Estimation and Inference in Over-identified Structural Factor-augmented VAR Models*

Xu Han†

This Version: November 2015

Abstract

During the past decade, factor-augmented VAR (FAVAR) models have been widely used for structural analysis in the literature, where the effects of structural shocks are often estimated under just-identifying restrictions. However, as the number of restrictions in the FAVAR setup can be large due to its high dimensionality, the structural shocks are over-identified. This paper develops a new estimator for the impulse response functions with a fixed number of over-identifying restrictions. The proposed identification scheme nests the conventional just-identified recursive scheme as a special case. We establish the asymptotic distributions of the new estimator and develop test statistics for the over-identifying restrictions. Simulation results show that adding a few more over-identifying restrictions can lead to a substantial improvement in estimation accuracy. We estimate the effects of the monetary policy shock based on a U.S. macroeconomic data set. The result shows that our over-identified scheme can help to improve statistical significance and eliminate incorrect restrictions that lead to spurious impulse responses.

Key words: High-dimensional Factor Models, Identification and Estimation, Structural Impulse Responses

JEL Classification: C33, C13, E32

*I would like to thank Xu Cheng, Atsushi Inoue, Arthur Lewbel, Bo Li, and Joris Pinkse for their insightful suggestions, and participants at 2015 Tsinghua International Conference on Econometrics for helpful comments.

†Department of Economics and Finance, City University of Hong Kong, Hong Kong SAR. E-mail: xuhan25@cityu.edu.hk

1 Introduction

In their seminal work, Bernanke et al. (2005) develop the factor-augmented VAR (FAVAR) model based on conventional VAR analysis and factor models (Sargent and Sims, 1977; Stock and Watson, 2002). They introduce a few common factors extracted from high-dimensional data into a VAR process. In contrast to a low-dimensional VAR model, FAVAR utilizes more information in high-dimensional data to identify the space spanned by the underlying structural shocks without the loss of parsimony. Due to such merits, the empirical literature related to FAVAR has rapidly grown in recent years (e.g., Stock and Watson, 2005; Boivin et al., 2009; Gilchrist et al., 2009; Bianchi et al., 2009; Forni and Gambetti, 2010; Koop and Korobilis, 2014; Ritschl and Sarferaz, 2014, among others).

As the structural shocks in FAVAR are not observed, additional conditions must be imposed for identification purposes. Most identification methods in conventional structural VAR models can also be applied to FAVAR. However, the numbers of restrictions used in FAVARs and structural VARs can be quite different. In practice, just-identified schemes are commonly used in structural VAR due to its limited number of variables and computational simplicity. In contrast, the number of restrictions in FAVAR can be quite large due to the high-dimensional nature of factor models. For example, Stock and Watson (2005) impose the assumption that the monetary policy shock does not contemporaneously affect 64 slow-moving variables, such as IP indexes, employment, sales, wages, and CPIs. Gilchrist et al. (2009) similarly impose dozens of short-run zero restrictions to identify the credit market shocks. As the number of factors is often set equal to a small number compared with the cross section dimension in a factor model, imposing all such restrictions leads to an over-identified system.

Using over-identifying restrictions in FAVAR is not only a natural result of the high dimensionality of the model. It also has at least two benefits from a theoretical perspective. First, it is well known that adding valid and informative restrictions leads to more efficient estimators. Second, over-identifying restrictions can be tested to let practitioners know whether the imposed conditions are satisfied for the given data. In the low-dimensional structural VAR literature, over-identified models are often estimated by maximum likelihood or Bayesian methods (e.g., Leeper et al., 1996; Sims and Zha, 1998; Canova and Pérez Forero, 2012; Kociecki, et al., 2012, among others). In general, such estimation procedures do not have closed-form solutions, and the computational cost can be very high, especially in a high-dimensional setup such as FAVAR. While Bai and Wang's (2015) recent study develops a Bayesian method under the dynamic factor model setup that allows for the possibility of over-identification, the literature related to frequentist estimation of over-identified structural

FAVAR models remains quite scarce.

This paper contributes to the literature by developing a new frequentist method for estimation and inference in over-identified structural FAVAR models. We focus on the identification method using contemporaneous zero restrictions, which have broad applications in the literature.[1] The zero restrictions, which are usually guided by economic theory, are imposed on variables that are not affected by the target structural shock within the same period. The just-identified recursive scheme studied by Forni and Gambetti (2010) and Bai et al. (2015) can be viewed as a special case of the identification scheme proposed in this paper. We prove that the new estimator for the impulse response function is consistent and establish its asymptotic distribution for the purpose of statistical inference.

The method developed in this paper has several advantages. First, the validity of the over-identifying restrictions can be tested. Based on Kleibergen and Paap's (2006) rank test, we propose a statistic that can jointly test two rank conditions used to identify the target structural shock. The new test can detect wrong identification conditions and thus avoid inconsistent estimators for the impulse response functions (IRFs). Second, since the conventional just-identified recursive scheme is nested as a special case under our framework, our estimator tends to be more efficient. Simulation results show that adding a few over-identifying restrictions can decrease the mean squared error substantially. Third, the new estimator under over-identifying restrictions has a closed-form solution and is easy to compute.

Some recent work has studied identification theory in factor models or FAVAR. Under the fundamentalness assumption, Forni et al. (2009) show that the identification of structural shocks in FAVAR reduces to the choice of the rotation matrix that links the reduced-form errors and structural shocks. Bai and Ng (2013) and Bai and Li (2012) develop several sets of conditions under which a factor model can be just-identified. Under similar just-identified schemes, Bai et al. (2015) develop estimators and excellent inferential theory for estimated IRFs in a FAVAR setup. However, it seems quite nontrivial to extend their estimators to allow for over-identifying restrictions. We also differentiate from Bai et al. (2015) in that our model distinguishes static and dynamic factors. Bai and Wang (2014) propose three sets of equivalent rank conditions under which the factors can be identified in static and dynamic factor models. While their theory allows for over-identification, this paper substantially

[1]As one of the most widely used methods, contemporaneous zero restrictions have been applied in FAVAR to identify various shocks including (but not limited to): the monetary policy shock (Bernanke et al., 2005; Stock and Watson, 2005; Boivin et al., 2009; Forni and Gambetti, 2010; Eickmeier et al., 2015), the credit market shocks (Gilchrist et al., 2009), the uncertainty shocks (Caggiano et al., 2014), the international shocks (Mumtaz and Surico, 2009), the banking shocks (Buch et al., 2014), and the news shocks (Forni et al., 2014).

differs from their work: we focus on the estimation and inference under an over-identified scheme, and they focus mainly on the conditions to ensure identification. Stock and Watson (2005) and Han (2015) consider an over-identified FAVAR where the number of restrictions is proportional to the cross-sectional dimension. Han's (2015) test for over-identifying restrictions is a test for the number of fast-moving shocks in Stock and Watson (2005). In contrast, the estimator in this paper is computed under a fixed number of restrictions, which decrease the possibility of misspecification yet still preserve the advantages of over-identified schemes. Thus, the estimation procedure, inferential theory, and asymptotics of the test statistics are all quite different from those presented by Stock and Watson (2005) and Han (2015). Gafarov's (2014) recent study also considers a FAVAR with many identifying restrictions, but the author mainly focuses on the possibility of point identification under sign restrictions.

In an empirical application, we estimate the effects of the monetary policy shock on the main U.S. macroeconomic variables via our estimators. The results show that the identification conditions rejected by our tests yield spurious IRFs with unexpected signs. In contrast, the identification conditions accepted by our tests lead to IRFs consistent with economic theory. In addition, the over-identified scheme improves the average statistical significance of IRFs compared with the conventional just-identified scheme.

The rest of the paper is organized as follows. In Section 2, we introduce our new estimator for the IRFs and discuss the identification conditions. Section 3 presents our assumptions, provides the main theorems about the asymptotics of our estimators, and illustrates how to implement our inferential theory in practice. We develop test statistics to determine the validity of the over-identifying restrictions. Section 4 investigates the finite-sample properties of the estimators and statistics via Monte Carlo experiments. Section 5 uses an empirical application to show the gains from using the proposed estimators under over-identifying restrictions. Section 6 concludes the paper.

Notations: Let $P_Z = Z(Z'Z)^{-1}Z'$ and $M_Z \equiv I - P_Z$ for any matrix Z. Z^+ denotes the MP inverse of Z. K_{ab} is the commutation matrix such that $K_{ab}\text{vec}(Z) = \text{vec}(Z')$ for any $a \times b$ matrix Z. Following the matrix algebra convention, we use the abbreviation K_a to denote K_{aa}. The Euclidean norm of a matrix Z is denoted as $\|Z\| = [\text{trace}(Z'Z)]^{1/2}$. We use $\to_p$ and $\to_d$ to denote convergences in probability and distribution, respectively.

2 The Identification of Structural Shocks in FAVAR Models

2.1 The Model Setup

We consider the following factor model for $t = 1, ..., T$,

$$X_t = \Lambda F_t + e_t, \tag{2.1}$$

$$F_t = \sum_{j=1}^{p} \Phi_j F_{t-j} + G\eta_t, \tag{2.2}$$

$$\eta_t = A\zeta_t, \tag{2.3}$$

where $X_t = [X_{1t}, ..., X_{Nt}]'$ is an N-dimensional vector, F_t is an $r-$dimensional unobserved static factor, Λ is an $N \times r$ factor loading matrix, $e_t = [e_{1t}, ..., e_{Nt}]'$ is an N-dimensional idiosyncratic error term, G is an $r \times q$ matrix of rank q, η_t is the $q-$dimensional reduced-form shock, ζ_t is the $q-$dimensional structural shock, $E(\zeta_t \zeta_t') = I_q$, and A is a $q \times q$ nonsingular matrix. We set $q \leq r$, so that our setup allows for dynamic factors. By assuming the stationarity of F_t, we have

$$(I_r - \Phi_1 L - ... - \Phi_p L^p)^{-1} = I_r + \sum_{s=1}^{\infty} \Psi_s L^s, \tag{2.4}$$

where Ψ_s is the coefficient matrix in the vector moving average representation of (2.2).

Let $\mathcal{F}_t = [F_{t-1}', ..., F_{t-p}']'$ and $\Phi = [\Phi_1, ..., \Phi_p]$. Plugging (2.2) and (2.3) into (2.1), we obtain

$$X_t = \Pi \mathcal{F}_t + \Theta \eta_t + e_t \tag{2.5}$$

$$= \Pi \mathcal{F}_t + \Gamma \zeta_t + e_t, \tag{2.6}$$

where $\Pi = \Lambda\Phi$, $\Theta = \Lambda G$ and $\Gamma = \Theta A$. The matrix representation of (2.6) is given by

$$\begin{aligned} X &= \mathcal{F}\Pi' + \eta\Theta' + e \\ &= \mathcal{F}\Pi' + \zeta\Gamma' + e, \end{aligned} \tag{2.7}$$

where $X = [X_{p+1}, ..., X_T]'$, $\mathcal{F} = [\mathcal{F}_{p+1}, ..., \mathcal{F}_T]'$, $\eta = [\eta_{p+1}, ..., \eta_T]'$, $\zeta = [\zeta_{p+1}, ..., \zeta_T]'$, and $e = [e_{p+1}, ..., e_T]'$. We use λ_i, θ_i and γ_i to denote the transpose of the i-th row of Λ, Θ and

Γ, respectively. In particular, we have

$$\Lambda = [\lambda_1, ..., \lambda_N]', \quad \Theta = [\theta_1, ..., \theta_N]', \quad \Gamma = [\gamma_1, ..., \gamma_N]'. \tag{2.8}$$

For the i-th cross-section, the factor model (2.1) can be represented as

$$\underline{\mathrm{X}}_i = F\lambda_i + \underline{\mathrm{e}}_i, \tag{2.9}$$

where $\underline{\mathrm{X}}_i = [X_{i1}, ..., X_{iT}]'$, $F = [F_1, ..., F_T]'$, and $\underline{\mathrm{e}}_i = [e_{i1}, ..., e_{iT}]'$.

Let $\hat{F}_t$ denote the standard principal component (PC) estimator for F_t, so that $T^{-1}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' = I_r$. The OLS estimators for λ_i and Λ are given by

$$\hat{\lambda}_i = \frac{1}{T}\sum_{t=1}^{T}\hat{F}_t X_{it},$$
$$\hat{\Lambda} = \frac{1}{T}\sum_{t=1}^{T} X_t \hat{F}_t'. \tag{2.10}$$

Let $\hat{\mathcal{F}}_t = [\hat{F}_{t-1}', ..., \hat{F}_{t-p}']'$, $\hat{\mathcal{F}} = [\hat{\mathcal{F}}_{p+1}, ..., \hat{\mathcal{F}}_T]'$, and

$$\hat{X} = M_{\hat{\mathcal{F}}} X. \tag{2.11}$$

Since $\hat{X}$ is an estimate for $\zeta\Gamma' + e$, the reduced-form shocks can be estimated by PC. Set the estimated reduced-form shocks, denoted as $\hat{\eta}$, equal to $\sqrt{T-p}$ times the eigenvectors corresponding to the q largest eigenvalues of the $(T-p)\times(T-p)$ matrix $\hat{X}\hat{X}'$. Note that $\hat{\mathcal{F}}$ and $\hat{\eta}$ are orthogonal by design. Thus, the OLS estimator for G in (2.2) can be computed as

$$\hat{G} = \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{F}_t\hat{\eta}_t', \tag{2.12}$$

where we use $\hat{\eta}'\hat{\eta}/(T-p) = I_q$. Furthermore, as $\Theta = \Lambda G$, the estimators for θ_i and Θ can be computed as

$$\hat{\theta}_i = \hat{G}'\hat{\lambda}_i, \quad \hat{\Theta} = \hat{\Lambda}\hat{G}. \tag{2.13}$$

The estimator for Φ is given by

$$\hat{\Phi} = \left(\sum_{t=p+1}^{T}\hat{F}_t\hat{\mathcal{F}}_t'\right)(\hat{\mathcal{F}}'\hat{\mathcal{F}})^{-1}. \tag{2.14}$$

Given $\hat{\Phi}$, we can compute $\hat{\Psi}_s$ for $s = 1, 2, ...$ by inverting the lag polynomial defined in (2.4).

2.2 Identification and Estimation

Suppose we are interested in the effects of the m-th structural shock. Given the preceding model setup, the IRF of X_t with respect to the m-th shock $\zeta_{m,t-s}$ is given by

$$\frac{\partial X_t}{\partial \zeta_{m,t-s}} = \Lambda \Psi_s G A_m, \tag{2.15}$$

where A_m denotes the m-th column of A for $m = 1, ..., q$. For $s = 0$, we set $\Psi_s = I_r$, so (2.15) becomes

$$\frac{\partial X_t}{\partial \zeta_{mt}} = \Lambda G A_m = \Theta A_m. \tag{2.16}$$

The estimators for Λ, Ψ_s, and G are defined in Section 2.1. We need only an appropriate estimator for A_m to compute the IRF in (2.15). It is well known that $\hat{\eta}$ can estimate its unobserved counterpart only up to some rotation. To identify A_m, additional restrictions must be imposed.

Assumption ID1 – $(T-p)^{-1}\zeta'\zeta = I_q$.

Assumption ID1 requires that the sample mean of $\zeta_t\zeta_t'$ is equal to its population mean $E\zeta_t\zeta_t' = I_q$. This assumption is commonly used in the structural FAVAR literature, such as in the identifying restrictions IC3 and IC5 of Bai and Li (2012), and PC2 of Bai and Ng (2013). Furthermore, recall that the principal component estimator $\hat{\eta}$ satisfies that $(T-p)^{-1}\hat{\eta}'\hat{\eta} = I_q$. Since Assumption ID1 implies that $(T-p)^{-1}\hat{\zeta}'\hat{\zeta} = I_q$, it follows that the estimator $\hat{A}$ must be orthonormal by Eq. (2.3).

We discuss the identification and estimation of A_m as follows. We begin with the case where $m = q$. We impose the short-run restrictions that ζ_{qt} does not affect $X_{1t}, ...X_{kt}$ contemporaneously for some $k \geq q-1$. This implies that Γ has the following structure:

$$\Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix}, \; \Gamma_1 = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1,q-1} & 0 \\ \vdots & \cdots & \vdots & \vdots \\ \gamma_{k1} & \cdots & \gamma_{k,q-1} & 0 \end{bmatrix}_{k\times q}, \; \Gamma_2 \text{ is } (N-k)\times q. \tag{2.17}$$

The structure of Γ_1 implies that

$$\Theta_1 A_q = 0_{k\times 1}, \tag{2.18}$$

where Θ_1 denotes the first k rows of Θ. Let $\hat{\Theta}_1$ and $\hat{\Lambda}_1$ denote the first k rows of $\hat{\Theta}$ and $\hat{\Lambda}$, respectively. Thus,

$$\hat{\Theta}_1 = \hat{\Lambda}_1 \hat{G}. \tag{2.19}$$

We make the following assumption on the rank of Γ_1 to ensure the identification of A_q.

Assumption ID2 – $\text{rank}(\Gamma_1) = q - 1$ *and the zero restrictions are satisfied for* Γ_1 *defined in Eq.* (2.17).

Assumption ID2 and Eq. (2.18) leads to a natural estimator for A_q, i.e.,

$$\hat{A}_q = \arg\min_{A_q \in \mathbb{R}^q} A_q' \hat{\Theta}_1' \hat{\Theta}_1 A_q \quad s.t.\ A_q' A_q = 1. \tag{2.20}$$

It is straightforward that $\hat{A}_q$ is the eigenvector associated with the smallest eigenvalue of $\hat{\Theta}_1' \hat{\Theta}_1$. The following procedure summarizes how to compute the IRFs with respect to ζ_{qt}.

Algorithm 1:
(1) Compute $\hat{A}_q$ as the eigenvector associated with the smallest eigenvalue of $\hat{\Theta}_1' \hat{\Theta}_1$, where $\hat{\Theta}_1$ denotes the first k rows of $\hat{\Theta}$ defined in (2.13).
(2) The contemporaneous effects of ζ_{qt} on X_t can be estimated as $\hat{\Theta}\hat{A}_q$, and the effects of $\zeta_{q,t-s}$ on X_t can be computed as $\hat{\Lambda}\hat{\Psi}_s \hat{G} \hat{A}_q$.

Next, consider the case where we are interested in the effects of the m-th shock ζ_{jt} for $1 \le m < q$. We assume that Γ_1 has the following structure:

$$\Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix}, \ \Gamma_1 = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1,m-1} & 0 & 0_{1\times(q-m)} \\ \vdots & \cdots & \vdots & \vdots & \vdots \\ \gamma_{\ell 1} & \cdots & \gamma_{\ell,m-1} & 0 & 0_{1\times(q-m)} \\ \gamma_{\ell+1,1} & \cdots & \gamma_{\ell+1,m-1} & \gamma_{\ell+1,m} & 0_{1\times(q-m)} \\ \vdots & \cdots & \vdots & \vdots & \vdots \\ \gamma_{k1} & \cdots & \gamma_{k,m-1} & \gamma_{k,m} & 0_{1\times(q-m)} \end{bmatrix}_{k\times q}, \ \Gamma_2 \text{ is } (N-k)\times q. \tag{2.21}$$

where $k \ge q-1$, $k > \ell \ge m-1$, and the m-th column of Γ_1 should have at least one nonzero element. The short-run restrictions in (2.21) have three implications: (1) the last $q-m$ shocks do not affect $X_{1t}, ..., X_{kt}$ contemporaneously; (2) the m-th shock does not affect $X_{1t}, ..., X_{\ell t}$ contemporaneously; and (3) the m-th shock could affect $X_{\ell+1,t}, ..., X_{kt}$ contemporaneously. We make the following assumption to ensure the identification of A_m.

Assumption ID3 – *Let* Γ_{11} *denote the first* ℓ *rows of* Γ_1 *in Eq. (2.21).* $\text{rank}(\Gamma_1) = m$, $\text{rank}(\Gamma_{11}) = m - 1$, *and the zero restrictions are satisfied for* Γ_1 *defined in Eq. (2.21).*

Furthermore, $\Gamma_1'\Gamma_1$ *has distinct nonzero eigenvalues.*

The assumption that $\Gamma_1'\Gamma_1$ has distinct nonzero eigenvalues ensures the differentiability of the eigenvector functions. This condition is used in Section 3 in the derivation of the asymptotic distribution of our estimators.

Given Eq. (2.21) and Assumption ID3, we propose the following procedure to estimate the IRFs with respect to ζ_{mt}.

Algorithm 2:
(1) Let $\hat{A}_{1:m}$ denote the eigenvectors associated with the first m eigenvalues of $\hat{\Theta}_1'\hat{\Theta}_1$ in descending order, where $\hat{\Theta}_1$ denotes the first k rows of $\hat{\Theta}$ defined in (2.13).
(2) We set $\tilde{A}_m$ equal to the eigenvector associated with the smallest eigenvalue of $\hat{A}_{1:m}'\hat{\Theta}_{11}'\hat{\Theta}_{11}\hat{A}_{1:m}$, where $\hat{\Theta}_{11}$ denotes the first ℓ rows of $\hat{\Theta}_1$. (Note that $\tilde{A}_m$ is $m \times 1$.) Hence, $\hat{A}_{1:m}\tilde{A}_m$ is an estimator for A_m.[2]
(3) The contemporaneous IRFs with respect to ζ_{mt} are computed as $\hat{\Theta}\hat{A}_{1:m}\tilde{A}_m$, and the IRFs with respect to $\zeta_{m,t-s}$ are computed as $\hat{\Lambda}\hat{\Psi}_s\hat{G}\hat{A}_{1:m}\tilde{A}_m$.

Remark 1: The identification of A_q in (2.17) or A_m in (2.21) is up to a sign change. To figure out the sign of A_m, we can use economic theory to impose a sign restriction on a nonzero element of the m-th column of Γ. In the rest of the paper, we assume that the sign of A_m has been pinned down unless otherwise stated.

Remark 2: In the case of $m = q$, we need at least q restrictions to identify A_q. The last column of Γ_1 provides k restrictions, and the constraint $A_q'A_q = 1$ in Eq. (2.20) provides an additional restriction. If $k \geq q$, then A_q is over-identified, and the identifying restrictions $\gamma_{1q} = ... = \gamma_{kq} = 0$ can be tested by examining whether $\text{rank}(\Gamma_1) = q - 1$. If $k = q - 1$, then A_q is just-identified and $\text{rank}(\Gamma_1)$ is at most $q - 1$, so the identifying restrictions are always satisfied.

In the case of $1 \leq m < q$, note that $\hat{A}_{1:m}$ is an estimator for the space spanned by $A_1, ..., A_m$, the identification of which requires the first m columns of Γ_1 to be of full column rank. When $k = m$, the smallest $q - m$ eigenvalues of $\hat{\Theta}_1'\hat{\Theta}_1$ are always zero, and the space spanned by $A_1, ..., A_m$ is just-identified. When $k > m$, $\text{rank}(\Gamma_1)$ may be greater than m if the last $q - m$ columns of Γ_1 are not zero, so the identifying restrictions can be tested by implementing a rank test on $\hat{\Theta}_1$. Once $\hat{\Theta}_{11}\hat{A}_{1:m}$ is obtained, the estimation of $\tilde{A}_m$ is similar to that of $\hat{A}_q$ in the case of $m = q$. If $\ell \geq m$, then $\tilde{A}_m$ is computed under over-identifying

[2] When $m = 1$, note that $\tilde{A}_m = 1$ and $\hat{A}_{1:m} = \hat{A}_1$.

restrictions, which can be tested by checking the rank of $\hat{\Theta}_{11}$. The tests for identifying restrictions are discussed in Section 3.4.

Remark 3 (relation to the just-identified recursive scheme): The conventional identification scheme usually imposes a recursive structure in the first q rows of Γ. To be specific, the upper $q \times q$ block of Γ is a lower triangular matrix of full rank, i.e.,

$$\Gamma = \begin{bmatrix} \gamma_{11} & 0 & \cdots & 0 \\ \gamma_{21} & \gamma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{q1} & \gamma_{q2} & \cdots & \gamma_{qq} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{N1} & \gamma_{N2} & \cdots & \gamma_{Nq} \end{bmatrix}. \tag{2.22}$$

This identification scheme is considered by Bai and Li (2012), Bai et al. (2015), Bai and Ng (2013), and Forni and Gambetti (2010) among others. In addition to the $q(q+1)/2$ restrictions given by ID1, Eq. (2.22) provides $q(q-1)/2$ restrictions, so matrix A is just-identified. Given $\hat{\Theta}$, the rotation matrix A can be computed by QR decomposition of the transpose of the first q rows of $\hat{\Theta}$. See, for example, PC2 in Bai and Ng (2013).

To see the link between (2.22) and our new identification scheme, let us start with the last column of A. Note that $\gamma_{1q} = ... = \gamma_{q-1,q} = 0$ provides $q-1$ restrictions, so the first $q-1$ rows of Γ in (2.22) correspond to Γ_1 in (2.17). Since $k = q-1$ and the first $q-1$ rows of Γ are of rank $q-1$, A_q is just-identified. Next, we consider A_{q-1}, i.e., $m = q-1$ in our notation. Since A_q is just-identified, the space spanned by $A_1, ..., A_{q-1}$ is also just-identified. The first ℓ rows of Γ in (2.22) correspond to Γ_{11} in (2.21). The $(q-1)$-th column of Γ has $q-2$ zero entries. Thus, $\ell = q-2$ and $\text{rank}(\Gamma_{11}) = q-2 = m-1$. Similar to A_q, the estimation of $\tilde{A}_{q-1}$ (defined in step 2 of Algorithm 2) is under just-identified restrictions. Thus, A_{q-1} (estimated by $\hat{A}_{1:q-1}\tilde{A}_{q-1}$) is just-identified. The remaining columns of A are also just-identified and can be estimated by implementing Algorithm 2 repeatedly. Hence, our new identification scheme nests the commonly used just-identified recursive scheme as a special case.

Remark 4: In contrast to the conventional scheme in (2.22), both (2.17) and (2.21) allow over-identification and have the following advantages. First, the number of restrictions in a high-dimensional FAVAR can be large. For example, the monetary policy shock is often assumed to have no contemporaneous effects on slow-moving variables such as industrial production, unemployment, and CPIs, all of which consist of numerous sub-aggregate series in a typical data set for factor models. Using every valid restriction leads to over-identification and is expected to yield more efficient IRFs than a just-identified scheme. Second, just-identifying restrictions can always lead to an estimator with a lower triangular structure in (2.22) even if the restrictions are wrong. In contrast, because the validity of over-identifying restrictions can be tested, incorrect restrictions can be avoided in our estimation of IRFs and the results are more reliable than those obtained under untestable just-identifying restrictions.

Remark 5: The identification schemes in this paper use a fixed number of restrictions. In contrast, the number of over-identifying restrictions in studies by Stock and Watson (2005) and Han (2015) is diverging and equal to N_S times the number of fast-moving shocks, where N_S is proportional to N and denotes the number of slow-moving variables. In practice, imposing a diverging number of restrictions increases the possibility of misspecification, so even the least restrictive setup (with only one fast-moving shock) may still involve incorrect restrictions due to a large N_S, and no candidate setup can pass the specification test. Thus, this paper considers a fixed number of restrictions, which decrease the possibility of misspecification yet still preserve the advantages of over-identified schemes.

3 Assumptions and Asymptotic Theory

3.1 Assumptions

We make the following assumptions.

Assumption 1 – *There exists a positive constant $M < \infty$ such that*:
(a) $E\|F_t\|^4 < M$, $T^{-1}\sum_{t=1}^{T} F_t F_t' \to_p \Sigma_F$, *and* $T^{-1}\sum_{t=p+1}^{T} \mathcal{F}_t \mathcal{F}_t' \to_p \Sigma_{\mathcal{F}}$ *for some positive definite matrices* Σ_F *and* $\Sigma_{\mathcal{F}}$.
(b) $E(\zeta_t \zeta_t') = I_q$, $E\|\zeta_t\|^4 < M$, $E(\zeta_s \zeta_t') = 0$ *for any* $s \neq t$, *and* $(T-p)^{-1}\sum_{p+1}^{T} \zeta_t \zeta_t' \to_p I_q$.
(c) $E\|T^{-1/2}\sum_{t=p+1}^{T} \zeta_t \mathcal{F}_t'\|^2 < M$.

Assumption 2 – *There exists a positive constant $M < \infty$ such that*:
(a) $E\|\lambda_i\|^4 \leq M$, *and* $\|\Lambda'\Lambda/N - \Sigma_\Lambda\| \to_p 0$ *for some $r \times r$ positive definite matrix* Σ_Λ.
(b) $\text{rank}(G) = q$, $\|G\| \leq M$, *and* $\|\Phi\| \leq M$. *A is orthonormal.*
(c) *All of the roots of* $|I_q - \Phi_1 L - ... - \Phi_p L^p| = 0$ *are outside the unit circle.*
(d) Both $\Sigma_F \Sigma_\Lambda$ and $G'\Sigma_\Lambda G$ have distinct eigenvalues.

Assumption 3 – *There exists a positive constant $M < \infty$ such that for all N and T*:
(a) $E(e_{it}) = 0$, $Ee_{it}^2 = \sigma_i^2$, *and* $E\,|e_{it}|^8 \leq M$ *for all i and t.*
(b) $E(e_s' e_t/N) = E(N^{-1}\sum_{i=1}^{N} e_{is} e_{it}) = \gamma_N(s, t)$, $|\gamma_N(s, s)| \leq M$ *for all s, and* $T^{-1}\sum_{s=1}^{T}\sum_{t=1}^{T}$

$|\gamma_N(s,\, t)| \leq M$.

(c) $E(e_{it}e_{h,t-j}) = \tau_{ih,t,j}$ *with* $|\tau_{ih,t,j}| \leq |\tau_{ih}|$ *for some* τ_{ih} *and for all* t *and* $j = 0, ..., p$. *In addition,* $N^{-1}\sum_{i=1}^{N}\sum_{h=1}^{N}|\tau_{ih}| \leq M$.

(d) $E(e_{it}e_{hs}) = \tau_{ih,ts}$, *and* $(NT)^{-1}\sum_{i=1}^{N}\sum_{h=1}^{N}\sum_{s=1}^{T}\sum_{t=1}^{T}|\tau_{ih,ts}| \leq M$.

(e) *For every* (t, s), $E\left|N^{-1/2}\sum_{i=1}^{N}[e_{is}e_{it} - E(e_{is}e_{it})]\right|^4 \leq M$.

Assumption 4 – *The variables*$\{\lambda_i\}$, $\{\zeta_t\}$ *and* $\{e_{it}\}$ *are three mutually independent groups.*

Assumption 5 – *There exists a positive constant* $M < \infty$ *such that for all* N *and* T, *and for every* $t \leq T$ *and* $i \leq N$:

(a) $\sum_{s=1}^{T}|\gamma_N(s,\, t)| \leq M$.

(b)$\sum_{h=1}^{N}|\tau_{ih}| \leq M$.

Assumption 6 – *There exists a positive constant* $M < \infty$ *such that for all* N *and* T:

(a) *For each* t, $E\left\|\frac{1}{\sqrt{NT}}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s[e_{is}e_{it} - E(e_{is}e_{it})]\right\|^2 \leq M$, $E\left\|\frac{1}{\sqrt{NT}}\sum_{s=p+1}^{T}\sum_{i=1}^{N}\zeta_s[e_{is}e_{it}\right.$ $\left.-E(e_{is}e_{it})]\right\|^2 \leq M$.

(b) $E\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}F_t\lambda_i' e_{it}\right\| \leq M$.

(c) $E\left\|\frac{1}{\sqrt{NT}}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\lambda_i e_{i,t-j}\mathcal{F}_t'\right\|^2 \leq M$ *and* $E\left\|\frac{1}{\sqrt{NT}}\sum_{t=p+1}^{T}\sum_{i=1}^{N}\lambda_i e_{i,t-j}\zeta_t'\right\|^2 \leq M$ *for* $j = 0, 1, ..., p$.

(d) *For each* t, $E\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\lambda_i e_{it}\right\|^2 \leq M$.

(e) *For each* i, $E\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}F_t e_{it}\right\|^4 \leq M$ *and* $E\left\|\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}\zeta_t e_{it}\right\|^4 \leq M$. *For each* i *and for* $j = 1, ..., p$, $E\left\|\frac{1}{\sqrt{T}}\sum_{t=p+1}^{T}F_{t-j}' e_{it}\right\|^4 \leq M$.

Assumptions 1–6 are either from or slight modifications of Assumptions A–G of Bai (2003). Assumption 1(a) regulates the moments of static factors. The positive definiteness of $\Sigma_{\mathcal{F}}$ is the same as Assumption A10 of Amengual and Watson (2007). Assumption 1(b) imposes conditions on the structural shocks, which are serially uncorrelated and have an identity covariance matrix. Assumption 1(c) is not restrictive, as it is commonly assumed that ζ_t and lags of F_t are uncorrelated in VAR models. Assumption 2(a) follows from Assumption B of Bai (2003). Furthermore, for any nonsingular matrix A, there exists a nonsingular matrix D_A such that $GA = \breve{G}\breve{A}$ with $\breve{A} = D_A A$ being orthonormal and $\breve{G} = GD_A^{-1}$. Thus, setting A to orthonormal does not result in a loss of generality. Assumption 2(d) is similar to Assumption G of Bai (2003). It ensures the existence of the probability limits of the rotation matrices H_F and H_η defined in (3.1) and (3.2). Assumptions 3 and 5 allow the idiosyncratic errors to be correlated in both time and cross section dimensions. Assumption 4 is similar to Assumption D of Bai and Ng (2004). Assumption 6 is not stringent because all of the sums in this assumption involve zero mean random variables. It is close to Assumption F of Bai (2003) and Assumption 6 of Han (2015).

It is well known that the principal components consistently estimate the original factors up to a rotation. Let $X^0 \equiv [X_1, ..., X_T]'$ be the full data matrix. We define the following rotation matrices for F_t and η_t,

$$H_F = \left(\frac{\Lambda'\Lambda}{N}\right)\left(\frac{F'\hat{F}}{T}\right)V_X^{-1}, \tag{3.1}$$

$$H_\eta = \left(\frac{\Theta'\Theta}{N}\right)\left(\frac{\eta'\hat{\eta}}{T-p}\right)\hat{V}, \tag{3.2}$$

where $F = [F_1, ..., F_T]'$, $\hat{F} = [\hat{F}_1, ..., \hat{F}_T]'$, V_X is the $r \times r$ diagonal matrix consisting of the first r largest eigenvalues of $X^0 X^{0\prime}/(NT)$ in decreasing order, and $\hat{V}$ is the $q \times q$ diagonal matrix consisting of the first q eigenvalues of $\hat{X}\hat{X}'/[N(T-p)]$ in descending order. By Lemma A3 and Proposition 1 of Bai (2003), the rotation matrices have well defined probability limits. Thus, we define

$$\begin{aligned} \bar{H}_F &= \text{plim}_{N,T\to\infty} H_F, \\ \bar{H}_\eta &= \text{plim}_{N,T\to\infty} H_\eta. \end{aligned} \tag{3.3}$$

Lemma A1 of Bai (2003) shows that $T^{-1}\sum_{t=1}^T \|\hat{F}_t - H_F' F_t\|^2 = O_p\left(\delta_{NT}^{-2}\right)$, where

$$\delta_{NT} = \min(\sqrt{N},\ \sqrt{T}). \tag{3.4}$$

We can analogously define $H_{\mathcal{F}}$ such that $\hat{\mathcal{F}}_t$ is a consistent estimate for $H_{\mathcal{F}}'\mathcal{F}_t$. Recall that $\hat{\mathcal{F}}_t = [\hat{F}_{t-1}', ..., \hat{F}_{t-p}']'$ and $\mathcal{F}_t = [F_{t-1}', ..., F_{t-p}']'$. Let

$$H_{\mathcal{F}} \equiv I_p \otimes H_F, \tag{3.5}$$

so that $\hat{\mathcal{F}}_t$ is a consistent estimator for $H_{\mathcal{F}}'\mathcal{F}_t$. The probability limit of $H_{\mathcal{F}}$ is defined as

$$\bar{H}_{\mathcal{F}} = I_p \otimes \bar{H}_F. \tag{3.6}$$

3.2 Asymptotic Distributions of the IRFs

This subsection presents the results of the asymptotic distributions of the estimated IRFs by Algorithms 1 and 2. We derive the asymptotic representations of $\hat{\Lambda}$, $\hat{\Psi}_s$, $\hat{G}$, $\hat{\Theta}$, $\hat{A}_q$, $\hat{A}_{1:m}$, and

$\tilde{A}_m$, and then combine these results to obtain the asymptotic distributions of the IRFs. Let

$$e^{(1)} = [e_1^{(1)}, ..., e_T^{(1)}]' \text{ and } e_t^{(1)} = [e_{1t}, ..., e_{kt}]'. \tag{3.7}$$

Thus, $e_t^{(1)}$ denotes the idiosyncratic errors for the first k variables used for the identification conditions in (2.17) and (2.21).

The following assumption is required to derive the asymptotic distributions of the proposed estimators.

Assumption 7 – *Central Limit Theorem*
For any $i = 1, ..., N$,

$$\frac{1}{\sqrt{T}} \begin{bmatrix} \text{vec}\left(e^{(1)\prime} F\right) \\ F' \underline{e}_i \\ \text{vec}(\mathcal{F}'\eta) \end{bmatrix} \to_d N\left(0_{(kr+r+rpq)\times 1}, \Omega_i\right). \tag{3.8}$$

Assumption 7 is simply the central limit theorem and can be confirmed under primitive assumptions. See, for example, White's (1984) Theorem 5.15 or Brockwell and Davis's (1991) Theorem 7.1.2.

Proposition 1 – *Under Assumptions 1–7, if $\sqrt{T}/N \to 0$ as $N, T \to \infty$, then*

$$\begin{aligned} \sqrt{T}\text{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1\prime}) &= \frac{1}{\sqrt{T}}(H_F' \otimes I_k)\text{vec}\left(e^{(1)\prime} F\right) + o_p(1) \\ \to_d N\left(0_{(k+1)r\times 1},\ (\bar{H}_F' \otimes I_k)\Omega^{(1)}(\bar{H}_F \otimes I_k)\right), \end{aligned} \tag{3.9}$$

where $\Omega^{(1)}$ denotes the $rk \times rk$ upper-left block of Ω_i, i.e., the asymptotic variance of $\frac{1}{\sqrt{T}}\text{vec}\left(e^{(1)\prime}F\right)$.

Proposition 2
(a) Under Assumptions 1–6 and ID1, $\hat{G} - H_F' G H_\eta = O_p(\delta_{NT}^{-2})$.
(b) Under Assumptions 1–7 and ID1, if $\sqrt{T}/N \to 0$ as $N, T \to \infty$, then

$$\begin{aligned} \sqrt{T}\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1\prime}) &= (\hat{G}' H_F' \otimes I_k)\frac{1}{\sqrt{T}}\text{vec}\left(e^{(1)\prime}F\right) + o_p(1) \\ &\to_d N\left(0_{(k+1)q\times 1},\ S\Omega^{(1)}S'\right), \end{aligned} \tag{3.10}$$

where $S = \bar{H}_\eta' G' \bar{H}_F \bar{H}_F' \otimes I_k$.

Proposition 3 – *Under Assumptions 1–7, if* $\sqrt{T}/N \to 0$ *as* $N, T \to \infty$*, then for any* $i = 1, ..., N$,

$$\sqrt{T}\begin{bmatrix} \mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ \hat{\lambda}_i - H_F^{-1}\lambda_i \\ \mathrm{vec}(\hat{\Psi}_s' - H_F^{-1}\Psi_s' H_F) \end{bmatrix} = W\frac{1}{\sqrt{T}}\begin{bmatrix} \mathrm{vec}\left(e^{(1)'}F\right) \\ F'\underline{e}_i \\ \mathrm{vec}(\mathcal{F}'\eta) \end{bmatrix} + o_p(1),$$
$$\to_d N\left(0_{(kq+r+r^2)\times 1}, W\Omega_i W'\right), \tag{3.11}$$

where $W = \begin{bmatrix} (\bar{H}_F' \otimes I_k) & 0_{k\times r} & 0_{k\times rpq} \\ 0_{r\times kr} & \bar{H}_F' & 0_{r\times rpq} \\ 0_{r^2\times kr} & 0_{r^2\times r} & \bar{R}_s\left[\bar{H}_F' G \otimes \left(\Sigma_{\mathcal{F}}\bar{H}_{\mathcal{F}}\right)^{-1}\right] \end{bmatrix}$ *and* $\bar{R}_s = \sum_{j=1}^{s}(\bar{H}_F^{-1}\Psi_{j-1}\bar{H}_F \otimes [\bar{H}_F^{-1}\Psi_{s-j}'\bar{H}_F,\ \bar{H}_F^{-1}\Psi_{s-j-1}'\bar{H}_F, ..., \bar{H}_F^{-1}\Psi_{s-j-p+1}'\bar{H}_F])$ *with* $\Psi_0 = I_r$ *and* $\Psi_s = 0_{r\times r}$ *for* $s < 0$.

Proposition 3 directly implies the asymptotic distribution of $[\mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'})', (\hat{\lambda}_i - H_F^{-1}\lambda_i)']'$, which is summarized by the following corollary.

Corollary 1 – *Under the Assumptions of Proposition 3,*

$$\sqrt{T}\begin{bmatrix} \mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ \hat{\lambda}_i - H_F^{-1}\lambda_i \end{bmatrix} \to_d N\left(0_{(kr+r)\times 1}, W_1\Omega_i W_1'\right), \tag{3.12}$$

where $W_1 = \begin{bmatrix} (\bar{H}_F' \otimes I_k) & 0_{k\times r} & 0_{k\times rpq} \\ 0_{r\times kr} & \bar{H}_F' & 0_{r\times rpq} \end{bmatrix}$.

Propositions 1–3 provide the asymptotic distributions of $\hat{\Lambda}_1$, $\hat{\lambda}_i$, $\hat{\Theta}_1$, $\hat{G}$, and $\hat{\Psi}_s$. Proposition 2(a) shows that $\sqrt{T}(\hat{G} - H_F' G H_\eta)$ has a degenerate limit distribution if $\sqrt{T}/N \to 0$, so that $H_F' G H_\eta$ can be replaced with $\hat{G}$ as if $\hat{G}$ is observed when $\sqrt{T}$ is much smaller than N. The distribution in Proposition 2(b) is useful for obtaining the asymptotic representations of $\hat{A}_q$ and $\hat{A}_{1:m}\tilde{A}_m$, both of which are continuous functions of $\hat{\Theta}_1$. The results in Proposition 3 and Corollary 1 are applied to obtain the asymptotic distributions of the IRFs over time. The following theorem presents the result under the identification scheme specified by (2.17).

Theorem 1 – *Under Assumptions ID1, ID2, and 1–7, and the condition that* $\sqrt{T}/N \to 0$ *as* N *and* $T \to \infty$, *(a)* $\hat{A}_q$ *is a consistent estimator for* $H_\eta' A_q$ *and*

$$\sqrt{T}(\hat{A}_q - H_\eta' A_q) = \bar{B}_1 \sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1).$$

For the first k *variables,*

$$\begin{aligned}\sqrt{T}(\hat{\Theta}_1 \hat{A}_q - \Theta_1 A_q) &= \bar{B}_2 \sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1) \\ &\to_d N(0_{k\times 1}, \bar{B}_2 S \Omega^{(1)} S' \bar{B}_2'),\end{aligned} \tag{3.13}$$

where $\bar{B}_1 = A_q' \bar{H}_\eta \otimes \left[(-\bar{H}_\eta^{-1}\Theta_1'\Theta_1 \bar{H}_\eta^{-1'})^+ \bar{H}_\eta^{-1}\Theta_1'\right]$ *with* $\mathrm{rank}(\bar{B}_1) = q-1$*,* $\bar{B}_2 = \Theta_1 \bar{H}_\eta^{-1'} \bar{B}_1 + (A_q'\bar{H}_\eta) \otimes I_k$ *with* $\mathrm{rank}(\bar{B}_2) = k - q + 1$*, and* S *is defined in Proposition 2(b).*
(b) For $i = k+1, ..., N$*,*

$$\begin{aligned}\sqrt{T}(\hat{\theta}_i' \hat{A}_q - \theta_i' A_q) &= \sqrt{T} \bar{B}_3^{(i)} \begin{bmatrix} \mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ (\hat{\lambda}_i - H_\eta^{-1}\lambda_i) \end{bmatrix} + o_p(1) \\ &\to_d N\left(0,\ \bar{B}_3^{(i)} W_1 \Omega_i W_1' \bar{B}_3^{(i)'}\right),\end{aligned} \tag{3.14}$$

where $C_1 = [I_{kr} \vdots 0_{kr\times r}]$*,* $C_2 = [0_{r\times kr} \vdots I_r]$*, and* $\bar{B}_3^{(i)} = \theta_i' \bar{H}_\eta^{-1'} \bar{B}_1 (\bar{H}_\eta' G' \bar{H}_F \otimes I_k) C_1 + A_q' G' \bar{H}_F C_2$*.*
(c) For the IRFs of X_{it} *to* $\zeta_{q,t-s}$ $(s \geq 1)$*,*

$$\begin{aligned}\sqrt{T}(\hat{\lambda}_i' \hat{\Psi}_s \hat{G} \hat{A}_q - \lambda_i' \Psi_s G A_q) &= \sqrt{T} \bar{B}_4^{(i)} \begin{bmatrix} \mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ \hat{\lambda}_i - H_F^{-1}\lambda_i \\ \mathrm{vec}(\hat{\Psi}_s' - H_F^{-1}\Psi_s' H_F) \end{bmatrix} + o_p(1) \\ &\to_d N(0,\ \bar{B}_4^{(i)} W \Omega_i W' \bar{B}_4^{(i)'}),\end{aligned} \tag{3.15}$$

where $C_3 = [I_{rk} \vdots 0_{rk\times(r+1)r}]$*,* $C_4 = [0_{r\times rk} \vdots I_r \vdots 0_{r\times r^2}]$*,* $C_5 = [0_{r^2\times(rk+r)} \vdots I_{r^2}]$*, and* $\bar{B}_4^{(i)} = \lambda_i' \Psi_s G \bar{H}_\eta$ $\bar{B}_1 (\bar{H}_\eta' G' \bar{H}_F \otimes I_k) C_3 + A_q' G' \Psi_s' \bar{H}_F C_4 + (\lambda_i' \bar{H}_F^{-1'} \otimes A_q' G' \bar{H}_F) C_5$*.*

Next, we derive the asymptotic results for the IRFs computed in Algorithm 2. Let $A_{1:m}$ denote the eigenvectors associated with the first m eigenvalues of $\Theta_1'\Theta_1$, and let $\bar{A}_m^*$ be the eigenvector associated with the smallest eigenvalue of $A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m}$, where Θ_{11} denotes the first ℓ rows of Θ_1. Let

$$C_0 = [I_\ell \vdots 0_{\ell\times(k-\ell)}], \tag{3.16}$$

so $\Theta_{11} = C_0\Theta_1$. Based on Assumption ID3, we know that $\mathrm{rank}(\Theta_1) = m$ and $\mathrm{rank}(\Theta_{11}) = m-1$. Hence, $A_{1:m}$ spans the space of $[A_1, ..., A_m]$, and the smallest eigenvalue of $A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m}$ is zero, which implies that $\bar{A}_m^{*'} A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m}\bar{A}_m^* = 0$. If a sign is appropriately selected for

$\bar{A}^*_m$, then it follows that

$$A_m = A_{1:m}\bar{A}^*_m. \tag{3.17}$$

Let α_j and $\hat{\alpha}_j$ denote the j-th eigenvalue of $\Theta_1'\Theta_1$ and $\hat{\Theta}_1'\hat{\Theta}_1$, respectively ($j \leq m$), and let $\hat{A}_j$ denote the j-th column of $\hat{A}_{1:m}$. Define

$$\begin{aligned}\bar{B}_5 &= (A'_{1:m}\bar{H}_\eta \otimes I_k) + (I_m \otimes \Theta_1 \bar{H}_\eta^{-1'})\bar{Q}\left[(K_q + I_{q^2})(I_q \otimes \bar{H}_\eta^{-1}\Theta_1')\right] \\ \bar{B}_6 &= \bar{A}^{*'}_m \otimes (-A'_{1:m}\Theta'_{11}\Theta_{11}A_{1:m})^{+}A'_{1:m}\Theta'_{11},\end{aligned} \tag{3.18}$$

where $\bar{Q} = \begin{bmatrix} \bar{Q}_1 \\ \vdots \\ \bar{Q}_m \end{bmatrix}$ with $\bar{Q}_j = \text{plim}_{N,T\to\infty}\left(\hat{A}_j'\right) \otimes (\alpha_j I_q - H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^{+}$ for $j = 1, ..., m$.

Theorem 2 – *Under Assumptions ID1, ID3, and 1–7, and the condition that $\sqrt{T}/N \to 0$ as N and $T \to \infty$,*
(a) $\tilde{A}_m$ consistently estimates $\bar{A}^*_m$. The IRFs of the first k variables of X_t with respect to ζ_{mt} ($m < q$) have the following asymptotic distribution

$$\begin{aligned}\sqrt{T}(\hat{\Theta}_1\hat{A}_{1:m}\tilde{A}_m - \Theta_1 A_m) =& \bar{B}_7\sqrt{T}\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1) \\ &\to_d N(0_{k\times 1}, \bar{B}_7 S\Omega^{(1)}S'\bar{B}_7')\end{aligned} \tag{3.19}$$

where $\bar{B}_7 = \left[(\bar{A}^{*'}_m \otimes I_k) + \Theta_1 A_{1:m}\bar{B}_6(I_m \otimes C_0)\right]\bar{B}_5$ with C_0 defined in (3.16).
(b) $\hat{A}_{1:m}\tilde{A}_m$ consistently estimates $H_\eta' A_m$ and

$$\sqrt{T}\text{vec}(\hat{A}_{1:m}\tilde{A}_m - H_\eta' A_m) = \bar{B}_8\sqrt{T}\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1),$$

where $\bar{B}_8 = (\bar{A}^{*'}_m \otimes I_q)\bar{Q}\left[(K_q + I_{q^2})(I_q \otimes \bar{H}_\eta^{-1}\Theta_1')\right] + \bar{H}_\eta' A_{1:m}\bar{B}_6(I_m \otimes C_0)\bar{B}_5$.

For $i = k+1, ..., N$, the IRFs of the i-th variable of X_t with respect to ζ_{mt} ($m < q$) have the following asymptotic representation

$$\begin{aligned}\sqrt{T}(\hat{\theta}_i'\hat{A}_{1:m}\tilde{A}_m - \theta_i' A_m) = \bar{B}_9^{(i)}\sqrt{T}&\begin{bmatrix} \text{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ (\hat{\lambda}_i - H_\eta^{-1}\lambda_i) \end{bmatrix} + o_p(1) \\ &\to_d N\left(0,\ \bar{B}_9^{(i)} W_1 \Omega_i W_1' \bar{B}_9^{(i)'}\right),\end{aligned} \tag{3.20}$$

where $\bar{B}_9^{(i)} = \theta_i' H_\eta^{-1'}\bar{B}_8(\bar{H}_\eta' G'\bar{H}_F \otimes I_k)C_1 + A_m' G'\bar{H}_F C_2$, with C_1 and C_2 defined in Theorem 1(b).

(c) For the IRFs of X_{it} with respect to $\zeta_{m,t-s}$ ($m < q$, $s \geq 1$), we have

$$\sqrt{T}\left(\hat{\lambda}_i'\hat{\Psi}_s\hat{G}\hat{A}_{1:m}\tilde{A}_m - \lambda_i'\Psi_s G A_m\right) = \bar{B}_{10}^{(i)}\sqrt{T}\begin{bmatrix} \mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1\prime}) \\ \hat{\lambda}_i - H_F^{-1}\lambda_i \\ \mathrm{vec}(\hat{\Psi}_s' - H_F^{-1}\Psi_s' H_F) \end{bmatrix} + o_p(1)$$
$$\to_d N(0, \bar{B}_{10}^{(i)} W \Omega_i W' \bar{B}_{10}^{(i)\prime}), \tag{3.21}$$

where $\bar{B}_{10}^{(i)} = \lambda_i'\Psi_s G\bar{H}_\eta \bar{B}_8(\bar{H}_\eta' G'\bar{H}_F \otimes I_k)C_3 + A_m' G'\Psi_s'\bar{H}_F C_4 + (\lambda_i'\bar{H}_F^{-1\prime} \otimes A_m' G'\bar{H}_F)C_5$, with C_3, C_4, and C_5 defined in Theorem 1(c).

Note that the asymptotic variance of $\sqrt{T}(\hat{\Theta}_1\hat{A}_q - \Theta_1 A_q)$ in Theorem 1(a) is not of full rank because $q-1$ of k zero restrictions are used for the identification of A_q. When $k = q-1$, A_q is just-identified, so $\hat{\Theta}_1\hat{A}_q$ and its asymptotic variance are exactly equal to zero. The asymptotic variance of $\sqrt{T}(\hat{\Theta}_1\hat{A}_{1:m}\tilde{A}_m - \Theta_1 A_m)$ in Theorem 2(a) is also not of full rank. More specifically, the first ℓ variables provide $m-1$ restrictions for the identification of $\bar{A}_m^*$. Hence, the rank of the variance of $\sqrt{T}(\hat{\Theta}_{11}\hat{A}_{1:m}\tilde{A}_m - \Theta_{11}A_m)$ (i.e., contemporaneous IRFs of the first ℓ variables) is no greater than $\ell - m + 1$. This result is not listed in Theorem 2 due to space limitations. However, it is included in the proof for Theorem 2(a).

Theorems 1 and 2 establish the consistency and asymptotic normality for the dynamic IRFs computed by Algorithms 1 and 2. Note that our estimators can consistently estimate the IRFs without any rotation, although $\hat{A}_q$ and $\hat{A}_{1:m}\tilde{A}_m$ are consistent estimators for $H_\eta' A_q$ and $H_\eta' A_m$, respectively. One does not need H_η or H_F to estimate the asymptotic variances of the estimated IRFs. Hence, Theorems 1 and 2 are very useful for frequentist inference in empirical analysis. The next subsection provides practical guidance for the implementation of Theorems 1 and 2.

3.3 Implementation Guidance for the Inference of Estimated IRFs

To conduct statistical inference for the estimated IRFs, we propose the following estimators to compute the asymptotic variances in Theorems 1 and 2. The idiosyncratic errors can be consistently estimated as follows:

$$\hat{e}_t = X_t - \hat{\Lambda}\hat{F}_t. \tag{3.22}$$

Let $\hat{e}_t^{(1)}$ and $\hat{e}_{it}$ denote the first k and the i-th elements of $\hat{e}_t$, respectively. Let

$$\hat{\Sigma}^{(1)} = \frac{1}{T-r}\sum_{t=1}^{T} \hat{e}_t^{(1)}\hat{e}_t^{(1)\prime}. \tag{3.23}$$

Since F and e are independent by Assumption 4, we propose the following estimator for $S\Omega^{(1)}S'$:

$$\widehat{S\Omega^{(1)}S'} = (\hat{G}'\hat{G}) \otimes \hat{\Sigma}^{(1)}. \tag{3.24}$$

For the inference of IRFs in Theorems 1(b), 1(c), 2(b), and 2(c), we compute the following estimator for $W\Omega_i W'$:

$$\widehat{W\Omega_i W'} = \frac{1}{T-r}\sum_{t=p+1}^{T} \xi_t^{(i)}\xi_t^{(i)\prime}, \tag{3.25}$$

where

$$\xi_t^{(i)} = \begin{bmatrix} \mathrm{vec}(\hat{e}_t^{(1)}\hat{F}_t') \\ \hat{F}_t\hat{e}_{it} \\ \hat{R}_s\left(\hat{G} \otimes \left(\frac{\hat{\mathcal{F}}'\hat{\mathcal{F}}}{T-p}\right)^{-1}\right)\mathrm{vec}(\hat{\mathcal{F}}_t\hat{\eta}_t') \end{bmatrix},$$

and

$$\hat{R}_s = \sum_{j=1}^{s}(\hat{\Psi}_{j-1} \otimes [\hat{\Psi}'_{s-j}, \hat{\Psi}'_{s-j-1}, ..., \hat{\Psi}'_{s-j-p+1}])$$

with $\hat{\Psi}_0 = I_r$ and $\hat{\Psi}_s = 0_{r\times r}$ for $s < 0$. The consistency of $\hat{\Sigma}^{(1)}$ and $\widehat{W\Omega_i W'}$ holds if e_t is serially uncorrelated.[3] When e_t is serially correlated, the HAC estimators for the asymptotic variances can be readily constructed following Bai (2003) and Han and Inoue (2014).[4]

The constant matrices $\bar{Q}$, $\bar{B}_s$ for $s = 1, 2, 5, 6, 7, 8$, and $\bar{B}_s^{(i)}$ for $s = 3, 4, 9, 10$ can be readily estimated by replacing the unknown parameters with their finite-sample analogs. Based on (3.3), Propositions 1–3, and Theorems 1–2, we know that $\hat{\Theta}_1$, $\hat{\theta}_i$, $\hat{\lambda}_i$, $\hat{\Psi}_s$, $\hat{G}$, $\hat{A}_q$, $\hat{A}_{1:m}$, $\tilde{A}_m$ consistently estimate $\Theta_1\bar{H}_\eta^{-1\prime}$, $\bar{H}_\eta^{-1}\theta_i$, $\bar{H}_F^{-1}\lambda_i$, $\bar{H}_F'\Psi_s\bar{H}_F^{-1\prime}$, $\bar{H}_F'G\bar{H}_\eta$, $\bar{H}_\eta'A_q$, $\bar{H}_\eta'A_{1:m}$, $\bar{A}_m^*$, respectively. Furthermore, $\bar{H}_\eta$ is orthonormal by Lemma 4 in the appendix. Hence, we propose the following estimators for the constant matrices used in Theorems 1 and 2:
$\hat{Q} = \left[\hat{Q}_1', ..., \hat{Q}_m'\right]'$ with $\hat{Q}_j = \hat{A}_j' \otimes (\hat{\alpha}_j I_q - \hat{\Theta}_1'\hat{\Theta}_1)^+$;
$\hat{B}_1 = \left[\hat{A}_q' \otimes (-\hat{\Theta}_1'\hat{\Theta}_1)^+\hat{\Theta}_1'\right]$, $\quad \hat{B}_2 = \hat{\Theta}_1\hat{B}_1 + \hat{A}_q' \otimes I_k$, $\quad \hat{B}_3^{(i)} = \hat{\theta}_i'\hat{B}_1(\hat{G}' \otimes I_k)C_1 + \hat{A}_q'\hat{G}'C_2$;
$\hat{B}_4^{(i)} = \hat{\lambda}_i'\hat{\Psi}_s\hat{G}\hat{B}_1(\hat{G}' \otimes I_k)C_3 + \hat{A}_q'\hat{G}'\hat{\Psi}_s'C_4 + (\hat{\lambda}_i' \otimes \hat{A}_q'\hat{G}')C_5$;

[3] The proof for the consistency of the variance estimators in (3.24) and (3.25) mainly follows Lemma B3 of Han (2015).

[4] The consistency of HAC estimators involving estimated factors is proved by, for example, Theorem 2 of Han and Inoue (2014).

$$\hat{B}_5 = (\hat{A}'_{1:m} \otimes I_k) + (I_m \otimes \hat{\Theta}_1)\hat{Q}\left[(K_q + I_{q^2})(I_q \otimes \hat{\Theta}'_1)\right];$$
$$\hat{B}_6 = \tilde{A}'_m \otimes (-\hat{A}'_{1:m}\hat{\Theta}'_{11}\hat{\Theta}_{11}\hat{A}_{1:m})^{+}\hat{A}'_{1:m}\hat{\Theta}'_{11};$$
$$\hat{B}_7 = \left[(\tilde{A}'_m \otimes I_k) + \hat{\Theta}_1\hat{A}_{1:m}\hat{B}_6(I_m \otimes C_0)\right]\hat{B}_5;$$
$$\hat{B}_8 = (\tilde{A}'_m \otimes I_q)\hat{Q}\left[(K_q + I_{q^2})(I_q \otimes \hat{\Theta}'_1)\right] + \hat{A}_{1:m}\hat{B}_6(I_m \otimes C_0)\hat{B}_5;$$
$$\hat{B}_9^{(i)} = \hat{\theta}'_i\hat{B}_8(\hat{G}' \otimes I_k)C_1 + \tilde{A}'_m\hat{A}'_{1:m}\hat{G}'C_2;$$
$$\hat{B}_{10}^{(i)} = \hat{\lambda}'_i\hat{\Psi}_s\hat{G}\hat{B}_8(\hat{G}' \otimes I_k)C_3 + \tilde{A}'_m\hat{A}'_{1:m}\hat{G}'\hat{\Psi}'_sC_4 + (\hat{\lambda}'_i \otimes \tilde{A}'_m\hat{A}'_{1:m}\hat{G}')C_5.$$

3.4 Testing the Over-identifying Restrictions

One major advantage of our estimators is that they allow us to test the validity of over-identifying restrictions in the factor loading matrix. To test the over-identifying restrictions in Γ_1 in (2.17) when $k > q-1$, we can check the rank of the estimator $\hat{\Theta}_1$. Consider the following hypotheses:

$$\begin{aligned} \mathbb{H}_0^{(d)} &: \ \text{rank}(\Theta_1) = d, \\ \mathbb{H}_1^{(d)} &: \ \text{rank}(\Theta_1) > d. \end{aligned} \tag{3.26}$$

Given the asymptotic distribution of $\hat{\Theta}_1$ (see Proposition 1), we can implement Kleibergen and Paap's test (2006, hereafter KP test). Define

$$\hat{\Xi} = \hat{\Sigma}^{(1)^{-1/2}}\hat{\Theta}_1(\hat{G}'\hat{G})^{-1/2}. \tag{3.27}$$

Via singular value decomposition, we decompose $\hat{\Xi}$ as

$$\hat{\Xi} = \mathcal{U}\mathcal{D}\mathcal{V}' = \begin{bmatrix} \mathcal{U}_{11} & \mathcal{U}_{12} \\ \mathcal{U}_{21} & \mathcal{U}_{22} \end{bmatrix} \begin{bmatrix} \mathcal{D}_1 & \\ & \mathcal{D}_2 \end{bmatrix} \begin{bmatrix} \mathcal{V}'_{11} & \mathcal{V}'_{21} \\ \mathcal{V}'_{12} & \mathcal{V}'_{22} \end{bmatrix}, \tag{3.28}$$

where $\mathcal{U}'\mathcal{U} = I_k$, $\mathcal{V}'\mathcal{V} = I_q$, $\mathcal{D}$ is a $k \times q$ matrix consisting of the singular values of $\hat{\Xi}$ on its main diagonal and is equal to zero elsewhere; $\mathcal{U}_{11}$, $\mathcal{D}_1$, and $\mathcal{V}_{11}$ are $d \times d$ matrices; $\mathcal{U}_{12}$ and $\mathcal{U}'_{21}$ are $d \times (k-d)$ matrices; $\mathcal{V}_{12}$ and $\mathcal{V}'_{21}$ are $d \times (q-d)$ matrices; $\mathcal{U}_{22}$ is a $(k-d) \times (k-d)$ matrix; $\mathcal{V}_{22}$ is a $(q-d) \times (q-d)$ matrix; and $\mathcal{D}_2$ is a $(k-d) \times (q-d)$ matrix. Define

$$\begin{aligned} \hat{M}_{d,1} &= \begin{bmatrix} \mathcal{U}_{12} \\ \mathcal{U}_{22} \end{bmatrix} \mathcal{U}_{22}^{-1}(\mathcal{U}_{22}\mathcal{U}'_{22})^{1/2}, \\ \hat{M}_{d,2} &= (\mathcal{V}_{22}\mathcal{V}'_{22})^{1/2}(\mathcal{V}'_{22})^{-1}\left[\mathcal{V}'_{12} \vdots \mathcal{V}'_{22}\right], \\ \hat{\rho}_d &= (\hat{M}_{d,2} \otimes \hat{M}'_{d,1})\text{vec}(\hat{\Xi}). \end{aligned} \tag{3.29}$$

Given the results in Proposition 2, we obtain the following theorem.

Theorem 3 – *Under $\mathbb{H}_0^{(d)}$, Assumptions 1–7, ID1, and the condition that $\sqrt{T}/N \to 0$ as N and $T \to \infty$, it follows that*

$$\begin{aligned} \sqrt{T}\hat{\rho}_d &\to_d \ N(0,\ I_{(q-d)(k-d)}), \\ w_d = T\hat{\rho}_d'\hat{\rho}_d &\to_d \ \chi^2_{(k-d)(q-d)}. \end{aligned} \tag{3.30}$$

Hence, if ID2 holds, then $d = q-1$ and under the null hypothesis $\mathbb{H}_0^{(q-1)}$, (3.30) implies

$$w_{q-1} \to_d \chi^2_{k-q+1}. \tag{3.31}$$

Remark 6: An alternative way to test $\mathbb{H}_0^{(q-1)}$ is to use the asymptotic distribution of $\sqrt{T}(\hat{\Theta}_1\hat{A}_q - \Theta_1 A_q)$. Based on Theorem 1(a), it follows that

$$T\hat{A}_q'\hat{\Theta}_1' \left[\hat{B}_2\left((\hat{G}'\hat{G}) \otimes \hat{\Sigma}^{(1)}\right)\hat{B}_2'\right]^+ \hat{\Theta}_1\hat{A}_q \to_d \ \chi^2_{k-q+1}. \tag{3.32}$$

The statistic in (3.32) is asymptotically equivalent to w_{q-1} under $\mathbb{H}_0^{(q-1)}$, but cannot be applied to test $\mathbb{H}_0^{(d)}$ when $d < q-1$. Thus, we consider only the more general test w_d in the rest of the paper.

Next, we consider testing the over-identifying restrictions in (2.21). Recall that ID3 imposes two rank conditions: $\text{rank}(\Gamma_1) = m$ and $\text{rank}(\Gamma_{11}) = m-1$. We could test the rank of $\hat{\Theta}_1$ and $\hat{\Theta}_{11}$ separately using the KP test defined in (3.30). However, this method would not control the size for jointly testing two rank conditions. Thus, we propose a joint test for the following hypotheses:

$$\begin{aligned} \mathbb{H}_0^{(d)'} &: \ \text{rank}(\Theta_1) = d \text{ and } \text{rank}(\Theta_{11}) = d-1 \\ \mathbb{H}_1^{(d)'} &: \ \text{rank}(\Theta_1) > d \text{ or } \text{rank}(\Theta_{11}) > d-1. \end{aligned} \tag{3.33}$$

Note that $\text{rank}(\Theta_{11}A_{1:m}) = d-1$ under $\mathbb{H}_0^{(d)'}$, so our test statistic checks the rank of $\hat{\Theta}_{11}\hat{A}_{1:m}$ in addition to $\hat{\Theta}_1$. Define

$$\hat{\Upsilon} = \hat{\Sigma}_{11}^{(1)^{-1/2}}\hat{\Theta}_{11}\hat{A}_{1:m}, \tag{3.34}$$

where $\hat{\Sigma}_{11}^{(1)}$ denotes the upper-left $\ell \times \ell$ corner of $\hat{\Sigma}^{(1)}$, corresponding to errors of the first ℓ

variables. Via singular value decomposition, we decompose $\hat{\Upsilon}$ as

$$\hat{\Upsilon} = \mathcal{U}^{(\ell)}\mathcal{D}^{(\ell)}\mathcal{V}^{(\ell)\prime} = \begin{bmatrix} \mathcal{U}_{11}^{(\ell)} & \mathcal{U}_{12}^{(\ell)} \\ \mathcal{U}_{21}^{(\ell)} & \mathcal{U}_{22}^{(\ell)} \end{bmatrix} \begin{bmatrix} \mathcal{D}_{1}^{(\ell)} & \\ & \mathcal{D}_{2}^{(\ell)} \end{bmatrix} \begin{bmatrix} \mathcal{V}_{11}^{(\ell)\prime} & \mathcal{V}_{21}^{(\ell)\prime} \\ \mathcal{V}_{12}^{(\ell)\prime} & \mathcal{V}_{22}^{(\ell)\prime} \end{bmatrix},$$

where $\mathcal{U}^{(\ell)\prime}\mathcal{U}^{(\ell)} = I_\ell$, $\mathcal{V}^{(\ell)\prime}\mathcal{V}^{(\ell)} = I_m$, $\mathcal{D}$ is an $\ell \times m$ matrix consisting of the singular values of $\hat{\Upsilon}$ on its main diagonal and is equal to zero elsewhere; $\mathcal{U}_{11}^{(\ell)}$, $\mathcal{D}_{1}^{(\ell)}$, and $\mathcal{V}_{11}^{(\ell)}$ are $(d-1)\times(d-1)$ matrices; $\mathcal{U}_{12}$ and $\mathcal{U}_{21}'$ are $(d-1)\times(\ell-d+1)$ matrices; $\mathcal{V}_{12}$ and $\mathcal{V}_{21}'$ are $(d-1)\times(m-d+1)$ matrices; $\mathcal{U}_{22}$ is a $(\ell-d+1)\times(\ell-d+1)$ matrix; $\mathcal{V}_{22}$ is a $(m-d+1)\times(m-d+1)$ matrix; and $\mathcal{D}_2$ is a $(\ell-d+1)\times(m-d+1)$ matrix. We define

$$\begin{aligned}
\hat{M}_{d,1}^{(\ell)} &= \begin{bmatrix} \mathcal{U}_{12}^{(\ell)} \\ \mathcal{U}_{22}^{(\ell)} \end{bmatrix} \left(\mathcal{U}_{22}^{(1)}\right)^{-1} \left(\mathcal{U}_{22}^{(\ell)}\mathcal{U}_{22}^{(\ell)\prime}\right)^{1/2}, \\
\hat{M}_{d,2}^{(\ell)} &= \left(\mathcal{V}_{22}^{(\ell)}\mathcal{V}_{22}^{(\ell)\prime}\right)^{1/2} (\mathcal{V}_{22}^{(\ell)\prime})^{-1} \left[\mathcal{V}_{12}^{(\ell)\prime} \vdots \mathcal{V}_{22}^{(\ell)\prime}\right], \\
\hat{\tau}_d &= (\hat{M}_{d,2}^{(\ell)} \otimes \hat{M}_{d,1}^{(\ell)\prime})\text{vec}(\hat{\Upsilon}), \\
B_d &= \begin{bmatrix} \left(\hat{M}_{d,2} \otimes \hat{M}_{d,1}'\right)\left((\hat{G}'\hat{G})^{-1/2} \otimes \hat{\Sigma}^{(1)^{-1/2}}\right) \\ \left(\hat{M}_{d,2}^{(\ell)} \otimes \hat{M}_{d,1}^{(\ell)\prime}\hat{\Sigma}_{11}^{(1)^{-1/2}} C_0\right)\hat{B}_5 \end{bmatrix}.
\end{aligned} \tag{3.35}$$

The joint test is then computed as

$$w_d^{joint} = T\,[\hat{\rho}_d', \hat{\tau}_d']\left(B_d\left((\hat{G}'\hat{G}) \otimes \hat{\Sigma}^{(1)}\right)B_d'\right)^{-1}\begin{bmatrix} \hat{\rho}_d \\ \hat{\tau}_d \end{bmatrix}. \tag{3.36}$$

The term $\hat{\tau}_d$ is constructed in a similar manner to $\hat{\rho}_d$ and designed to test the rank of $\hat{\Theta}_{11}\hat{A}_{1:m}$. The joint statistic essentially combines the two tests for $\text{rank}(\hat{\Theta}_1) = d$ and $\text{rank}(\hat{\Theta}_{11}\hat{A}_{1:m}) = d-1$ given that both of their distributions depend on that of $\hat{\Theta}_1$. The matrix B_d is a bridge connecting the distributions of $(\hat{\rho}_d', \hat{\tau}_d')'$ and $\text{vec}(\hat{\Theta}_1)$. To establish the theoretical result for the joint test, we make the following assumption.

Assumption 8 – *The matrix $B_d(S\Omega^{(1)}S')B_d'$ is asymptotically nonsingular.*

This assumption is similar to Assumption 2 of Kleibergen and Paap (2006) and ensures the invertibility of the covariance matrix. Theorem 4 presents the asymptotic distribution of w_d^{joint}.

Theorem 4 – *Under $\mathbb{H}_0^{(d)\prime}$, Assumption ID1 and 1–8, if $\sqrt{T}/N \to 0$ and $\Gamma_1'\Gamma_1$ has distinct* *nonzero eigenvalues, then*

$$w_d^{joint} \to_d \chi^2_{(k-d)(q-d)+(\ell-d+1)}. \tag{3.37}$$

Thus under the identification assumption ID3, the following is true for the joint statistic

$$w_m^{joint} \to_d \chi^2_{(k-m)(q-m)+(\ell-m+1)}. \tag{3.38}$$

In practice, the true value of m is unknown. We propose the following test procedure to determine the value of m.

Algorithm 3:

Start with a value $0 \le \underline{m} \le q-1$ and test the null hypothesis $\mathbb{H}_0^{(\underline{m})'}$: $\text{rank}(\Theta_1) = \underline{m}$ and $\text{rank}(\Theta_{11}) = \underline{m} - 1$ versus the alternative hypothesis $\mathbb{H}_1^{(\underline{m})'}$: $\text{rank}(\Theta_1) > \underline{m}$ or $\text{rank}(\Theta_{11}) > \underline{m} - 1$. If the null is not rejected at the level α, then we set $m = \underline{m}$. If the null is rejected at the level α, then test the null $\mathbb{H}_0^{(\underline{m}+1)'}$ versus the alternative $\mathbb{H}_1^{(\underline{m}+1)'}$. Repeat the testing procedure until $\mathbb{H}_0^{(j)'}$ is not rejected at the level α and set $m = j$.

Algorithm 3 can select the true m with a probability approaching $1-\alpha$. If the null hypothesis $m = q-1$ is rejected, then it implies that the identification assumption is incorrect. One should consider changing the variables used for identification.

Remark 7: Consider the case of $k = \ell + 1$ in (2.21). Assume that the true data generating process is such that all of the zero restrictions are satisfied except those in the last $q-m$ columns of the $(\ell+1)$-th row, i.e.,

$$\Gamma_1 = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1,m-1} & 0 & 0_{1\times(q-m)} \\ \vdots & \cdots & \vdots & \vdots & \vdots \\ \gamma_{\ell 1} & \cdots & \gamma_{\ell,m-1} & 0 & 0_{1\times(q-m)} \\ \gamma_{\ell+1,1} & \cdots & \gamma_{\ell+1,m-1} & \gamma_{\ell+1,m} & *_{1\times(q-m)} \end{bmatrix}_{k\times q}, \tag{3.39}$$

where the asterisk denotes the nonzero entries. In such a scenario, the rank conditions in ID3 continue to hold, but the zero restrictions in ID3 do not. The incorrect restrictions denoted by the asterisk in (3.39) cannot be tested, and the estimated IRFs are inconsistent in general. Recall that the just-identified scheme in (2.22) always has $k = \ell + 1$ (see Remark 3). This again demonstrates the advantage of using over-identifying restrictions. In practice, we recommend using $k > \ell + 1$, if possible, to avoid the untestable restrictions (denoted by the asterisk) in (3.39).

4 Simulations

This section investigates the finite-sample performance of the proposed estimators. We consider the following data generating processes (DGPs). Similar to Bai et al. (2015), we specify the following VAR process for the dynamic factors

$$f_t = \phi I \cdot f_{t-1} + \zeta_t, \tag{4.1}$$

where the structural shocks $\zeta_t \sim N(0_{q\times 1}, I_q)$ and $\phi = 0.7$. The static factors are generated as $F_t = [f_t', f_{1,t-1}, ..., f_{r-q,t-1}]'$ with $r \leq 2q$ for $t = 1, ..., T$, so some lags of f_t are included as static factors. The idiosyncratic errors are generated by $e_{it} \overset{i.i.d.}{\sim} N(0,\ r)$ for $i = 1, ..., N$ and $t = 1, ..., T$. Eq. (4.1) implies that

$$F_t = \Phi_1 F_{t-1} + GA\zeta_t,$$

where $\Phi_1 = \begin{bmatrix} \phi I & 0_{q\times(r-q)} \\ 0_{(r-q)\times q} & 0_{(r-q)\times(r-q)} \end{bmatrix}$ and $GA = \begin{bmatrix} I_q \\ 0_{(r-q)\times 1} \end{bmatrix}$. We set $r = 7$ and $q = 5$. The number of replications is 5000.

Recall that $\Gamma = \Lambda GA$. By the design on GA, Γ is equal to the first q columns of Λ. The elements in the last $r - q$ columns in Λ are drawn from i.i.d. $N(0, 1)$. The first q columns of Λ (i.e., Γ) are generated according to the structure specified as follows:

$$\Gamma = \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \end{bmatrix},\ \Gamma_1 = \begin{bmatrix} * & \cdots & * & 0 & 0_{1\times(q-m)} \\ \vdots & \cdots & \vdots & \vdots & \vdots \\ * & \cdots & * & 0 & 0_{1\times(q-m)} \\ * & \cdots & * & \iota & 0_{1\times(q-m)} \\ \vdots & \cdots & \vdots & \vdots & \vdots \\ * & \cdots & * & * & 0_{1\times(q-m)} \end{bmatrix}_{k\times q},\Gamma_2 = \begin{bmatrix} * & \cdots & * & \iota \\ * & \cdots & * & * \\ \vdots & \cdots & \vdots & \vdots \\ * & \cdots & * & * \end{bmatrix}_{(N-k)\times q} \tag{4.2}$$

where the asterisks are nonzero entries drawn from i.i.d. $N(0, 1)$ and ι is a fixed constant whose sign is assumed to be known to pin down the signs of $\hat{A}_q$ and $\hat{A}_{1:m}\tilde{A}_m$. Hence, the observables are generated as

$$X_t = \Lambda F_t + e_t = [\Gamma \vdots \Lambda_{r-q+1:r}]F_t + e_t, \tag{4.3}$$

where $\Lambda_{r-q+1:r}$ denotes the last $r - q$ columns of Λ. We set $m \in \{3, 4\}$, $(k, \ell) \in \{(5, 4),\ (6, 4),\ (6, 5),\ (7, 4),\ (7, 5)\}$, and $\iota = 1.5$ in the simulations. For $m = 4$, all of the $k - \ell$ combinations considered in our simulations lead to over-identification of A_q and A_m. For $m = 3$, A_m is over-identified, but A_q is not identified.

We investigate the estimation accuracy based on just-identifying and over-identifying restrictions. For $q = 5$ and $m = 4$, we would need $(k, \ell) = (4, 3)$ to achieve just-identification for A_q and A_m by Remark 2. Under our DGP with $m = 4$, it is always possible to use a subset of valid restrictions so that A_q and A_m are just-identified. [5] Table 1 reports the ratio of root mean square errors (RMSEs) of $\hat{\Theta}\hat{A}_q$ and $\hat{\Theta}\hat{A}_{1:m}\tilde{A}_m$ under all of the available restrictions in Γ_1 to the RMSEs of $\hat{\Theta}\hat{A}_q$ and $\hat{\Theta}\hat{A}_{1:m}\tilde{A}_m$ under a subset of just-identifying restrictions. It is clear that all of the ratios in Table 1 are much less than one, indicating that adding a few over-identifying restrictions can substantially improve the estimation accuracy of the estimators (i.e., the contemporaneous IRFs). For instance, using $k = 5$ (i.e., the degree of over-identification is one) can decrease the RMSE of $\hat{\Theta}\hat{A}_q$ by more than 30% compared with the just-identified scheme for $N = 125$ and $T = 250$. The improvement in efficiency increases as N and T increase. In addition, $\hat{\Theta}\hat{A}_q$ is more accurate when k increases, and $\hat{\Theta}\hat{A}_{1:m}\tilde{A}_m$ is more accurate when either k or ℓ increases. Thus, Table 1 shows that the proposed estimators based on over-identifying restrictions can significantly improve the estimation accuracy compared with the conventional estimators under the just-identified scheme.

Tables 2 and 3 focus on the finite-sample performance of the asymptotic distributions established in Theorems 1 and 2. The frequencies with which the t-statistics lie outside the interval $[-1.96, 1.96]$ are reported. Table 2 summarizes the results for the IRFs with respect to the q-th shock and m-th shock when $m = 4$, and Table 3 summarizes the results for the IRFs with respect to the m-th shock when $m = 3$. We consider the horizon $s = 5$ for the dynamic IRFs. In general, the asymptotic approximation works well in finite samples and the rejection frequency approaches 5% as N and T increase. In addition, the numbers in Table 2 ($m = 4$) are even closer to 5% than those in Table 3. This is consistent with the following intuition: a smaller m means more zero eigenvalues of $\Theta_1'\Theta_1$, but the eigenvalues of its finite-sample analog $\hat{\Theta}_1'\hat{\Theta}_1$ are never exactly zero, so the a smaller m introduces larger estimation errors (than the prediction based on asymptotic distributions) in small samples. However, as N and T increase, all the rejection frequencies in Table 3 improve and converge to the nominal level.

We also study the finite-sample performance of w_d and w_d^{joint} proposed in Section (3.4). We use w_m and w_m^{joint} to test the null hypothesis that $\text{rank}(\Theta_m) = m$ and the joint null hypothesis that $\text{rank}(\Theta_1) = m$ and $\text{rank}(\Theta_{11}) = m - 1$, respectively. The nominal size is 5%. It is not surprising that the magnitude of the size distortion is positively correlated with

[5] For example, we only use restrictions in the 3rd – 6th rows of Γ_1 for the case $(k, \ell) = (7, 5)$.

the degrees of freedom of the asymptotic χ^2 distribution. For $m = 3$ and $(k, \ell) = (7, 5)$, the joint test follows a χ^2_{11} asymptotically, and its rejection frequency is 10.6% when $(N, T) = (125,\ 250)$ and 7.5% when $(N, T) = (250,\ 500)$. Thus, the size distortion vanishes as N and T increase. In general, the effective size of the proposed tests is acceptable for the sample sizes considered in the simulations.

In the last experiment, we investigate the power of the w_m and w_m^{joint} tests. The data are generated in the same way as before, except in this case all of the zero entries of Γ_1 in (4.2) are replaced with random draws from i.i.d. $N(0, \beta^2)$ with $\beta \in \{0.1, ..., 1\}$. Under this revised DGP, the matrix Γ_1 is of full rank q. We compare the power of w_m and w_m^{joint} with the power of the infeasible rank tests, assuming that both $X_t - \Lambda_{r-q+1:r}[f_{1,t-1}, ..., f_{r-q,t-1}]'$ and f_t are observed.[6] For the null hypothesis $\text{rank}(\Theta_m) = m$, we use the KP test as the infeasible analog. For the joint null hypothesis, we construct a joint test in a similar manner to the feasible w_d^{joint}. We consider the size-adjusted power for the feasible and infeasible tests. Figure 1 shows the size-adjusted power when $(N, T) = (125, 500)$ and $(k, \ell) = (6, 4)$. The solid lines with circles denote the size-adjusted power of w_d and w_d^{joint}. The solid lines with asterisks denote the size-adjusted power of the infeasible KP test and joint rank test. The upper-left (lower-left) panel shows the power of w_d and the KP test against the null hypothesis that $\text{rank}(\Theta_1) = 4$ ($\text{rank}(\Theta_1) = 3$). The upper-right (lower-right) panel shows the power of w_d^{joint} and the infeasible joint test against the joint null hypothesis that $\text{rank}(\Theta_1) = 4$ and $\text{rank}(\Theta_{11}) = 3$ ($\text{rank}(\Theta_1) = 3$ and $\text{rank}(\Theta_{11}) = 2$). It is clear that the size-adjusted power of w_d and w_d^{joint} is very close to that of their infeasible analogs, which confirms that the proposed tests are powerful against the false null hypothesis.[7] The power increases in the value of β and number of restrictions tested. We also run simulations for other values of $k - \ell$ combinations with different sample sizes. The results exhibit very similar patterns to what is shown in Figure 1.

5 Empirical Application

This section presents the empirical applications of the proposed method in practice. The data set is an updated version of that used by Stock and Watson (2005). It consists of monthly observations of 124 U.S. macroeconomic time series from 1960:1 through 2010:12. The series are properly transformed so that they are approximately stationary. The transformation is similar to those seen in studies by Stock and Watson (2005), Bernanke et al. (2005), and

[6] Note that $X_t - \Lambda_{r-q+1:r}[f_{1,t-1}, ..., f_{r-q,t-1}]' = \Gamma f_t + e_t$ by our DGP.

[7] The power of the proposed test could be even higher than that of the infeasible ones if no size adjustment is conducted.

Forni and Gambetti (2010). See the supplement appendix for the full list of variables and transformation codes.

5.1 Model Specification

We use Bai and Ng's (2002) information criteria and detect 12 static factors. Hence, we set $r = 12$, which seems to be large enough to recover the space spanned by the static factors. We also try $r = 10$ and the results are qualitatively the same. BIC suggests that F_t follows a VAR(1) process, so we set $p = 1$ in our analysis. Following Amengual and Watson (2007), we apply Bai and Ng's (2002) information criteria on $\hat{X}$ and detect five dynamic factors. Thus, we set $q = 5$ in our benchmark model.

We consider the identification of the monetary policy shock. In a conventional structural VAR analysis, it is common to assume that slow-moving variables such as real output and price levels are not affected by the monetary policy shock contemporaneously. In addition, some other fast-moving shocks do not affect the monetary policy instruments within the same month, i.e., the central bank is not aware of the contemporaneous information about these fast shocks when it makes decisions related to the monetary policy. Therefore, it is reasonable to consider a setup where the factor loading Γ follows a structure similar to (2.21). For Γ_{11}, we include the following four variables in our benchmark setup: IP index, CPI, PCE deflator, and commodity price. For Γ_1, we add a few more variables: the federal funds rate (FFR), M1, M2, and total reserve, all of which are closely related to the monetary policy instruments adopted by Christiano, Eichenbaum, and Evans (1998). Hence, the benchmark model has $k = 8$ and $\ell = 4$.

Before estimating the IRFs, we implement the proposed tests in Section (3.4) to check the model specification. Based on Algorithm 3, we find the following results: the p-value for the null hypothesis $\text{rank}(\Theta_1) = 3$ and $\text{rank}(\Theta_{11}) = 2$ is 2×10^{-10}, and the p-value for the null hypothesis $\text{rank}(\Theta_1) = 4$ and $\text{rank}(\Theta_{11}) = 3$ is 0.276. This implies $m = 4$ for our benchmark specification. We also implement the w^d test to double check the ranks of Θ_1 and Θ_{11} separately. The p-values are reported in Table 5 and are consistent with the joint test results.

5.2 The Effects of a Monetary Policy Shock

First, we compare the IRFs based on two different setups for m: (1) $m = 4$, the benchmark setup selected by our tests; and (2) $m = 3$, the setup rejected by our over-identification tests. Figure 2 presents the cumulative IRFs of various macroeconomic variables after a contractionary monetary policy shock. It is remarkable that the results based on $m = 3$ (marked by dashed curves) are quite contradictory to economic theory. For $m = 3$, the real output and employment undergo substantial increases after the shock. The price puzzle is also evident: all the price indexes (CPI, PCE deflator, and PPI) remain above their pre-shock levels even three years after the shock. Many other variables also exhibit a wrong sign in their IRFs, such as consumer expectation, inventories, unemployment, and capacity utilization.

In contrast, the IRFs under $m = 4$ are very consistent with what is expected to occur after a contractionary monetary policy shock. The real output, consumption, employment, orders, M2, consumer expectation, and all price indexes start to decline after the shock. No price puzzle appears. In addition, both the JPY/USD exchange rate and the SP500 respond very quickly after the shock, undergoing a significant shift and then remaining at a constant level. Hence, there is no delayed overshooting puzzle for exchange rate (Eichenbaum and Evans, 1995), which is consistent with the findings of Forni and Gambetti (2010). In summary, the comparison in Figure 2 shows the advantage of using testable over-identifying restrictions to uncover the effects of a structural shock. The setup selected by our tests ($m = 4$) yields more reasonable results than the rejected setup ($m = 3$). With a conventional just-identified scheme, we are unable to test the validity of the restrictions and could obtain very misleading estimates for the IRFs.

Next, we compare the inferential results for the IRFs under over- and just-identifying restrictions. A subset of restrictions in our benchmark specification can be used so that the model is just-identified. Under the just-identification scheme, the variables in Γ_{11} are IP index, CPI, and commodity price, and Γ_1 includes one additional variable: FFR. The contemporaneous IRFs of the variables in Γ_{11} are exactly zero. The IRFs under the just-identified scheme are similar to those shown in Figure 2 with $m = 4$ and thus not reported. Instead, we focus on whether the over-identification scheme can improve the inferential results for the IRFs.

Overall, for the variables considered in Figure 2, our benchmark (over-identified) setup leads to 434, 382, and 282 t-ratios that are significant at the 10%, 5%, and 1% levels, respectively, whereas the just-identified setup leads to 384, 351, and 261 t-ratios that are significant at the 10%, 5%, and 1% levels, respectively.[8] Moreover, Table 6 reports the p-values of the t-ratios of the estimated IRFs under just- and over-identified schemes. The major patterns of the t-ratios are similar across these two schemes, but adding over-identifying restrictions does lead to improvements in statistical significance. For example, although

[8]The total number of comparable t-ratios are $20 \times 37 - 3 = 737$ for 20 variables over $s = 0, 1, ..., 36$ with 3 unavailable t-ratios excluded under the just identified scheme.

employment does not respond significantly to the monetary policy shock under the just-identified scheme, it responds to the shock significantly at the 10% level 18 months after the shock under our over-identified scheme. In addition, the contemporaneous response of the JPY/USD exchange rate has a p-value equal to 1.2% under the over-identified scheme versus a p-value equal to 10.5% under the just-identified scheme.

6 Conclusion

This paper develops new estimators for the impulse response functions in structural FAVAR models under over-identifying restrictions. Compared with the untestable just-identified schemes commonly used in the literature, our framework allows practitioners to test the validity of the identification restrictions. We establish the asymptotic distributions of the new estimators and develop test statistics for the over-identifying restrictions. A simulation study confirms that the estimated impulse response functions tend to be more accurate under an over-identified scheme than under a just-identified scheme. An empirical application with U.S. macroeconomic data shows that our over-identified scheme can help to improve statistical significance and eliminate the use of incorrect restrictions that lead to spurious impulse responses.

Appendix

A Lemmas

The following lemmas are useful in the proofs of main theoretical results in the paper. To save space, the proofs of the following lemmas are provided in the supplement appendix.

Lemma 1: under Assumptions 1–4

(a) $T^{-1}\sum_{t=1}^{T}\|\hat{F}_t - H_F' F_t\|^2 = O_p\left(\delta_{NT}^{-2}\right)$ and $T^{-1}\sum_{t=p+1}^{T}\|\hat{\mathcal{F}}_t - H_{\mathcal{F}}'\mathcal{F}_t\|^2 = O_p(\delta_{NT}^{-2})$.

(b) $V_X \to_p V$, where V is the diagonal matrix consisting of the eigenvalues of $\Sigma_F\Sigma_\Lambda$.

(c) H_F and $H_{\mathcal{F}}$ are $O_p(1)$ and nonsingular as N and $T \to \infty$.

Lemma 2: under Assumptions 1–6

(a) $T^{-1}\sum_{t=1}^{T}(\hat{F}_t - H_F' F_t)[F_t', e_{it}, \eta_t'] = O_p(\delta_{NT}^{-2})$ for any given $i = 1, ..., N$.

(b) $T^{-1}(\hat{\mathcal{F}} - \mathcal{F}H_{\mathcal{F}})'[\mathcal{F}\vdots\eta] = O_p(\delta_{NT}^{-2})$.

Lemma 3: under Assumptions 1–6

$T^{-1}\sum_{t=p+1}^{T}\|\hat{\eta}_t - H_\eta'\eta_t\|^2 = O_p(\delta_{NT}^{-2})$.

Lemma 4: under Assumptions 1–6

(a) $T^{-1}(\hat{\eta} - \eta H_\eta)'\eta = O_p(\delta_{NT}^{-2})$.

(b) $H_\eta \to_p \bar{H}_\eta$, where $\bar{H}_\eta$ is orthonormal.

B Proofs for Propositions

Proof of Proposition 1:

Consider the first k variables of X_t. Let Λ_1 denote the first k rows of Λ and $X_t^{(1)} = [X_{1t}, ..., X_{kt}]'$. Recall that $e_t^{(1)} = [e_{1t}, ..., e_{kt}]'$ and $e^{(1)} = [e_1^{(1)}, ..., e_T^{(1)}]'$. Hence,

$$X_t^{(1)} = \Lambda_1 F_t + e_t^{(1)}, \tag{B.1}$$

Since $T^{-1}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' = I_r$, the OLS estimator for Λ_1 is

$$\hat{\Lambda}_1 = \frac{1}{T}\sum_{t=1}^{T} X_t^{(1)}\hat{F}_t' = \frac{1}{T}\sum_{t=1}^{T}(\Lambda_1 F_t + e_t^{(1)})\hat{F}_t' \tag{B.2}$$

Note that

$$\frac{1}{T}\sum_{t=1}^{T}\Lambda_1 F_t\hat{F}_t' = \Lambda_1 H_F^{-1'}\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' + \Lambda_1 H_F^{-1'}\frac{1}{T}\sum_{t=1}^{T}(H_F'F_t - \hat{F}_t)\hat{F}_t'$$
$$= \Lambda_1 H_F^{-1'} + O_p(\delta_{NT}^{-2})$$

where the term $O_p(\delta_{NT}^{-2})$ follows Lemma B3 of Bai (2003). Now, we can rewrite (B.2) as

$$\begin{aligned}\hat{\Lambda}_1 &= \Lambda_1 H_F^{-1'} + \frac{1}{T}\sum_{t=1}^{T} e_t^{(1)}F_t'H_F + \frac{1}{T}\sum_{t=1}^{T} e_t^{(1)}(\hat{F}_t - H_F'F_t)' + O_p(\delta_{NT}^{-2}) \\ &= \Lambda_1 H_F^{-1'} + \frac{1}{T}e^{(1)'}FH_F + O_p(\delta_{NT}^{-2}), \end{aligned} \tag{B.3}$$

where $T^{-1}\sum_{t=1}^{T} e_t^{(1)}(\hat{F}_t - H_F'F_t)' = O_p(\delta_{NT}^{-2})$ by Lemmas 2(a). Hence, using the condition $\sqrt{T}/N \to 0$, we obtain

$$\sqrt{T}\text{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) = \frac{1}{\sqrt{T}}(H_F' \otimes I_k)\text{vec}\left(e^{(1)'}F\right) + o_p(1). \tag{B.4}$$

The asymptotic normality directly follows from Assumption 7 and (3.3).
Q.E.D.

Proof of Proposition 2:
(a) Note that

$$\begin{aligned}\hat{G} &= \frac{1}{T-p}\sum_{t=p+1}^{T}\hat{F}_t\hat{\eta}_t' \\ &= \frac{1}{T-p}\sum_{t=p+1}^{T}H_F'F_t\hat{\eta}_t' + \frac{1}{T-p}\sum_{t=p+1}^{T}(\hat{F}_t - H_F'F_t)(\hat{\eta}_t - H_\eta'\eta_t)' + \frac{1}{T-p}\sum_{t=p+1}^{T}(\hat{F}_t - H_F'F_t)\eta_t'H_\eta \\ &= \frac{1}{T-p}\sum_{t=p+1}^{T}\left(H_F'\Phi H_{\mathcal{F}}^{-1'}H_{\mathcal{F}}'\mathcal{F}_{t-1}\hat{\eta}_t' + H_F'G\eta_t\hat{\eta}_t'\right) + O_p(\delta_{NT}^{-2}) \\ &= \frac{1}{T-p}\sum_{t=p+1}^{T}\left(H_F'\Phi H_{\mathcal{F}}^{-1'}H_{\mathcal{F}}'\mathcal{F}_{t-1}\hat{\eta}_t' + H_F'G\eta_t\eta_t'H_\eta\right) + O_p(\delta_{NT}^{-2}),\end{aligned}$$

where the last two terms in the 2nd line are $O_p(\delta_{NT}^{-2})$ by Lemmas 1(a), 2(a), 3, and Cauchy-Schwarz inequality, and the 4th line follows from Lemma 4(a). Recall the identification restriction $\zeta'\zeta/(T-p) = I_q$ in Assumption ID1, we have $\eta'\eta/(T-p) = I_q$ since A is orthonormal. By fact that $\hat{\mathcal{F}}'\hat{\eta} = 0$, we have

$$\begin{aligned}\hat{G} - H_F' G H_\eta &= \frac{1}{T-p} H_F' \Phi H_{\mathcal{F}}^{-1'} (\mathcal{F} H_{\mathcal{F}} - \hat{\mathcal{F}})'\hat{\eta} + O_p(\delta_{NT}^{-2}) \\ &= \frac{H_F' \Phi H_{\mathcal{F}}^{-1'} [(\mathcal{F} H_{\mathcal{F}} - \hat{\mathcal{F}})'(\hat{\eta} - \eta H_\eta) + (\mathcal{F} H_{\mathcal{F}} - \hat{\mathcal{F}})'\eta H_\eta]}{T-p} + O_p(\delta_{NT}^{-2}) \\ &= O_p(\delta_{NT}^{-2}), \end{aligned} \tag{B.5}$$

by Cauchy-Schwarz inequality and Lemmas 1(a), 3 and 2(a).

(b) Consider the distribution of $\hat{\Theta}_1$. Note that Lemmas 3 and 4(a) imply

$$\frac{\hat{\eta}'\hat{\eta} - H_\eta'\eta'\eta H_\eta}{T-p} = \frac{(\hat{\eta} - \eta H_\eta)'(\hat{\eta} - \eta H_\eta) + (\hat{\eta} - \eta H_\eta)'\eta H_\eta + H_\eta'\eta'(\hat{\eta} - \eta H_\eta)}{T-p} = O_p(\delta_{NT}^{-2}).$$

Thus, we have

$$(\hat{\eta}'\hat{\eta} - H_\eta'\eta'\eta H_\eta)/(T-p) = I_q - H_\eta'\eta'\eta H_\eta/(T-p) = O_p(\delta_{NT}^{-2}). \tag{B.6}$$

By Assumptions ID1 and 2(b), we know $\eta'\eta/(T-p) = I_q$, and thus (B.6) implies that $I_q = H_\eta' H_\eta + O_p(\delta_{NT}^{-2})$, which means

$$H_\eta' = H_\eta^{-1} + O_p(\delta_{NT}^{-2}). \tag{B.7}$$

Now, we can derive the asymptotic distribution of $\sqrt{T}\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'})$. Since $\hat{\Theta}_1 = \hat{\Lambda}_1\hat{G}$, we can obtain

$$\begin{aligned}\sqrt{T}\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) &= \sqrt{T}\text{vec}(\hat{\Lambda}_1\hat{G} - \Lambda_1 H_F^{-1'} H_F' G H_\eta) + O_p(\sqrt{T}\delta_{NT}^{-2}) \\ &= (H_\eta' G' H_F \otimes I_k)\sqrt{T}\text{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) + O_p(\sqrt{T}\delta_{NT}^{-2}), \end{aligned} \tag{B.8}$$

where the 1st line uses the result in (B.7) and 2nd line uses Proposition 2(a). The desired result is obtained by combining (B.4) and (B.8).

Q.E.D.

Proof of Proposition 3:

First, consider the distributions of $\hat{\Phi}$ and $\hat{\Psi}_s$. Note that Lemma 2 implies the following:

$$T^{-1}\sum_{t=p+1}^{T} (\hat{\mathcal{F}}_t - H_{\mathcal{F}}'\mathcal{F}_t)F_t' = T^{-1}\sum_{t=p+1}^{T} (\hat{\mathcal{F}}_t - H_{\mathcal{F}}'\mathcal{F}_t)(\mathcal{F}_t'\Phi' + \eta_t' G') = O_p(\delta_{NT}^{-2}), \tag{B.9}$$

$$T^{-1}\sum_{t=p+1}^{T} \Phi\mathcal{F}_t(\hat{F}_t - H_F' F_t)' = T^{-1}\sum_{t=p+1}^{T} (F_t - G\eta_t)(\hat{F}_t - H_F' F_t)' = O_p(\delta_{NT}^{-2}). \tag{B.10}$$

Hence, we have

$$
\begin{aligned}
&T^{-1}\sum_{t=p+1}^{T}\hat{\mathcal{F}}_t(\hat{F}_t-H_F'F_t)' \\
=&T^{-1}\sum_{t=p+1}^{T}(\hat{\mathcal{F}}_t-H_{\mathcal{F}}'\mathcal{F}_t)(\hat{F}_t-H_F'F_t)'+T^{-1}\sum_{t=p+1}^{T}H_{\mathcal{F}}'\mathcal{F}_t(\hat{F}_t-H_F'F_t)'=O_p(\delta_{NT}^{-2}), \quad \text{(B.11)}
\end{aligned}
$$

where the 1st term is $O_p(\delta_{NT}^{-2})$ by Cauchy-Schwarz inequality and Lemma 1(a), and the 2nd term is $O_p(\delta_{NT}^{-2})$ by (B.10).

The model (2.2) can be rewritten as

$$
\begin{aligned}
F_t'H_F &= \sum_{j=1}^{p}F_{t-j}'H_FH_F^{-1}\Phi_j'H_F+\eta_t'G'H_F \\
&= \mathcal{F}_{t-1}'H_{\mathcal{F}}H_{\mathcal{F}}^{-1}\Phi'H_F+\eta_t'G'H_F. \qquad \text{(B.12)}
\end{aligned}
$$

The OLS estimate for Φ is given by

$$
\begin{aligned}
\hat{\Phi}' &= \left(\frac{\hat{\mathcal{F}}'\hat{\mathcal{F}}}{T-p}\right)^{-1}\frac{\sum_{t=p+1}^{T}\hat{\mathcal{F}}_t\hat{F}_t'}{T-p} \\
&= \left(\frac{\hat{\mathcal{F}}'\hat{\mathcal{F}}}{T-p}\right)^{-1}\frac{\sum_{t=p+1}^{T}\left[H_{\mathcal{F}}'\mathcal{F}_tF_t'H_F+(\hat{\mathcal{F}}_t-H_{\mathcal{F}}'\mathcal{F}_t)F_t'H_F+\hat{\mathcal{F}}_t(\hat{F}_t-H_F'F_t)'\right]}{T-p}, \qquad \text{(B.13)}
\end{aligned}
$$

where the last two terms in the square brackets are $O_p(\delta_{NT}^{-2})$ by (B.9) and (B.11). Hence,

$$
\hat{\Phi}' = \left(\frac{H_{\mathcal{F}}'\mathcal{F}'\mathcal{F}H_{\mathcal{F}}}{T-p}\right)^{-1}\frac{\sum_{t=p+1}^{T}H_{\mathcal{F}}'\mathcal{F}_tF_t'H_F}{T-p}+O_p(\delta_{NT}^{-2}), \qquad \text{(B.14)}
$$

where we use the fact that

$$
\frac{\hat{\mathcal{F}}'\hat{\mathcal{F}}}{T-p}-\frac{H_{\mathcal{F}}'\mathcal{F}'\mathcal{F}H_{\mathcal{F}}}{T-p}=\frac{(\hat{\mathcal{F}}-\mathcal{F}H_{\mathcal{F}})'(\hat{\mathcal{F}}-\mathcal{F}H_{\mathcal{F}})}{T-p}+\frac{H_{\mathcal{F}}'\mathcal{F}'(\hat{\mathcal{F}}-\mathcal{F}H_{\mathcal{F}})}{T-p}+\frac{(\hat{\mathcal{F}}-\mathcal{F}H_{\mathcal{F}})'\mathcal{F}H_{\mathcal{F}}}{T-p}=O_p(\delta_{NT}^{-2}) \qquad \text{(B.15)}
$$

by Lemmas 1(a) and 2(b). Plugging (B.12) into (B.14) yields

$$
\begin{aligned}
\hat{\Phi}' &= \left(\frac{H_{\mathcal{F}}'\mathcal{F}'\mathcal{F}H_{\mathcal{F}}}{T-p}\right)^{-1}\frac{\sum_{t=p+1}^{T}H_{\mathcal{F}}'\mathcal{F}_t(\mathcal{F}_t'H_{\mathcal{F}}H_{\mathcal{F}}^{-1}\Phi'H_F+\eta_t'G'H_F)}{T-p}+O_p(\delta_{NT}^{-2}) \\
\sqrt{T}\text{vec}(\hat{\Phi}'-H_{\mathcal{F}}^{-1}\Phi'H_F) &= \left[H_F'G\otimes\left(\frac{H_{\mathcal{F}}'\mathcal{F}'\mathcal{F}H_{\mathcal{F}}}{T-p}\right)^{-1}H_{\mathcal{F}}'\right]\frac{\sqrt{T}\sum_{t=p+1}^{T}\text{vec}(\mathcal{F}_t\eta_t')}{T-p}+O_p\left(\frac{\sqrt{T}}{\delta_{NT}^2}\right). \qquad \text{(B.16)}
\end{aligned}
$$

Combining (2.2) and (2.4) gives

$$F_t' H_F = \eta_t' G' H_F + \sum_{s=1}^{\infty} \eta_t' G' H_F H_F^{-1} \Psi_s' H_F.$$

(B.16) implies that $\hat{\Phi}_j'$ is a consistent estimated for $H_F^{-1}\Phi_j' H_F$, so $\hat{\Psi}_s'$ is a consistent estimate for $H_F^{-1}\Psi_s' H_F$ in the VMA representation. By (11.7.1) to (11.7.5) of Hamilton (1994), we have

$$\sqrt{T}\text{vec}(\hat{\Psi}_s' - H_F^{-1}\Psi_s' H_F) = R_s \sqrt{T}\text{vec}(\hat{\Phi}' - H_{\mathcal{F}}^{-1}\Phi' H_F) + o_p(1), \tag{B.17}$$

where $\bar{R}_s = \sum_{j=1}^{s}(\bar{H}_F^{-1}\Psi_{j-1}\bar{H}_F \otimes [\bar{H}_F^{-1}\Psi_{s-j}'\bar{H}_F,\ \bar{H}_F^{-1}\Psi_{s-j-1}'\bar{H}_F, ..., \bar{H}_F^{-1}\Psi_{s-j-p+1}'\bar{H}_F])$ with $\Psi_0 = I_r$ and $\Psi_s = 0_{r\times r}$ for $s < 0$.

Finally, for the OLS estimator $\hat{\lambda}_i$, it follows that

$$\hat{\lambda}_i - H_F^{-1}\lambda_i = \frac{1}{T}\bar{H}_F' F' \underline{e}_i + O_p(\delta_{NT}^{-2}). \tag{B.18}$$

Combining the results in (B.4), (B.18), and (B.17) gives

$$\sqrt{T}\begin{bmatrix} \text{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ \hat{\lambda}_i - H_F^{-1}\lambda_i \\ \text{vec}(\hat{\Psi}_s' - H_F^{-1}\Psi_s' H_F) \end{bmatrix} = \frac{1}{\sqrt{T}} W \begin{bmatrix} \text{vec}\left(e^{(1)'}F\right) \\ F'\underline{e}_i \\ \text{vec}(\mathcal{F}'\eta) \end{bmatrix} + o_p(1).$$

The proof of Proposition 3 is completed.
Q.E.D.

C Proofs for Theorems

Proof of Theorem 1:

(a) First, we derive the asymptotics of $\hat{A}_q$. Recall that α_q and $\hat{\alpha}_q$ denotes the smallest eigenvalue of $\Theta_1'\Theta_1$ and $\hat{\Theta}_1'\hat{\Theta}_1$, respectively, so we have

$$\hat{\Theta}_1'\hat{\Theta}_1\hat{A}_q = \hat{\alpha}_q\hat{A}_q.$$

Assumption ID2 implies $\alpha_q = 0$. Since $\Theta_1 A_q = 0_{k\times 1}$ by (2.17) and $A_q' A_q = 1$, we know that A_q is the eigenvector associated with the smallest eigenvalue of $\Theta_1'\Theta_1$, i.e.,

$$\begin{aligned}\Theta_1'\Theta_1 A_q &= A_q\alpha_q,\\ H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'} H_\eta' A_q &= H_\eta^{-1} A_q\alpha_q,\end{aligned} \tag{C.1}$$

Combining (B.7) and (C.1), we obtain

$$H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'} H_\eta' A_q = H_\eta' A_q\alpha_q + O_p(\delta_{NT}^{-2}). \tag{C.2}$$

Let A_q^* denote the eigenvector associated with the smallest eigenvalue of $H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'}$. Since $\alpha_q = 0$ by Assumption ID2, we have

$$\begin{aligned}H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'} A_q^* &= 0\\ H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'} H_\eta' A_q &= O_p(\delta_{NT}^{-2}).\end{aligned} \tag{C.3}$$

Since $H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'}$ has only one zero eigenvalue, (C.3) implies $A_q^* \pm H_\eta' A_q = O_p(\delta_{NT}^{-2})$. By implementing an appropriate sign restriction on A_q^*, we obtain

$$A_q^* = H_\eta' A_q + O_p(\delta_{NT}^{-2}). \tag{C.4}$$

Since $\alpha_q = 0$ is unique, the eigenvector associated with the zero eigenvalue of $H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'}$ is a continuously differentiable function in the neighborhood of $\Theta_1 H_\eta^{-1'}$. Also, Proposition 2 shows that $\hat{\Theta}_1$ consistently estimates $\Theta_1 H_\eta^{-1'}$, so $\hat{A}_q$ consistently estimates A_q^* by continuous mapping theorem. By Theorem 7 of Magnus and Neudecker (1999), we know that,

$$\begin{aligned}\mathrm{d}\hat{A}_q &= (-H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^+(\mathrm{d}\hat{\Theta}_1'\hat{\Theta}_1 + \hat{\Theta}_1'\mathrm{d}\hat{\Theta}_1)A_q^*\\ &= \left[A_q^{*'} \otimes (-H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^+\right]\left[(\hat{\Theta}_1' \otimes I_q)\mathrm{vec}(\mathrm{d}\hat{\Theta}_1') + (I_q \otimes \hat{\Theta}_1')\mathrm{vec}(\mathrm{d}\hat{\Theta}_1)\right]\\ &= \left[A_q^{*'} \otimes (-H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^+\right]\left[(\hat{\Theta}_1' \otimes I_q)K_{kq} + (I_q \otimes \hat{\Theta}_1')\right]\mathrm{vec}(\mathrm{d}\hat{\Theta}_1)\\ &= B_1\mathrm{vec}(\mathrm{d}\hat{\Theta}_1).\end{aligned} \tag{C.5}$$

Since $\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'} \to_p 0$ and $\Theta_1 H_\eta^{-1'} A_q^* = 0_{k\times 1}$ by the definition of A_q^*, it follows that $A_q^{*'}\hat{\Theta}_1' \to_p 0_{1\times k}$. Thus, we have

$$B_1 \to_p \bar{B}_1 \equiv A_q'\bar{H}_\eta \otimes \left[(-\bar{H}_\eta^{-1}\Theta_1'\Theta_1\bar{H}_\eta^{-1'})^+\bar{H}_\eta^{-1}\Theta_1'\right], \text{ with } \mathrm{rank}(\bar{B}_1) = q-1,$$

because $\mathrm{rank}(\Theta_1) = q-1$ by Assumption ID2. By Delta method and asymptotic distribution in Proposition 1, we obtain

$$\sqrt{T}(\hat{A}_q - A_q^*) = \bar{B}_1\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1)$$
$$\to_d N(0_{q\times 1}, \bar{B}_1 S\Omega^{(1)}S'\bar{B}_1'). \qquad \text{(C.6)}$$

Under the condition that $\sqrt{T}/N \to 0$, (C.4) and (C.6) imply that

$$\sqrt{T}(\hat{A}_q - H_\eta^{'}A_q) = \bar{B}_1\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1), \qquad \text{(C.7)}$$

which has the same asymptotic distribution as (C.6).

Next, we obtain the contemporaneous IRFs of the first k variables,

$$\begin{aligned}\sqrt{T}(\hat{\Theta}_1\hat{A}_q - \Theta_1 A_q) &= \sqrt{T}\hat{\Theta}_1(\hat{A}_q - H_\eta^{'}A_q) + \sqrt{T}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'})H_\eta^{'}A_q \\ &= \{\hat{\Theta}_1\bar{B}_1 + [(A_q'H_\eta)\otimes I_k]\}\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1) \\ &= \bar{B}_2\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1) \\ &\to_d N(0_{k\times 1}, \bar{B}_2 S\Omega^{(1)}S'\bar{B}_2'),\end{aligned}$$

where

$$\begin{aligned}\bar{B}_2 &= \Theta_1\bar{H}_\eta^{-1'}\bar{B}_1 + (A_q'\bar{H}_\eta)\otimes I_k \\ &= A_q'\bar{H}_\eta \otimes \left[I_k - \Theta_1\bar{H}_\eta^{-1'}(\bar{H}_\eta^{-1}\Theta_1'\Theta_1\bar{H}_\eta^{-1'})^+\bar{H}_\eta^{-1}\Theta_1'\right].\end{aligned} \qquad \text{(C.8)}$$

Note that $\Theta_1\bar{H}_\eta^{-1'}(\bar{H}_\eta^{-1}\Theta_1'\Theta_1\bar{H}_\eta^{-1'})^+\bar{H}_\eta^{-1}\Theta_1'$ is symmetric and idempotent and its rank is equal torank$(\Theta_1) = q-1$. Hence,

$$\begin{aligned}\mathrm{rank}(\bar{B}_2) &= 1\times \mathrm{trace}\left[I_k - \Theta_1\bar{H}_\eta^{-1'}(\bar{H}_\eta^{-1}\Theta_1'\Theta_1\bar{H}_\eta^{-1'})^+\bar{H}_\eta^{-1}\Theta_1'\right] \\ &= k - \mathrm{rank}(\Theta_1) = k - q + 1,\end{aligned}$$

where we use the fact that the rank and the trace are equal for symmetric and idempotent matrices.

(b) To obtain the asymptotic representation of $\hat{\theta}_i'\hat{A}_q$ for $i = k+1, ..., N$, note that

$$\mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) = [I_{kr} \vdots 0_{kr\times r}]\begin{bmatrix}\mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ (\hat{\lambda}_i - H_\eta^{-1}\lambda_i)\end{bmatrix} + O_p(\delta_{NT}^{-2}),$$
$$\hat{\lambda}_i - H_\eta^{-1}\lambda_i = [0_{r\times kr} \vdots I_r]\begin{bmatrix}\mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ (\hat{\lambda}_i - H_\eta^{-1}\lambda_i)\end{bmatrix} + O_p(\delta_{NT}^{-2}).$$

Let $C_1 = [I_{kr} \vdots 0_{kr\times r}]$ and $C_2 = [0_{r\times kr} \vdots I_r]$. Then we have

$$\begin{aligned}
\sqrt{T}(\hat{\theta}_i'\hat{A}_q - \theta_i' A_q) &= \sqrt{T}\hat{\theta}_i'(\hat{A}_q - H_\eta' A_q) + \sqrt{T}(\hat{\theta}_i' - \theta_i' H_\eta^{-1'})H_\eta' A_q \\
&= \sqrt{T}\hat{\theta}_i'\bar{B}_1 \mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + \sqrt{T}A_q' H_\eta(\hat{\theta}_i - H_\eta^{-1}\theta_i) + o_p(1) \\
&= \sqrt{T}(\hat{\theta}_i'\bar{B}_1(\bar{H}_\eta' G'\bar{H}_F \otimes I_k)C_1 + A_q' G'\bar{H}_F C_2)\begin{bmatrix} \mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ (\hat{\lambda}_i - H_\eta^{-1}\lambda_i) \end{bmatrix} + o_p(1) \\
&\to_d N\left(0,\ \bar{B}_3^{(i)} W_1 \Omega_i W_1' \bar{B}_3^{(i)'}\right),
\end{aligned}$$

where the third line uses (B.8), and $\bar{B}_3^{(i)} = \theta_i'\bar{H}_\eta^{-1'}\bar{B}_1(\bar{H}_\eta' G'\bar{H}_F \otimes I_k)C_1 + A_q' G'\bar{H}_F C_2$.

(c) Next, consider the IRFs of X_{it} to $\zeta_{q,t-s}$ ($s \geq 1$). By the condition $\sqrt{T}/N \to 0$, and results in (B.5), (B.7), and Proposition 2, we have

$$\begin{aligned}
&\sqrt{T}(\hat{\lambda}_i'\hat{\Psi}_s\hat{G}\hat{A}_q - \lambda_i' H_F^{-1'} H_F'\Psi_s H_F^{-1'} H_F' G H_\eta H_\eta^{-1} A_q) \\
=&\sqrt{T}\hat{\lambda}_i'\hat{\Psi}_s\hat{G}(\hat{A}_q - H_\eta' A_q) + \sqrt{T}(\hat{\lambda}_i - H_F^{-1}\lambda_i)'\hat{\Psi}_s H_F' G A_q + \sqrt{T}\lambda_i' H_F^{-1'}(\hat{\Psi}_s - H_F'\Psi_s H_F^{-1'})H_F' G A_q + o_p(1) \\
=&\sqrt{T}\hat{\lambda}_i'\hat{\Psi}_s\hat{G}\bar{B}_1\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + \sqrt{T}A_q' G' H_F\hat{\Psi}_s'(\hat{\lambda}_i - H_F^{-1}\lambda_i) \\
&+ (\lambda_i' H_F^{-1'} \otimes A_q' G' H_F)\sqrt{T}\mathrm{vec}(\hat{\Psi}_s' - H_F^{-1}\Psi_s' H_F) + o_p(1) \\
=&\left[\hat{\lambda}_i'\hat{\Psi}_s\hat{G}\bar{B}_1(\bar{H}_\eta' G'\bar{H}_F \otimes I_k)C_3 + A_q' G' H_F\hat{\Psi}_s' C_4 + (\lambda_i' H_F^{-1'} \otimes A_q' G' H_F)C_5\right]\sqrt{T}\begin{bmatrix} \mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H_F^{-1'}) \\ \hat{\lambda}_i - H_F^{-1}\lambda_i \\ \mathrm{vec}(\hat{\Psi}_s' - H_F^{-1}\Psi_s' H_F) \end{bmatrix} + o_p(1),
\end{aligned}$$

where $C_3 = [I_{rk} \vdots 0_{rk\times(r+1)r}]$, $C_4 = [0_{r\times rk} \vdots I_r \vdots 0_{r\times r^2}]$, and $C_5 = [0_{r^2\times(rk+r)} \vdots I_{r^2}]$, and the second last line follows from (C.7). Hence,

$$\sqrt{T}(\hat{\lambda}_i'\hat{\Psi}_s\hat{G}\hat{A}_q - \lambda_i'\Psi_s G A_q) \to_d N(0,\ \bar{B}_4^{(i)} W\Omega_i W'\bar{B}_4^{(i)'}),$$

where $\bar{B}_4^{(i)} = \lambda_i'\Psi_s G\bar{H}_\eta\bar{B}_1(\bar{H}_\eta' G'\bar{H}_F \otimes I_k)C_3 + A_q' G'\Psi_s'\bar{H}_F C_4 + (\lambda_i'\bar{H}_F^{-1'} \otimes A_q' G'\bar{H}_F)C_5$.
Q.E.D.

Proof of Theorem 2:

(a) Recall that $\hat{A}_{1:m}$ are the eigenvectors associated with first m eigenvalues of $\hat{\Theta}_1'\hat{\Theta}_1$ in descending order. Recall that $C_0 = [I_\ell \vdots 0_{\ell\times(k-\ell)}]$, so $\hat{\Theta}_{11} = C_0\hat{\Theta}_1$. By the definition of $\hat{A}_{1:m}$,

$$\hat{A}_{1:m}'\hat{\Theta}_1' C_0' C_0\hat{\Theta}_1\hat{A}_{1:m}\tilde{A}_m = \tilde{A}_m\tilde{\alpha}_m, \tag{C.9}$$

where $\tilde{\alpha}_m$ is the smallest eigenvalue of $\hat{A}_{1:m}'\hat{\Theta}_{11}'\hat{\Theta}_{11}\hat{A}_{1:m}$. To find the asymptotic distribution of $\tilde{A}_m$, we first investigate the asymptotic representation of $\hat{\Theta}_{11}\hat{A}_{1:m}$. By design of $\hat{A}_{1:m}$, we have

$$\hat{\Theta}_1'\hat{\Theta}_1\hat{A}_{1:m} = \hat{A}_{1:m}\hat{Z},$$

where $\hat{Z}$ is a diagonal matrix consisting of the first m eigenvalues of $\hat{\Theta}_1'\hat{\Theta}_1$ in descending order. Similar to (C.1), let Z be the diagonal matrix consisting of the first m eigenvalues of $\Theta_1'\Theta_1$ and let $A_{1:m}$ denote the first m columns of A, so we have

$$\begin{aligned} H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'} H_\eta' A_{1:m} &= H_\eta^{-1} A_{1:m} Z \\ H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'} H_\eta' A_{1:m} &= H_\eta' A_{1:m} Z + O_p(\delta_{NT}^{-2}), \end{aligned} \tag{C.10}$$

where the last line of (C.10) uses (B.7). By Assumption ID3, $\Theta_1'\Theta_1 = A\Gamma_1'\Gamma_1 A'$ has distinct eigenvalues, so $H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'}$ also has distinct nonzero eigenvalues because H_η is orthonormal asymptotically by Lemma 4(b). Hence, the first m eigenvectors of $H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'}$ are continuously differentiable functions. Recall that $A_{1:m}^*$ denotes the eigenvectors associated with the first m eigenvalues of $H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'}$, so (C.10) implies that

$$A_{1:m}^* = H_\eta' A_{1:m} + O_p(\delta_{NT}^{-2}), \tag{C.11}$$

if the signs of each column of $A_{1:m}^*$ are properly adjusted. Let A_j^* and $\hat{A}_j$ denote the j-th column of $A_{1:m}^*$ and $\hat{A}_{1:m}$, respectively ($1 \le j \le m$); let α_j and $\hat{\alpha}_j$ denote the j-th diagonal element of Z and $\hat{Z}$, respectively. Also, recall that $\hat{\Theta}_1$ consistently estimates $\Theta_1 H_\eta^{-1'}$, so $\hat{A}_{1:m}$ consistently estimates $A_{1:m}^*$ and $H_\eta' A_{1:m}$ (up to a sign). By Theorem 7 of Magnus and Neudecker (1999), we have

$$\begin{aligned} \mathrm{d}\hat{A}_j &= (\alpha_j I_q - H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^+ (\mathrm{d}\hat{\Theta}_1'\hat{\Theta}_1 + \hat{\Theta}_1'\mathrm{d}\hat{\Theta}_1) A_j^* \\ &= \left[A_j^{*'} \otimes (\alpha_j I_q - H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^+\right] \left[(\hat{\Theta}_1' \otimes I_q)\mathrm{vec}(\mathrm{d}\hat{\Theta}_1') + (I_q \otimes \hat{\Theta}_1')\mathrm{vec}(\mathrm{d}\hat{\Theta}_1\right]) \\ &= \left[A_j^{*'} \otimes (\alpha_j I_q - H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^+\right] \left[(\hat{\Theta}_1' \otimes I_q)K_{kq} + (I_q \otimes \hat{\Theta}_1')\right] \mathrm{vec}(\mathrm{d}\hat{\Theta}_1) \\ &= \left[A_j^{*'} \otimes (\alpha_j I_q - H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^+\right] \left[(K_q + I_{q^2})(I_q \otimes \hat{\Theta}_1')\right] \mathrm{vec}(\mathrm{d}\hat{\Theta}_1) \end{aligned}$$

for $j = 1, ..., m$. Let $Q_j = A_j^{*'} \otimes (\alpha_j I_q - H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1'})^+$. Hence,

$$\mathrm{dvec}(\hat{A}_{1:m}) = Q\left[(K_q + I_{q^2})(I_q \otimes \hat{\Theta}_1')\right] \mathrm{vec}(\mathrm{d}\hat{\Theta}_1), \tag{C.12}$$

where $Q = \begin{bmatrix} Q_1 \\ \vdots \\ Q_m \end{bmatrix}$. For the first k variables, by (C.12) and (C.11) we have

$$\begin{aligned}
&\sqrt{T}\text{vec}(\hat{\Theta}_1\hat{A}_{1:m} - \Theta_1 H_\eta^{-1\prime} H_\eta' A_{1:m}) \\
=&\text{vec}\left[\sqrt{T}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1\prime})\hat{A}_{1:m} + \sqrt{T}\Theta_1 H_\eta^{-1\prime}(\hat{A}_{1:m} - H_\eta' A_{1:m})\right] \\
=&\sqrt{T}(\hat{A}_{1:m}' \otimes I_k)\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1\prime}) + (I_m \otimes \Theta_1 H_\eta^{-1\prime})\sqrt{T}\text{vec}(\hat{A}_{1:m} - H_\eta' A_{1:m}) \\
=&B_5\sqrt{T}\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1\prime}) + o_p(1) \qquad \text{(C.13)}
\end{aligned}$$

where $B_5 = (\hat{A}_{1:m}' \otimes I_k) + (I_m \otimes \Theta_1 H_\eta^{-1\prime})Q\left[(K_q + I_{q^2})(I_q \otimes \hat{\Theta}_1')\right]$. The probability limit of B_5 is given by

$$\bar{B}_5 = (A_{1:m}'\bar{H}_\eta \otimes I_k) + (I_m \otimes \Theta_1 \bar{H}_\eta^{-1\prime})\bar{Q}\left[(K_q + I_{q^2})(I_q \otimes \bar{H}_\eta^{-1}\Theta_1')\right],$$

where $\bar{Q} = \begin{bmatrix} \bar{Q}_1 \\ \vdots \\ \bar{Q}_m \end{bmatrix}$ with $Q_j = \text{plim}\left(\hat{A}_j'\right) \otimes (\alpha_j I_q - H_\eta^{-1}\Theta_1'\Theta_1 H_\eta^{-1\prime})^+$ for $j = 1, ..., m$.

For the first ℓ variables, we have

$$\begin{aligned}
&\sqrt{T}\text{vec}(C_0\hat{\Theta}_1\hat{A}_{1:m} - C_0\Theta_1 H_\eta^{-1\prime} H_\eta' A_{1:m}) \\
=&(I_m \otimes C_0)\bar{B}_5\sqrt{T}\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1\prime}) + o_p(1). \qquad \text{(C.14)}
\end{aligned}$$

Next, we consider the asymptotic properties of $\tilde{A}_m$. Recall that $\bar{\alpha}_m$ is the smallest eigenvalue of $A_{1:m}'\Theta_1' C_0' C_0 \Theta_1 A_{1:m}$ and $\bar{A}_m^*$ is the corresponding eigenvector, so

$$A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m}\bar{A}_m^* = \bar{A}_m^*\bar{\alpha}_m^*. \qquad \text{(C.15)}$$

Let

$$\tilde{\Theta}_{11} = \hat{\Theta}_{11}\hat{A}_{1:m}. \qquad \text{(C.16)}$$

Since $\hat{\Theta}_{11}$ is a consistent estimator for $\Theta_{11}H_\eta^{-1\prime}$ and $\hat{A}_{1:m}$ is consistent estimator for $H_\eta' A_{1:m}$, it follows that $\tilde{\Theta}_{11} - \Theta_{11}A_{1:m} \to_p 0$ and

$$\tilde{\Theta}_{11}'\tilde{\Theta}_{11} - A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m} \to_p 0. \qquad \text{(C.17)}$$

By the identification condition in Assumption ID3, we have $\bar{\alpha}_m^* = 0$ being the unique zero eigenvalue of $A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m}$, so $\tilde{A}_m - \bar{A}_m^* \to_p 0$. Similar to (C.5), we have

$$\begin{aligned}
\mathrm{d}\tilde{A}_m &= \left[\bar{A}_m^{*\prime} \otimes (-A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m})^+\right] \mathrm{vec}(\mathrm{d}\tilde{\Theta}_{11}'\tilde{\Theta}_{11} + \tilde{\Theta}_{11}'\mathrm{d}\tilde{\Theta}_{11}) \\
&= \left[\bar{A}_m^{*\prime} \otimes (-A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m})^+\right] \left[(\tilde{\Theta}_{11}' \otimes I_m)K_{\ell m} + (I_m \otimes \tilde{\Theta}_{11}')\right] \mathrm{vec}(\mathrm{d}\tilde{\Theta}_{11}) \\
&= B_6 \mathrm{vec}(\mathrm{d}\tilde{\Theta}_{11}),
\end{aligned}$$

where $B_6 = \left[\bar{A}_m^{*\prime} \otimes (-A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m})^+\right] \left[(\tilde{\Theta}_{11}' \otimes I_m)K_{\ell m} + (I_m \otimes \tilde{\Theta}_{11}')\right]$. Since $\Theta_{11}A_{1:m}\bar{A}_m^* = 0_{\ell\times 1}$ by definition of $\bar{A}_m^*$ in (C.15) and $\tilde{\Theta}_{11} - \Theta_{11}A_{1:m} \to_p 0$, it follows that $\tilde{\Theta}_{11}\bar{A}_m^* \to_p 0_{\ell\times 1}$. Thus, we have

$$B_6 \to_p \bar{B}_6 \equiv \left[\bar{A}_m^{*\prime} \otimes (-A_{1:m}'\Theta_{11}'\Theta_{11}A_{1:m})^+ A_{1:m}'\Theta_{11}'\right], \text{ with } \mathrm{rank}(\bar{B}_6) = m-1,$$

because rank $(\Theta_{11}A_{1:m}) = m-1$ by ID3. By Delta method, we have

$$\sqrt{T}(\tilde{A}_m - \bar{A}_m^*) = \sqrt{T}\bar{B}_6 \mathrm{vec}(\tilde{\Theta}_{11} - \Theta_{11}A_{1:m}) + o_p(1). \tag{C.18}$$

Next, we can derive the following asymptotic representation for the contemporaneous IRFs of the first k variables in X_t with respect to ζ_{mt},

$$\begin{aligned}
&\sqrt{T}(\hat{\Theta}_1\hat{A}_{1:m}\tilde{A}_m - \Theta_1 A_{1:m}\bar{A}_m^*) \\
=&\sqrt{T}(\hat{\Theta}_1\hat{A}_{1:m} - \Theta_1 A_{1:m})\tilde{A}_m + \Theta_1 A_{1:m}\sqrt{T}(\tilde{A}_m - \bar{A}_m^*) \\
=&\sqrt{T}(\tilde{A}_m' \otimes I_k)\mathrm{vec}(\hat{\Theta}_1\hat{A}_{1:m} - \Theta_1 A_{1:m}) + \sqrt{T}\Theta_1 A_{1:m}\bar{B}_6 \mathrm{vec}(\tilde{\Theta}_{11} - \Theta_{11}A_{1:m}) + o_p(1) \\
=&\left[(\tilde{A}_m' \otimes I_k) + \Theta_1 A_{1:m}\bar{B}_6(I_m \otimes C_0)\right] \bar{B}_5\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1\prime}) + o_p(1),
\end{aligned}$$

where we use the results in (C.18) and (C.14). By the distribution of $\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1\prime})$ in Proposition 1 and (3.17), we obtain

$$\sqrt{T}(\hat{\Theta}_1\hat{A}_{1:m}\tilde{A}_m - \Theta_1 A_m) \to_d N(0_{k\times 1}, \bar{B}_7 S\Omega^{(1)}S'\bar{B}_7'),$$

where $\bar{B}_7 = \left[(\bar{A}_m^{*\prime} \otimes I_k) + \Theta_1 A_{1:m}\bar{B}_6(I_m \otimes C_0)\right] \bar{B}_5$.

For the IRFs of the first ℓ variables in X_t with respect to ζ_{mt},

$$\begin{aligned}
&\sqrt{T}(\hat{\Theta}_{11}\hat{A}_{1:m}\tilde{A}_m - \Theta_{11}A_{1:m}\bar{A}_m^*) \\
=&\sqrt{T}(\hat{\Theta}_{11}\hat{A}_{1:m} - \Theta_1 A_{1:m})\tilde{A}_m + \Theta_{11}A_{1:m}\sqrt{T}(\tilde{A}_m - \bar{A}_m^*) \\
=&\bar{B}_{7\ell}\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1\prime}) + o_p(1),
\end{aligned}$$

where $\bar{B}_{7\ell} = \left[(\bar{A}^{*\prime}_m \otimes I_\ell) + \Theta_{11} A_{1:m}\bar{B}_6\right](I_m \otimes C_0)\bar{B}_5$. Similar to (C.8),

$$(\bar{A}^{*\prime}_m \otimes I_\ell) + \Theta_{11} A_{1:m}\bar{B}_6 = \bar{A}^{*\prime}_m \otimes \left[I_\ell - \Theta_{11}A_{1:m}(A'_{1:m}\Theta'_{11}\Theta_{11}A_{1:m})^{+}A'_{1:m}\Theta'_{11}\right],$$

which is of rank $\ell - m + 1$.

(b) Now, consider the contemporaneous IRFs of i-th variable for $i = \ell+1, ..., N$. Note that

$$\begin{aligned}\sqrt{T}\mathrm{vec}(\hat{A}_{1:m}\tilde{A}_m - H'_\eta A_{1:m}\bar{A}^*_m) &= \mathrm{vec}\left[\sqrt{T}(\hat{A}_{1:m} - H'_\eta A_{1:m})\tilde{A}_m + H'_\eta A_{1:m}\sqrt{T}(\tilde{A}_m - \bar{A}^*_m)\right] \\ &= (\tilde{A}'_m \otimes I_q)\sqrt{T}\mathrm{vec}(\hat{A}_{1:m} - H'_\eta A_{1:m}) + H'_\eta A_{1:m}\sqrt{T}(\tilde{A}_m - \bar{A}^*_m) \\ &= \bar{B}_8\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H^{-1\prime}_\eta) + o_p(1), \qquad \text{(C.19)}\end{aligned}$$

where we use (C.12), (C.14), (C.18) and $\bar{B}_8 = (\bar{A}^{*\prime}_m \otimes I_q)\bar{Q}\left[(K_{qq} + I_{q^2})(I_q \otimes \bar{H}^{-1}_\eta \Theta'_1)\right] + \bar{H}'_\eta A_{1:m}\bar{B}_6(I_m \otimes C_0)\bar{B}_5$.

For $\hat{\theta}'_i\hat{A}_{1:m}\tilde{A}_m$ for $i = k+1, ..., N$, by (B.8) and (C.19) we can obtain

$$\begin{aligned}&\sqrt{T}(\hat{\theta}'_i\hat{A}_{1:m}\tilde{A}_m - \theta'_i A_{1:m}\bar{A}^*_m) \\ =&\hat{\theta}'_i\sqrt{T}(\hat{A}_{1:m}\tilde{A}_m - H'_\eta A_{1:m}\bar{A}^*_m) + \sqrt{T}(\hat{\theta}_i - H^{-1}_\eta\theta_i)'H'_\eta A_{1:m}\bar{A}^*_m \\ =&\hat{\theta}'_i\bar{B}_8\sqrt{T}\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H^{-1\prime}_\eta) + \bar{A}^{*\prime}_m A'_{1:m}H_\eta\hat{G}'\sqrt{T}(\hat{\lambda}_i - H^{-1}_\eta\lambda_i) + o_p(1) \\ =&\bar{B}^{(i)}_9\sqrt{T}\begin{bmatrix}\mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H^{-1\prime}_F) \\ (\hat{\lambda}_i - H^{-1}_\eta\lambda_i)\end{bmatrix} + o_p(1)\end{aligned}$$

where $\bar{B}^{(i)}_9 = \theta'_i H^{-1\prime}_\eta \bar{B}_8(\bar{H}'_\eta G'\bar{H}_F \otimes I_k)C_1 + A'_m G'\bar{H}_F C_2$.

(c) For IRFs of X_{it} with respect to $\zeta_{m,t-s}$ $(s \geq 1)$, we have

$$\begin{aligned}&\sqrt{T}\left(\hat{\lambda}'_i\hat{\Psi}_s\hat{G}\hat{A}_{1:m}\tilde{A}_m - \lambda'_i H^{-1\prime}_F H'_F\Psi_s H^{-1\prime}_F H'_F G H_\eta H^{-1}_\eta A_{1:m}\bar{A}^*_m\right) \\ =&\sqrt{T}\hat{\lambda}'_i\hat{\Psi}_s\hat{G}(\hat{A}_{1:m}\tilde{A}_m - H'_\eta A_{1:m}\bar{A}^*_m) + \sqrt{T}\hat{\lambda}'_i(\hat{\Psi}_s - H'_F\Psi_s H^{-1\prime}_F)H'_F G A_{1:m}\bar{A}^*_m \\ &+ \sqrt{T}(\hat{\lambda}_i - H^{-1}_F\lambda_i)'H'_F\Psi_s G A_{1:m}\bar{A}^*_m + o_p(1) \\ =&\sqrt{T}\hat{\lambda}'_i\hat{\Psi}_s\hat{G}\bar{B}_8\mathrm{vec}(\hat{\Theta}_1 - \Theta_1 H^{-1\prime}_\eta) + \bar{A}^{*\prime}_m A'_{1:m}G'\Psi'_s H_F\sqrt{T}(\hat{\lambda}_i - H^{-1}_F\lambda_i) \\ &+ \sqrt{T}(\hat{\lambda}'_i \otimes \bar{A}^{*\prime}_m A'_{1:m}G'H_F)\mathrm{vec}(\hat{\Psi}'_s - H^{-1}_F\Psi'_s H_F) + o_p(1) \\ =&\bar{B}^{(i)}_{10}\sqrt{T}\begin{bmatrix}\mathrm{vec}(\hat{\Lambda}_1 - \Lambda_1 H^{-1\prime}_F) \\ \hat{\lambda}_i - H^{-1}_F\lambda_i \\ \mathrm{vec}(\hat{\Psi}'_s - H^{-1}_F\Psi'_s H_F)\end{bmatrix} + o_p(1)\end{aligned}$$

where the 2nd equality uses (C.19), the last line uses (B.8), and $\bar{B}^{(i)}_{10} = \lambda'_i\Psi_s G\bar{H}_\eta\bar{B}_8(\bar{H}'_\eta G'\bar{H}_F \otimes$

$I_k)C_3 + \bar{A}^{*\prime}_m A'_{1:m} G' \Psi'_s \bar{H}_F C_4 + (\lambda'_i \bar{H}_F^{-1'} \otimes \bar{A}^{*\prime}_m A'_{1:m} G' \bar{H}_F) C_5$.
Q.E.D.

Proof of Theorem 3:

The proof mainly follows that of Theorem 1 of Kleibergen and Paap (2006). Let $\Xi = \hat{\Sigma}^{(1)^{-1/2}} \Theta_1 H_\eta^{-1'} (\hat{G}'\hat{G})^{-1/2}$, so we have

$$\begin{aligned}\sqrt{T}\text{vec}(\hat{\Xi} - \Xi) &= \left[(\hat{G}'\hat{G})^{-1/2} \otimes \Sigma^{(1)^{-1/2}}\right] \sqrt{T}\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) \\ &\to_d N(0_{kq\times 1},\ I_{kq})\end{aligned} \tag{C.20}$$

Note that $\hat{M}_{d,2}\hat{M}'_{d,2}$ and $\hat{M}'_{d,1}\hat{M}_{d,1}$ are identity matrices, so we have

$$\sqrt{T}\hat{\rho}_d = \underbrace{\sqrt{T}(\hat{M}_{d,2} \otimes \hat{M}'_{d,1})\text{vec}(\hat{\Xi} - \Xi)}_{\to_d N(0_{kq\times 1},\ I_{kq})} + \sqrt{T}\text{vec}(\hat{M}'_{d,1}\Xi\hat{M}'_{d,2}). \tag{C.21}$$

We next show that $\sqrt{T}\text{vec}(\hat{M}'_{d,1}\Xi\hat{M}'_{d,2}) \to_p 0$. By the decomposition suggested by Kleibergen and Paap (2006), it follows that $\Xi = K_{d,1}K_{d,2}$ under $\mathbb{H}_0^{(d)}$, where $K_{d,1}$ is $k \times d$, $K_{d,2}$ is $d \times q$, $\hat{M}'_{d,1}K_{d,1} \to_p 0$, and $K_{d,2}\hat{M}'_{d,2} \to_p 0$. Kleibergen and Paap show that $\hat{M}_{d,1}$ and $\hat{M}_{d,2}$ converge to their probability limits at the same rate as $\hat{\Theta}_1$. Hence, we have $\hat{M}'_{d,1}K_{d,1} = O_p(T^{-1/2}) + O_p(\delta_{NT}^{-2})$ and $K_{d,2}\hat{M}'_{d,2} = O_p(T^{-1/2}) + O_p(\delta_{NT}^{-2})$ by (B.4). Under the condition that $\sqrt{T}/N \to 0$ as $N,\ T \to \infty$, we have $\sqrt{T}\text{vec}(\hat{M}'_{d,1}\Xi\hat{M}'_{d,2}) \to_p 0$.
Q.E.D.

Proof of Theorem 4:
Recall that Theorem 3 has shown that $\sqrt{T}\hat{\rho}_d$ is asymptotically normal if $\text{rank}(\hat{\Theta}_1) = d$. The term $\hat{\tau}_d$ is designed to test the rank of $\hat{\Theta}_{11}\hat{A}_{1:m}$ and we only need to find its asymptotic distribution. Note that

$$\sqrt{T}\hat{\tau}_d = \sqrt{T}(\hat{M}^{(\ell)}_{d,2} \otimes \hat{M}^{(\ell)\prime}_{d,1})\text{vec}\left[\hat{\Sigma}_{11}^{(1)^{-1/2}}\left(\hat{\Theta}_{11}\hat{A}_{1:m} - \Theta_{11}A_{1:m}\right)\right] + \sqrt{T}\text{vec}\left(\hat{M}^{(\ell)\prime}_{d,1}\hat{\Sigma}_{11}^{(1)^{-1/2}}\Theta_{11}A_{1:m}\hat{M}^{(\ell)\prime}_{d,2}\right), \tag{C.22}$$

where the second term in (C.22) is negligible for the same reason as the second term in (C.21). By (C.14), the first term in (C.22) can be reduced to

$$\begin{aligned}&\sqrt{T}(\hat{M}^{(\ell)}_{d,2} \otimes \hat{M}^{(\ell)\prime}_{d,1})\text{vec}\left[\hat{\Sigma}_{11}^{(1)^{-1/2}}\left(\hat{\Theta}_{11}\hat{A}_{1:m} - \Theta_{11}A_{1:m}\right)\right] \\ =&\sqrt{T}(\hat{M}^{(\ell)}_{d,2} \otimes \hat{M}^{(\ell)\prime}_{d,1}\hat{\Sigma}_{11}^{(1)^{-1/2}})(I_m \otimes C_0)\bar{B}_5\text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1).\end{aligned} \tag{C.23}$$

This gives the second row of matrix B_d. Combining with result for $\hat{\rho}_d$, we obtain

$$\sqrt{T}\begin{bmatrix}\hat{\rho}_d \\ \hat{\tau}_d\end{bmatrix} = B_d \text{vec}(\hat{\Theta}_1 - \Theta_1 H_\eta^{-1'}) + o_p(1). \tag{C.24}$$

Under Assumption 8, it follows that $w_d^{joint} \to_d \chi^2_{(k-d)(q-d)+(\ell-d+1)}$.
Q.E.D.

References

[1] Amengual, D. and M. Watson, 2007. Consistent Estimation of the Number of Dynamic Factors in a Large *N* And *T* Panel, *Journal of Business and Economic Statistics*, 25(1): 91-96.

[2] Bai, J., 2003. Inferential Theory for Factor Models of Large Dimensions, *Econometrica* 71, 135-172.

[3] Bai, J. and K.P. Li, 2012. Statistical analysis of factor models of high dimension, *Annals of Statistics*, 40(1), 436-465.

[4] Bai, J., K.P. Li, and L. Lu, 2015. Estimation and inference of FAVAR models, *Journal of Business and Economic Statistics*, forthcoming.

[5] Bai, J. and S. Ng, 2002. Determining the Number of Factors in Approximate Factor Models, *Econometrica*, 70, 191-221.

[6] Bai, J. and S. Ng, 2004. A Panic Attack on Unit Roots and Cointegration, *Econometrica*, 72(4), 1127-1177.

[7] Bai, J. and S. Ng, 2013. Principal Components Estimation and Identification of Static Factors, *Journal of Econometrics*, 176, 18-29.

[8] Bai, J. and P. Wang, 2014. Identification Theory for High Dimensional Static and Dynamic Factor Models, *Journal of Econometrics*, 178(2), 794-804.

[9] Bai, J. and P. Wang, 2015. Identification and Bayesian Estimation of Dynamic Factor Models, *Journal of Business and Economic Statistics*, 33(2), 221-240.

[10] Bernanke, B.S., J. Boivin, and P. Eliasz, 2005. Measuring the Effects of Monetary Policy: a Factor-augmented Vector Autoregressive (FAVAR) Approach, *Quarterly Journal of Economics* 120, 387–422.

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

[12] Boivin, J. and M. Giannoni, and I. Mihov, 2009. Sticky Prices and Monetary Policy: Evidence from Disaggregated US Data, *American Economic Review*, 99:1, 350-384.

[13] Brockwell, P. and R. Davis, 1991. Time Series: Theory and Methods (2d ed.), Berlin: Springer-Verlag.

[14] Buch, C. M., S. Eickmeier, and E. Prieto, 2014. Macroeconomic Factors and Microlevel Bank Behavior, *Journal of Money, Credit and Banking*, 41, Vol. 46, Issue 4, 715-751.

[15] Canova, F. and Pérez Forero F.J., 2012. Estimating Overidentified, Non-recursive, Time Varying Coefficients Structural VARs, Barcelona GSE Working Paper Series, 637.

[16] Caggiano, G., E. Castelnuovo, and N. Groshenny, 2014. Uncertainty Shocks and Unemployment Dynamics in U.S. Recessions, *Journal of Monetary Economics*, 67, 78-92.

[17] Christiano, L., M. Eichenbaum, and C. Evans 1998. Monetary Policy Shocks: What Have We Learned and to What End? NBER working paper, 6400.

[18] Eichenbaum, M. and C. Evans 1995. Some Empirical Evidence of Shocks to Monetary Policy on Exchange Rates, *Quarterly Journal of Economics*, 110, 975–1010.

[19] Eickmeier, S., W. Lemke, and M. Marcellino, 2015. Classical Time Varying Factor-Augmented Vector Autoregressive Models–Estimation, Forecasting and Structural Analysis, *Journal of Royal Statistical Society: Series A*, 178, Part(3), 493-533.

[20] Forni, M., D. Giannone, M. Lippi, and L. Reichlin, 2009. Opening the Black Box: Structural Factor Models with Large Cross Sections, *Econometric Theory*, 25, 1319-1347.

[21] Forni, M. and L. Gambetti, 2010. The Dynamic Effects of Monetary Policy: A Structural Factor Model Approach, *Journal of Monetary Economics*, 57(2), 203-216.

[22] Forni, M., L. Gambetti, and L. Sala, 2014. No News in Business Cycles, *The Economic Journal*, 124 (December), 1168-1191.

[23] Gafarov, B., 2014. Identification in Dynamic Models Using Sign Restrictions, Manuscript, Pennsylvania State University.

[24] Gilchrist, S., V. Yankov, and E. Zakraisek. Credit Market Shocks and Economic Fluctuations: Evidence from Corporate Bond and Stock Markets, *Journal of Monetary Economics*, 56, 471-493.

[25] Hamilton, J., 1994. Time Series Analysis, Princeton University Press.

[26] Han, X., 2015. Tests for Overidentifying Restrictions in Factor-Augmented VAR Models, *Journal of Econometrics*, 184(2), 394-419.

[27] Han, X. and A. Inoue, 2014. Tests for Parameter Instability in Dynamic Factor Models, *Econometric Theory*, forthcoming.

[28] Kleibergen, F and R. Paap, 2006. Generalized Reduced Rank Tests Using the Singular Value Decomposition, *Journal of Econometrics*, 133(1), 97-126.

[29] Kociecki, A., M. Rubaszek, and M. Ca'Zorz, 2012. Bayesian Analysis of Recursive SVAR Models with Overidentified Restrictions, European Central Bank Working Paper Series, No. 1492.

[30] Koop, G. and D. Korobilis, 2014. A New Index of Financial Conditions, *European Economic Review*, Vol. 71, 101-116.

[31] Leeper, E., C. Sims, and T. Zha, 1996. What Does Monetary Policy Do? Brookings Papers on Economic Activity 2.

[32] Magnus, J.R. and h. Neudecker, 1999. Matrix Differential Calculus with Applications in Statistics and Econometrics, John Wiley & Sons.

[33] Mumtaz, H. and P. Surico, 2009. The Transmission of International Shocks: A Factor-Augmented VAR Approach, *Journal of Money, Credit and Banking*, Vol. 41, Supplement 1, 71-100.

[34] Ritschl, A. and S. Sarferaz, 2014. Currency versus Banking in the Financial Crisis of 1931, *International Economic Review*, Vol. 55, No. 2, 349-373.

[35] Sargent, T.J. and C.A. Sims, 1977. Business Cycle Modelling without Pretending to Have Too Much a-priori Economic Theory, in: Sims *et al.*, eds., New Methods in Business Cycle Research (Federal Reserve Bank of Minneapolis, Minneapolis).

[36] Sims, C. and T. Zha, 1998. Bayesian Methods for Dynamic Multivariate Models, *International Economic Review* 39(4), 949-968.

[37] Stock, J. and M. Watson, 2002. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association*, 97, 1167–1179.

[38] Stock, J. and M. Watson, 2005. Implications of Dynamic Factor Models for VAR Analysis, NBER Working Paper No. 11467.

[39] White, H. 1984. Asymptotic Theory for Econometricians, New York: Academic.

Table 1: RMSE ratios between estimators under over- and just-identifying restrictions for $m = 4$

	$\hat{\Theta}\hat{A}_q$				
	(k, ℓ)				
N, T	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.668	0.472	0.517	0.389	0.403
125, 500	0.620	0.403	0.425	0.313	0.321
250, 250	0.671	0.491	0.498	0.383	0.393
250, 500	0.602	0.402	0.419	0.305	0.317

	$\hat{\Theta}\hat{A}_{1:m}\tilde{A}_m$				
	(k, ℓ)				
N, T	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.700	0.691	0.570	0.684	0.545
125, 500	0.645	0.632	0.464	0.616	0.432
250, 250	0.711	0.704	0.548	0.697	0.533
250, 500	0.640	0.632	0.463	0.627	0.440

Notes: the ratios between the RMSEs of estimators under all of the available over-identifying restrictions and the RMSEs of estimators under just-identifying restrictions. $\hat{\Theta}\hat{A}_q$ is an estimator for the q-th column of Γ, which is the contemporaneous IRFs with respect to the q-th structural shock. $\hat{\Theta}\hat{A}_{1:m}\tilde{A}_m$ is an estimator for the m-th column of Γ, which is the contemporaneous IRFs with respect to the m-th structural shock.

Table 2: Size properties of the estimated IRFs when $m = 4$

	$\hat{\Theta}_1 \hat{A}_q$					$\hat{\Theta}_1 \hat{A}_{1:m} \tilde{A}_m$				
	(k, ℓ)					(k, ℓ)				
$N,\ T$	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.045	0.058	0.057	0.063	0.062	0.064	0.066	0.067	0.071	0.070
125, 500	0.057	0.065	0.065	0.068	0.068	0.070	0.072	0.072	0.075	0.075
250, 250	0.038	0.050	0.047	0.052	0.052	0.048	0.050	0.053	0.054	0.055
250, 500	0.049	0.055	0.055	0.055	0.055	0.051	0.054	0.053	0.057	0.054

	$\hat{\theta}_i' \hat{A}_q$					$\hat{\theta}_i' \hat{A}_{1:m} \tilde{A}_m$				
	(k, ℓ)					(k, ℓ)				
$N,\ T$	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.092	0.082	0.087	0.081	0.085	0.083	0.083	0.081	0.085	0.081
125, 500	0.090	0.090	0.088	0.082	0.084	0.076	0.075	0.079	0.079	0.079
250, 250	0.081	0.068	0.074	0.063	0.068	0.065	0.067	0.069	0.067	0.072
250, 500	0.077	0.066	0.073	0.064	0.073	0.071	0.070	0.067	0.072	0.069

	$\hat{\lambda}_i' \hat{\Psi}_s \hat{G} \hat{A}_q$					$\hat{\lambda}_i' \hat{\Psi}_s \hat{G} \hat{A}_{1:m} \tilde{A}_m$				
	(k, ℓ)					(k, ℓ)				
$N,\ T$	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.063	0.064	0.057	0.060	0.058	0.061	0.060	0.059	0.061	0.059
125, 500	0.064	0.061	0.061	0.062	0.062	0.063	0.063	0.062	0.065	0.064
250, 250	0.056	0.054	0.055	0.053	0.053	0.049	0.052	0.052	0.053	0.051
250, 500	0.065	0.061	0.059	0.061	0.060	0.062	0.062	0.060	0.065	0.063

Notes: The normal size is 5%. $\hat{\Theta}_1 \hat{A}_q$ and $\hat{\Theta}_1 \hat{A}_{1:m} \tilde{A}_m$ are the contemporaneous IRFs of the first k variables with respect to the q-th shock and m-th shock, respectively. $\hat{\theta}_i' \hat{A}_q$ and $\hat{\theta}_i' \hat{A}_{1:m} \tilde{A}_m$ are the contemporaneous IRFs of the i-th variable $(i > k)$ with respect to the q-th shock and m-th shock, respectively. $\hat{\lambda}_i' \hat{\Psi}_s \hat{G} \hat{A}_q$ and $\hat{\lambda}_i' \hat{\Psi}_s \hat{G} \hat{A}_{1:m} \tilde{A}_m$ are the dynamic IRFs of the i-th variable with respect to the q-th and m-th shock, respectively. The horizon s is set equal to 5.

Table 3: Size properties of the estimated IRFs when $m = 3$

	$\hat{\Theta}_1 \hat{A}_{1:m} \tilde{A}_m$				
	(k, ℓ)				
$N,\ T$	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.077	0.073	0.078	0.072	0.075
125, 500	0.083	0.080	0.082	0.081	0.081
250, 250	0.067	0.062	0.062	0.062	0.060
250, 500	0.063	0.062	0.063	0.061	0.061

	$\hat{\theta}_i' \hat{A}_{1:m} \tilde{A}_m$				
	(k, ℓ)				
$N,\ T$	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.110	0.101	0.106	0.094	0.099
125, 500	0.117	0.099	0.107	0.099	0.103
250, 250	0.086	0.085	0.094	0.082	0.085
250, 500	0.089	0.079	0.083	0.069	0.074

	$\hat{\lambda}_i' \hat{\Psi}_s \hat{G} \hat{A}_{1:m} \tilde{A}_m$				
	(k, ℓ)				
$N,\ T$	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.062	0.059	0.064	0.058	0.061
125, 500	0.067	0.065	0.065	0.065	0.064
250, 250	0.056	0.055	0.056	0.054	0.055
250, 500	0.057	0.055	0.057	0.055	0.057

Notes: The normal size is 5%. $\hat{\Theta}_1 \hat{A}_{1:m} \tilde{A}_m$ is the contemporaneous IRF of the first k variables with respect to the m-th shock. $\hat{\theta}_i' \hat{A}_{1:m} \tilde{A}_m$ is the contemporaneous IRF of the i-th variable ($i > k$) with respect to the m-th shock. $\hat{\lambda}_i' \hat{\Psi}_s \hat{G} \hat{A}_{1:m} \tilde{A}_m$ is the dynamic IRF of the i-th variable with respect to the m-th shock. The horizon s is set equal to 5.

Table 4: Size properties of the tests for over-identifying restrictions specified in (2.17) and (2.21)

	w_m					w_m^{joint}				
$m = 4$	(k, ℓ)					(k, ℓ)				
$N,\ T$	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.045	0.060	0.059	0.073	0.073	0.052	0.060	0.070	0.072	0.079
125, 500	0.055	0.068	0.066	0.076	0.080	0.062	0.069	0.075	0.079	0.086
250, 250	0.038	0.050	0.047	0.056	0.055	0.039	0.048	0.053	0.054	0.057
250, 500	0.046	0.054	0.055	0.057	0.060	0.051	0.052	0.059	0.057	0.063

	w_m					w_m^{joint}				
$m = 3$	(k, ℓ)					(k, ℓ)				
$N,\ T$	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)	(5, 4)	(6, 4)	(6, 5)	(7, 4)	(7, 5)
125, 250	0.072	0.088	0.084	0.097	0.099	0.083	0.093	0.100	0.103	0.106
125, 500	0.081	0.089	0.091	0.104	0.100	0.089	0.095	0.103	0.110	0.110
250, 250	0.049	0.059	0.062	0.072	0.069	0.060	0.065	0.069	0.070	0.074
250, 500	0.058	0.063	0.068	0.071	0.073	0.061	0.066	0.071	0.073	0.075

Notes: The normal size is 5%. w_m tests the null hypothesis $\mathbb{H}_0^{(m)}$: $\text{rank}(\Theta_1) = m$, and w_m^{joint} tests the joint null hypothesis $\mathbb{H}_0^{(m)'}$: $\text{rank}(\Theta_1) = m$ and $\text{rank}(\Theta_{11}) = m - 1$.

Table 5: p-values of tests for over-identifying restrictions

Joint $\mathbb{H}_0^{(d)'}$	p-value	$\text{rank}(\Theta_{11}) = d$	p-value	$\text{rank}(\Theta_1) = d$	p-value
$d = 2$	2×10^{-10}	$d = 2$	6×10^{-11}	$d = 3$	0.001
$d = 3$	0.276	$d = 3$	0.188	$d = 4$	0.502

Table 6: p-values of IRFs: over-identification versus just identification

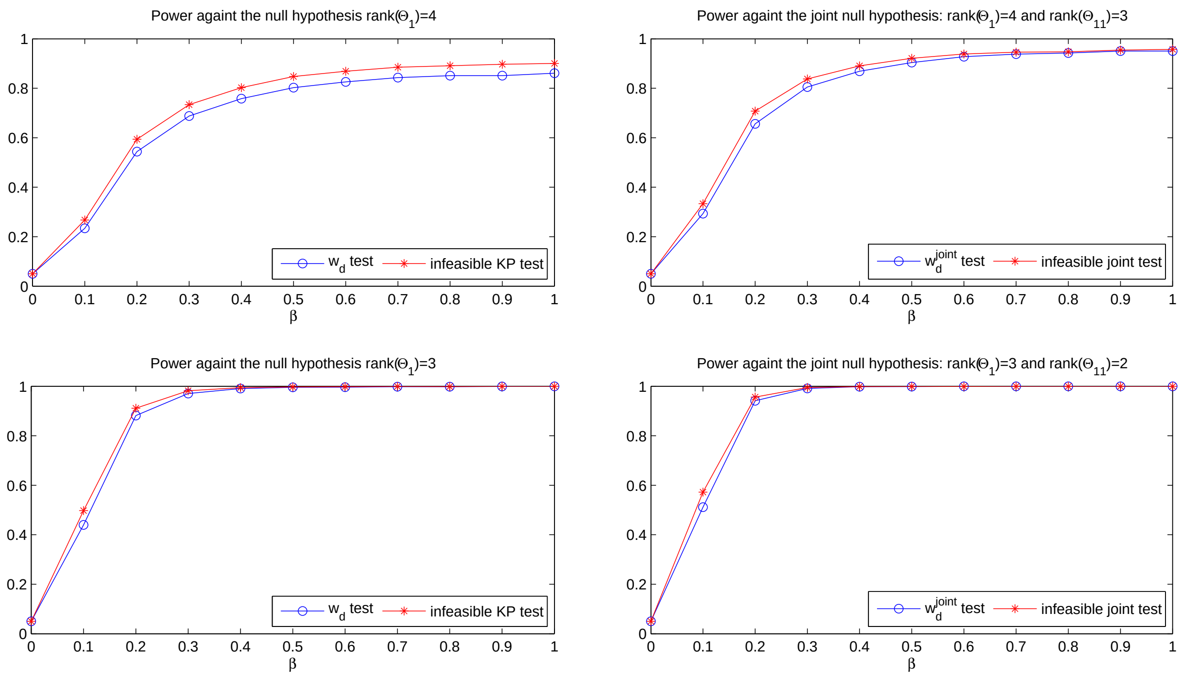
p-values of IRFs under just-identifying restrictions

mths after shock	0	1	2	3	6	9	12	18	24	30	36
FFR	0.002	0.019	0.067	0.119	0.265	0.398	0.528	0.762	0.938	0.942	0.863
IP	NA	0.265	0.604	0.896	0.641	0.454	0.371	0.339	0.365	0.411	0.467
CPI	NA	0.535	0.242	0.091	0.015	0.013	0.015	0.014	0.007	0.002	0.001
Consumption	0.091	0.005	0.003	0.002	0.002	0.002	0.002	0.003	0.004	0.006	0.006
Cap. Utilization	0.661	0.420	0.771	0.932	0.499	0.367	0.335	0.376	0.474	0.612	0.776
Unemployment	0.100	0.278	0.445	0.570	0.837	0.992	0.880	0.766	0.732	0.732	0.744
Employment	0.559	0.658	0.775	0.915	0.767	0.567	0.438	0.302	0.243	0.214	0.197
PCE Deflator	0.339	0.071	0.021	0.004	0.000	0.000	0.001	0.001	0.001	0.000	0.000
Earning	0.475	0.183	0.044	0.011	0.001	0.000	0.000	0.000	0.000	0.000	0.000
Housing Starts	0.981	0.219	0.058	0.024	0.007	0.005	0.004	0.004	0.005	0.005	0.006
Orders	0.922	0.466	0.294	0.160	0.040	0.019	0.014	0.013	0.018	0.029	0.047
Inventories	0.575	0.617	0.270	0.210	0.466	0.759	0.989	0.647	0.488	0.415	0.379
M2	0.752	0.263	0.069	0.026	0.004	0.002	0.001	0.000	0.000	0.000	0.000
Reserves	0.097	0.700	0.581	0.269	0.070	0.058	0.065	0.090	0.115	0.128	0.132
Consumer Credits	0.016	0.035	0.094	0.227	0.883	0.631	0.384	0.198	0.138	0.111	0.096
SP500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ex Rate Yen	0.105	0.072	0.107	0.118	0.148	0.185	0.226	0.310	0.387	0.450	0.498
PPI	0.015	0.057	0.060	0.040	0.018	0.016	0.018	0.015	0.009	0.004	0.001
Commodity Price	NA	0.088	0.129	0.128	0.110	0.104	0.108	0.130	0.159	0.199	0.250
Cons. Expectation	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.015	0.031

p-values of IRFs under over-identifying restrictions (the benchmark specification)

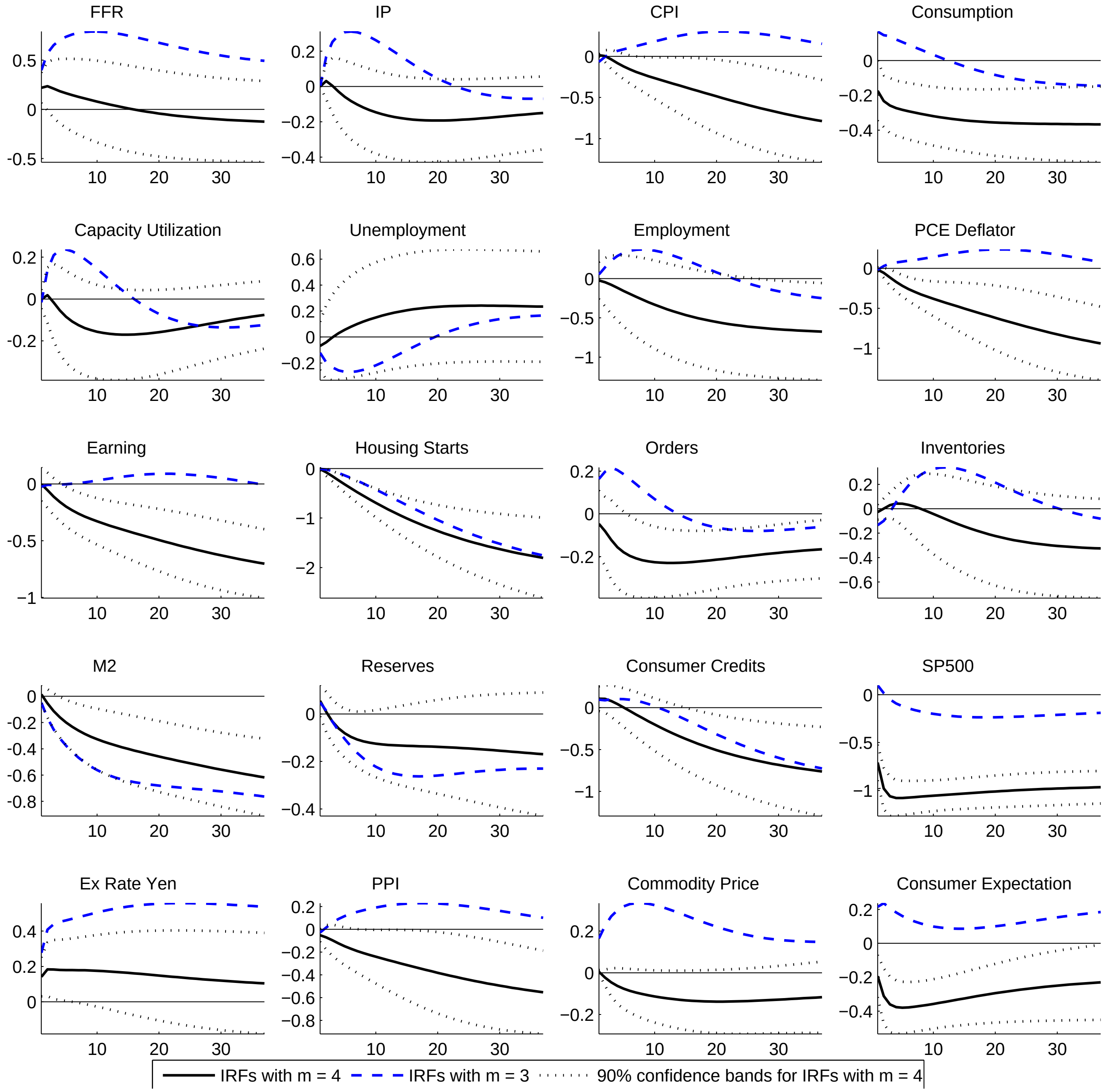
mths after shock	0	1	2	3	6	9	12	18	24	30	36
FFR	0.006	0.061	0.166	0.270	0.518	0.710	0.872	0.881	0.719	0.617	0.555
IP	0.255	0.554	0.961	0.755	0.371	0.224	0.154	0.109	0.109	0.128	0.155
CPI	0.255	0.914	0.542	0.243	0.050	0.040	0.042	0.035	0.019	0.007	0.002
Consumption	0.048	0.002	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.001
Cap. Utilization	0.971	0.788	0.862	0.610	0.278	0.171	0.130	0.122	0.160	0.241	0.358
Unemployment	0.475	0.781	0.987	0.874	0.627	0.486	0.395	0.304	0.275	0.272	0.280
Employment	0.841	0.770	0.677	0.577	0.371	0.246	0.167	0.086	0.054	0.039	0.032
PCE Deflator	0.255	0.079	0.028	0.006	0.000	0.001	0.001	0.003	0.002	0.000	0.000
Earning	0.981	0.497	0.185	0.065	0.006	0.002	0.001	0.000	0.000	0.000	0.000
Housing Starts	0.487	0.040	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Orders	0.553	0.327	0.200	0.109	0.024	0.008	0.004	0.002	0.003	0.008	0.018
Inventories	0.397	0.974	0.576	0.546	0.894	0.800	0.574	0.308	0.193	0.141	0.117
M2	0.706	0.321	0.099	0.043	0.011	0.006	0.004	0.001	0.000	0.000	0.000
Reserves	0.100	0.831	0.498	0.223	0.074	0.083	0.110	0.163	0.196	0.205	0.201
Consumer Credits	0.160	0.206	0.406	0.686	0.555	0.210	0.085	0.022	0.010	0.006	0.005
SP500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ex Rate Yen	0.012	0.017	0.032	0.038	0.057	0.090	0.133	0.235	0.333	0.413	0.473
PPI	0.082	0.196	0.170	0.112	0.051	0.045	0.046	0.039	0.023	0.010	0.003
Commodity Price	0.255	0.233	0.184	0.150	0.094	0.076	0.070	0.074	0.091	0.126	0.178
Cons. Expectation	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.020	0.041

Figure 1: Size-adjusted power of w_d and w_d^{joint}



Notes: The solid lines with circles denote the size-adjusted power of w_d and w_d^{joint}. The solid lines with asterisks denote the size-adjusted power of the infeasible KP test and joint rank test. β specifies the strength of the violation against the null hypothesis. The sample size $(N, T) = (125, 500)$. The upper-left (lower-left) panel shows the power of w_d and KP test against the null hypothesis that $\text{rank}(\Theta_1) = 4$ ($\text{rank}(\Theta_1) = 3$); the upper-right (lower-right) panel shows the power of w_d^{joint} and the infeasible joint test against the joint null hypothesis that $\text{rank}(\Theta_1) = 4$ and $\text{rank}(\Theta_{11}) = 3$ ($\text{rank}(\Theta_1) = 3$ and $\text{rank}(\Theta_{11}) = 2$).

Figure 2: Cumulative IRFs after a contractionary monetary policy shock
$m = 4$ versus $m = 3$



Notes: The solid curves are the cumulative IRFs after a contractionary monetary policy shock under the benchmark setup $m = 4$, which passes our specification tests. The dashed curves are the cumulative IRFs under $m = 3$, which does not pass our specification tests. The dotted lines are the 90% confidence bands for the cumulative IRFs with $m = 4$.