of $\hat{\gamma}_i - \gamma_i$ are asymptotically independent. The asymptotic independence follows from the absence of correlation between $v_t\varepsilon_t'$ and $\eta_t e_{it}$, which is similar to the case in (4.1) under IRa.

Corollary 2. Under the assumptions of Theorem 2, we have

$$\sqrt{T}(\hat{\gamma}_i - \gamma_i) \xrightarrow{d} \mathcal{N}\left(0, \left(\lambda_i'\Omega_{\varepsilon\varepsilon}\lambda_i\right)\Omega_{\boldsymbol{v}\boldsymbol{v}}^{-1} + \sigma_i^2\mathbf{\Delta}_{\eta\eta}^{-1}\right),$$

where $\Omega_{\varepsilon\varepsilon}$ and $\Omega_{\boldsymbol{v}\boldsymbol{v}}$ are defined in (2.4).

After deriving the asymptotic result of loadings, we consider the estimation of the unobservable factors $\hat{f}_t$. The asymptotic result of $\hat{f}_t - f_t$ involves both V and W matrices, which is stated in the following theorem.

Theorem 3. Let $\rho = N/T$. Under Assumptions A–D, when $N, T \to \infty$, we have

$$\sqrt{N}(\hat{f}_t - f_t) = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\lambda_i\lambda_i'\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\lambda_i e_{it}\right) - \sqrt{\rho}\left(\sqrt{T}V'f_t + \sqrt{T}W'g_t\right) + o_p(1).$$

In the asymptotic representation of $\hat{f}_t - f_t$, the first term, V and W are asymptotically independent with each other. To see this, first consider V and W. Under IRa, notice $E[\mathrm{vec}(V)\mathrm{vec}(W)'|\varepsilon] = 0$ where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$. Thus, $E[\mathrm{vec}(V)\mathrm{vec}(W)'] = E[E(\mathrm{vec}(V)\mathrm{vec}(W)'|\varepsilon)] = 0$. Under IRc, we also have the same result since $E[\mathrm{vec}(V)\mathrm{vec}(W)'|u] = 0$ where $u = (u_1, \ldots, u_T)'$. Combining the above two results under IRa and IRc, we have $E(\mathrm{vec}(V)\mathrm{vec}(W)'] = 0$ under IRb in view of the expression of V. We next show the first term is asymptotically independent with both V and W. Since W only involves u while the first term only involves e, they are independent under all IRs. Under IRa, V is independent with the first term for the same reason. Under IRc, V involves $\xi_t\phi_t'$ over t where $\xi_t = (e_{1t}, \ldots, e_{r_1t})$, while the first term involves summations of e_{it} over i, so they are uncorrelated because ϕ_t and e_{it} are independent from the assumption that u_t is independent of e_{it}. The previous two cases imply that under IRb, V is also asymptotically uncorrelated with the first term. Given the above analysis, we derive the limiting distribution in the following corollary.

Corollary 3. Under the assumptions of Theorem 3, we have

Under IRa:

$$\sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} \mathcal{N}\Big(0, \mathbf{Q}^{-1} + \rho\Big[\mathcal{F}_t'\mathbb{B}_{\mathbf{Q}}^{-1}\mathbb{P}_1\Big(2(D_{r_1}'D_{r_1})^{-1} + \mathcal{G}\Big) \times \mathbb{P}_1'\mathbb{B}_{\mathbf{Q}}^{-1\prime}\mathcal{F}_t + g_t'\Omega_{\upsilon\upsilon}^{-1}g_t\Omega_{\varepsilon\varepsilon}\Big]\Big),$$

under IRb:

$$\sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} \mathcal{N}\Big(\Big[\mathcal{F}_t'\Big(\mathbb{D}_2(2I_{r_1^2} + \mathcal{H})\mathbb{D}_2' + \mathbb{D}_3 \times [(\Lambda_1'\Sigma_{\xi\xi}^{-1}\Lambda_1) \otimes \mathbf{\Delta}_{\phi\phi}]^{-1}\mathbb{D}_3'\Big)\mathcal{F}_t + g_t'\Omega_{\upsilon\upsilon}^{-1}g_t\Omega_{\varepsilon\varepsilon}\Big]\Big),$$

under IRc:

$$\sqrt{N}(\hat{f}_t - f_t) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{Q}^{-1} + \rho\left[f_t'\mathbf{\Delta}_{\phi\phi}^{-1}f_t\Sigma_{\xi\xi} + g_t'\Omega_{\upsilon\upsilon}^{-1}g_t\Omega_{\varepsilon\varepsilon}\right]\right),$$

where $\mathcal{F}_t = I_{r_1} \otimes f_t$ and $\mathbf{Q} = \lim_{N\to\infty}(\Lambda'\Sigma_{ee}^{-1}\Lambda)/N$.

Remark 8. In the absence of g_t, the term $\sqrt{T}Wg_t$ drops out. Furthermore, in the absence of factor dynamics (so that $h_t = f_t = \varepsilon_t$), Theorem 3 characterizes the MLE of f_t under the population restriction $E(f_tf_t') = I_r$ when IRa or IRb is used. Especially, the principal components estimator of f_t (with $\hat{\Sigma}_{ee}$ replaced by an identity matrix), has the following representation,

$$\sqrt{N}(\hat{f}_t - f_t) = \left(\frac{1}{N}\sum_{i=1}^N \lambda_i\lambda_i'\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^N \lambda_i e_{it}\right) - \sqrt{\rho}\sqrt{T}V'f_t + o_p(1).$$

Here in the definition of V, the matrix $Q = \Lambda'\Sigma_{ee}^{-1}\Lambda/N$ is replaced by $\Lambda'\Lambda/N$.

For estimator $\hat{\sigma}_i^2$, we have the following theorem and corollary.

Theorem 4. Under Assumptions A–D, when $N, T \to 0$,

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

In addition, we have

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

where κ_i is the excess kurtosis of e_{it}. With the normality of e_{it}, the limiting distribution reduces to $\mathcal{N}(0, 2\sigma_i^4)$.

Notice e_{it} does not have the identification problem. Consequently its asymptotic representation does not depend on the identification restrictions. We then consider the asymptotic representation of $\hat{\Phi}_k - \Phi_k$, which is stated in the following theorem.

Theorem 5. Under Assumptions A–D, when $N, T \to 0$ and $\sqrt{T}/N \to 0$, we have

$$\sqrt{T}(\hat{\Phi}_k - \Phi_k) = \Big(\frac{1}{\sqrt{T}}\sum_{t=\bar{K}}^T u_t\psi_t'\Big)\Big(\frac{1}{T}\sum_{t=\bar{K}}^T \psi_t\psi_t'\Big)^{-1}(i_k \otimes I_r) - \sqrt{T}B'\Phi_k + \sqrt{T}\Phi_kB' + o_p(1),$$

where $\psi_t = (h_{t-1}', h_{t-2}', \ldots, h_{t-K}')'$ and B is defined as $B = [V, 0; W, 0]$.

If the factors f_t were observed, the asymptotic representation of $\sqrt{T}(\hat{\Phi}_k - \Phi_k)$ would be

$$\left(\frac{1}{\sqrt{T}}\sum_{t=\bar{K}}^T u_t\psi_t'\right)\left(\frac{1}{T}\sum_{t=\bar{K}}^T \psi_t\psi_t'\right)^{-1}(i_k \otimes I_r) + o_p(1).$$

However, f_t is unobservable, the asymptotic representation of $\sqrt{T}(\hat{\Phi}_k - \Phi_k)$ then has two extra terms, $-\sqrt{T}B'\Phi_k + \sqrt{T}\Phi_kB'$. Theorem 5 shows that the inferential theory of the standard VAR models cannot be applied to the FAVAR model.

Given Theorem 5, we have the following corollary.

Corollary 4. Under the assumptions of Theorem 5, we have

$$\sqrt{T}\mathrm{vec}(\hat{\Phi}_k - \Phi_k) \xrightarrow{d} \mathcal{N}(0, \mathscr{V}_k \otimes \Omega + \mathbb{D}_6J\mathbb{D}_6'),$$

where $\mathscr{V}_k$ denotes the (k,k)th $r \times r$ submatrix of $[E(\psi_t\psi_t')]^{-1}$ and J is the limiting variance of $\sqrt{T}\mathrm{vec}(B)$ and defined as

Under IRa:

$$J = \mathbb{D}_4\mathbb{B}_{\mathbf{Q}}^{-1}\mathbb{P}_1\Big[2(D_{r_1}'D_{r_1})^{-1} + \mathcal{G}\Big]\mathbb{P}_1'\mathbb{B}_{\mathbf{Q}}^{-1\prime}\mathbb{D}_4' + \mathbb{D}_5(\Omega_{\varepsilon\varepsilon} \otimes \Omega_{\upsilon\upsilon}^{-1})\mathbb{D}_5.$$

Under IRb:

$$J = \mathbb{D}_4\Big(\mathbb{D}_2(2I_{r_1^2} + \mathcal{H})\mathbb{D}_2' + \mathbb{D}_3[(\Lambda_1'\Sigma_{\xi\xi}^{-1}\Lambda_1) \otimes \mathbf{\Delta}_{\phi\phi}]^{-1}\mathbb{D}_3'\Big) \times \mathbb{D}_4' + \mathbb{D}_5(\Omega_{\varepsilon\varepsilon} \otimes \Omega_{\upsilon\upsilon}^{-1})\mathbb{D}_5.$$

Under IRc:

$$J = \mathbb{D}_4(\Sigma_{\xi\xi} \otimes \mathbf{\Delta}_{\phi\phi}^{-1})\mathbb{D}_4' + \mathbb{D}_5(\Omega_{\varepsilon\varepsilon} \otimes \Omega_{\upsilon\upsilon}^{-1})\mathbb{D}_5,$$

where $\mathbb{D}_4$ and $\mathbb{D}_5$ are respective $r^2 \times r_1^2$ and $r^2 \times r_1r_2$ matrices such that $\mathrm{vec}(B) = \mathbb{D}_4\mathrm{vec}(V) + \mathbb{D}_5\mathrm{vec}(W)$; $\mathbb{D}_6 = (I_r \otimes \Phi_k - \Phi_k' \otimes I_r)K_r$ with K_r the r-dimensional commutation matrix.

Remark 9. This article assumes h_t follows a finite order AR(K) process. This may not be appropriate for certain settings. For example, if the idiosyncratic errors contain unit roots, then differencing the data is necessary. But the factor process h_t may be over-differenced, and a finite lag VAR for h_t will not be appropriate. Our model does not apply to this case. A possibility is to allow K to increase slowly with T as in Berk (1974), but the analysis will be much more complicated. When the errors e_{it} are all stationary AR(1), there are two methods to proceed. Method 1 is to model the AR(1) process, and the likelihood function needs to be modified. Method 2 is to ignore the serial dependence, and the same likelihood function is used. It can be shown that both approaches give consistency for the factor loadings and the factors. But the limiting distributions will be different. In a non-FAVAR context, Bai and Li (2016) studied some related issues.

5. IMPULSE RESPONSE FUNCTION

Impulse response function plays an important role in the VAR analysis. In this section, we construct the confidence intervals for impulse response function of model (2.1). Let

$\Phi = (\Phi_1, \Phi_2, \ldots, \Phi_K)$. Theorem 5 gives

$$\sqrt{T}\text{vec}(\hat{\Phi}' - \Phi') = \Big[I_r \otimes \Big(\frac{1}{T}\sum_{t=1}^{T}\psi_t\psi_t'\Big)\Big]^{-1}\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t \otimes \psi_t)\Big] - \sqrt{T}(I_r \otimes \Phi')\text{vec}(B) + \sqrt{T}(\Phi \otimes I_{Kr})\text{vec}(I_K \otimes B) + o_p(1).$$

Let $\mathbb{D}_9$ be a $K^2r^2 \times r^2$ matrix satisfying that $\text{vec}(I_K \otimes B) = \mathbb{D}_9\text{vec}(B)$. Given this result, we have

$$\sqrt{T}\text{vec}(\hat{\Phi}' - \Phi') = \Big[I_r \otimes \Big(\frac{1}{T}\sum_{t=1}^{T}\psi_t\psi_t'\Big)\Big]^{-1}\Big[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t \otimes \psi_t)\Big] + [(\Phi \otimes I_{Kr})\mathbb{D}_9 - (I_r \otimes \Phi')]\sqrt{T}\text{vec}(B) + o_p(1).$$

By definition, it is seen that $\sqrt{T}\text{vec}(B)$ is asymptotically independent with $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t \otimes \psi_t)$. Let $\mathbb{D}_{10} = (\Phi \otimes I_{Kr})\mathbb{D}_9 - (I_r \otimes \Phi')$. Then we have

$$\sqrt{T}\text{vec}(\hat{\Phi}' - \Phi') \xrightarrow{d} \mathcal{N}\Big(0, \Omega \otimes [E(\psi_t\psi_t')]^{-1} + \mathbb{D}_{10}J\mathbb{D}_{10}'\Big),$$

where J is the limiting variance of $\sqrt{T}\text{vec}(B)$ and $\Omega = E(u_t u_t')$.

Under the assumption of stationarity of the process h_t, model (2.1) has a vector MA(∞) expression

$$h_t = u_t + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + \cdots \qquad (5.1)$$

Given the asymptotic results of $\sqrt{T}\text{vec}(\hat{\Phi}' - \Phi')$, the limiting distribution of $\hat{\Psi}_s - \Psi_s$ for all s can be derived in the standard way (see Hamilton 1994, p. 336). The limiting result is stated in the following theorem.

Theorem 6. Under Assumptions A–D, when $N, T \to \infty$ and $\sqrt{T}/N \to 0$,

$$\sqrt{T}\text{vec}(\hat{\Psi}_s' - \Psi_s') \xrightarrow{d} \mathcal{N}\Big(0, \Upsilon_s\Big[\Omega \otimes [E(\psi_t\psi_t')]^{-1} + \mathbb{D}_{10}J\mathbb{D}_{10}'\Big]\Upsilon_s'\Big),$$

where Υ_s is defined recursively by

$$\Upsilon_s = \sum_{i=1}^{s}\Psi_{i-1} \otimes [\Psi_{s-i}'\Psi_{s-i-1}' \ldots \Psi_{s-i-K+1}']$$

with $\Psi_0 = I_r$ and $\Psi_s = 0$ for $s < 0$.

We notice that the above impulse response functions are derived from the nonorthogonal shocks. In the analysis of some structural models, the impulse response functions for orthogonal shocks are required. For this, we consider decomposing $\Omega = \text{var}(u_t)$. Let $\mathcal{P}$ be the lower triangular matrix, which is obtained by the Cholesky decomposition of Ω. And let ω_t be the corresponding structural shocks with the relation that $u_t = \mathcal{P}\omega_t$. Then the moving average expression (5.1) can be written as

$$h_t = \mathcal{P}\omega_t + \Psi_1\mathcal{P}\omega_{t-1} + \Psi_2\mathcal{P}\omega_{t-2} + \cdots = \mathbb{C}_0\omega_t + \mathbb{C}_1\omega_{t-1} + \mathbb{C}_2\omega_{t-2} + \cdots \qquad (5.2)$$

with $\mathbb{C}_s = \Psi_s\mathcal{P}$ being the impulse response function corresponding to the structural shocks ω_t.

Remark 10. There are some cases in which no Cholesky decomposition is needed. For instance, in the application of Bernanke, Boivin, and Eliasz (2005), g_t is a scalar that is the federal fund rate. Then $\Omega_{\boldsymbol{vv}}$ is a scalar and hence a diagonal matrix. So under IRa and IRb, $\hat{\Omega}$ is diagonal implying that the innovations u_t are mutually orthogonal and hence can be interpreted as structural shocks. But under IRc, $\hat{\Omega}$ is not diagonal due to the nondiagonal matrix $\hat{\Omega}_{\varepsilon\varepsilon}$.

Next we aim to derive the limiting distribution of $\sqrt{T}\text{vec}(\hat{\mathbb{C}}_s - \mathbb{C}_s)$, on which basis the confidence intervals of the impulse response function can be constructed.

By definition, $\mathbb{C}_s$ is related to both Ψ_s and $\mathcal{P}$. The limiting distribution of $\hat{\Psi}_s - \Psi_s$ is given in Theorem 6. The limiting distribution of $\hat{\mathcal{P}} - \mathcal{P}$ can be derived based on the following theorem, since by definition, $\mathcal{P}$ is related to $\Omega_{\varepsilon\varepsilon}$ and $\Omega_{\boldsymbol{vv}}$.

Theorem 7. Under Assumption A–D, when $N, T \to \infty$, the estimator $\tilde{\Omega}_{\boldsymbol{vv}}$ is consistent for $\Omega_{\boldsymbol{vv}}$. With $\sqrt{T}/N \to 0$, under IRa, IRb, or IRc, we have

$$\sqrt{T}\text{vech}(\tilde{\Omega}_{\boldsymbol{vv}} - \Omega_{\boldsymbol{vv}}) \xrightarrow{d} \mathcal{N}\Big(0, D_{r_2}^{+}\Big[2\Omega_{\boldsymbol{vv}} \otimes \Omega_{\boldsymbol{vv}} + (\Omega_{\boldsymbol{vv}}^{1/2} \otimes \Omega_{\boldsymbol{vv}}^{1/2})(\mathscr{D}_{r_2}\mathcal{W}_{\boldsymbol{v}}\mathscr{D}_{r_2}')(\Omega_{\boldsymbol{vv}}^{1/2} \otimes \Omega_{\boldsymbol{vv}}^{1/2})\Big]D_{r_2}^{+\prime}\Big),$$

where $D_{r_2}^{+}$ is the Moore–Penrose inverse of an r_2-dimensional duplication matrix, $\mathscr{D}_r$ is defined in Corollary 1, and $\mathcal{W}_{\boldsymbol{v}}$ is an r_2-dimensional diagonal matrix with its ith element $\kappa_i^{\dagger} - 3$ where $\kappa_i^{\dagger}$ is the fourth moment of its ith element of $\Omega_{\boldsymbol{vv}}^{-1/2}\boldsymbol{v}_t$. In addition, under IRc, we also have

$$\sqrt{T}\text{vech}(\hat{\Omega}_{\varepsilon\varepsilon} - \Omega_{\varepsilon\varepsilon}) \xrightarrow{d} \mathcal{N}\Big(0, D_{r_1}^{+}\Big[2\Omega_{\varepsilon\varepsilon} \otimes \Omega_{\varepsilon\varepsilon} + (\Omega_{\varepsilon\varepsilon}^{1/2} \otimes \Omega_{\varepsilon\varepsilon}^{1/2})(\mathscr{D}_{r_1}\mathcal{W}_{\varepsilon}\mathscr{D}_{r_1}')(\Omega_{\varepsilon\varepsilon}^{1/2} \otimes \Omega_{\varepsilon\varepsilon}^{1/2}) + 4S_{r_1}(\Sigma_{\xi\xi} \otimes \boldsymbol{\Delta}_{\phi\phi}^{-1})S_{r_1}'\Big]D_{r_1}^{+\prime}\Big),$$

where $\mathcal{W}_{\varepsilon}$ is defined similarly as $\mathcal{W}_{\boldsymbol{v}}$.

Further, based on Theorems 6 and 7, we can derive the limiting distribution of $\sqrt{T}\text{vec}(\hat{\mathbb{C}}_s - \mathbb{C}_s)$ as in the following theorem.

Theorem 8. Under Assumptions A–D, when $N, T \to \infty$ and $\sqrt{T}/N \to 0$, we have:

Under IRa and IRb,

$$\sqrt{T}\text{vec}(\hat{\mathbb{C}}_s - \mathbb{C}_s) \xrightarrow{d} \mathcal{N}\Big(0, (\mathcal{P}' \otimes I_r)K_r\Upsilon_s\mathbb{J}_1\Upsilon_s'K_r'(\mathcal{P} \otimes I_r) + (I_r \otimes \Psi_s)\mathbb{D}_7\mathbb{J}_2\mathbb{D}_7'(I_r \otimes \Psi_s')\Big).$$

Under IRc,

$$\sqrt{T}\text{vec}(\hat{\mathbb{C}}_s - \mathbb{C}_s) \xrightarrow{d} \mathcal{N}\Big(0, (\mathcal{P}' \otimes I_r)K_r\Upsilon_s\mathbb{J}_1\Upsilon_s'K_r'(\mathcal{P} \otimes I_r) + (I_r \otimes \Psi_s)\mathbb{J}_3(I_r \otimes \Psi_s')\Big)$$

with $\mathbb{J}_1 = \Omega \otimes [E(\psi_t \psi_t')]^{-1} + \mathbb{D}_{10} J \mathbb{D}_{10}'$, $\mathbb{J}_3 = \mathbb{D}_8 \mathbb{J}_4 \mathbb{D}_8' + \mathbb{D}_7 \mathbb{J}_2 \mathbb{D}_7'$ and

$$\mathbb{J}_2 = \mathbb{W}_2 S_{r_2}\Big[2\Omega_{vv} \otimes \Omega_{vv} + (\Omega_{vv}^{1/2} \otimes \Omega_{vv}^{1/2})(\mathscr{D}_{r_2} \mathcal{W}_v \mathscr{D}_{r_2}')(\Omega_{vv}^{1/2} \otimes \Omega_{vv}^{1/2})\Big] S_{r_2}' \mathbb{W}_2',$$

where

$$\mathbb{J}_4 = \mathbb{W}_1 S_{r_1}\Big[2\Omega_{\varepsilon\varepsilon} \otimes \Omega_{\varepsilon\varepsilon} + (\Omega_{\varepsilon\varepsilon}^{1/2} \otimes \Omega_{\varepsilon\varepsilon}^{1/2}) \times (\mathscr{D}_{r_1} \mathcal{W}_\varepsilon \mathscr{D}_{r_1}')(\Omega_{\varepsilon\varepsilon}^{1/2} \otimes \Omega_{\varepsilon\varepsilon}^{1/2}) + 4 S_{r_1}(\Sigma_{\xi\xi} \otimes \mathbf{\Delta}_{\phi\phi}^{-1}) S_{r_1}'\Big] S_{r_1}' \mathbb{W}_1'.$$

$\mathbb{D}_7$ and $\mathbb{D}_8$ are transformation matrices such that for any $M_{r\times r} = [M_1, 0; 0, M_2]$ where M_1 is $r_1 \times r_1$ and M_2 is $r_2 \times r_2$ and both are lower-triangular matrices, $\text{vec}(M) = \mathbb{D}_8 \text{vech}(M_1) + \mathbb{D}_7 \text{vech}(M_2)$; $\mathbb{W}_1$ and $\mathbb{W}_2$ are defined in Appendix E (online supplement).

Then based on (2.2) and (5.2), the impulse response function of the observable variables z_t with respect to the structural shocks ω_t is

$$\frac{\partial z_{i,t+k}}{\partial \omega_t} = \mathbb{C}_k' \begin{bmatrix} \lambda_i \\ \gamma_i \end{bmatrix}$$

for each i and for all $k \geq 0$. Then note that

$$\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}} - \frac{\partial z_{i,t+k}}{\partial \omega_t} = (\hat{\mathbb{C}}_k - \mathbb{C}_k)' \begin{bmatrix} \lambda_i \\ \gamma_i \end{bmatrix} + \mathbb{C}_k' \begin{bmatrix} \hat{\lambda}_i - \lambda_i \\ \hat{\gamma}_i - \gamma_i \end{bmatrix}$$

has two components, which arise from estimating the loadings (λ_i, γ_i) and the MA(∞) coefficients $\mathbb{C}_k$. From the asymptotic representations of $(\hat{\lambda}_i - \lambda_i)$, $(\hat{\gamma}_i - \gamma_i)$, and $(\hat{\mathbb{C}}_k - \mathbb{C}_k)$, taking into account their covariances, we obtain the following theorem on the impulse response function.

Theorem 9 (Impulse Response Function). Under Assumptions A–D, when $N, T \to \infty$ and $\sqrt{T}/N \to 0$, under IRa, IRb, or IRc, we have

$$\sqrt{T}\left(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}} - \frac{\partial z_{i,t+k}}{\partial \omega_t}\right) \xrightarrow{d} \mathcal{N}\left(0, \text{Avar}\Big(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}}\Big)\right),$$

with

$$\text{Avar}\Big(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}}\Big) = [(\lambda_i', \gamma_i') \otimes K_r] \cdot \text{Avar}(\text{vec}(\hat{\mathbb{C}}_k)) \cdot [(\lambda_i', \gamma_i')' \otimes K_r'] + \mathbb{C}_k' \cdot \text{Avar}(\hat{\lambda}_i', \hat{\gamma}_i') \cdot \mathbb{C}_k + [(\lambda_i', \gamma_i') \otimes K_r] \cdot (\mathcal{P}' \otimes I_r) K_r \Upsilon_k \mathbb{D}_{10} J [(\lambda_i', \gamma_i')' \otimes \mathbb{C}_k] + [(\lambda_i', \gamma_i') \otimes \mathbb{C}_k'] J \mathbb{D}_{10}' \Upsilon_k' K_r' (\mathcal{P} \otimes I_r) \cdot [(\lambda_i', \gamma_i')' \otimes K_r'],$$

where K_r is the commutation matrix defined as in Section 4; J is the limiting variance of $\sqrt{T}\text{vec}(B)$ defined as in Corollary 4 and Υ_k is defined in Theorem 6. In addition, $\text{Avar}(\hat{\lambda}_i', \hat{\gamma}_i') = \text{diag}(\text{Avar}(\hat{\lambda}_i), \text{Avar}(\hat{\gamma}_i))$; $\text{Avar}(\hat{\lambda}_i)$, $\text{Avar}(\hat{\gamma}_i)$, and $\text{Avar}(\text{vec}(\hat{\mathbb{C}}_k))$ are the asymptotic variances of $\hat{\lambda}_i$, $\hat{\gamma}_i$, and $\hat{\mathbb{C}}_k$, respectively, and are given in Corollary 1, Corollary 2, and Theorem 8, respectively.

Once estimators for $\text{Avar}(\hat{\lambda}_i)$, $\text{Avar}(\hat{\gamma}_i)$, and $\text{Avar}(\text{vec}(\hat{\mathbb{C}}_k))$ are obtained, the confidence intervals for the impulse response function can be easily constructed. For example, the 95% confidence interval for the impulse response function $\frac{\partial z_{i,t+k}}{\partial \omega_t}$ is

$$\left(\Big(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}}\Big) - \frac{1.96}{\sqrt{T}}\Big[\text{diag}\Big\{\widehat{\text{Avar}}\Big(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}}\Big)\Big\}\Big]^{1/2}, \Big(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}}\Big) + \frac{1.96}{\sqrt{T}}\Big[\text{diag}\Big\{\widehat{\text{Avar}}\Big(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}}\Big)\Big\}\Big]^{1/2}\right),$$

where $(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}}) = \hat{\mathbb{C}}_k' [\begin{smallmatrix} \hat{\lambda}_i \\ \hat{\gamma}_i \end{smallmatrix}]$, and $\text{diag}\{\cdot\}$ stacks the diagonal elements of the argument into a column vector, and

$$\widehat{\text{Avar}}\Big(\widehat{\frac{\partial z_{i,t+k}}{\partial \omega_t}}\Big) = (\hat{\lambda}_i', \hat{\gamma}_i') \otimes I_r \cdot K_r \cdot \widehat{\text{Avar}}(\text{vec}(\hat{\mathbb{C}}_k)) \cdot K_r' \cdot (\hat{\lambda}_i', \hat{\gamma}_i')' \otimes I_r + \hat{\mathbb{C}}_k' \cdot \widehat{\text{Avar}}(\hat{\lambda}_i', \hat{\gamma}_i') \cdot \hat{\mathbb{C}}_k + (\hat{\mathcal{P}}' \otimes I_r) K_r \hat{\Upsilon}_k \hat{\mathbb{D}}_{10} \hat{J} [(\hat{\lambda}_i', \hat{\gamma}_i')' \otimes \hat{\mathbb{C}}_k] + [(\hat{\lambda}_i', \hat{\gamma}_i') \otimes \hat{\mathbb{C}}_k'] \hat{J} \hat{\mathbb{D}}_{10}' \hat{\Upsilon}_k' K_r' (\hat{\mathcal{P}} \otimes I_r)$$

with $\widehat{\text{Avar}}(\hat{\lambda}_i', \hat{\gamma}_i')$ being the estimate of $\text{Avar}(\hat{\lambda}_i', \hat{\gamma}_i')$ and $\widehat{\text{Avar}}(\text{vec}(\hat{\mathbb{C}}_k))$ being the estimate of $\text{Avar}(\text{vec}(\hat{\mathbb{C}}_k))$; $\hat{\mathcal{P}}$, $\hat{\Upsilon}_k$, $\hat{\mathbb{D}}_{10}$, and $\hat{J}$ are the estimates of $\mathcal{P}$, Υ_k, $\mathbb{D}_{10}$, and J, respectively.

6. FINITE SAMPLE PROPERTIES

In this section, we run Monte Carlo simulations to investigate the finite sample properties of the two-step estimators. For the sake of space, we only consider IRb and IRc, which are of more practical relevance. In factor analysis literature, many studies, such as Bai and Li (2012, 2016), Doz, Giannone, and Reichlin (2012), investigate the finite sample properties of the QMLE, that is, $\tilde{\Lambda}$, $\tilde{F}$, $\tilde{\Sigma}_{ee}$. Consequently, in this article we instead focus on the performance of the estimator $\hat{\Phi}$. Notice that $\hat{\Phi}$ has a close relation with the impulse response function, which in many occasions is the primary tool of the economic analysis. Hence, the finite sample properties of $\hat{\Phi}$ deserves our special attention.

The factors are assumed to follow VAR(1) and are generated according to

$$h_t = \Phi h_{t-1} + u_t,$$

where $h_t = (f_t', g_t')'$ and u_t is an iid $\mathcal{N}(0, \Omega)$ process. Matrix Ω is restricted by the identification IRb and IRc and exhibits the form, respectively,

$$\text{(IRb)}\begin{bmatrix} I_{r_1} & 0 \\ 0 & \Omega_{22} \end{bmatrix}, \qquad \text{(IRc)}\begin{bmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{bmatrix},$$

where Ω_{11} and Ω_{22} are both symmetric positive definite matrices. The symmetric positive matrix is generated according to $\Omega = \mathbb{M} D \mathbb{M}'$, where $\mathbb{M} = M(M'M)^{-1/2}$ with M being any $r \times r$ standard normal random matrix and D is a diagonal matrix with all its diagonal elements drawn from $(1 + \mathcal{U}[0, 1])^2$. Throughout the simulation, the number of unknown factors and known factors, r_1 and r_2, are set to 2 and 1 (so $r = r_1 + r_2 = 3$). In addition, the parameter Φ is fixed to $0.7 I_r$.

All the factor loadings are generated independently from $\mathcal{N}(0, 1)$ (where Λ is $N \times 2$ and Γ is $N \times 1$). To make the underlying factor loadings satisfy the identification restrictions, we set the (1, 2)th element of Λ to be 0 under IRb and the upper

Table 1. The RMSEs of all the elements of $\hat{\Phi}$ under IRb

N	T	Φ_{11}	Φ_{12}	Φ_{13}	Φ_{21}	Φ_{22}	Φ_{23}	Φ_{31}	Φ_{32}	Φ_{33}	Ave
50	50	0.1360	0.1261	0.0902	0.1318	0.1474	0.0925	0.1804	0.1748	0.1170	0.1329
100	50	0.1307	0.1177	0.0919	0.1252	0.1351	0.0939	0.1799	0.1722	0.1209	0.1297
200	50	0.1335	0.1208	0.0894	0.1230	0.1320	0.0885	0.1736	0.1625	0.1218	0.1272
50	100	0.0933	0.0803	0.0569	0.0846	0.0891	0.0591	0.1179	0.1185	0.0810	0.0867
100	100	0.0841	0.0788	0.0563	0.0783	0.0848	0.0563	0.1113	0.1133	0.0811	0.0827
200	100	0.0844	0.0794	0.0554	0.0798	0.0879	0.0538	0.1106	0.1147	0.0816	0.0831
50	200	0.0611	0.0534	0.0380	0.0560	0.0614	0.0369	0.0787	0.0833	0.0529	0.0580
100	200	0.0572	0.0563	0.0379	0.0532	0.0600	0.0396	0.0795	0.0799	0.0559	0.0577
200	200	0.0557	0.0519	0.0370	0.0528	0.0547	0.0369	0.0836	0.0797	0.0551	0.0564
50	500	0.0373	0.0326	0.0236	0.0328	0.0380	0.0235	0.0495	0.0514	0.0336	0.0358
100	500	0.0343	0.0335	0.0229	0.0327	0.0350	0.0233	0.0509	0.0495	0.0349	0.0352
200	500	0.0335	0.0321	0.0233	0.0324	0.0341	0.0235	0.0505	0.0484	0.0322	0.0344

Table 2. The empirical size of the t-test (nominal 5%) for all the elements of Φ under IRb

N	T	Φ_{11}	Φ_{12}	Φ_{13}	Φ_{21}	Φ_{22}	Φ_{23}	Φ_{31}	Φ_{32}	Φ_{33}	Ave
50	50	0.068	0.060	0.049	0.069	0.086	0.062	0.058	0.052	0.055	0.0621
100	50	0.063	0.055	0.055	0.062	0.071	0.066	0.063	0.055	0.055	0.0606
200	50	0.079	0.069	0.049	0.049	0.065	0.058	0.049	0.055	0.074	0.0608
50	100	0.085	0.046	0.044	0.065	0.072	0.060	0.042	0.067	0.054	0.0594
100	100	0.059	0.055	0.055	0.051	0.072	0.046	0.044	0.041	0.066	0.0543
200	100	0.060	0.045	0.050	0.055	0.074	0.040	0.034	0.055	0.060	0.0526
50	200	0.069	0.042	0.052	0.056	0.072	0.039	0.047	0.062	0.050	0.0543
100	200	0.058	0.057	0.048	0.058	0.069	0.052	0.042	0.040	0.071	0.0550
200	200	0.057	0.042	0.043	0.043	0.056	0.055	0.056	0.053	0.058	0.0514
50	500	0.070	0.043	0.055	0.043	0.081	0.058	0.049	0.064	0.053	0.0573
100	500	0.054	0.053	0.053	0.046	0.054	0.052	0.049	0.046	0.064	0.0523
200	500	0.055	0.047	0.056	0.049	0.046	0.053	0.057	0.048	0.048	0.0510

2×2 matrix of Λ to be the identity matrix under IRc. After the factor loadings are obtained, the data are generated by

$$z_t = \Lambda f_t + \Gamma g_t + e_t,$$

where $e_t = (e_{1t}, e_{2t}, \dots, e_{Nt})'$ with $e_{it} \sim \mathcal{N}(0, \sigma_i^2)$, where $\sigma_i^2 \sim 1 + \mathcal{U}[0, 1]$.

After the data (Z, G) are constructed, we need to determine the number of unknown factors r_1 before estimation. There are two approaches to determine r_1. One approach is to first estimate the total number of factors based on Z (denoted as $\hat{r}$) by the information criterion proposed by Bai and Ng (2002), and then get $\hat{r}_1 = \hat{r} - r_2$ where r_2 is the number of known factors. A better approach is to directly estimate the number of unknown factors $\hat{r}_1$ through the transformed data $Z\mathbb{M}_G$. The second approach is adopted in simulations. Once r_1 is determined, we use the method described in the previous section to estimate the parameters.

The identification conditions IRb have so-called sign problem. (See Bai and Li (2012) for an illustration on the sign problem.) To eliminate this problem, after the estimated factors $\hat{F}$ are obtained, we calculate the correlation coefficients between each column of $\hat{F}$ and the corresponding column

Table 3. The RMSEs of all the elements of $\hat{\Phi}$ under IRc

N	T	Φ_{11}	Φ_{12}	Φ_{13}	Φ_{21}	Φ_{22}	Φ_{23}	Φ_{31}	Φ_{32}	Φ_{33}	Ave
50	50	0.1407	0.1397	0.1336	0.1354	0.1383	0.1258	0.1359	0.1281	0.1323	0.1344
100	50	0.1405	0.1403	0.1349	0.1365	0.1443	0.1299	0.1319	0.1419	0.1284	0.1365
200	50	0.1347	0.1290	0.1350	0.1394	0.1406	0.1295	0.1312	0.1323	0.1261	0.1331
50	100	0.0878	0.0868	0.0800	0.0897	0.0897	0.0849	0.0842	0.0859	0.0806	0.0855
100	100	0.0871	0.0869	0.0877	0.0810	0.0843	0.0838	0.0831	0.0871	0.0812	0.0847
200	100	0.0827	0.0840	0.0874	0.0813	0.0857	0.0852	0.0879	0.0878	0.0843	0.0851
50	200	0.0586	0.0562	0.0600	0.0571	0.0583	0.0568	0.0538	0.0571	0.0528	0.0567
100	200	0.0559	0.0566	0.0592	0.0558	0.0560	0.0586	0.0564	0.0567	0.0547	0.0567
200	200	0.0534	0.0583	0.0554	0.0577	0.0558	0.0565	0.0579	0.0562	0.0541	0.0562
50	500	0.0348	0.0343	0.0346	0.0334	0.0353	0.0364	0.0346	0.0348	0.0312	0.0344
100	500	0.0338	0.0337	0.0353	0.0344	0.0347	0.0353	0.0347	0.0359	0.0324	0.0344
200	500	0.0332	0.0339	0.0347	0.0341	0.0341	0.0369	0.0351	0.0370	0.0328	0.0346

Table 4. The empirical size of the t-test (nominal 5%) for all the elements of Φ under IRc

N	T	Φ_{11}	Φ_{12}	Φ_{13}	Φ_{21}	Φ_{22}	Φ_{23}	Φ_{31}	Φ_{32}	Φ_{33}	Ave
50	50	0.079	0.064	0.069	0.081	0.081	0.051	0.076	0.073	0.076	0.0722
100	50	0.088	0.078	0.065	0.071	0.089	0.069	0.074	0.078	0.087	0.0777
200	50	0.069	0.059	0.061	0.069	0.089	0.064	0.067	0.069	0.077	0.0693
50	100	0.078	0.064	0.051	0.070	0.074	0.058	0.057	0.050	0.062	0.0627
100	100	0.073	0.067	0.054	0.045	0.069	0.053	0.053	0.056	0.047	0.0574
200	100	0.055	0.055	0.070	0.058	0.068	0.066	0.058	0.066	0.065	0.0623
50	200	0.058	0.048	0.060	0.055	0.056	0.046	0.043	0.051	0.049	0.0518
100	200	0.060	0.046	0.067	0.050	0.059	0.058	0.064	0.064	0.066	0.0593
200	200	0.049	0.060	0.052	0.059	0.052	0.037	0.055	0.057	0.055	0.0529
50	500	0.055	0.046	0.046	0.046	0.066	0.057	0.058	0.060	0.048	0.0536
100	500	0.051	0.045	0.050	0.054	0.057	0.050	0.049	0.056	0.052	0.0516
200	500	0.051	0.051	0.045	0.050	0.058	0.067	0.047	0.060	0.047	0.0529

of F. If the coefficient is negative, then multiply -1 to that column of $\hat{F}$ and the corresponding column of $\hat{\Lambda}$. In practice, this treatment is not feasible. However, sign problem can be fixed by other means, see Stock and Watson (2005). We consider a combination of $N = 50, 100, 200$ and $T = 50, 100, 200, 500$. All the results are obtained in 1000 repetitions.

Table 1 reports the root of mean square error (RMSE) of all elements of Φ. The last element of each row is the average of the left nine elements under IRb. On the whole, we can see that the RMSE decreases as the sample size becomes larger. More concretely, Table 1 shows that the RMSE is closely linked with the time length T and little related to the cross-sectional size N. Take Φ_{11} as an example. When $T = 200$ and $N = 50, 100, 200$, the corresponding three RMSEs are 0.0611, 0.0572, and 0.0557, which are roughly equal. However, when T increases to 500, the corresponding three RMSEs are 0.0373, 0.0343, and 0.0335, which are still roughly equal but dramatically lower in comparison with those of $T = 200$. This result is consistent with results in Theorem 5.

Aside from the consistency, we are also concerned about the limiting distribution of $\sqrt{T}\text{vec}(\hat{\Phi}' - \Phi')$, which, as seen in the last section, has a direct effect on the confidence interval of the impulse response function. To this end, we calculate the size of t-test for every Φ_{ij} in each simulation and count the number of times that the absolute value of t-statistics is greater than the critical value of the 5% significance level for the standard normal distribution (i.e., 1.96) in 1000 repetitions. Table 2 reports the actual significance level that corresponds to 5% nominal size for every Φ_{ij}. As in Table 1, we average the result of the nine elements of Φ and report the result in the last column. From Table 2, we find that, unlike in Table 1, the actual significance level is related to both N and T. When the sample size is small, say $N = 50, T = 50$, the size distortion is a little larger, for Φ_{22}, the actual significance level is 0.086. However, when the sample becomes larger, the distortion gradually decreases (see the last column). When $N = 200, T = 500$, we can see that all the elements of Φ have a satisfactory size.

The results under IRc are similar to those under IRb and are reported in Tables 3 and 4. We do not repeat the detailed analysis.

7. CONCLUDING REMARKS

This article considers the identification, estimation, and inferential theory of the FAVAR model. Three sets of identification restrictions are considered. We propose a likelihood-based two-step method to estimate the parameters. Consistency, convergence rates, asymptotic representations, and the limiting distributions have been established. The impulse response function and its confidence intervals are also provided. An important result from our theory is that if the identification conditions are imposed on the population variance rather than on the sample variance of the factor process, an additional term, which arises from the distance between the sample variance and the population variance, would enter the final asymptotic representations. Consequently the limiting variances of the estimators are larger. We studied the ways in which this distance affects the limiting distributions. The finite sample Monte Carlo simulation confirms our theoretical results.

The analysis of this article assumes constant parameters. In empirical applications with a long time span, it is likely that a structural change occurs, either in the dynamics of h_t, or in the factor loadings (Λ, Γ). It is of interest to develop inference procedures allowing for this possibility, as in Chen, Dolado, and Gonzalo (2014), Cheng, Liao, and Schorfheide (2013), and Han and Inoue (2015).

APPENDIX: TECHNICAL MATERIALS FOR THE ASYMPTOTIC RESULTS

In this appendix, we provide the detailed derivations for the asymptotic results under IRa. The derivations for the asymptotic results under IRb and IRc as well as the theorems in Section 5 are delegated to the online supplement. Throughout the appendix, we use $\bar{K}$ to denote $K + 1$ and $\bar{T}$ to denote $T - K - 1$. To facilitate the analysis, we introduce the following auxiliary identification condition (an intermediate step analysis).

AU1: The underlying parameter values $\theta^* = (\Lambda^*, \Gamma^*, F^*, \Phi^*, \Sigma_{ee})$ satisfy: $\frac{1}{N}\Lambda^{*\prime}\Sigma_{ee}^{-1}\Lambda^* = Q^*$, $\frac{1}{T}\sum_{t=1}^{T} f_t^* f_t^{*\prime} = I_{r_1}$, and $\frac{1}{T}\sum_{t=1}^{T} f_t^* g_t' = 0$, where Q^* is a diagonal matrix, whose diagonal elements are distinct and arranged in descending order.

APPENDIX A: THE ASYMPTOTIC RESULTS OF THE QMLE

In this appendix, we show that the QMLE $\tilde{\lambda}_i, \tilde{\sigma}_i^2, \tilde{f}_t$ and $\tilde{\Phi}_k$ are respectively consistent estimator of $\lambda_i^*, \sigma_i^2, f_t^*$ and Φ_k^* under AU1. We also derive their asymptotic representations.

Proposition A.1. Under Assumptions A-D, together with AU1,

$$\tilde{\lambda}_i - \lambda_i^* = \left(\frac{1}{T}\sum_{t=1}^T f_t^* f_t^{*\prime}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T f_t^* e_{it}\right) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}), \tag{A.1}$$

$$\tilde{\gamma}_i - \gamma_i^* = \left(\frac{1}{T}\sum_{t=1}^T g_t g_{t'}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T g_t e_{it}\right), \tag{A.2}$$

$$\tilde{f}_t - f_t^* = \left(\frac{1}{N}\sum_{i=1}^T \frac{1}{\sigma_i^2}\lambda_i^*\lambda_i^{*\prime}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i^* e_{it}\right) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}), \tag{A.3}$$

$$\tilde{\sigma}_i^2 - \sigma_i^2 = \frac{1}{T}\sum_{t=1}^T (e_{it}^2 - \sigma_i^2) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}). \tag{A.4}$$

Proof of Proposition A.1. Write $z_t = \Lambda^* f_t^* + \Gamma^* g_t + e_t$ into matrix form,

$$Z = \Lambda^* F^{*\prime} + \Gamma^* G' + e. \tag{A.5}$$

Post-multiplying $\mathbb{M}_G = I_T - G(G'G)^{-1}G'$ on both sides, together with $F^{*\prime}G = 0$ by AU1, we have

$$Z\mathbb{M}_G = \Lambda^* F^{*\prime} + e\mathbb{M}_G.$$

Let $Y = Z\mathbb{M}_G$ and y_t denotes the t-th column of Y. The above equation is equivalent to

$$y_t = \Lambda^* f_t^* + e_t - eG(G'G)^{-1}g_t \tag{A.6}$$

Bai and Li (2012) derive the asymptotic representations of $\tilde{\lambda}_i, \tilde{f}_t, \tilde{\sigma}_i^2$ under the case that $g_t \equiv 1$. However, when g_t is a general random variable, as like in the present context, the derivation is the same since term $eG(G'G)^{-1}g_t$ is essentially negligible. Using the arguments of Bai and Li (2012) under IC3, we obtain (A.1), (A.3) and (A.4). Consider (A.2). Substituting $z_{it} = \lambda_i^{*\prime} f_t^* + \gamma_i^{*\prime} g_t + e_{it}$ into $\tilde{\gamma}_i = (\sum_{t=1}^T g_t g_t')^{-1}(\sum_{t=1}^T g_t(z_{it} - \tilde{\lambda}_i'\tilde{f}_t))$, we have

$$\begin{aligned}\tilde{\gamma}_i - \gamma_i^* &= \left(\sum_{t=1}^T g_t g_t'\right)^{-1}\left(\sum_{t=1}^T g_t e_{it}\right) \\ &\quad - \left(\sum_{t=1}^T g_t g_t'\right)^{-1}\left(\sum_{t=1}^T g_t f_t^{*\prime}\right)(\tilde{\lambda}_i - \lambda_i^*) \\ &\quad - \left(\sum_{t=1}^T g_t g_t'\right)^{-1}\left(\sum_{t=1}^T g_t(\tilde{f}_t - f_t^*)'\right)\tilde{\lambda}_i\end{aligned}$$

The second term of the right hand side is zero by $\sum_{t=1}^T g_t f_t^{*\prime} = G'F^* = 0$. Consider the third term. Notice

$$\begin{aligned}\tilde{f}_t &= (\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}y_t \\ &= (\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\left[z_t - \left(\sum_{s=1}^T z_s g_s'\right)\left(\sum_{s=1}^T g_s g_s'\right)^{-1} g_t\right] \\ &= (\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\left[\Lambda^* f_t^* + e_t - \left(\sum_{s=1}^T e_s g_s'\right)\left(\sum_{s=1}^T g_s g_s'\right)^{-1} g_t\right].\end{aligned}$$

Then it follows

$$\begin{aligned}\tilde{f}_t - f_t^* &= -A^{*\prime} f_t^* + (\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}e_t \\ &\quad - (\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\left(\sum_{s=1}^T e_s g_s'\right)\left(\sum_{s=1}^T g_s g_s'\right)^{-1} g_t\end{aligned} \tag{A.7}$$

where $A^* = (\tilde{\Lambda} - \Lambda^*)'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}(\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}$.

Given the above expression, together with $\sum_{t=1}^T g_t f_t^{*\prime} = 0$, we have

$$\frac{1}{T}\sum_{t=1}^T g_t(\tilde{f}_t - f_t^*)' = 0. \tag{A.8}$$

Then (A.2) follows. This completes the proof of Proposition A.1. □

Lemma A.1 Under Assumptions A-D,

(a) $\frac{1}{\bar{T}}\sum_{t=\bar{K}}^T \tilde{h}_{t-p}\tilde{h}_{t-q}' - \frac{1}{\bar{T}}\sum_{t=\bar{K}}^T h_{t-p}^* h_{t-q}^{*\prime} = O_p(N^{-1}) + O_p(T^{-1})$,

for $p, q = 0, \ldots, K$

(b) $\frac{1}{\bar{T}}\sum_{t=\bar{K}}^T (\tilde{h}_{t-p} - h_{t-p}^*)\tilde{h}_{t-q}' = O_p(N^{-1}) + O_p(T^{-1})$,

for $p, q = 0, 1, \ldots, K$

(c) $\frac{1}{\bar{T}}\sum_{t=\bar{K}}^T u_t^*\tilde{h}_{t-p}' - \frac{1}{\bar{T}}\sum_{t=\bar{K}}^T u_t^* h_{t-p}^{*\prime} = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$,

for $p = 1, \ldots, K$,

where $\tilde{h}_t = (\tilde{f}_t', g_t')'$ and $h_t^* = (f_t^{*\prime}, h_t')'$.

Proof of Lemma A.1. Consider (a). By the definitions of $\tilde{h}_t$ and h_t^*, the left hand side of (a) is equal to

$$\begin{bmatrix} J_{11} & J_{12} \\ J_{21} & 0 \end{bmatrix}$$

where

$$J_{11} = \frac{1}{\bar{T}}\sum_{t=\bar{K}}^T (\tilde{f}_{t-p} - f_{t-p}^*)(\tilde{f}_{t-q} - f_{t-q}^*)' + \frac{1}{\bar{T}}\sum_{t=\bar{K}}^T (\tilde{f}_{t-p} - f_{t-p}^*)f_{t-q}^{*\prime} + \frac{1}{\bar{T}}\sum_{t=\bar{K}}^T f_{t-p}^*(\tilde{f}_{t-q} - f_{t-q}^*)';$$

$$J_{12} = \frac{1}{\bar{T}}\sum_{t=\bar{K}}^T (\tilde{f}_{t-p} - f_{t-p}^*)g_{t-q}'; \qquad J_{21} = \frac{1}{\bar{T}}\sum_{t=\bar{K}}^T g_{t-p}(\tilde{f}_{t-q} - f_{t-q}^*)'.$$

The first term of J_{11} is $O_p(N^{-1}) + O_p(T^{-2})$, as shown in Bai an Li (2012). Consider the second term. By (A.7), $\frac{1}{\bar{T}}\sum_{t=\bar{K}}^T (\tilde{f}_{t-p} - f_{t-p}^*)f_{t-q}^{*\prime}$ is equal to

$$-A^{*\prime}\frac{1}{\bar{T}}\sum_{t=\bar{K}}^T f_{t-p}^* f_{t-q}^{*\prime} + (\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}\frac{1}{\bar{T}}\sum_{t=\bar{K}}^T \tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}e_{t-p}f_{t-q}^{*\prime}$$

$$-(\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}\left(\frac{1}{T}\sum_{s=1}^T \tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}e_s g_s'\right)\left(\frac{1}{T}\sum_{s=1}^T g_s g_s'\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^T g_{t-p}f_{t-q}^{*\prime}\right)$$

The first term of the above expression is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by $\frac{1}{\bar{T}}\sum_{t=\bar{K}}^T f_{t-p}^* f_{t-q}^{*\prime} = O_p(1)$ and $A = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$, as shown in Bai and Li (2012). The second and third terms are also $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$, which can be proved similarly as Lemma C.1(e) of Bai and Li (2012). Given these results, the second term of J_{11} is $O_p(N^{-1}) + O_p(T^{-1})$. The last term can be proved to be the same magnitude by the similar arguments. Summarizing these results, we have $J_{11} = O_p(N^{-1}) + O_p(T^{-1})$. Terms J_{12} and J_{21} can be proved to be $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ similarly as J_{11}. Then (a) follows.

Consider (b). The left hand side of (b) is equal to

$$\begin{bmatrix} \frac{1}{T}\sum_{t=\bar{K}}^{T}(\tilde{f}_{t-p} - f_{t-p}^*)\tilde{f}_{t-q}' & \frac{1}{T}\sum_{t=\bar{K}}^{T}(\tilde{f}_{t-p} - f_{t-p}^*)g_{t-q}' \\ 0 & 0 \end{bmatrix}$$

The two none-zero terms of the above are $O_p(N^{-1}) + O_p(T^{-1})$, which are shown in (a). Then (b) follows.

Consider (c). The left hand side of (c) is equal to

$$\begin{bmatrix} \frac{1}{T}\sum_{t=\bar{K}}^{T} u_t^*(\tilde{f}_{t-p} - f_{t-p}^*)' \\ 0 \end{bmatrix}.$$

So it suffices to consider term $\frac{1}{T}\sum_{t=\bar{K}}^{T} u_t^*(\tilde{f}_{t-p} - f_{t-p}^*)'$, which, by (A.7), can be written as

$$-\frac{1}{T}\sum_{t=\bar{K}}^{T} u_t^* f_{t-p}^{*\prime} A^* + \frac{1}{T}\sum_{t=\bar{K}}^{T} u_t^* e_{t-p}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}(\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}$$

$$-\frac{1}{T}\sum_{t=\bar{K}}^{T} u_t^* g_{t-p}'\left(\sum_{s=1}^{T} g_s g_s'\right)^{-1}\sum_{s=1}^{T} g_s e_s'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}(\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1}$$

Both $\frac{1}{NT}\sum_{t=\bar{K}}^{T} u_t^* e_{t-p}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}$ and $\frac{1}{NT}\sum_{s=1}^{T} g_s e_s'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}$ can be proved to be $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ similarly as Lemma C.1(e) of Bai and Li (2012). Given these results, together with $A = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ and $(\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda})^{-1} = O_p(N^{-1})$, we have

$$\frac{1}{T}\sum_{t=\bar{K}}^{T} u_t^*(\tilde{f}_t - f_t^*)' = O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}).$$

Then (c) follows. This completes the proof of Lemma A.1. □

Proposition A.2. Under Assumptions A-D, together with the identification condition AU1, for each $k = 1, 2, \ldots, K$, we have

$$\tilde{\Phi}_k - \Phi_k^* = \left(\sum_{t=\bar{K}}^{T} u_t^*\psi_t^{*\prime}\right)\left(\sum_{t=\bar{K}}^{T} \psi_t^*\psi_t^{*\prime}\right)^{-1}(i_k \otimes I_r) + O_p(N^{-1}) + O_p(T^{-1})$$

where $\psi_t^* = (h_{t-1}^{*\prime}, h_{t-2}^{*\prime}, \ldots, h_{t-K}^{*\prime})'$ and i_k is the k-th column of the $K \times K$ identity matrix.

Proof of Proposition A.2. Let $\Phi^* = (\Phi_1^*, \Phi_2^*, \ldots, \Phi_K^*)$ and $\tilde{\Phi}$ be defined similarly. Notice $\tilde{\Phi}$ is obtained by running the regression

$$\tilde{h}_t = \Phi_1\tilde{h}_{t-1} + \Phi_2\tilde{h}_{t-2} + \cdots + \Phi_K\tilde{h}_{t-K} + error$$

So we have

$$\tilde{\Phi} = \left(\sum_{t=\bar{K}}^{T} \tilde{h}_t\tilde{\psi}_t'\right)\left(\sum_{t=\bar{K}}^{T} \tilde{\psi}_t\tilde{\psi}_t'\right)^{-1}$$

where $\tilde{\psi}_t = (\tilde{h}_{t-1}', \tilde{h}_{t-2}', \ldots, \tilde{h}_{t-K}')'$. By $h_t^* = \Phi^*\psi_t^* + u_t^*$,

$$\begin{aligned}\tilde{\Phi} - \Phi^* &= \left[\sum_{t=\bar{K}}^{T}\big(u_t^* + (\tilde{h}_t - h_t^*) - \Phi^*(\tilde{\psi}_t - \psi_t^*)\big)\tilde{\psi}_t'\right]\left[\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right]^{-1} \\ &= \left[\frac{1}{T}\sum_{t=\bar{K}}^{T}\big(u_t^* + (\tilde{h}_t - h_t^*) - \Phi^*(\tilde{\psi}_t - \psi_t^*)\big)\tilde{\psi}_t'\right] \\ &\quad\times\left[\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right]^{-1}\end{aligned}$$

By Lemma A.1(a) and (b),

$$\left[\frac{1}{T}\sum_{t=\bar{K}}^{T}(\tilde{h}_t - h_t^*)\tilde{\psi}_t'\right]\left[\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right]^{-1} = O_p(N^{-1}) + O_p(T^{-1})$$

$$\left[\frac{1}{T}\sum_{t=\bar{K}}^{T}(\tilde{\psi}_t - \psi_t^*)\tilde{\psi}_t'\right]\left[\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right]^{-1} = O_p(N^{-1}) + O_p(T^{-1})$$

By Lemma A.1(a) and (c),

$$\begin{aligned}&\left[\frac{1}{T}\sum_{t=\bar{K}}^{T}u_t^*\tilde{\psi}_t'\right]\left[\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right]^{-1} \\ &= \left[\frac{1}{T}\sum_{t=\bar{K}}^{T}u_t^*\psi_t^{*\prime}\right]\left[\frac{1}{T}\sum_{t=\bar{K}}^{T}\psi_t^*\psi_t^{*\prime}\right]^{-1} + O_p(N^{-1}) + O_p(T^{-1})\end{aligned}$$

Given this result, we have

$$\tilde{\Phi} - \Phi^* = \left(\sum_{t=\bar{K}}^{T}u_t^*\psi_t^{*\prime}\right)\left(\frac{1}{T}\sum_{t=\bar{K}}^{T}\psi_t^*\psi_t^{*\prime}\right)^{-1} + O_p(N^{-1}) + O_p(T^{-1})$$

Post-multiplying $i_k \otimes I_r$ on both sides gives Proposition A.2. □

Now we consider the following auxiliary identification restrictions (denoted by AU2), in which the loading restrictions are the same as AU1 but factor restrictions are imposed on the population.

AU2 The underlying parameter values $\theta^\star = (\Lambda^\star, \Gamma^\star, F^\star, \Phi^\star, \Sigma_{ee})$ satisfy: $\frac{1}{N}\Lambda^{\star\prime}\Sigma_{ee}^{-1}\Lambda^\star = Q^\star$, $E(f_t^\star f_t^{\star\prime}) = I_{r_1}$ and $E(f_t^\star g_t') = 0$, where $Q^\star$ is a diagonal matrix, whose diagonal elements are distinct and arranged in descending order.

Note that the superscript "stars" in $\theta^\star$ and θ^* are different. Different identification restrictions imply different rotations. Because AU1 and AU2 are asymptotically the same (the former with sample moment restriction $\frac{1}{T}\sum_t f_t f_t' = I_{r_1}$ and the latter with population moment restriction $E(f_t f_t') = I_{r_1}$), $\theta^\star$ and θ^* are also asymptotically the same. That is why the deviation of MLE from $\theta^\star$ also converges to zero in probability, which will be proved below.

The following lemma is useful to our analysis.

Lemma A.2 Let Q be an $r \times r$ matrix satisfying

$$QQ' = I_r$$
$$Q'VQ = D$$

where V is an $r \times r$ diagonal matrix with strictly positive and distinct elements, arranged in decreasing order, and D is also diagonal. Then Q must be a diagonal matrix with elements either -1 or 1 and $V = D$.

Lemma A.2 is proved in Bai and Li (2012). The following Propositions A.3 and A.4 summarize the asymptotic results of the QMLE under AU2. These results show that the limiting distributions under AU2 are different.

Proposition A.3. Under Assumptions A-D, together with the identification condition AU2, when $N, T \to \infty$, we have

$$\tilde{\lambda}_i - \lambda_i^\star = V^\star\lambda_i^\star + \mathbf{\Delta}_{ff}^{\star-1}\left(\frac{1}{T}\sum_{t=1}^{T}f_t^\star e_{it}\right) + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1}) \quad \text{(A.9)}$$

$$\tilde{\gamma}_i - \gamma_i^\star = W^\star\lambda_i^\star + \mathbf{\Delta}_{gg}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}g_t e_{it}\right) + O_p(T^{-1}) \quad \text{(A.10)}$$

$$\tilde{f}_t - f_t^\star = -V^{\star\prime}f_t^\star - W^{\star\prime}g_t + \left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\lambda_i^\star\lambda_i^{\star\prime}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\lambda_i^\star e_{it}\right) + O_p(N^{-1}) + O_p(T^{-1}) \quad \text{(A.11)}$$

where $W^\star = \mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf}^\star$ with $\mathbf{\Delta}_{gf}^\star = E(g_t f_t^{\star\prime})$; $\mathbf{\Delta}_{ff}^\star = E(f_t^\star f_t^{\star\prime})$; $V^\star$ is an $r_1 \times r_1$ matrix, which is $O_p(T^{-1/2})$.

Proof of Proposition A.3. Notice that

$$\tilde{\lambda}_i - \lambda_i^\star = (\tilde{\lambda}_i - \lambda_i^*) + (\lambda_i^* - \lambda_i^\star).$$

We show λ_i^* and $\lambda_i^\star$ are close to each other because AU1 and AU2 are asymptotically the same. Different identification restrictions imply different rotations. Let $R^\star$ be the rotation matrix, which transform $(\lambda_i^{*\prime}, \gamma_i^{*\prime})'$ to $(\lambda_i^{\star\prime}, \gamma_i^{\star\prime})'$. Then we have

$$\begin{aligned} z_t &= \Lambda^\star f_t^\star + \Gamma^\star g_t + e_t = [\Lambda^\star, \Gamma^\star]\begin{bmatrix} f_t^\star \\ g_t \end{bmatrix} + e_t \\ &= [\Lambda^*, \Gamma^*]\begin{bmatrix} R_{11}^{\star\prime} & R_{21}^{\star\prime} \\ R_{12}^{\star\prime} & R_{22}^{\star\prime} \end{bmatrix}\begin{bmatrix} R_{11}^{\star\prime} & R_{21}^{\star\prime} \\ R_{12}^{\star\prime} & R_{22}^{\star\prime} \end{bmatrix}^{-1}\begin{bmatrix} f_t^* \\ g_t \end{bmatrix} + e_t \qquad \text{(A.12)} \end{aligned}$$

As mentioned in the main text, due to the fact that the factors g_t are observed, matrix $R_{12}^{\star\prime}$ is fixed to 0 and matrix $R_{22}^{\star\prime}$ is fixed to I_{r_2}. So equation (A.12) reduces to

$$\begin{aligned} z_t &= [\Lambda^\star, \Gamma^\star]\begin{bmatrix} f_t^\star \\ g_t \end{bmatrix} + e_t \\ &= [\Lambda^*, \Gamma^*]\begin{bmatrix} R_{11}^{\star\prime} & R_{21}^{\star\prime} \\ 0 & I_{r_2} \end{bmatrix}\begin{bmatrix} R_{11}^{\star\prime-1} & -R_{11}^{\star\prime-1}R_{21}^{\star\prime} \\ 0 & I_{r_2} \end{bmatrix}\begin{bmatrix} f_t^* \\ g_t \end{bmatrix} + e_t \end{aligned}$$

This gives

$$\lambda_i^\star = R_{11}^\star \lambda_i^*, \;\; \gamma_i^\star = R_{21}^\star \lambda_i^* + \gamma_i^*, \;\; f_t^\star = R_{11}^{\star\prime-1} f_t^* - R_{11}^{\star\prime-1} R_{21}^{\star\prime} g_t. \quad \text{(A.13)}$$

The last equation of (A.13) can also be written as

$$f_t^* = R_{11}^{\star\prime} f_t^\star + R_{21}^{\star\prime} g_t. \qquad \text{(A.14)}$$

Post-multiplying g_t' on both sides and taking summation over t, by $\sum_{t=1}^T g_t f_t^{*\prime} = 0$, we have

$$R_{21}^\star = -\left[\sum_{t=1}^T g_t g_t'\right]^{-1}\left[\sum_{t=1}^T g_t f_t^{\star\prime}\right] R_{11}^\star, \qquad \text{(A.15)}$$

Substituting (A.15) into (A.14),

$$f_t^* = R_{11}^{\star\prime}\left(f_t^\star - \left[\sum_{t=1}^T f_t^\star g_t'\right]\left[\sum_{t=1}^T g_t g_t'\right]^{-1} g_t\right).$$

By $T^{-1}\sum_{t=1}^T f_t^* f_t^{*\prime} = I_{r_1}$, the preceding equation implies

$$\left(\frac{1}{T}\sum_{t=1}^T f_t^\star f_t^{\star\prime}\right) - \left(\frac{1}{T}\sum_{t=1}^T f_t^\star g_t'\right)\left(\frac{1}{T}\sum_{t=1}^T g_t g_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T g_t f_t^{\star\prime}\right) = R_{11}^{\star\prime-1} R_{11}^{\star-1}. \qquad \text{(A.16)}$$

The first equation of (A.13) shows $\Lambda^\star = \Lambda^* R_{11}^{\star\prime}$. So we have

$$\begin{aligned} R_{11}^{\star-1} Q^\star R_{11}^{\star\prime-1} &= R_{11}^{\star-1}\left(\frac{1}{N}\Lambda^{\star\prime}\Sigma_{ee}^{-1}\Lambda^\star\right) R_{11}^{\star\prime-1} \\ &= \frac{1}{N}\Lambda^{*\prime}\Sigma_{ee}^{-1}\Lambda^* = \text{diag}. \qquad \text{(A.17)} \end{aligned}$$

Consider (A.16). By $E(f_t^\star g_t') = 0$, we have

$$\left(\frac{1}{T}\sum_{t=1}^T f_t^\star g_t'\right)\left(\frac{1}{T}\sum_{t=1}^T g_t g_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T g_t f_t^{\star\prime}\right) = O_p(T^{-1}) \qquad \text{(A.18)}$$

The left hand side of (A.16) converges to I_{r_1} in probability. Thus $R_{11}^{\star\prime-1}R_{11}^{\star-1} \xrightarrow{p} I_{r_1}$. Applying Lemma A.2 to $R_{11}^{\star\prime-1}R_{11}^{\star-1} \xrightarrow{p} I_{r_1}$ and $R_{11}^{\star-1}Q^\star R_{11}^{\star\prime-1} = \frac{1}{N}\Lambda^{*\prime}\Sigma_{ee}^{-1}\Lambda^*$ with $Q = R_{11}^{\star\prime-1}$, $V = Q^\star$ and $D = \frac{1}{N}\Lambda^{*\prime}\Sigma_{ee}^{-1}\Lambda^*$, we have $R_{11}^{\star-1}$ converges to a matrix whose diagonal elements either 1 or -1. Since we assume that the sign problem is precluded in our analysis, it follows $R_{11}^{\star-1} \xrightarrow{p} I_{r_1}$. Let

$$U^\star = R_{11}^{\star-1} - I_{r_1}. \qquad \text{(A.19)}$$

Apparently, $U^\star \xrightarrow{p} 0$. Then (A.16) is equivalent to

$$\begin{aligned} &\left(\frac{1}{T}\sum_{t=1}^T [f_t^\star f_t^{\star\prime} - I_{r_1}]\right) - \left(\frac{1}{T}\sum_{t=1}^T f_t^\star g_t'\right)\left(\frac{1}{T}\sum_{t=1}^T g_t g_t'\right)\left(\frac{1}{T}\sum_{t=1}^T g_t f_t^{\star\prime}\right) \\ &= U^\star + U^{\star\prime} + U^{\star\prime}U^\star. \qquad \text{(A.20)} \end{aligned}$$

Also, (A.17) is equivalent to

$$\text{Ndg}\left(U^\star Q^\star + Q^\star U^{\star\prime} + U^\star Q^\star U^{\star\prime}\right) = 0 \qquad \text{(A.21)}$$

where Ndg$\{\cdot\}$ denotes the non-diagonal elements of the argument. Neglecting the terms $U^\star Q^\star U^{\star\prime}$ and $U^{\star\prime}U^\star$ since they are of smaller order than $U^\star$, we can uniquely determine matrix $U^\star$ by solving the equation system (A.20) and (A.21). Let $V^\star$ be the leading term of $U^\star$. It is easy to see that $U^\star = O_p(T^{-1/2})$, $V^\star = O_p(T^{-1/2})$ and $U^\star = V^\star + O_p(T^{-1})$. This result gives $R_{11}^{\star-1} = I_{r_1} + O_p(T^{-1/2})$ by (A.19), which, together with (A.15), implies

$$\begin{aligned} R_{21}^\star &= -\left[\sum_{t=1}^T g_t g_t'\right]^{-1}\left[\sum_{t=1}^T g_t f_t^{\star\prime}\right] + O_p(T^{-1}) \\ &= -\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf}^\star + O_p(T^{-1/2}) \\ &\triangleq -W^\star + O_p(T^{-1}) = O_p(T^{-1/2}) \qquad \text{(A.22)} \end{aligned}$$

Now consider the asymptotic representation of $\tilde{\lambda}_i - \lambda_i^\star$. Notice

$$\tilde{\lambda}_i - \lambda_i^\star = \tilde{\lambda}_i - R_{11}^\star \lambda_i^* = (\tilde{\lambda}_i - \lambda_i^*) - (R_{11}^\star - I_{r_1})\lambda_i^*$$

By (A.1), the above result is equivalent to

$$\begin{aligned} \tilde{\lambda}_i - \lambda_i^\star &= \left[\frac{1}{T}\sum_{t=1}^T f_t^* f_t^{*\prime}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T f_t^* e_{it}\right] - (R_{11}^\star - I_{r_1})\lambda_i^* \\ &\quad + O_p(T^{-1}) + O_p(N^{-1/2}T^{-1/2}) \qquad \text{(A.23)} \end{aligned}$$

By (A.14), we have

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^T f_t^* f_t^{*\prime} &= \frac{1}{T}\sum_{t=1}^T f_t^\star f_t^{\star\prime} + o_p(1) = \mathbf{\Delta}_{ff}^\star + o_p(1), \\ \frac{1}{T}\sum_{t=1}^T f_t^* e_{it} &= \frac{1}{T}\sum_{t=1}^T f_t^\star e_{it} + O_p(T^{-1}). \end{aligned}$$

Notice $R_{11}^\star = (I_{r_1} + U^\star)^{-1} = I_{r_1} - U^\star(I_{r_1} + V^\star)^{-1} = I_{r_1} - U^\star R_{11}^\star$. Then it follows

$$-(R_{11}^\star - I_{r_1})\lambda_i^* = U^\star \lambda_i^\star.$$

Given the above three results, together with $U^\star = V^\star + O_p(T^{-1})$ and (A.23), we have

$$\begin{aligned} \tilde{\lambda}_i - \lambda_i^\star &= V^\star \lambda_i^\star + \left(\frac{1}{T}\sum_{t=1}^T f_t^\star f_t^{\star\prime}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^T f_t^\star e_{it}\right) + O_p(T^{-1}) \\ &\quad + O_p(N^{-1/2}T^{-1/2}). \qquad \text{(A.24)} \end{aligned}$$

We then consider $\tilde{\gamma}_i - \gamma_i$. By (A.13), we have $\tilde{\gamma}_i - \gamma_i^\star = \tilde{\gamma}_i - \gamma_i^* - R_{21}^\star \lambda_i^*$. Then, by (A.3),

$$\tilde{\gamma}_i - \gamma_i^\star = -R_{21}^\star \lambda_i^* + \left(\sum_{t=1}^T g_t g_t'\right)^{-1}\left(\sum_{t=1}^T g_t e_{it}\right).$$

Substituting (A.15) into the above equation and noticing $\lambda_i^\star = R_{11}^\star \lambda_i^*$, we have

$$\begin{aligned} \tilde{\gamma}_i - \gamma_i^\star &= \left(\sum_{t=1}^T g_t g_t'\right)^{-1}\left(\sum_{t=1}^T g_t (e_{it} + f_t^{\star\prime}\lambda_i^\star)\right) \\ &= W^\star \lambda_i^\star + \mathbf{\Delta}_{gg}^{-1}\left(\sum_{t=1}^T g_t e_{it}\right) + O_p(T^{-1}). \qquad \text{(A.25)} \end{aligned}$$

Now consider $\tilde{f}_t - f_t^\star$. By (A.13),

$$\tilde{f}_t - f_t^\star = \tilde{f}_t - R_{11}^{\star\prime-1} f_t^* + R_{11}^{\star\prime-1} R_{21}^{\star\prime} g_t$$

By $R_{11}^{\star\prime-1} = I_{r_1} + U^{\star\prime}$, the above equation is equal to

$$\tilde{f}_t - f_t^\star = (\tilde{f}_t - f_t^*) - U^{\star\prime} f_t^* + R_{11}^{\star\prime-1} R_{21}^{\star\prime} g_t$$

Substituting (A.3) into the above result, we have

$$\tilde{f}_t - f_t^\star = -U^{\star\prime} f_t^* + R_{11}^{\star\prime-1} R_{21}^{\star\prime} g_t + \left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i^*\lambda_i^{*\prime}\right)^{-1}\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i^* e_{it}'\right) + O_p(N^{-1}) + O_p(T^{-1}) \quad \text{(A.26)}$$

However, by (A.13) together with $R_{11}^\star = (I_{r_1} + U^\star)^{-1}$ and $U^\star = O_p(T^{-1/2})$, we have

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i^*\lambda_i^{*\prime} = \frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i^\star\lambda_i^{\star\prime} + o_p(1)$$

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i^* e_{it} = \frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i^\star e_{it} + O_p(N^{-1/2}T^{-1/2})$$

In addition, by (A.14), (A.12) and $U^\star = V^\star + O_p(T^{-1})$, we have

$$U^{\star\prime} f_t^* = V^{\star\prime} f_t^\star + O_p(T^{-1})$$

$$R_{11}^{\star\prime-1} R_{21}^{\star\prime} g_t = -\left(\sum_{t=1}^T f_t^\star g_t'\right)\left(\sum_{t=1}^T g_t g_t'\right)^{-1} g_t = -W^{\star\prime} g_t + O_p(T^{-1})$$

Given the above results, by (A.26), we have the last expression of Proposition A.3. This completes the proof of Proposition A.3. □

The asymptotic result for $\tilde{\Phi}_k$ under AU2 is given in the following proposition.

Proposition A.4. Under Assumptions A-D, together with the identification condition AU2, we have

$$\tilde{\Phi}_k - \Phi_k^\star = \left(\sum_{t=\bar{K}}^T u_t^\star \psi_t^{\star\prime}\right)\left(\sum_{t=\bar{K}}^T \psi_t^\star \psi_t^{\star\prime}\right)^{-1}(i_k \otimes I_r) - B^{\star\prime}\Phi_k^\star + \Phi_k^\star B^{\star\prime} + O_p(N^{-1}) + O_p(T^{-1})$$

where $B^\star$ is defined as

$$B^\star = \begin{bmatrix} V^\star & 0 \\ W^\star & 0 \end{bmatrix}.$$

Proof of Proposition A.4. Notice

$$h_t^* = \Phi_1^* h_{t-1}^* + \Phi_2^* h_{t-2}^* + \cdots + \Phi_K^* h_{t-K}^* + u_t^*,$$

and

$$h_t^\star = \Phi_1^\star h_{t-1}^\star + \Phi_2^\star h_{t-2}^\star + \cdots + \Phi_K^\star h_{t-K}^\star + u_t^\star.$$

By $h_t^\star = R^{\star\prime-1} h_t^*$, it follows that $\Phi_k^\star = R^{\star\prime-1}\Phi_k^* R^{\star\prime}$. Thus,

$$\tilde{\Phi}_k - \Phi_k^\star = \tilde{\Phi}_k - R^{\star\prime-1}\Phi_k^* R^{\star\prime} \quad \text{(A.27)}$$

However, by $R_{11}^{\star-1} = I_{r_1} + V^\star + O_p(T^{-1})$ and $R_{21}^\star = -W^\star + O_p(T^{-1})$, we have

$$R^{\star\prime-1} = \begin{bmatrix} R_{11}^{\star\prime-1} & -R_{11}^{\star\prime-1}R_{21}^{\star\prime} \\ 0 & I_{r_2} \end{bmatrix} = I_r + \begin{bmatrix} V^{\star\prime} & W^{\star\prime} \\ 0 & 0 \end{bmatrix} + O_p(T^{-1})$$
$$= I_r + B^{\star\prime} + O_p(T^{-1}) \quad \text{(A.28)}$$

Given the above result, we have $R^{\star\prime} = I_r - B^{\star\prime} + O_p(T^{-1})$. Substituting the preceding two results into (A.27), we have

$$\tilde{\Phi}_k - \Phi_k^\star = \tilde{\Phi}_k - \Phi_k^* - B^{\star\prime}\Phi_k^* + \Phi_k^* B^{\star\prime} + O_p(T^{-1}).$$

By Proposition A.2, we can rewrite the above result as

$$\tilde{\Phi}_k - \Phi_k^* = \left(\sum_{t=\bar{K}}^T u_t^* \psi_t^{*\prime}\right)\left(\sum_{t=\bar{K}}^T \psi_t^* \psi_t^{*\prime}\right)^{-1}(i_k \otimes I_r) - B^{\star\prime}\Phi_k^* + \Phi_k^* B^{\star\prime} + O_p(N^{-1}) + O_p(T^{-1}).$$

By $h_t^\star = R^{\star\prime-1}h_t^*$, we have $h_t^* = R^{\star\prime}h_t^\star = h_t^\star + (R^\star - I_r)'h_t^\star$. Given this result, together with the fact that $R^\star - I_r = O_p(T^{-1/2})$, we have

$$\left(\sum_{t=\bar{K}}^T u_t^* \psi_t^{*\prime}\right)\left(\sum_{t=\bar{K}}^T \psi_t^* \psi_t^{*\prime}\right)^{-1} = \left(\sum_{t=\bar{K}}^T u_t^\star \psi_t^{\star\prime}\right)\left(\sum_{t=\bar{K}}^T \psi_t^\star \psi_t^{\star\prime}\right)^{-1} + O_p(T^{-1})$$

and

$$B^{\star\prime}\Phi_k^* = B^{\star\prime}\Phi_k^\star + O_p(T^{-1}), \qquad \Phi_k^* B^{\star\prime} = \Phi_k^\star B^{\star\prime} + O_p(T^{-1})$$

Given these results, we have

$$\tilde{\Phi}_k - \Phi_k^\star = \left(\sum_{t=\bar{K}}^T u_t^\star \psi_t^{\star\prime}\right)\left(\sum_{t=\bar{K}}^T \psi_t^\star \psi_t^{\star\prime}\right)^{-1}(i_k \otimes I_r) - B^{\star\prime}\Phi_k^\star + \Phi_k^\star B^{\star\prime} + O_p(N^{-1}) + O_p(T^{-1})$$

This completes the proof of Proposition A.4. □

APPENDIX B: THE ASYMPTOTIC RESULTS AND THEIR PROOFS UNDER IRa

As in the main text, we use (Λ, Γ, F) to denote the underlying parameters satisfying IRa. Let R be the rotation matrix which transforms $(\lambda_i^{\star\prime}, \gamma_i^{\star\prime})'$ into $(\lambda_i', \gamma_i')'$. Then we have

$$z_t = \Lambda f_t + \Gamma g_t + e_t = [\Lambda, \Gamma]\begin{bmatrix} f_t \\ g_t \end{bmatrix} + e_t$$
$$= [\Lambda^\star, \Gamma^\star]\begin{bmatrix} R_{11}' & R_{21}' \\ 0 & I_{r_2} \end{bmatrix}\begin{bmatrix} R_{11}'^{-1} & -R_{11}'^{-1}R_{21}' \\ 0 & I_{r_2} \end{bmatrix}\begin{bmatrix} f_t^\star \\ g_t \end{bmatrix} + e_t \quad \text{(B.1)}$$

Then we have

$$\lambda_i = R_{11}\lambda_i^\star, \ \gamma_i = \gamma_i^\star + R_{21}\lambda_i^\star, \ f_t = R_{11}'^{-1} f_t^\star - R_{11}'^{-1}R_{21}' g_t. \quad \text{(B.2)}$$

The last equation in (B.2) can be written as

$$f_t^\star = R_{11}' f_t + R_{21}' g_t. \quad \text{(B.3)}$$

Note that the rotation matrix R is nonrandom. To see this, both AU2 and IRa impose restrictions on the loadings and the covariance of h_t. So the rotation matrix R, which transform the underlying parameters from AU2 to IRa, only involves loadings and covariance of h_t. Thus it is nonrandom. This is in contrast with $R^\star$, which is random since AU1 involves f_t.

Post-multiplying g_t' on both sides and taking the expectation, by $E(f_t^\star g_t') = 0$, we have

$$R_{21} = -\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf} R_{11}.$$

Define $\phi_t = R_{11}'^{-1} f_t^\star$. From the above results, ϕ has an alternative expression

$$\phi_t = f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1} g_t. \quad \text{(B.4)}$$

The following lemmas will be used in the subsequent proof.

Lemma B.1. For any compatible matrices $\mathcal{A}$ and $\mathcal{B}$ and their corresponding estimates $\hat{\mathcal{A}}$ and $\hat{\mathcal{B}}$, we have

$$\hat{\mathcal{A}}\hat{\mathcal{B}}^{-1}\hat{\mathcal{A}}' - \mathcal{A}\mathcal{B}^{-1}\mathcal{A}' = (\hat{\mathcal{A}} - \mathcal{A})\mathcal{B}^{-1}\mathcal{A}' + \mathcal{A}\mathcal{B}^{-1}(\hat{\mathcal{A}} - \mathcal{A})'$$
$$-\mathcal{A}\mathcal{B}^{-1}(\hat{\mathcal{B}} - \mathcal{B})\mathcal{B}^{-1}\mathcal{A}' + R$$

where

$$R = -(\hat{\mathcal{A}} - \mathcal{A})\hat{\mathcal{B}}^{-1}(\hat{\mathcal{B}} - \mathcal{B})\mathcal{B}^{-1}\mathcal{A}' + (\hat{\mathcal{A}} - \mathcal{A})\hat{\mathcal{B}}^{-1}(\hat{\mathcal{A}} - \mathcal{A})' + \mathcal{A}\hat{\mathcal{B}}^{-1}(\hat{\mathcal{B}}-\mathcal{B})\mathcal{B}^{-1}(\hat{\mathcal{B}}-\mathcal{B})\mathcal{B}^{-1}\hat{\mathcal{A}}' - \mathcal{A}\hat{\mathcal{B}}^{-1}(\hat{\mathcal{B}}-\mathcal{B})\mathcal{B}^{-1}(\hat{\mathcal{A}}-\mathcal{A})'.$$

Lemma B.1 can be proved easily by matrix algebra.

Lemma B.2 Under Assumptions A-D, we have

$$(a)\ \frac{1}{T}\tilde{H}'M_{\tilde{\Psi}}\tilde{H} - \frac{1}{T}H^{*\prime}M_{\Psi^*}H^* = O_p(N^{-1}) + O_p(T^{-1})$$

$$(b)\ \frac{1}{T}H^{\star\prime}M_{\Psi^\star}H^\star - \frac{1}{T}H^{*\prime}M_{\Psi^*}H^* = B^{\star\prime}\Omega^\star + \Omega^\star B^\star + O_p(T^{-1})$$

$$(c)\ \frac{1}{T}\tilde{H}'M_{\tilde{\Psi}}\tilde{H} - \frac{1}{T}H^{\star\prime}M_{\Psi^\star}H^\star = -B^{\star\prime}\Omega^\star - \Omega^\star B^\star + O_p(N^{-1}) + O_p(T^{-1})$$

where $\frac{1}{T}\tilde{H}'M_{\tilde{\Psi}}\tilde{H}$ is defined as

$$\frac{1}{T}\tilde{H}'M_{\tilde{\Psi}}\tilde{H} = \frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{h}_t' - \left(\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{\psi}_t'\right)\left(\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{h}_t'\right),$$

and $\frac{1}{T}H^{*\prime}M_{\Psi^*}H^*$ and $\frac{1}{T}H^{\star\prime}M_{\Psi^\star}H^\star$ are defined similarly.

Proof of Lemma B.2. Consider (a). By Lemma A.1(a), we have

$$\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{h}_t' - \frac{1}{T}\sum_{t=\bar{K}}^{T}h_t^*h_t^{*\prime} = O_p(N^{-1}) + O_p(T^{-1})$$

$$\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{\psi}_t' - \frac{1}{T}\sum_{t=\bar{K}}^{T}h_t^*\psi_t^{*\prime} = O_p(N^{-1}) + O_p(T^{-1})$$

$$\frac{1}{T}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t' - \frac{1}{T}\sum_{t=\bar{K}}^{T}\psi_t^*\psi_t^{*\prime} = O_p(N^{-1}) + O_p(T^{-1})$$

Given the above results, together with Lemma B.1, we have (a).

Consider (b). By $h_t^\star = R^{\star\prime-1}h_t^*$, we have $\psi_t^\star = (I_K \otimes R^{\star\prime-1})\psi_t^*$. This gives

$$\frac{1}{T}H^{\star\prime}M_{\Psi^\star}H^\star = \frac{1}{T}H^{\star\prime}M_{\Psi^*}H^\star$$

By $H^\star = H^*R^{\star-1}$, we have

$$\frac{1}{T}H^{\star\prime}M_{\Psi^*}H^\star = R^{\star\prime-1}\left(\frac{1}{T}H^{*\prime}M_{\Psi^*}H^*\right)R^{\star-1}$$

However, (A.28) shows that $R^{\star\prime-1} = I_r + B^{\star\prime} + O_p(T^{-1})$. Thus, we have

$$\frac{1}{T}H^{\star\prime}M_{\Psi^\star}H^\star - \frac{1}{T}H^{*\prime}M_{\Psi^*}H^* = R^{\star\prime-1}\left(\frac{1}{T}H^{*\prime}M_{\Psi^*}H^*\right)R^{\star-1} - \frac{1}{T}H^{*\prime}M_{\Psi^*}H^* = B^{\star\prime}\left(\frac{1}{T}H^{*\prime}M_{\Psi^*}H^*\right) + \left(\frac{1}{T}H^{*\prime}M_{\Psi^*}H^*\right)B^\star + O_p(T^{-1}) \quad \text{(B.5)}$$

Now consider $\frac{1}{T}H^{*\prime}M_{\Psi^*}H^*$, which is equivalent to

$$\frac{1}{T}\sum_{t=\bar{K}}^{T}u_t^*u_t^{*\prime} - \left(\frac{1}{T}\sum_{t=\bar{K}}^{T}u_t^*\psi_t^{*\prime}\right)\left(\frac{1}{T}\sum_{t=\bar{K}}^{T}\psi_t^*\psi_t^{*\prime}\right)^{-1}\left(\frac{1}{T}\sum_{t=\bar{K}}^{T}\psi_t^*u_t^{*\prime}\right)$$

The second term is $O_p(T^{-1})$. The first term, by $u_t^\star = R^{\star\prime-1}u_t^*$ and $R^{\star\prime} = I_r + O_p(T^{-1/2})$, is equal to

$$R^{\star\prime}\left(\frac{1}{T}\sum_{t=\bar{K}}^{T}u_t^\star u_t^{\star\prime}\right)R^\star = \frac{1}{T}\sum_{t=\bar{K}}^{T}u_t^\star u_t^{\star\prime} + O_p(T^{-1/2}) = \Omega^\star + O_p(T^{-1/2}).$$

Then it follows

$$\frac{1}{T}H^{*\prime}M_{\Psi^*}H^* = \Omega^\star + O_p(T^{-1/2}) \quad \text{(B.6)}$$

Substituting (B.6) into (B.5), we have (b).

Result (c) is a direct result of (a) and (b). This completes the proof of Lemma B.2. □

Note that Theorems 1–3 under IRa are implied by the following results:

Proposition B.1. Under Assumption A-D, together with the identification condition IRa, we have

$$(a)\ \hat{\lambda}_i - \lambda_i = V\lambda_i + \mathbf{\Delta}_{\phi\phi}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\phi_t e_{it}\right) + O_p(N^{-1}) + O_p(T^{-1})$$

$$(b)\ \hat{\gamma}_i - \gamma_i = W\lambda_i + \mathbf{\Delta}_{\eta\eta}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it}\right) + O_p(N^{-1}) + O_p(T^{-1})$$

$$(c)\ \hat{f}_t - f_t = \left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\lambda_i\lambda_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\lambda_i e_{it}\right) - V'f_t - W'g_t + O_p(N^{-1}) + O_p(T^{-1})$$

where $\text{vec}(V) = \mathbb{B}_Q^{-1}\mathbb{P}_1 D_{r_1}^{+}\frac{1}{T}\sum_{t=\bar{K}}^{T}[\varepsilon_t \otimes \varepsilon_t - \text{vec}(I_{r_1})]$; $\phi_t = f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}g_t$; $\mathbf{\Delta}_{\phi\phi} = E(\phi_t\phi_t')$; $W = \Omega_{\boldsymbol{vv}}^{-1}\frac{1}{T}\sum_{t=\bar{K}}^{T}\boldsymbol{v}_t\varepsilon_t'$; $\eta_t = g_t - \mathbf{\Delta}_{gf}\mathbf{\Delta}_{ff}^{-1}f_t$; $\mathbf{\Delta}_{\eta\eta} = E(\eta_t\eta_t')$.

Proof of Proposition B.1. Consider the VAR expression under AU2:

$$h_t^\star = \Phi_1^\star h_{t-1}^\star + \Phi_2^\star h_{t-2}^\star + \cdots + \Phi_K^\star h_{t-K}^\star + u_t^\star.$$

Pre-multiplying R'^{-1} gives

$$h_t = \left(R'^{-1}\Phi_1^\star R'\right)h_{t-1} + \cdots + \left(R'^{-1}\Phi_K^\star R'\right)h_{t-K} + R'^{-1}u_t^\star.$$

So we have $\Phi_i = R'^{-1}\Phi_i^\star R'$ for $i = 1, 2, \ldots, K$ and $u_t = R'^{-1}u_t^\star$. Then we have

$$\varepsilon_t = R_{11}'^{-1}\varepsilon_t^\star - R_{11}'^{-1}R_{21}'\boldsymbol{v}_t^\star,$$
$$\boldsymbol{v}_t = \boldsymbol{v}_t^\star. \quad \text{(B.7)}$$

Post-multiplying $\boldsymbol{v}_t'$ on both sides and taking the expectation, by $E(\varepsilon_t\boldsymbol{v}_t') = 0$, we have

$$R_{21} = \Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^\star, \quad \text{(B.8)}$$

Substituting the proceeding result into (B.7), by $E(\varepsilon_t\varepsilon_t') = I_{r_1}$, we have

$$\Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\star = \Omega_{\varepsilon\varepsilon}^\star - \Omega_{\varepsilon\boldsymbol{v}}^\star\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^\star = R_{11}'R_{11}. \quad \text{(B.9)}$$

where $\Omega_{\varepsilon\varepsilon}^\star = E(\varepsilon_t^\star\varepsilon_t^{\star\prime})$, $\Omega_{\boldsymbol{vv}}^\star = E(\boldsymbol{v}_t^\star\boldsymbol{v}_t^{\star\prime})$ and $\Omega_{\varepsilon\boldsymbol{v}}^\star = E(\varepsilon_t^\star\boldsymbol{v}_t^{\star\prime})$. In addition, the identification condition also requires that

$$Q = \frac{1}{N}\Lambda'\Sigma_{ee}^{-1}\Lambda = R_{11}\left(\frac{1}{N}\Lambda^{\star\prime}\Sigma_{ee}^{-1}\Lambda^\star\right)R_{11}'.$$

This is equivalent to

$$Q^\star = \frac{1}{N}\Lambda^{\star\prime}\Sigma_{ee}^{-1}\Lambda^\star = R_{11}^{-1}QR_{11}'^{-1}. \quad \text{(B.10)}$$

However, our estimation procedure implies that the estimators of R_{11}, R_{21}, denoted by $\hat{R}_{11}$, $\hat{R}_{21}$, satisfy

$$\hat{R}_{21} = \tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon} \quad \text{(B.11)}$$

$$\hat{R}_{11}'\hat{R}_{11} = \tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}} = \tilde{\Omega}_{\varepsilon\varepsilon} - \tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon} \tag{B.12}$$

$$\hat{R}_{11}\left(\frac{1}{N}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}\right)\hat{R}_{11}' = \text{diag} \tag{B.13}$$

where $\tilde{\Omega}_{\varepsilon\varepsilon}, \tilde{\Omega}_{\boldsymbol{vv}}, \tilde{\Omega}_{\varepsilon\boldsymbol{v}}$ are submatrices of $\tilde{\Omega}$, which is defined as

$$\tilde{\Omega} = \frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{u}_t\tilde{u}_t'$$

with $\tilde{u}_t$ being the residuals of the regression

$$\tilde{h}_t = \Phi_1\tilde{h}_{t-1} + \Phi_2\tilde{h}_{t-2} + \cdots + \Phi_K\tilde{h}_{t-K} + error$$

Let $\tilde{\psi}_t = (\tilde{h}_{t-1}', \tilde{h}_{t-2}'\ldots, \tilde{h}_{t-K}')'$. Thus

$$\tilde{u}_t = \tilde{h}_t - \left(\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{\psi}_t'\right)\left(\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right)^{-1}\tilde{\psi}_t$$

So we have

$$\tilde{\Omega} = \frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{h}_t' - \left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{\psi}_t'\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{h}_t'\right)$$

The above result can be rewritten as

$$\begin{aligned}\tilde{\Omega} - \Omega^\star = &\Bigg\{\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{h}_t' - \left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{h}_t\tilde{\psi}_t'\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{\psi}_t'\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\tilde{\psi}_t\tilde{h}_t'\right)\\ &-\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}h_t^\star h_t^{\star\prime} + \left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}h_t^\star\psi_t^{\star\prime}\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\psi_t^\star\psi_t^{\star\prime}\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\psi_t^\star h_t^{\star\prime}\right)\Bigg\}\\ &+\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(u_t^\star u_t^{\star\prime} - \Omega^\star) - \left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}u_t^\star\psi_t^{\star\prime}\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\psi_t^\star\psi_t^{\star\prime}\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\psi_t^\star u_t^{\star\prime}\right)\end{aligned} \tag{B.14}$$

where $\psi_t^\star = (h_{t-1}^{\star\prime}, h_{t-2}^{\star\prime}, \ldots, h_{t-K}^{\star\prime})'$. The expression in bracket is given in Lemma B.2(c). Given this result, together with

$$\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}u_t^\star\psi_t^{\star\prime}\right)\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\psi_t^\star\psi_t^{\star\prime}\right)^{-1}\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\psi_t^\star u_t^{\star\prime}\right) = O_p(T^{-1}),$$

we have

$$\begin{aligned}\tilde{\Omega} - \Omega^\star &= -B^{\star\prime}\Omega^\star - \Omega^\star B^\star + \frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(u_t^\star u_t^{\star\prime} - \Omega^\star)\\ &\quad + O_p(N^{-1}) + O_p(T^{-1}).\end{aligned} \tag{B.15}$$

The above result implies

$$\begin{aligned}\tilde{\Omega}_{\varepsilon\varepsilon} - \Omega_{\varepsilon\varepsilon}^\star &= -V^{\star\prime}\Omega_{\varepsilon\varepsilon}^\star - W^{\star\prime}\Omega_{\boldsymbol{v}\varepsilon}^\star - \Omega_{\varepsilon\varepsilon}^\star V^\star - \Omega_{\varepsilon\boldsymbol{v}}^\star W^\star\\ &\quad+\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\varepsilon_t^\star\varepsilon_t^{\star\prime} - \Omega_{\varepsilon\varepsilon}^\star) + O_p(N^{-1}) + O_p(T^{-1});\end{aligned} \tag{B.16}$$

$$\begin{aligned}\tilde{\Omega}_{\varepsilon\boldsymbol{v}} - \Omega_{\varepsilon\boldsymbol{v}}^\star &= -V^{\star\prime}\Omega_{\varepsilon\boldsymbol{v}}^\star - W^{\star\prime}\Omega_{\boldsymbol{vv}}^\star + \frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\varepsilon_t^\star\boldsymbol{v}_t^{\star\prime} - \Omega_{\varepsilon\boldsymbol{v}}^\star)\\ &\quad+ O_p(N^{-1}) + O_p(T^{-1});\end{aligned} \tag{B.17}$$

$$\tilde{\Omega}_{\boldsymbol{vv}} - \Omega_{\boldsymbol{vv}}^\star = \frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\boldsymbol{v}_t^\star\boldsymbol{v}_t^{\star\prime} - \Omega_{\boldsymbol{vv}}^\star) + O_p(N^{-1}) + O_p(T^{-1}). \tag{B.18}$$

By (B.15), we have $\tilde{\Omega} - \Omega^\star \xrightarrow{p} 0$. Then it follows $\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}} - \Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\star \xrightarrow{p} 0$, where $\tilde{\Omega}_{\varepsilon\varepsilon\cdot\boldsymbol{v}}$ and $\Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\star$ are defined in (B.9) and (B.12). Thus

$$\hat{R}_{11}'\hat{R}_{11}R_{11}^{-1}R_{11}'^{-1} \xrightarrow{p} I_{r_1},$$

which, by the fact that $AB = I$ then $BA = I$, leads to

$$(\hat{R}_{11}R_{11}^{-1})'(\hat{R}_{11}R_{11}^{-1}) \xrightarrow{p} I_{r_1} \tag{B.19}$$

Furthermore, by (B.13), we have

$$(\hat{R}_{11}R_{11}^{-1})\left[R_{11}\left(\frac{1}{N}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}\right)R_{11}'\right](\hat{R}_{11}R_{11}^{-1})' = \text{diag}$$

By $\frac{1}{N}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda} - \frac{1}{N}\Lambda^{\star\prime}\Sigma_{ee}^{-1}\Lambda^\star = o_p(1)$ and $R_{11}\frac{1}{N}\Lambda^{\star\prime}\Sigma_{ee}^{-1}\Lambda^\star R_{11}' = \frac{1}{N}\Lambda'\Sigma_{ee}^{-1}\Lambda = Q$, we have

$$(\hat{R}_{11}R_{11}^{-1})Q(\hat{R}_{11}R_{11}^{-1})' = \text{diag} \tag{B.20}$$

Notice Q is a diagonal matrix by identification. Applying Lemma A.2 to (B.19) and (B.20), we have $\hat{R}_{11}R_{11}^{-1}$ converges to a diagonal matrix whose diagonal elements are either 1 or -1. However, the possibility of -1 is precluded by our sign restrictions. Given this result, we have $\hat{R}_{11} - R_{11} \xrightarrow{p} 0$. Henceforth, we use $\widehat{\Delta R}_{11}$ to denote $\hat{R}_{11} - R_{11}$. Apparently $\widehat{\Delta R}_{11} \xrightarrow{p} 0$. By (B.9) and (B.12), we have

$$\hat{R}_{11}'\hat{R}_{11} - R_{11}'R_{11} = \tilde{\Omega}_{\varepsilon\varepsilon} - \Omega_{\varepsilon\varepsilon}^\star - (\tilde{\Omega}_{\varepsilon\boldsymbol{v}}\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon} - \Omega_{\varepsilon\boldsymbol{v}}^\star\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^\star)$$

Substituting (B.16)-(B.18) into the above equation, together with Lemma B.1, we have

$$\begin{aligned}&\widehat{\Delta R}_{11}'R_{11} + R_{11}'\widehat{\Delta R}_{11} + \widehat{\Delta R}_{11}'\widehat{\Delta R}_{11} = -V^{\star\prime}\Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\star - \Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\star V^\star\\ &+\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\left[(\varepsilon_t^\star - \Omega_{\varepsilon\boldsymbol{v}}^\star\Omega_{\boldsymbol{vv}}^{\star-1}\boldsymbol{v}_t^\star)(\varepsilon_t^\star - \Omega_{\varepsilon\boldsymbol{v}}^\star\Omega_{\boldsymbol{vv}}^{\star-1}\boldsymbol{v}_t^\star)' - \Omega_{\varepsilon\varepsilon\cdot\boldsymbol{v}}^\star\right]\\ &+O_p(N^{-1}) + O_p(T^{-1}).\end{aligned}$$

However, by (B.7) and (B.8), we have $R_{11}'\varepsilon_t = \varepsilon_t^\star - \Omega_{\varepsilon\boldsymbol{v}}^\star\Omega_{\boldsymbol{vv}}^{\star-1}\boldsymbol{v}_t^\star$. Given this result, together with (B.9), we have

$$\begin{aligned}&\widehat{\Delta R}_{11}'R_{11} + R_{11}'\widehat{\Delta R}_{11} + \widehat{\Delta R}_{11}'\widehat{\Delta R}_{11} = -V^{\star\prime}R_{11}'R_{11} - R_{11}'R_{11}V^\star\\ &+R_{11}'\left[\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\varepsilon_t\varepsilon_t' - I_{r_1}\right]R_{11} + O_p(N^{-1}) + O_p(T^{-1}).\end{aligned} \tag{B.21}$$

Pre-multiplying $R_{11}'^{-1}$ and post-multiplying R_{11}^{-1} on both sides, and neglecting the smaller order term $R_{11}'^{-1}\widehat{\Delta R}_{11}'\widehat{\Delta R}_{11}R_{11}^{-1}$, we have

$$\begin{aligned}&\left(\widehat{\Delta R}_{11}R_{11}^{-1} + R_{11}V^\star R_{11}^{-1}\right) + \left(\widehat{\Delta R}_{11}R_{11}^{-1} + R_{11}V^\star R_{11}^{-1}\right)'\\ &= \frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\varepsilon_t\varepsilon_t' - I_{r_1}) + O_p(N^{-1}) + O_p(T^{-1}).\end{aligned} \tag{B.22}$$

Now consider

$$\begin{aligned}&\frac{1}{N}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda} - \frac{1}{N}\Lambda^{\star\prime}\Sigma_{ee}^{-1}\Lambda^\star = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}(\tilde{\lambda}_i - \lambda_i^\star)\tilde{\lambda}_i'\\ &+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}\tilde{\lambda}_i(\tilde{\lambda}_i - \lambda_i^\star)' - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\tilde{\sigma}_i^2}(\tilde{\lambda}_i - \lambda_i^\star)(\tilde{\lambda}_i - \lambda_i^\star)'\\ &+\frac{1}{N}\sum_{i=1}^{N}\lambda_i^\star\lambda_i^{\star\prime}\left(\frac{1}{\tilde{\sigma}_i^2} - \frac{1}{\sigma_i^2}\right).\end{aligned}$$

The last term is $O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ which is shown in Bai and Li (2012). The third term is $O_p(T^{-1})$. The first two terms are $V^\star Q^\star + Q^\star V^{\star\prime} + O_p(N^{-1/2}T^{-1/2}) + O_p(T^{-1})$ by Proposition A.3.

Then it follows

$$\frac{1}{N}\tilde{\Lambda}'\tilde{\Sigma}_{ee}^{-1}\tilde{\Lambda}-\frac{1}{N}\Lambda^{\star\prime}\Sigma_{ee}^{-1}\Lambda^{\star}=V^{\star}Q^{\star}+Q^{\star}V^{\star\prime}+O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1}). \quad (B.23)$$

Given the above results, (B.13) is equivalent to

$$\text{Ndg}\left\{\hat{R}_{11}(Q^{\star}+V^{\star}Q^{\star}+Q^{\star}V^{\star\prime})\hat{R}_{11}'\right\}=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1}).$$

Substituting (B.10) into the proceeding equation, we have

$$\text{Ndg}\left\{\hat{R}_{11}(R_{11}^{-1}QR_{11}'^{-1}+V^{\star}R_{11}^{-1}QR_{11}'^{-1}+R_{11}^{-1}QR_{11}'^{-1}V^{\star\prime})\hat{R}_{11}'\right\}=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1}).$$

Replace $\hat{R}_{11}=\widehat{\Delta R}_{11}+R_{11}$, the left hand side is (neglecting Ndg)

$$Q+\widehat{\Delta R}_{11}R_{11}^{-1}Q+Q(\widehat{\Delta R}_{11}R_{11}^{-1})'+\widehat{\Delta R}_{11}R_{11}^{-1}Q(\widehat{\Delta R}_{11}R_{11}^{-1})'Q+R_{11}V^{\star}R_{11}^{-1}Q+\widehat{\Delta R}_{11}V^{\star}R_{11}^{-1}Q+\hat{R}_{11}V^{\star}R_{11}^{-1}Q(\widehat{\Delta R}_{11}R_{11}^{-1})'+QR_{11}'^{-1}V^{\star\prime}R_{11}'+QR_{11}'^{-1}V^{\star\prime}\widehat{\Delta R}_{11}'+(\widehat{\Delta R}_{11}R_{11}^{-1})QR_{11}'^{-1}V^{\star\prime}\hat{R}_{11}'$$

By neglecting the terms of smaller magnitude and noticing that $\text{Ndg}(Q)=0$, we have

$$\text{Ndg}\left\{\left(\widehat{\Delta R}_{11}R_{11}^{-1}+R_{11}V^{\star}R_{11}^{-1}\right)Q+Q\left(\widehat{\Delta R}_{11}R_{11}^{-1}+R_{11}V^{\star}R_{11}^{-1}\right)'\right\}=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1}). \quad (B.24)$$

Let $\mathbb{V}=\widehat{\Delta R}_{11}R_{11}^{-1}+R_{11}V^{\star}R_{11}^{-1}$. Taking the half-vectorization operation $\text{vech}(\cdot)$ which stacks the elements on and below the diagonal of the argument into a vector on both sides of (B.22), we get

$$\text{vech}(\mathbb{V}+\mathbb{V}')=\text{vech}\left[\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\varepsilon_t\varepsilon_t'-I_{r_1})\right]+O_p(N^{-1})+O_p(T^{-1}).$$

By the definitions of duplication matrix D_{r_1} and its Moore-Penrose inverse $D_{r_1}^{+}$, and symmetrizer matrix $S_{r_1}=(I_{r_1^2}+K_{r_1})/2$, the left hand side of the above equation can be written as

$$\text{vech}(\mathbb{V}+\mathbb{V}')=D_{r_1}^{+}\text{vec}(\mathbb{V}+\mathbb{V}')=2D_{r_1}^{+}S_{r_1}\text{vec}(\mathbb{V})=2D_{r_1}^{+}\text{vec}(\mathbb{V}),$$

where the last equation is due to $D_{r_1}^{+}S_{r_1}=D_{r_1}^{+}$, we have

$$2D_{r_1}^{+}\text{vec}(\mathbb{V})=D_{r_1}^{+}\text{vec}\left[\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\varepsilon_t\varepsilon_t'-I_{r_1})\right]+O_p(N^{-1})+O_p(T^{-1}). \quad (B.25)$$

Let $\text{veck}(M)$ be the operation which stacks the elements below the diagonal into a vector. Let $\mathbb{D}_1$ be the matrix such that $\text{veck}(M)=\mathbb{D}_1\text{vec}(M)$ for any symmetric matrix M. By (B.24), we have

$$\text{veck}(\mathbb{V}Q+Q\mathbb{V}')=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1}).$$

implying

$$\mathbb{D}_1[Q\otimes I_{r_1}+(I_{r_1}\otimes Q)K_{r_1}]\text{vec}(\mathbb{V})=O_p(N^{-1/2}T^{-1/2})+O_p(T^{-1}). \quad (B.26)$$

The preceding two equations imply

$$\begin{bmatrix}2D_{r_1}^{+}\\ \mathbb{D}_1[Q\otimes I_{r_1}+(I_{r_1}\otimes Q)K_{r_1}]\end{bmatrix}\text{vec}(\mathbb{V})=\begin{bmatrix}D_{r_1}^{+}\text{vec}\left(\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\varepsilon_t\varepsilon_t'-I_{r_1})\right)\\ 0\end{bmatrix}+O_p(N^{-1})+O_p(T^{-1}).$$

Let $\mathbb{B}_Q$ be the matrix before $\text{vec}(\mathbb{V})$ and $\mathbb{P}_1=[I_p,0_{p\times q}]'$, then the above result is equivalent to

$$\text{vec}(\mathbb{V})=\mathbb{B}_Q^{-1}\mathbb{P}_1D_{r_1}^{+}\text{vec}\left[\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\varepsilon_t\varepsilon_t'-I_{r_1})\right]+O_p(N^{-1})+O_p(T^{-1}).$$

Define V by

$$\text{vec}(V)=\mathbb{B}_Q^{-1}\mathbb{P}_1D_{r_1}^{+}\text{vec}\left[\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}(\varepsilon_t\varepsilon_t'-I_{r_1})\right].$$

Then by the definition of $\mathbb{V}$,

$$\widehat{\Delta R}_{11}R_{11}^{-1}+R_{11}V^{\star}R_{11}^{-1}=V+O_p(N^{-1})+O_p(T^{-1}). \quad (B.27)$$

Post-multiplying R_{11} on both sides of (B.27), we have

$$\widehat{\Delta R}_{11}=-R_{11}V^{\star}+VR_{11}+O_p(N^{-1})+O_p(T^{-1})=O_p(T^{-1/2})+O_p(N^{-1}) \quad (B.28)$$

since $V^{\star}=O_p(T^{-1/2})$ and $V=O_p(T^{-1/2})$.

Now consider $\hat{\lambda}_i-\lambda_i$. By $\hat{\lambda}_i=\hat{R}_{11}\tilde{\lambda}_i$ and $\lambda_i=R_{11}\lambda_i^{\star}$, we have

$$\hat{\lambda}_i-\lambda_i=\hat{R}_{11}\tilde{\lambda}_i-R_{11}\lambda_i^{\star}=\widehat{\Delta R}_{11}\lambda_i^{\star}+R_{11}(\tilde{\lambda}_i-\lambda_i^{\star})+\widehat{\Delta R}_{11}(\tilde{\lambda}_i-\lambda_i^{\star}).$$

The last term of right hand side is $O_p(T^{-1})+O_p(N^{-2})$ by $\tilde{\lambda}_i-\lambda_i^{\star}=O_p(T^{-1/2})+O_p(N^{-1})$ and $\widehat{\Delta R}_{11}=O_p(T^{-1/2})+O_p(N^{-1})$. By (B.28) and (A.24), together with $\lambda_i=R_{11}\lambda_i^{\star}$, we have

$$\hat{\lambda}_i-\lambda_i=V\lambda_i+R_{11}\left(\frac{1}{T}\sum_{t=1}^{T}f_t^{\star}f_t^{\star\prime}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}f_t^{\star}e_{it}\right)+O_p(N^{-1})+O_p(T^{-1}).$$

Using (B.4), the above expression can be rewritten as

$$\hat{\lambda}_i-\lambda_i=V\lambda_i+\left(\frac{1}{T}\sum_{t=1}^{T}\phi_t\phi_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\phi_te_{it}\right)+O_p(N^{-1})+O_p(T^{-1})=(\lambda_i'\otimes I_{r_1})\text{vec}(V)+\mathbf{\Delta}_{\phi\phi}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\phi_te_{it}\right)+O_p(N^{-1})+O_p(T^{-1}).$$

To derive the remaining asymptotic results, we first consider $\widehat{\Delta R}_{21}=\hat{R}_{21}-R_{21}$. Notice

$$\hat{R}_{21}-R_{21}=\tilde{\Omega}_{\boldsymbol{vv}}^{-1}\tilde{\Omega}_{\boldsymbol{v}\varepsilon}-\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^{\star}=-\Omega_{\boldsymbol{vv}}^{\star-1}(\tilde{\Omega}_{\boldsymbol{vv}}-\Omega_{\boldsymbol{vv}}^{\star})\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^{\star}+\Omega_{\boldsymbol{vv}}^{\star-1}(\tilde{\Omega}_{\boldsymbol{v}\varepsilon}-\Omega_{\boldsymbol{v}\varepsilon}^{\star})-(\tilde{\Omega}_{\boldsymbol{vv}}^{-1}-\Omega_{\boldsymbol{vv}}^{\star-1})(\tilde{\Omega}_{\boldsymbol{vv}}-\Omega_{\boldsymbol{vv}}^{\star})\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^{\star}+(\tilde{\Omega}_{\boldsymbol{vv}}^{-1}-\Omega_{\boldsymbol{vv}}^{\star-1})(\tilde{\Omega}_{\boldsymbol{v}\varepsilon}-\Omega_{\boldsymbol{v}\varepsilon}^{\star}).$$

The last two terms of the right hand side are $O_p(N^{-2})+O_p(T^{-1})$. Substituting (B.17) and (B.18) into the above result, we have

$$\widehat{\Delta R}_{21}=\Omega_{\boldsymbol{vv}}^{\star-1}\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\boldsymbol{v}_t^{\star}(\varepsilon_t^{\star}-\Omega_{\varepsilon\boldsymbol{v}}^{\star}\Omega_{\boldsymbol{vv}}^{\star-1}\boldsymbol{v}_t^{\star})'-W^{\star}-\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^{\star}V^{\star}+O_p(N^{-1})+O_p(T^{-1}).$$

However, by (B.7) and (B.8), we have $R_{11}'\varepsilon_t=\varepsilon_t^{\star}-\Omega_{\varepsilon\boldsymbol{v}}^{\star}\Omega_{\boldsymbol{vv}}^{\star-1}\boldsymbol{v}_t^{\star}$ and $\boldsymbol{v}_t=\boldsymbol{v}_t^{\star}$. Given these results, we have

$$\widehat{\Delta R}_{21}=\Omega_{\boldsymbol{vv}}^{\star-1}\left[\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\boldsymbol{v}_t\varepsilon_t'\right]R_{11}-W^{\star}-\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^{\star}V^{\star}+O_p(N^{-1})+O_p(T^{-1}).$$

Notice that $R_{21}=\Omega_{\boldsymbol{vv}}^{\star-1}\Omega_{\boldsymbol{v}\varepsilon}^{\star}$ by (B.8) and $\frac{1}{\bar{T}}\sum_{t=\bar{K}}^{T}\boldsymbol{v}_t\boldsymbol{v}_t'=E(\boldsymbol{v}_t\boldsymbol{v}_t')+O_p(T^{-1/2})=\Omega_{\boldsymbol{vv}}+O_p(T^{-1/2})=\Omega_{\boldsymbol{vv}}^{\star}+O_p(T^{-1/2})$, where the last equality is due to $\boldsymbol{v}_t=\boldsymbol{v}_t^{\star}$ by (B.7). Thus

$$\widehat{\Delta R}_{21}=WR_{11}-W^{\star}-R_{21}V^{\star}+O_p(N^{-1})+O_p(T^{-1}), \quad (B.29)$$

where $W = \Omega_{\upsilon\upsilon}^{-1}(\frac{1}{T}\sum_{t=\bar{K}}^{T} \upsilon_t \varepsilon_t')$. Notice $\hat{\gamma}_i = \tilde{\gamma}_i + \hat{R}_{21}\tilde{\lambda}_i$ and $\gamma_i = \gamma_i^\star + R_{21}\lambda_i^\star$. Then

$$\hat{\gamma}_i - \gamma_i = (\tilde{\gamma}_i - \gamma_i^\star) + (\hat{R}_{21}\tilde{\lambda}_i - R_{21}\lambda_i^\star)$$
$$= (\tilde{\gamma}_i - \gamma_i^\star) + \widehat{\Delta R}_{21}\lambda_i^\star + R_{21}(\tilde{\lambda}_i - \lambda_i^\star) + \widehat{\Delta R}_{21}(\tilde{\lambda}_i - \lambda_i^\star).$$

The last term of the right hand side of the above equation is $O_p(T^{-1}) + O_p(N^{-2})$. Substituting (A.25), (A.24) and (B.29) into the above result, we have

$$\hat{\gamma}_i - \gamma_i = \left[\frac{1}{T}\sum_{t=1}^{T} g_t g_t'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} g_t e_{it}\right] + W\lambda_i + R_{21}\left[\frac{1}{T}\sum_{t=1}^{T} f_t^\star f_t^{\star\prime}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} f_t^\star e_{it}\right] + O_p(N^{-1}) + O_p(T^{-1}), \tag{B.30}$$

by $\lambda_i = R_{11}\lambda_i^\star$. Consider the third expression, which, by (B.4), is equal to

$$R_{21}R_{11}^{-1}R_{11}\left(\frac{1}{T}\sum_{t=1}^{T} f_t^\star f_t^{\star\prime}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} f_t^\star e_{it}\right) = R_{21}R_{11}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\phi_t\phi_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\phi_t e_{it}\right).$$

Consider the last equation of (B.2). Post-multiplying g_t' on both sides and taking expectation, by $E(f_t^\star g_t') = 0$, we have

$$R_{21}R_{11}^{-1} = -\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf}.$$

The preceding two results imply that the third expression of (B.30) is equal to

$$-\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf}\mathbf{\Delta}_{\phi\phi}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\phi_t e_{it}\right] + O_p(T^{-1}).$$

Let $\Xi_t = \mathbf{\Delta}_{gf}\mathbf{\Delta}_{\phi\phi}^{-1}\phi_t$. Given the above result, the asymptotic representation of $\hat{\gamma}_i - \gamma_i$ can be rewritten as

$$\hat{\gamma}_i - \gamma_i = \mathbf{\Delta}_{gg}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}(g_t - \Xi_t)e_{it}\right] + W\lambda_i + O_p(N^{-1}) + O_p(T^{-1}). \tag{B.31}$$

The above asymptotic representation has an alternative expression. First, we define

$$\eta_t = g_t - E(g_t f_t')[E(f_t f_t')]^{-1} f_t = g_t - \mathbf{\Delta}_{gf}\mathbf{\Delta}_{ff}^{-1} f_t. \tag{B.32}$$

which implies that

$$\mathbf{\Delta}_{\eta\eta} = \mathbf{\Delta}_{gg} - \mathbf{\Delta}_{gf}\mathbf{\Delta}_{ff}^{-1}\mathbf{\Delta}_{fg}$$

By the Woodbury formula, we have

$$\mathbf{\Delta}_{gg}^{-1} = \mathbf{\Delta}_{\eta\eta}^{-1} - \mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}(\mathbf{\Delta}_{ff} + \mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf})^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1} \tag{B.33}$$

With (B.33) and the relation that $g_t - \Xi_t = \eta_t + \mathbf{\Delta}_{gf}\mathbf{\Delta}_{ff}^{-1} f_t - \mathbf{\Delta}_{gf}\mathbf{\Delta}_{\phi\phi}^{-1}\phi_t$, we can rewrite the first term of the right hand side of (B.31) as

$$\mathbf{\Delta}_{gg}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}(g_t - \Xi_t)e_{it}\right] =$$
$$\mathbf{\Delta}_{\eta\eta}^{-1}\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it} + \mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{\Delta}_{ff}^{-1} f_t - \mathbf{\Delta}_{\phi\phi}^{-1}\phi_t)e_{it}$$
$$-\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}(\mathbf{\Delta}_{ff} + \mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf})^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it}$$
$$-\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}(\mathbf{\Delta}_{ff} + \mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf})^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}\frac{1}{T}\sum_{t=1}^{T}(\mathbf{\Delta}_{ff}^{-1} f_t - \mathbf{\Delta}_{\phi\phi}^{-1}\phi_t)e_{it}$$

Consider the term $(\mathbf{\Delta}_{ff}^{-1} f_t - \mathbf{\Delta}_{\phi\phi}^{-1}\phi_t)$. From the definition of $\phi_t = f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1} g_t$, we have

$$\mathbf{\Delta}_{\phi\phi} = \mathbf{\Delta}_{ff} - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf} \tag{B.34}$$

which can be used to derive

$$\phi_t = f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}(\eta_t + \mathbf{\Delta}_{gf}\mathbf{\Delta}_{ff}^{-1} f_t) = \mathbf{\Delta}_{\phi\phi}\mathbf{\Delta}_{ff}^{-1} f_t - \mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}\eta_t$$

Then

$$(\mathbf{\Delta}_{ff}^{-1} f_t - \mathbf{\Delta}_{\phi\phi}^{-1}\phi_t) = \mathbf{\Delta}_{\phi\phi}^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}\eta_t$$

With the above equation, the first term of the right hand side of (B.31) can be further rewritten as

$$\mathbf{\Delta}_{gg}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}(g_t - \Xi_t)e_{it}\right] = \mathbf{\Delta}_{\eta\eta}^{-1}\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it} + \mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}\mathbf{\Delta}_{\phi\phi}^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it} \tag{B.35}$$
$$-\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}(\mathbf{\Delta}_{ff} + \mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf})^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it}$$
$$-\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}(\mathbf{\Delta}_{ff} + \mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf})^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}\mathbf{\Delta}_{\phi\phi}^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it}$$

From the two basic facts that

$$\Delta_{\phi\phi}^{-1} = \mathbf{\Delta}_{ff}^{-1} + \mathbf{\Delta}_{ff}^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}\mathbf{\Delta}_{ff}^{-1},$$

and

$$\mathbf{\Delta}_{ff}^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{\eta\eta}^{-1} = \mathbf{\Delta}_{\phi\phi}^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}.$$

we can rewrite the 2nd, 3rd and 4th terms on the right hand side of (B.35) as

$$\mathbf{\Delta}_{\eta\eta}^{-1}\mathbf{\Delta}_{gf}\left(\mathbf{\Delta}_{\phi\phi}^{-1} - \mathbf{\Delta}_{ff}^{-1} - \mathbf{\Delta}_{ff}^{-1}\mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}\mathbf{\Delta}_{gf}^{-1}\mathbf{\Delta}_{\phi\phi}^{-1}\right)\mathbf{\Delta}_{fg}\mathbf{\Delta}_{gg}^{-1}\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it}$$

which equals zero by (B.34). So we can alternatively write the asymptotic representation of $\hat{\gamma}_i - \gamma_i$ as

$$\hat{\gamma}_i - \gamma_i = \mathbf{\Delta}_{\eta\eta}^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\eta_t e_{it}\right] + W\lambda_i + O_p(N^{-1}) + O_p(T^{-1}).$$

We proceed to consider $\hat{f}_t - f_t$. Notice $\hat{f}_t = \hat{R}_{11}'^{-1}\tilde{f}_t - \hat{R}_{11}'^{-1}\hat{R}_{21}'g_t$ and $f_t = R_{11}'^{-1}f_t^\star - R_{11}'^{-1}R_{21}'g_t$. Then

$$\begin{aligned}
\hat{f}_t - f_t &= \hat{R}_{11}'^{-1}\tilde{f}_t - \hat{R}_{11}'^{-1}\hat{R}_{21}'g_t - R_{11}'^{-1}f_t^\star - R_{11}'^{-1}R_{21}'g_t\\
&= -R_{11}'^{-1}(\hat{R}_{11}' - R_{11}')R_{11}'^{-1}f_t^\star + R_{11}'^{-1}(\tilde{f}_t - f_t^\star) - R_{11}'^{-1}(\hat{R}_{21}' - R_{21}')g_t\\
&\quad + R_{11}'^{-1}(\hat{R}_{11}' - R_{11}')R_{11}'^{-1}R_{21}'g_t\\
&\quad - (\hat{R}_{11}'^{-1} - R_{11}'^{-1})(\hat{R}_{11}' - R_{11}')R_{11}'^{-1}f_t^\star + (\hat{R}_{11}'^{-1} - R_{11}'^{-1})(\tilde{f}_t - f_t^\star)\\
&\quad - (\hat{R}_{11}'^{-1} - R_{11}'^{-1})(\hat{R}_{21} - R_{21})g_t\\
&\quad + (\hat{R}_{11}'^{-1} - R_{11}'^{-1})(\hat{R}_{11}' - R_{11}')R_{11}'^{-1}R_{21}'g_t
\end{aligned}$$

The last four terms of the above expression are $O_p(N^{-2}) + O_p(T^{-1})$. Given this result, by (B.28), (B.29) and Proposition A.3, we have

$$\hat{f}_t - f_t = -V'(R_{11}'^{-1}f_t^\star - R_{11}'^{-1}R_{21}'g_t) - W'g_t + \left[\frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i\lambda_i'\right]^{-1}\left(\frac{1}{N}\sum_{i=1}^T \frac{1}{\sigma_i^2}\lambda_i e_{it}\right) + O_p(N^{-1}) + O_p(T^{-1})$$

By $f_t = R_{11}'^{-1}f_t^\star - R_{11}'^{-1}R_{21}'g_t$, we have

$$\hat{f}_t - f_t = \left(\frac{1}{N}\sum_{i=1}^N \frac{1}{\sigma_i^2}\lambda_i\lambda_i'\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^T \frac{1}{\sigma_i^2}\lambda_i e_{it}\right) - V'f_t - W'g_t + O_p(N^{-1}) + O_p(T^{-1}).$$

This completes the proof of Proposition B.1. □

Proof of Theorems 1–3 under IRa. Theorem 1 follows from Proposition B.1(a), Theorem 2 follows from Proposition B.1(b), and Theorem 3 follows from Proposition B.1(c). Theorem 4 follows from result (A.4) in Proposition A.1, since σ_i^2 does not have the identification problem and the intermediate estimator and the final estimator are the same. □

Proof of Theorem 5 under IRa. This theorem is implied by the following:

$$\hat{\Phi}_k - \Phi_k = \left(\sum_{t=\bar{K}}^T u_t\psi_t'\right)\left(\sum_{t=\bar{K}}^T \psi_t\psi_t'\right)^{-1}(i_k \otimes I_r) - B'\Phi_k + \Phi_k B' + O_p(N^{-1}) + O_p(T^{-1})$$

We shall prove the above equation. Notice $\hat{\Phi}_k = R'^{-1}\hat{\Phi}_k\hat{R}'$ and $\Phi_k = R'^{-1}\Phi_k^\star R'$. Thus

$$\begin{aligned}
\hat{\Phi}_k - \Phi_k &= R'^{-1}\hat{\Phi}_k\hat{R}' - R'^{-1}\Phi_k^\star R'\\
&= R'^{-1}\Phi_k^\star\widehat{\Delta R}' - R'^{-1}\widehat{\Delta R}'R'^{-1}\Phi_k^\star R'\\
&\quad + R'^{-1}(\hat{\Phi}_k - \Phi_k^\star)R' + V
\end{aligned}$$

where

$$\begin{aligned}
V &= (R'^{-1} - R'^{-1})\hat{\Phi}_k\widehat{\Delta R}' + (\hat{R}'^{-1} - R'^{-1})(\hat{\Phi}_k - \Phi_k^\star)R'\\
&\quad - (\hat{R}'^{-1} - R'^{-1})\widehat{\Delta R}'R'^{-1} + R'^{-1}(\hat{\Phi}_k - \Phi_k^\star)\widehat{\Delta R}'
\end{aligned}$$

However, notice

$$\begin{aligned}
\widehat{\Delta R} = \hat{R} - R &= \begin{bmatrix} \widehat{\Delta R}_{11} & 0 \\ \widehat{\Delta R}_{21} & 0 \end{bmatrix} = \begin{bmatrix} -R_{11}V^\star + VR_{11} & 0 \\ WR_{11} - W^\star - R_{21}V^\star & 0 \end{bmatrix}\\
&\quad + O_p(N^{-1}) + O_p(T^{-1})\\
&= BR - RB^\star + O_p(N^{-1}) + O_p(T^{-1}) \qquad \text{(B.36)}
\end{aligned}$$

where

$$B = \begin{bmatrix} V & 0 \\ W & 0 \end{bmatrix}, \quad B^\star = \begin{bmatrix} V^\star & 0 \\ W^\star & 0 \end{bmatrix}$$

and $W = (\sum_{t=1}^T \upsilon_t\upsilon_t')^{-1}(\sum_{t=1}^T \upsilon_t\varepsilon_t')$. Then $\widehat{\Delta R}$ is $O_p(T^{-1/2})$ since both B and $B^\star$ are $O_p(T^{-1/2})$. This result together with $\hat{\Phi}_k - \Phi_k^\star = O_p(T^{-1/2}) + O_p(N^{-1})$ implies $V = O_p(N^{-2}) + O_p(T^{-1})$. Given this result, together with $\Phi_k = R'^{-1}\Phi_k^\star R'$, we have

$$\begin{aligned}
\hat{\Phi}_k - \Phi_k &= \Phi_k R'^{-1}\widehat{\Delta R}' - R'^{-1}\widehat{\Delta R}'\Phi_k + R'^{-1}(\hat{\Phi}_k - \Phi_k^\star)R'\\
&\quad + O_p(N^{-2}) + O_p(T^{-1}) \qquad \text{(B.37)}
\end{aligned}$$

Substituting (B.36) into the above equation, together with $u_t = R'^{-1}u_t^\star$, $h_t = R'^{-1}h_t^\star$ and Proposition A.4, we have

$$\hat{\Phi}_k - \Phi_k = \left(\sum_{t=\bar{K}}^T u_t\psi_t'\right)\left(\sum_{t=\bar{K}}^T \psi_t\psi_t'\right)^{-1}(i_k \otimes I_r) - B'\Phi_k + \Phi_k B' + O_p(N^{-1}) + O_p(T^{-1})$$

This completes the proof of Theorem 5. □

SUPPLEMENTARY MATERIALS

The supplementary materials include Appendices C–E and detailed derivations for the asymptotic results under IRb in Appendix C, and results under IRc in Appendix D. We also provide the derivations for the asymptotic results of the impulse response function in Appendix E.

ACKNOWLEDGMENTS

The authors thank an associate editor and two anonymous referees for their constructive comments. Jushan Bai's research is supported by the NSF (SES1357598). Kunpeng Li's research is supported by the NSFC No. 71571122, No. 71201031 and the fund from Research and Innovation Centre of Metropolis Economic and Social Development, CUEB.

[Received December 2014. Revised May 2015.]

REFERENCES

Anderson, T. W. (2003), *An Introduction to Multivariate Statistical Analysis*, Hoboken, NJ: Wiley. [620,622]

Anderson, T. W., and Rubin, H. (1956), "Statistical Inference in Factor Analysis," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability: Contributions to the Theory of Statistics* (Vol. 5), ed. J. Neyman, Berkeley, CA: University of California Press, pp. 111–150. [623]

Bai, J. (2003), "Inferential Theory for Factor Models of Large Dimensions," *Econometrica*, 71, 135–171. [625]

Bai, J., and Li, K. (2012), "Statistical Analysis of Factor Models of High Dimension," *The Annals of Statistics*, 40, 436–465. [620,623,625,629,630]

Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191–221. [630]

——— (2013), "Principal Components Estimation and Identification of Static Factors," *Journal of Econometrics*, 176, 18–29. [623]

Bai, J., and Li, K. (2016), "Maximum Likelihood Estimation and Inference for Approximate Factor Models of High Dimension," *The Review of Economics and Statistics*, 98, 298–309. [621,624,627,629]

Berk, K. N. (1974), "Consistent Autoregressive Spectral Estimates," *Annals of Statistics*, 2, 489–502. [627]

Bernanke, B. S., and Boivin, J. (2003), "Monetary Policy in a Data-Rich Environment," *Journal of Monetary Economics*, 50, 525–546. [620]

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005), "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120, 387–422. [620,621,624,628]

Bianchi, F., Mumtaz, H., and Surico, P. (2009), "The Great Moderation of the Term Structure of U.K. Interest Rates," *Journal of Monetary Economics*, 56, 856–871. [620]

Boivin, J., Giannoni, M. P., and Mihov, I. (2009), "Sticky Prices and Monetary Policy: Evidence From Disaggregated US Data," *American Economic Review*, 99, 350–384. [620,621,624]

Chamberlain, G., and Rothschild, M. (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica*, 51, 1281–1304. [623]

Chen, L., Dolado, J. J., and Gonzalo, J. (2014), "Detecting Big Structural Breaks in Large Factor Models," *Journal of Econometrics*, 180, 30–48. [631]

Cheng, X., Liao, Z., and Schorfheide, F. (2013), "Shrinkage Estimation of High-Dimensional Factor Models With Structural Instabilities," NBER Working Paper No. 19792, Cambridge, MA. [631]

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999), "Monetary Policy Shocks: What Have We Learned and to What End?," in *Handbook of Macroeconomics* (Vol. 1), eds. J. B. Taylor and M. Woodford, Amsterdam: North Holland, pp. 65–148. [620]

Doan, T., Litterman, R. B., and Sims, C. A. (1984), "Forecasting and Policy Analysis Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100. [620]

Doz, C., Giannone, D., and Reichlin, L. (2011), "A Two-Step Estimator for Large Approximate Dynamic Factor Models Based on Kalman Filtering," *Journal of Econometrics*, 164, 188–205. [624]

——— (2012), "A Quasi-Maximum Likelihood Approach for Large Approximate Dynamic Factor Models," *Review of Economics and Statistics*, 94, 1014–1024. [624,629]

Fan, J., Liao, Y., and Mincheva, M. (2011), "High Dimensional Covariance Matrix Estimation in Approximate Factor Models," *The Annals of Statistics*, 39, 3320–3356. [620]

——— (2013), "Large Covariance Estimation by Thresholding Principal Orthogonal Complements," *Journal of the Royal Statistical Society*, Series B, 75, 603–680. [620]

Forni, M., and Gambetti, L. (2010), "The Dynamic Effects of Monetary Policy: A Structural Factor Model Approach," *Journal of Monetary Economics*, 57, 203–216. [620]

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000), "The Generalized Dynamic-Factor Model: Identification and Estimation," *Review of Economics and Statistics*, 82, 540–554. [621]

Geweke, J., and Zhou, G. (1996), "Measuring the Pricing Error of the Arbitrage Pricing Theory," *The Review of Financial Studies*, 9, 557–587. [623]

Goncalves, S., and Perron, B. (2014), "Bootstrapping Factor-Augmented Regression Models," *Journal of Econometrics*, 182, 156–173. [621]

Hamilton, J. (1994), *Time Series Analysis*, Princeton: Princeton University Press. [628]

Han, X. (2015), "Tests for Overidentifying Restrictions in Factor-Augmented VAR Models," *Journal of Econometrics*, 184, 394–419. [621]

Han, X., and Inoue, A. (2015), "Tests for Parameter Instability in Dynamic Factor Models," *Econometric Theory*, 31, 1117–1152. [631]

Lawley, D. N., and Maxwell, A. E. (1971), *Factor Analysis as a Statistical Method*, New York: American Elsevier Publishing Company. [623]

Leeper, E. M., Sims, C. A., and Zha, T. (1996), "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity*, 2, 1–63. [620]

Litterman, R. B. (1986), "Forecasting With Bayesian Vector Autoregressions: Five Years of Experience," *Journal of Business and Economic Statistics*, 4, 25–38. [620]

Ludvigson, S. C., and Ng, S. (2009), "Macro Factors in Bond Risk Premia," *Review of Financial Studies*, 22, 5027–5067. [620]

Moench, E. (2008), "Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach," *Journal of Econometrics*, 146, 26–43. [620]

Quah, D., and Sargent, T. (1992), "A Dynamic Index Model for Large Cross-Section," Discussion Paper 77, Federal Reserve Bank of Minneapolis. [624]

Shintani, M., and Guo, Y. (2011), "Finite Sample Performance of Principal Components Estimators for Dynamic Factor Models: Asymptotic vs. Bootstrap Approximations," unpublished manuscript, Vanderbilt University. [621]

Sims, C. A. (1980), "Macroeconomics and Reality," *Econometrica*, 48, 1–48. [620]

——— (1992), "Interpreting the Macroeconomic Time Series Facts: The Effects of Monetary Policy," *European Economic Review*, 36, 975–1000. [620]

——— (1993), "A Nine-Variable Probabilistic Macroeconomic Forecasting Model," in *Business Cycles, Indicators, and Forecasting*, eds. J. H. Stock and M. W. Watson, Chicago: University of Chicago Press for the NBER, 179–204. [620]

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. [620]

——— (2005), "Implications of Dynamic Factor Models for VAR Analysis," NBER Working Paper No. 11467, Cambridge, MA. [620,621,631]

Tsai, H., and Tsay, R. S. (2010), "Constrained Factor Models," *Journal of the American Statistical Association*, 105, 1593–1605. [620]

Watson, M. W., and Engle, R. F. (1983), "Alternative Algorithms for the Estimation of Dynamic Factor, Mimic and Varying Coefficient Regression Models," *Journal of Econometrics*, 23, 385–400. [624]

Yamamoto, Y. (2011), "Bootstrap Inference of Impulse Response Functions in Factor-Augmented Vector Regression," unpublished manuscript, Hitotsubashi University. [621]