ESSAYS ON IMPULSE RESPONSE INFERENCE IN VECTOR AUTOREGRESSIVE MODELS

By

Zizheng Luo

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

Aug 11, 2023

Nashville, Tennessee

Approved:

Atsushi Inoue, Ph.D.

Lutz Kilian, Ph.D.

Tong Li, Ph.D.

Sarah Quincy, Ph.D.

ACKNOWLEDGMENTS

Words cannot fully express my gratitude towards my advisor, Dr. Atsushi Inoue, as this dissertation would be impossible without his guidance and advice. He gave me intellectual freedom in my work, consistently engaged me with new ideas, and demanded a high quality of work in all my endeavors. In addition to his guidance as an advisor, his devotion to academic advancement and attention to detail also sets a bar that I will always aim to reach in my work.

I would also like to acknowledge and give my warmest thanks to Dr. Tong Li, who not only offered me significant knowledge and insight in research, but also care and guidance in life, which was instrumental for me in supporting myself through living in a foreign land away from family. I am also extremely grateful to Dr. Lutz Kilian and Dr. Sarah Quincy for their thoughtful suggestion and comments throughout all stages of this dissertation.

Finally, I am extremely grateful to my family and friends, without whom I would not have been able to finish the dissertation. My parents have always offered me unconditional love and support, regardless of the situation. And my friends, especially my fellow PhD students in the economics department at Vanderbilt University, have also created a friendly and comfortable environment that I feel very lucky to live and work in.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

Uniform inference of impulse response in vector autoregressive models

1.1 Introduction

Impulse response analysis based on autoregressive models is an important tool in empirical macroeconomics to describe dynamic interactions between variables (Kilian and Lütkepohl (2017)). However, it is well known that statistics of autoregressive models with unit root or near unit roots are no longer asymptotically Gaussian. In fact, as discussed in Phillips (2014) and Mikusheva (2007), standard inference procedures on impulse responses might fail in the presence of near unit roots, while pointwise local to unity methods might fail as roots are far enough from unity. Such phenomenon calls for a uniform solution to inference of impulse responses, as discussed in Park and Phillips (1988), Park and Phillips (1989), Mikusheva (2007), Mikusheva (2012), Inoue and Kilian (2020) and Montiel Olea and Plagborg-Møller (2021).

While uniform inference methods in autoregressions has been studied in the aforementioned studies, the existing studies either focus on univariate autoregressions instead of vector autoregressions, or have trouble accommodating longer horizons. Mikusheva (2007) established uniform approximation to LR statistics and considered grid bootstrap methods in univariate autoregressive models, and Mikusheva (2012) extended the results to a class of vector autoregression models, although the discussion was restricted to existence of at most one unit root. VAR impulse response inference has been shown to be uniform at short horizons in Inoue and Kilian (2020), but the result breaks down at longer horizons. For local projections, Montiel Olea and Plagborg-Møller (2021) provides analogous results that remain valid at somewhat longer horizons, but also break down as horizons increase, as in many empirical studies.

In this chapter, we provide a more general result for VAR inference that applies to all horizons, in a wide class of vector autoregression models with potentially multiple unit roots. We suggest constructing and performing inference on impulse responses via a Wald test, similar to Lütkepohl et al. (2015). We show that when using lag augmentation, the Wald test has uniformly valid size, and thus the confidence intervals constructed by inverting the Wald test and calculating impulse responses has uniformly controlled size. A simulation study is also conducted to evaluate the finite sample performance of our method, and to compare it to that of some recent uniform inference methods, especially the lag-augmented local projection discussed in Montiel Olea and Plagborg-Møller (2021) and Xu (2023).

Our main contribution to the literature is twofold.

First, as mentioned above, we propose a method of estimation and inference of that is uniformly valid in a large class of vector autoregression models, allowing for multiple potential unit roots. The method is not only uniformly valid over the parameter space, but also across horizons. In addition, since our method is based on inverting a Wald statistic, in principle we could construct confidence interval with uniformly controlled coverage rate for not only impulse responses, but also any function of VAR parameters.

Second, we provide a simulation study involving vector autoregressions with various degrees of persistence that confirms the uniform validity of our method in finite sample. We also compare these results with uniform inference methods via local projections, which sheds light on the general comparison between local projection and VAR methods of impulse response inference. Our results show that in general the relative performance is largely dependent on the persistence level and horizon. Our method provides better coverage rate and shorter confidence intervals when the persistence level is very high or exact unit root exists. Meanwhile, our method also outperforms local projection methods at low persistence level unless the horizon is very short. On the contrary, local projection methods tend to perform well when the roots are mildly large, but not very close to 1.

The remainder of the paper is organized as follows: In section 1.2, we establish our model and assumptions. Section 1.3 establishes the uniform validity of the Wald statistic constructed after running a lag-augmented regression, and thus provides the foundation of our proposed inference method. Section 1.4 contains the simulation results that examines the accuracy of our method in finite sample and compares it to that of other methods in the literature. Finally, section 1.5 summarizes and concludes the chapter.

1.2 Model setup

1.2.1 Assumptions

Throughout this chapter, we consider the following model:

$$\begin{aligned} y_t &= Qy_t^\dagger + d^\dagger \\ (I - \Psi L)y_t^\dagger &= u_t \end{aligned} \tag{1.1}$$

$y_t, y_t^\dagger$ are N dimensional vectors, Q and Ψ are $N \times N$ matrices. We assume that Q is invertible and Psi is a diagonal matrix. u_t is a stationary process such that $\Phi(L)u_t = \varepsilon_t$ where $\Phi(L)$ is of order $p-1$. ε_t is an $N \times 1$ vector of martingale difference sequence.

Suppose $y_0^\dagger = y_{-1}^\dagger = \cdots = y_{-p+1}^\dagger = 0$. This setup implies a VAR(p) model for y_t:

$$\Pi(L)y_t = d + Q\varepsilon_t,\ \Pi(L) = Q\Phi(L)(I - \Psi L)Q^{-1},\ d = \Pi(1)d^\dagger \tag{1.2}$$

so that $QPhi(L)(I - \Psi L)Q^{-1}$ can be interpreted as a decomposition of the original lag polynomial $Pi(L)$.

We present the details of our assumption on the parameter space in the following Assumption 1.1:

Assumption 1.1. *Assume that*

$$E\left\{\begin{bmatrix} \varepsilon_t \\ vech(\varepsilon_t\varepsilon_t' - \Sigma) \end{bmatrix}\begin{bmatrix} \varepsilon_t \\ vech(\varepsilon_t\varepsilon_t' - \Sigma) \end{bmatrix}'\right\} = \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma_4 \end{bmatrix} \tag{1.3}$$

and $E[\|\varepsilon_t\|^6] < K$ *for some constant K. We consider the following parameter space:*

$$\Theta = \Big\{ \Big(diag(\Psi)', vec(Q)', vec(\Phi)', d', vech(\Sigma), vech(\Sigma_4) \Big)' : \\ \Psi = diag(\psi_1, \ldots, \psi_N), \\ Q \in \mathscr{Q} \subset \mathbb{R}^{N \times N}, \\ \Phi = (\Phi_1, \ldots, \Phi_{p-1}) \in \mathscr{P} \subset \mathbb{R}^{N \times (Np-N)}, \\ d \in \mathscr{D} \subset R^N, \\ \Sigma \in \mathscr{S} \subset \mathbb{R}^{N \times N}, \Sigma_4 \in \mathscr{S}_4 \subset \mathbb{R}^{\frac{N(N+1)}{2} \times \frac{N(N+1)}{2}} \Big\} \tag{1.4}$$

where $\mathscr{Q}, \mathscr{P}, \mathscr{D}, \mathscr{S}, \mathscr{S}_4$ *are compact sets, and there exists constants* $K_2 \geq K_1 > 0$ *and* $\delta \in (0,1)$ *such that*

(1) $\psi_j \in [-\delta, 1]$ *for all* $j = \{1, 2, \ldots, N\}$

(2) Σ *is positive definite and any eigenvalue* λ_Σ *of* Σ *satisfies* $K_1 \leq \lambda_\Sigma \leq K_2$

(3) Σ_4 *is positive definite and any eigenvalue* λ_{Σ_4} *of* Σ_4 *satisfies* $K_1 \leq \lambda_{\Sigma_4} \leq K_2$

(4) For any $Q \in \mathscr{Q}$*, Q is invertible and any eigenvalue* λ_Q *of Q satisfies* $K_1 \leq |\lambda_Q| \leq K_2$

(5) For any $(\Phi_1, \ldots, \Phi_{p-1}) \in \mathscr{P}$*,* $D(L) = \sum_{j=0}^{\infty} D_j L^j := \Phi(L)^{-1}$ *exists and* $\|D_j\| \leq K_2 \delta^j$ *for all j*

The assumption of diagonality on Ψ is similar to that in Elliott (1998), Mikusheva (2012) and Montiel Olea and Plagborg-Møller (2021). This assumption rules out explosive processes or processes with integration order larger than 1, but includes all stationary processes, and many I(1) or near I(1) processes. Compared to the literature, our inclusion of an arbitrary invertible matrix Q further expands the parameter space we cover. However, we do restrict each process $y_{1,t}, \ldots, y_{N,t}$ to have at most one unit root, as in Mikusheva (2012), Inoue and Kilian (2020) and Montiel Olea and Plagborg-Møller (2021).

1.2.2 Lag-augmented regression

To obtain a uniformly valid Wald statistic, we utilize lag augmentation and augment the model by one lag.

Let $\Pi = [-\Pi_1, \ldots, -\Pi_p]$ denote the slope coefficient estimates used to construct the Wald statistic, and let $x_t^* = [1, y'_{t-1}, \ldots, y'_{t-p}, y'_{t-p-1}]'$ denote the augmented regressors which includes one additional lag. By the Frisch-Waugh theorem, let $x_{1,t}^* = [1, y'_{t-p-1}]'$ and $x_{2,t}^* = [y'_{t-1}, \ldots, y'_{t-p}]'$, we know that the estimator for Π, $\hat{\Pi}$ can be written as:

$$\begin{aligned}\hat{\Pi} - \Pi &= \left(\sum_{t=1}^{T} Q\varepsilon_t x_{2,t}^{*'}\right)\left[\sum_{t=1}^{T} x_{2,t}^* x_{2,t}^{*'} - \left(\sum_{t=1}^{T} x_{2,t}^* x_{1,t}^{*'}\right)\left(\sum_{t=1}^{T} x_{1,t}^* x_{1,t}^{*'}\right)^{-1}\left(\sum_{t=1}^{T} x_{1,t}^* x_{2,t}^{*'}\right)\right]^{-1} \\ &\equiv \left(\sum_{t=1}^{T} Q\varepsilon_t x_{2,t}^{*'}\right) M_x^*\end{aligned} \tag{1.5}$$

As a theoretical device, similar to Toda and Yamamoto (1995), let $x_t = [1, \Delta y_{t-1}^{\dagger'}, \ldots, \Delta y_{t-p}^{\dagger'}, y_{t-p-1}^{\dagger'}]'$. Then $x_{1,t}^*$ and $x_{2,t}^*$ can be written as

$$x_{1,t}^* = \begin{bmatrix} 1 & 0_{1\times Np} \ldots & 0_{1\times N} \\ d^\dagger & 0_{N\times Np} & Q \end{bmatrix} x_t \equiv P_1 x_t \tag{1.6}$$

where P_1 is a $(N+1)\times(Np+N+1)$ matrix, and

$$x_{2,t}^* = \begin{bmatrix} d^\dagger & Q & Q & \ldots & Q & Q \\ d^\dagger & 0 & Q & \ldots & Q & Q \\ \ddots & \ddots & \ddots & \vdots & \ddots & \ddots \\ d^\dagger & 0 & 0 & \ldots & 0 & Q \end{bmatrix} x_t \equiv P_2 x_t \tag{1.7}$$

where P_2 is an $Np\times(Np+N+1)$ matrix. These notations will be helpful for deriving our theoretical results in section 1.3. Appendix A includes additional notations used throughout this chapter.

1.3 Main results

Eventually, we seek to perform inference on the structural impulse response functions, which we achieve by inverting the Wald test for candidate parameters and using these to construct confidence sets for impulse responses. To establish the uniform validity of such method, we follow Andrews et al. (2020), which provides a framework of obtaining uniform size through pointwise asymptotics under specific parameter sequences.

1.3.1 Theoretical device and additional assumptions

Following Andrews et al. (2020), we consider sequences $\{\theta_T\} \subset \Theta$ such that

$$\mathbf{c}_T(\theta_T) = \begin{bmatrix} T(1-\psi_{1,T}) \\ T(1-\psi_{2,T}) \\ \vdots \\ T(1-\psi_{N,T}) \\ \theta_T \end{bmatrix} \to \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \\ \theta \end{bmatrix} \in [0,+\infty]^N \times \Theta \tag{1.8}$$

By investigating the pointwise asymptotic of statistics under these parameter sequences, we can obtain the asymptotic size of the Wald test, and by extension the confidence sets of impulse responses, in a uniform sense.

Without lost of generality, for a given parameter sequence satisfying (1.8), we may assume that there exists an integer $1 \le k \le N$ such that $c_j = +\infty$ for the for all $j \le N-k$, and $c_j \in [0,+\infty)$ for all $j > N-k$. Such a sequence can be interpreted as the last k roots being near or exact unit root and the first $N-k$ roots being stationary.

Corresponding to the above characterization, let $y^{\dagger}_{1,t}$ be a vector of the first $N-k$ elements of $y^{\dagger}_t$ and $y^{\dagger}_{2,t}$ a vector of the rest so that $y^{\dagger}_t = (y^{\dagger\prime}_{1,t}, y^{\dagger\prime}_{2,t})'$, then we have

$$x_t = \begin{bmatrix} 1 \\ x_{3,t} \\ y^{\dagger}_{1,t-p-1} \\ y^{\dagger}_{2,t-p-1} \end{bmatrix} \tag{1.9}$$

where all the first differences are collected into $x_{3,t}$.

Now we introduce some additional notations. Let $C = diag(c_1,\ldots,c_N)$. In order to uniformly bound the decaying rate of moving average coefficients, let $\psi^*_j = \max\left\{|\psi_j|, \frac{1+\delta}{2}\right\}$ and $\Psi^* = diag(\psi^*_1,\ldots,\psi^*_N)$. This ensures that power of roots ψ_j are by construction bounded by the same power of the corresponding ψ^*_j.

Corresponding to x_t, we partition the matrices C, Ψ and Ψ^* as

$$C = \begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix}, \quad \Psi = \begin{bmatrix} \Psi_1 & 0 \\ 0 & \Psi_2 \end{bmatrix}, \quad \Psi^* = \begin{bmatrix} \Psi_1^* & 0 \\ 0 & \Psi_2^* \end{bmatrix} \tag{1.10}$$

where C_2, Ψ_2, Ψ_2^* are $k \times k$.

Let $\varepsilon_t^* = Q\varepsilon_t$, let $\Sigma^* = Q\Sigma Q'$ denote the variance of ε_t^*, and let

$$\Sigma_4^* = L_N(Q \otimes Q)D_N \Sigma_4 D_N'(Q' \otimes Q')L_N' \tag{1.11}$$

where L_N and D_N are elimination and duplication matrices defined in Appendix A.1.

In multiple occasions throughout the proof, we need to focus on specific rows/columns/elements of a matrix. To that end, let e_j denote a row vector of proper length with its j-th element being 1 and all other elements being 0.

Let Λ be a normalizing matrix:

$$\Lambda = \begin{bmatrix} \frac{1}{\sqrt{T}} I_{Np+1} & 0 \\ 0 & \Lambda^\dagger \end{bmatrix} \tag{1.12}$$

where

$$\begin{aligned} \Lambda^\dagger &= \begin{bmatrix} \max\left\{\frac{1}{T}, \frac{1}{\sqrt{T}}(1-\psi_1^{*2})^{1/2}\right\} & 0 & \dots & 0 \\ 0 & \max\left\{\frac{1}{T}, \frac{1}{\sqrt{T}}(1-\psi_2^{*2})^{1/2}\right\} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \max\left\{\frac{1}{T}, \frac{1}{\sqrt{T}}(1-\psi_N^{*2})^{1/2}\right\} \end{bmatrix} \\ &= \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} \end{aligned} \tag{1.13}$$

so that Λ_1, Λ_2 are $(N-k)\times(N-k)$ and $k \times k$ diagonal matrices respectively.

Let U_k be a diagonal matrix whose last k diagonal elements are 1 and the rest are 0, and let

$$J_{C_2}(x) = \int_0^x e^{(s-x)C_2} U_k D(1) dB(s) \tag{1.14}$$

where $B(s)$ is a Brownian motion with variance Σ^*. (Recall that $D(L) = \Phi(L)^{-1}$)

In addition to the model assumptions in previous section, we impose the following technical assumption so that the least square estimator is always uniformly well-defined:

Assumption 1.2.

$$\lim_{K\to\infty} \lim_{T\to\infty} \inf_{\theta\in\Theta} P\left(\lambda_{\min}\left(E\left[\sum_{t=1}^{T} \Lambda x_t x_t' \Lambda'\right]\right) \geq 1/K\right) = 1 \tag{1.15}$$

This assumption is similar in spirit to Assumption 3 in Montiel Olea and Plagborg-Møller (2021) and Assumption 4 in Xu (2023). For a more detailed discussion on the necessity and verification of this assumption, see Montiel Olea and Plagborg-Møller (2021).

1.3.2 Asymptotics under drifting parameter sequences

In this section, we present Lemmas 1.1, 1.2, 1.3 and 1.4, which establish the asymptotic distribution of the relevant statistics under an arbitrary parameter sequence satisfying (1.8). Lemma 1.1 establishes the convergence of $\sum_{t=1}^{T} x_t x_t'$ at the appropriate rate.

Lemma 1.1. *For sequences $\{\theta_T\} \subset \Theta$ satisfying (1.8), suppose there exists $k \in \{1,\ldots,N\}$ such that $c_j \in [0,+\infty)$ for all $j > N-k$, and $c_j = +\infty$ for all $j \leq N-k$, then*

$$\Lambda \sum_{t=1}^{T} x_t x_t' \Lambda' - \begin{bmatrix} 1 & 0 & 0 & \int_0^1 J_{C_2}(x)' dx \\ 0 & E[\frac{1}{T}\sum_{t=1}^{T} x_{3,t} x_{3,t}'] & E[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} x_{3,t} y_{1,t-p-1}^{\dagger'} \Lambda_1'] & 0 \\ 0 & E[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \Lambda_1 y_{1,t-p-1}^{\dagger} x_{3,t}'] & E[\Lambda_1 \sum_{t=1}^{T} y_{1,t-p-1}^{\dagger} y_{1,t-p-1}^{\dagger'} \Lambda_1'] & 0 \\ \int_0^1 J_{C_2}(x) dx & 0 & 0 & \int_0^1 J_{C_2}(x) J_{C_2}(x)' dx \end{bmatrix} \xrightarrow{p} 0 \tag{1.16}$$

Proof. We obtain the result following a block by block approach. The $(1,1)$ block is straightforward. The $(4,4)$ block and $(1,4)$ and $(4,1)$ blocks follow from Lemma 1(iii) in Elliott (1998).

(i) We start from the $(3,3)$ block. First, write $y_t^\dagger$ as

$$\begin{aligned} y_t^\dagger &= \sum_{j=0}^{t} \Psi^j u_{t-j} = \sum_{j=0}^{t} \Psi^j \left(\sum_{l=0}^{t-j} D_k \varepsilon_{t-j-j} \right) \\ &= \sum_{j=0}^{t} \left(\sum_{l=0}^{j} \Psi^{j-k} D_l \right) \varepsilon_{t-j} \equiv \sum_{j=0}^{t} F_j \varepsilon_{t-j} \end{aligned} \tag{1.17}$$

We have

$$\begin{aligned} \|e_m F_j\| &\le \sum_{l=0}^{j} \|e_m \Psi^{j-l}\| \|D_l\| \le \sum_{l=0}^{j} |\psi_m|^{j-l} \|D_l\| \\ &\le Const \cdot \psi_m^{*j} \sum_{l=0}^{j} \frac{\delta^l}{\psi_m^{*l}} \le Const \cdot \psi_m^{*j} \end{aligned} \tag{1.18}$$

The above equation provides us the basis to bound the moments of the relevant elements.

For $(3,3)$ block, we are dealing with "stationary" roots, or in other words, we are focusing on $m,n \le N-k$, for which we have $T(1-\psi_m^*), T(1-\psi_n^*) \to +\infty$. This guarantees that for T large enough, $\psi_m^*, \psi_n^* < 1$, so that $\frac{1}{1-\psi_m^*}$ and $\frac{1}{1-\psi_n^*}$ are well-defined, and

$$\begin{aligned} & \left| E\left[e_m y_t^\dagger y_t^{\dagger\prime} e_n' e_m y_{t+s}^\dagger y_{t+s}^{\dagger\prime} e_n' \right] - E\left[e_m y_t^\dagger y_t^{\dagger\prime} e_n' \right] E\left[e_m y_{t+s}^\dagger y_{t+s}^{\dagger\prime} e_n' \right] \right| \\ =& \left| \sum_{j_1=0}^{t} \sum_{j_2=0}^{t} \sum_{j_3=0}^{t+s} \sum_{j_4=0}^{t+s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon_{t+s-j_4}' F_{j_4}' e_n' \right] \right. \\ & \left. - E\left[\sum_{j_1=0}^{t} \sum_{j_2=0}^{t} e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon_{t-j_2}' F_{j_2}' e_n' \right] E\left[\sum_{j_3=0}^{t+s} \sum_{j_4=0}^{t+s} e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon_{t+s-j_4}' F_{j_4}' e_n' \right] \right| \\ \le& \left| \sum_{j_1=j_3-s\neq j_2=j_4-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon_{t+s-j_4}' F_{j_4}' e_n' \right] \right| \\ & + \left| \sum_{j_1=j_4-s\neq j_2=j_3-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon_{t+s-j_4}' F_{j_4}' e_n' \right] \right| \\ & + \left| \sum_{j_1=j_2=j_3-s=j_4-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon_{t+s-j_4}' F_{j_4}' e_n' \right] \right| \\ =& O\left(\sum_l \psi_m^{*2l+s} \sum_l \psi_n^{*2l+s} \right) + O\left(\sum_l \psi_m^{*l} \psi_n^{*l+s} \sum_l \psi_m^{*l+s} \psi_n^{*l} \right) + O\left(\sum_l \psi_m^{*2l+s} \psi_n^{*2l+s} \right) \\ =& O\left(\frac{\psi_m^{*s} \psi_n^{*s}}{(1-\psi_m^*)(1-\psi_n^*)} \right) + O\left(\frac{\psi_m^{*s} \psi_n^{*s}}{1-\psi_m^* \psi_n^*} \right) \end{aligned} \tag{1.19}$$

The above equation implies that

$$
\begin{aligned}
&\left|Var\left[\sum_{t=1}^{T} e_m \Lambda_1 y^{\dagger}_{1,t-p-1} y^{\dagger'}_{1,t-p-1} \Lambda_1' e_n'\right]\right| \\
\leq & \frac{2}{T}(1-\psi_m^{*2})(1-\psi_n^{*2}) \\
& \cdot \max_{1\leq t\leq T} \sum_{s=0}^{T-t+p+1} \left| Cov\left(e_m y^{\dagger}_{2,t-p-1} y^{\dagger'}_{2,t-p-1} e_n', e_m y^{\dagger}_{2,t+s-p-1} y^{\dagger'}_{2,t+s-p-1} e_n'\right)\right| \\
= & O\Big(T^{-1}\Big) \cdot O\Big((1-\psi_m^*)(1-\psi_n^*)\Big) \\
& \cdot \left(O\Big(\big[(1-\psi_m^*)(1-\psi_n^*)(1-\psi_m^*\psi_n^*)\big]^{-1}\Big) + O\Big((1-\psi_m^*\psi_n^*)^{-2}\Big)\right) \\
= & O\Big(T^{-1}(1-\psi_m^*)^{-1}\Big) + O(T^{-1}) = o(1)
\end{aligned} \tag{1.20}
$$

due to $T(1-\psi_m^*), T(1-\psi_n^*) \to +\infty$. Thus, by the multivariate Chebyshev inequality,

$$
P\left(\left\|vec\left(\Lambda_1 \sum_{t=1}^{T} y^{\dagger}_{1,t-p-1} y^{\dagger'}_{1,t-p-1} \Lambda_1'\right) - E\left[vec\left(\Lambda_1 \sum_{t=1}^{T} y^{\dagger}_{1,t-p-1} y^{\dagger'}_{1,t-p-1} \Lambda_1'\right)\right]\right\| > x\right) \to 0 \tag{1.21}
$$

(ii) Now consider the $(4,3)$ and $(3,4)$ blocks. We follow similar steps as in part (i). As argued in part (i), for $n < N-k$, $\psi_n^* < 1$ for large enough T, so similar to part (i), we can obtain

$$
\begin{aligned}
&\left|E\left[e_m y_t^\dagger y_t^{\dagger'} e_n' e_m y_{t+s}^\dagger y_{t+s}^{\dagger'} e_n'\right]\right| \\
=&\left|\sum_{j_1=0}^{t}\sum_{j_2=0}^{t}\sum_{j_3=0}^{t+s}\sum_{j_4=0}^{t+s} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j_4}' F_{j_4}' e_n'\right]\right| \\
\leq&\left|\sum_{j_1=j_3-s\neq j_2=j_4-s} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j_4}' F_{j_4}' e_n'\right]\right| \\
&+\left|\sum_{j_1=j_4-s\neq j_2=j_3-s} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j_4}' F_{j_4}' e_n'\right]\right| \\
&+\left|\sum_{j_1=j_2=j_3-s=j_4-s} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j_4}' F_{j_4}' e_n'\right]\right| \\
&+\left|\sum_{j_1=j_2\neq j_3-s=j_4-s} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j_2}' F_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j_4}' F_{j_4}' e_n'\right]\right| \\
=&O\left(\sum_l \psi_m^{*2l+s}\sum_l \psi_n^{*2l+s}\right)+O\left(\sum_l \psi_m^{*l}\psi_n^{*l+s}\sum_l \psi_m^{*l+s}\psi_n^{*l}\right) \\
&+O\left(\sum_l \psi_m^{*2l+s}\psi_n^{*2l+s}\right)+O\left((\sum_l \psi_m^{*l}\psi_n^{*l})^2\right) \\
=&O\left(\frac{T\psi_n^{*s}}{(1-\psi_n^*)}\right)+O\left(\frac{\psi_n^{*s}}{(1-\psi_n^*)^2}\right)+O\left(\frac{\psi_n^{*s}}{1-\psi_n^*}\right)+O\left(\frac{1}{(1-\psi_n^*)^2}\right)
\end{aligned}
\tag{1.22}
$$

where last two steps depend on $\psi_m^* \leq 1$ and for T large enough, $\psi_n^* < 1$.

Now notice that for large enough T, by construction, $\Lambda_2 = \frac{1}{T} I_k$. Then we have

$$
\begin{aligned}
&\left|E\left[\left(\frac{1}{T}\sum_{t=1}^{T} e_m y_{2,t-p-1}^\dagger y_{1,t-p-1}^{\dagger'}\Lambda_1' e_n'\right)^2\right]\right| \\
\leq&\frac{2}{T^2}(1-\psi_n^{*2})\max_{1\leq t\leq T}\sum_{s=0}^{T-t+1}\left|E\left(e_m y_{2,t-p-1}^\dagger y_{1,t-p-1}^{\dagger'} e_n' e_m y_{2,t+s-p-1}^\dagger y_{1,t+s-p-1}^{\dagger'} e_n'\right)\right| \\
=&O\Big(T^{-2}\Big)\cdot O\Big(1-\psi_n^*\Big) \\
&\cdot\left[O\left(\frac{T\psi_n^{*s}}{(1-\psi_n^*)}\right)+O\left(\frac{\psi_n^{*s}}{(1-\psi_n^*)^2}\right)+O\left(\frac{\psi_n^{*s}}{1-\psi_n^*}\right)+O\left(\frac{1}{(1-\psi_n^*)^2}\right)\right] \\
=&O\left(\frac{1}{T(1-\psi_n^*)}\right)+O\left(\frac{1}{T^2(1-\psi_n^*)^2}\right)+O\left(\frac{1}{T^2(1-\psi_n^*)}\right)+O\left(\frac{1}{T}\right) \\
=&o(1)
\end{aligned}
\tag{1.23}
$$

by $T(1-\psi_n^*) \to +\infty$ for $n \leq N-k$.

Thus, by the multivariate Chebyshev inequality,

$$P\left(\left\|vec\left(\Lambda_2\sum_{t=1}^{T}y^{\dagger}_{2,t-p-1}y^{\dagger'}_{1,t-p-1}\Lambda_1'\right)\right\|>x\right)\to 0 \tag{1.24}$$

(iii) Now consider the $(4,2)$ and $(2,4)$ blocks. First notice that for the difference terms,

$$\Delta y^{\dagger}_t = y^{\dagger}_t - y^{\dagger}_{t-1} = \sum_{j=1}^{t} G_j\varepsilon_{t-j} \tag{1.25}$$

where $G_j = F_j - F_{j-1}$ for $j\geq 1$ and $G_0 = F_0 = I$. For any $m = 1,\ldots,N$,

$$\begin{aligned}
\|e_mG_j\| &= \|\sum_{l=0}^{j}e_m\Psi^{j-l}D_l - \sum_{l=0}^{j-1}e_m\Psi^{j-l-1}D_l\| \\
&= \|e_mD_j + e_m(\Psi-I)\sum_{l=0}^{j-1}\Psi^{j-l-1}D_l\| \\
&\leq \|D_j\| + (1-\psi_m)\sum_{l=0}^{j}\psi_m^{j-l}\|D_l\| \\
&\leq Const\cdot\left(\delta^j + (1-\psi_m)\psi_m^{*j}\sum_{l=0}^{j}\frac{\delta^l}{\psi_m^{*l}}\right) \\
&\leq Const\cdot\left(\delta^j + (1-\psi_m)\psi_m^{*j}\right)
\end{aligned} \tag{1.26}$$

Then for $1 \leq j \leq p$,

$$
\begin{aligned}
&\left\| E\left[e_m y^{\dagger}_{t-p-1} \Delta y^{\dagger\prime}_{t-j} e_n' e_m y^{\dagger}_{t-p-1+s} \Delta y^{\dagger\prime}_{t-j+s} e_n'\right] \right\| \\
=&\left\| \sum_{j_1=0}^{t-p-1} \sum_{j_2=0}^{t-j} \sum_{j_3=0}^{t-p-1+s} \sum_{j_4=0}^{t-j+s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n'\right] \right\| \\
\leq &\left\| \sum_{j_1=j_2-(p-j+1)\neq j_3-s=j_4-(p-j+1)-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n'\right] \right\| \\
&+ \left\| \sum_{j_1=j_3-s\neq j_2-(p-j+1)+j=j_4-(p-j+1)-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n'\right] \right\| \\
&+ \left\| \sum_{j_1=j_4-(p-j+1)-s\neq j_2-(p-j+1)=j_3-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n'\right] \right\| \\
&+ \left\| \sum_{j_1=j_2-(p-j+1)=j_3-s=j_4-(p-j+1)-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n'\right] \right\| \\
=&O\left(\left[\sum_l \psi_m^{*l} \left((1-\psi_n) \psi_n^{*l+(p-j+1)} + \delta^{l+(p-j+1)} \right) \right]^2 \right) \\
&+ O\left(\left[\sum_l \psi_m^{*2l+s} \right] \left[\sum_l \left((1-\psi_n) \psi_n^{*l} + \delta^l \right) \left((1-\psi_n) \psi_n^{*l+s} + \delta^{l+s} \right) \right] \right) \\
&+ O\left(\left[\sum_l \psi_m^{*l} \left((1-\psi_n) \psi_n^{*l+s+(p-j+1)} + \delta^{l+s+(p-j+1)} \right) \right] \right. \\
&\qquad \left. \cdot \left[\sum_l \psi_m^{*l+s} \left((1-\psi_n) \psi_n^{*l+(p-j+1)} + \delta^{l+(p-j+1)} \right) \right] \right) \\
&+ O\left(\sum_l \psi_m^{*2l+s} \left((1-\psi_n) \psi_n^{*l+(p-j+1)} + \delta^{l+(p-j+1)} \right) \left((1-\psi_n) \psi_n^{*l+s+(p-j+1)} + \delta^{l+s+(p-j+1)} \right) \right)
\end{aligned} \tag{1.27}
$$

Let

$$
\kappa(\psi_n^*) = \begin{cases} \frac{1}{1-\psi_n^*} & \text{if } \psi_n^* < 1 \\ T & \text{if } \psi_n^* = 1 \end{cases} \tag{1.28}
$$

Notice that if $N-k+1 \leq n \leq N$, since

$$
(1-\psi_n) \frac{1}{1-\frac{1+\delta}{2}} \leq \frac{2}{1-\frac{1+\delta}{2}} = \frac{4}{1-\delta} \tag{1.29}
$$

we have

$$
(1-\psi_n)\kappa(\psi_n^*) \leq \max\left\{ \frac{4}{1-\delta}, T(1-\psi_n) \right\} = O(1) \tag{1.30}
$$

Meanwhile, if $1 \leq n \leq N-k$, then for large enough T, $\psi^*_{n,T} < 1$ since $T(1-\psi_{n,T}) \to +\infty$. Thus, for large enough T,

$$(1-\psi_{n,T})\kappa(\psi^*_{n,T}) = \frac{1-\psi_{n,T}}{1-\psi^*_{n,T}} = O(1) \tag{1.31}$$

Therefore $(1-\psi_n)\kappa(\psi^*_n) = O(1)$. Back to (1.27), for the first term,

$$\begin{aligned}
&\Big(\sum_l \psi_m^{*l}\Big((1-\psi_n)\psi_n^{*l+(p-j+1)} + \delta^{l+(p-j+1)}\Big)\Big)^2 \\
\leq&\Big((1-\psi_n)\kappa(\psi^*_n) + \frac{1}{1-\delta}\Big)^2 = O(1)
\end{aligned} \tag{1.32}$$

Similarly, for the second term,

$$\begin{aligned}
&\Big(\sum_l \psi_m^{*2l+s}\Big)\Big(\sum_l \Big((1-\psi_n)\psi_n^{*l} + \delta^l\Big)\Big((1-\psi_n)\psi_n^{*l+s} + \delta^{l+s}\Big)\Big) \\
\leq&T\Big(\frac{\delta^s}{1-\delta^2} + \frac{(1-\psi_n)(\delta^s + \psi_n^{*s})}{1-\delta} + (1-\psi_n)^2\kappa(\psi^*_n)\psi_n^{*s}\Big) \\
=&O\Big(T\delta^s\Big) + O\Big(T(1-\psi_n)\psi_n^{*s}\Big)
\end{aligned} \tag{1.33}$$

the third term,

$$\begin{aligned}
&\Big(\sum_l \psi_m^{*l}\Big((1-\psi_n)\psi_n^{*l+s+(p-j+1)} + \delta^{l+s+(p-j+1)}\Big)\Big) \\
&\qquad\Big(\sum_l \psi_m^{*l+s}\Big((1-\psi_n)\psi_n^{*l+(p-j+1)} + \delta^{l+(p-j+1)}\Big)\Big) \\
\leq&\Big((1-\psi_n)\psi_n^{*s}\kappa(\psi^*_n) + \frac{\delta^s}{1-\delta}\Big)\Big((1-\psi_n)\kappa(\psi^*_n) + \frac{1}{1-\delta}\Big) \\
=&O(\psi_n^{*s}) + O(\delta^s)
\end{aligned} \tag{1.34}$$

and the fourth term

$$\begin{aligned}
&\sum_l \psi_m^{*2l+s}\Big((1-\psi_n)\psi_n^{*l+(p-j+1)} + \delta^{l+(p-j+1)}\Big)\Big((1-\psi_n)\psi_n^{*l+s+(p-j+1)} + \delta^{l+s+(p-j+1)}\Big) \\
\leq&(1-\psi_n)^2\psi_n^{*s}\kappa(\psi^*_n) + \frac{1-\psi_n}{1-\delta}(\delta^s + \psi_n^{*s}) + \frac{\delta^s}{1-\delta} \\
=&O((1-\psi_n)\psi_n^{*s}) + O(\delta^s)
\end{aligned} \tag{1.35}$$

Therefore for all m, n such that $N-k+1 \leq m \leq N$ and $1 \leq n \leq N-k$,

$$
\begin{aligned}
&\left\| E\left[\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} e_m y^{\dagger}_{2,t-p-1}\Delta y^{\dagger'}_{t-j} e'_n\right)^2\right]\right\| \\
\leq &\frac{1}{T^2}\max_t \sum_{s=-t}^{T-t}\left\| E\left[e_m y^{\dagger}_{t-p-1}\Delta y^{\dagger'}_{t-j} e'_n e_m y^{\dagger}_{t-p-1+s}\Delta y^{\dagger'}_{t-j+s} e'_n\right]\right\| \\
= &O\Big(T^{-2}\Big)\Big(O\Big(T\Big)+O\Big(\frac{T}{1-\delta}\Big)+O\Big(T(1-\psi_n)\kappa(\psi^*_n)\Big) \\
&+O\Big(\kappa(\psi^*_n)\Big)+O\Big(\frac{1}{1-\delta}\Big)+O\Big((1-\psi_n)\psi^{*s}_n\Big)+O(\frac{1}{1-\delta})\Big) \\
= &o(1)
\end{aligned}
\tag{1.36}
$$

Since $(1-\psi_n)\kappa(\psi^*_n) = O(1)$ and $T^{-1}\kappa(\psi^*_n) = O(1)$. Thus, by the multivariate Chebyshev inequality,

$$
P\left(\left\| vec\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y^{\dagger}_{2,t-p-1}\Delta y^{\dagger'}_{t-j}\right)\right\| > x\right) \to 0 \tag{1.37}
$$

which implies that

$$
P\left(\left\| vec\left(\frac{1}{T^{3/2}}\sum_{t=1}^{T} y^{\dagger}_{2,t-p-1} x'_{3,t}\right)\right\| > x\right) \to 0 \tag{1.38}
$$

(iv) The proofs for the $(3,2)$ and $(2,3)$ blocks are similar to part (iii). For $m \leq N-k$, we know that $\psi^*_{m,T} < 1$ for large enough T since $T(1-\psi_m) \to +\infty$. Hence, we have

$$\begin{aligned}
&\left| Cov\left(e_m y^{\dagger}_{t-p-1} \Delta y^{\dagger\prime}_{t-j} e_n', e_m y^{\dagger}_{t-p-1+s} \Delta y^{\dagger\prime}_{t-j+s} e_n' \right) \right| \\
=&\left| \sum_{j_1=0}^{t-p-1} \sum_{j_2=0}^{t-j} \sum_{j_3=0}^{t-p-1+s} \sum_{j_4=0}^{t-j+s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n' \right] \right. \\
&\left. - \sum_{j_1=0}^{t-p-1} \sum_{j_2=0}^{t-j} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' \right] \sum_{j_3=0}^{t-p-1+s} \sum_{j_4=0}^{t-j+s} E\left[e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n' \right] \right| \\
\leq& \left| \sum_{j_1=j_3-s\neq j_2-(p-j+1)=j_4-(p-j+1)-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n' \right] \right| \\
&+ \left| \sum_{j_1=j_4-(p-j+1)-s\neq j_2-(p-j+1)=j_3-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n' \right] \right| \\
&+ \left| \sum_{j_1=j_2-(p-j+1)=j_3-s=j_4-(p-j+1)-s} E\left[e_m F_{j_1} \varepsilon_{t-j_1} \varepsilon'_{t-j-j_2} G'_{j_2} e_n' e_m F_{j_3} \varepsilon_{t+s-j_3} \varepsilon'_{t+s-j-j_4} G'_{j_4} e_n' \right] \right| \\
=&O\Big(\frac{\delta^s}{1-\psi^*_m} + \frac{(1-\psi_n)\psi^{*s}_n}{1-\psi^*_m} \Big) + O\Big(\psi^{*s}_m \Big) + O\Big(\psi^{*s}_m + \delta^s \Big)
\end{aligned} \tag{1.39}$$

and therefore

$$\begin{aligned}
&\left\| Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} e_m \Lambda_2 y^{\dagger}_{2,t-1} \Delta y^{\dagger\prime}_{t-j} e_n' \right) \right\| \\
\leq& \frac{1-\psi^{*2}_m}{T} \max_t \sum_{s=-t}^{T-t} \left\| Cov\left(e_m y^{\dagger}_{t-1} \Delta y^{\dagger\prime}_{t-j} e_n', e_m y^{\dagger}_{t+s-1} \Delta y^{\dagger\prime}_{t+s-j} e_n' \right) \right\| \\
=&O\Big(T^{-1} \Big) \cdot O\Big(1-\psi^*_m \Big) \cdot \Big(O\Big(\frac{1}{1-\psi^*_m} \Big) + O\Big(\frac{1}{1-\psi^*_m} \Big) + O\Big(\frac{1}{1-\psi^*_m} + \frac{1}{1-\delta} \Big) \Big) \\
=&o(1)
\end{aligned} \tag{1.40}$$

for all m,n such that $1 \leq m \leq k$ and $1 \leq n \leq N$. Thus, by the multivariate Chebyshev inequality,

$$P\left(\left\| vec\left(\frac{1}{\sqrt{T}} \Lambda_1 \sum_{t=1}^{T} y^{\dagger}_{1,t-p-1} \Delta y^{\dagger\prime}_{t-j} \right) - E\left[vec\left(\frac{1}{\sqrt{T}} \Lambda_1 \sum_{t=1}^{T} y^{\dagger}_{1,t-p-1} \Delta y^{\dagger\prime}_{t-j} \right) \right] \right\| > x \right) \to 0 \tag{1.41}$$

which implies that

$$P\left(\left\| vec\left(\frac{1}{\sqrt{T}} \Lambda_1 \sum_{t=1}^{T} y^{\dagger}_{1,t-p-1} x'_{3t} \right) - E\left[vec\left(\frac{1}{\sqrt{T}} \Lambda_1 \sum_{t=1}^{T} y^{\dagger}_{1,t-p-1} x'_{3t} \right) \right] \right\| > x \right) \to 0 \tag{1.42}$$

(v) Now consider the $(2,2)$ block. Similar as previous blocks, we have

$$
\begin{aligned}
&\left|Cov\left(e_m\Delta y_t^{\dagger}\Delta y_{t-j}^{\dagger'}e_n', e_m\Delta y_{t+s}^{\dagger}\Delta y_{t+s-j}^{\dagger'}e_n'\right)\right| \\
=&\left|\sum_{j_1=0}^{t}\sum_{j_2=0}^{t-j}\sum_{j_3=0}^{t+s}\sum_{j_4=0}^{t+s-j} E\left[e_m G_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m G_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j-j_4}' G_{j_4}' e_n'\right]\right. \\
&\left. -\sum_{j_1=0}^{t}\sum_{j_2=0}^{t-j} E\left[e_m G_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n'\right]\sum_{j_3=0}^{t+s}\sum_{j_4=0}^{t+s-j} E\left[e_m G_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j-j_4}' G_{j_4}' e_n'\right]\right| \\
\leq&\left|\sum_{j_1=j_3-s\neq j_2+j=j_4+j-s} E\left[e_m G_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m G_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j-j_4}' G_{j_4}' e_n'\right]\right| \\
&+\left|\sum_{j_1=j_4+j-s\neq j_2+j=j_3-s} E\left[e_m G_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m G_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j-j_4}' G_{j_4}' e_n'\right]\right| \\
&+\left|\sum_{j_1=j_2+j=j_3-s=j_4+j-s} E\left[e_m G_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m G_{j_3}\varepsilon_{t+s-j_3}\varepsilon_{t+s-j-j_4}' G_{j_4}' e_n'\right]\right| \\
=&O\bigg(\Big[\sum_l(\delta^l+(1-\psi_m)\psi_m^{*l})(\delta^{l+s}+(1-\psi_m)\psi_m^{*l+s})\Big] \\
&\qquad\Big[\sum_l(\delta^l+(1-\psi_n)\psi_n^{*l})(\delta^{l+s}+(1-\psi_n)\psi_n^{*l+s})\Big]\bigg) \\
&+O\bigg(\Big[\sum_l(\delta^l+(1-\psi_m)\psi_m^{*l})(\delta^{l+s-j}+(1-\psi_n)\psi_n^{*l+s-j})\Big] \\
&\qquad\Big[\sum_l(\delta^l+(1-\psi_n)\psi_n^{*l})(\delta^{l+s}+(1-\psi_m)\psi_m^{*l+s})\Big]\bigg) \\
&+O\bigg(\sum_l\Big[(\delta^{l+j}+(1-\psi_m)\psi_m^{*l+j})(\delta^l+(1-\psi_n)\psi_n^{*l}) \\
&\qquad(\delta^{l+j+s}+(1-\psi_m)\psi_m^{*l+j+s})(\delta^{l+s}+(1-\psi_n)\psi_n^{*l+s})\Big]\bigg)
\end{aligned}
\tag{1.43}
$$

The first term can be bounded as follows:

$$
\begin{aligned}
&\Big[\sum_l(\delta^l+(1-\psi_m)\psi_m^{*l})(\delta^{l+s}+(1-\psi_m)\psi_m^{*l+s})\Big] \\
&\qquad\Big[\sum_l(\delta^l+(1-\psi_n)\psi_n^{*l})(\delta^{l+s}+(1-\psi_n)\psi_n^{*l+s})\Big] \\
\leq&\Big(\frac{\delta^s}{1-\delta}+\frac{(1-\psi_m)(\psi_m^{*s}+\delta^s)}{1-\delta}+(1-\psi_m)^2\psi_m^{*s}\kappa(\psi_m^*)\Big) \\
&\qquad\Big(\frac{\delta^s}{1-\delta}+\frac{(1-\psi_n)(\psi_n^{*s}+\delta^s)}{1-\delta}+(1-\psi_n)^2\psi_n^{*s}\kappa(\psi_n^*)\Big) \\
=&O\Big(\delta^s\Big)+O\Big((1-\psi_m)\psi_m^{*s}\Big)
\end{aligned}
\tag{1.44}
$$

Similarly, for the second term we have

$$
\begin{aligned}
&\Big[\sum_l (\delta^l + (1-\psi_m)\psi_m^{*l})(\delta^{l+s-j} + (1-\psi_n)\psi_n^{*l+s-j})\Big] \\
&\qquad\qquad \Big[\sum_l (\delta^l + (1-\psi_n)\psi_n^{*l})(\delta^{l+s} + (1-\psi_m)\psi_m^{*l+s})\Big] \\
\leq &\Big(\frac{\delta^{s-j}}{1-\delta} + \frac{(1-\psi_m)\delta^{s-j}}{1-\delta} + \frac{(1-\psi_n)\psi_n^{*s-j}}{1-\delta} + (1-\psi_m)(1-\psi_n)\psi_n^{*s-j}\kappa(\psi_m^*)\Big) \\
&\qquad\qquad \Big(\frac{\delta^s}{1-\delta} + \frac{(1-\psi_m)\psi_m^{*s}}{1-\delta} + \frac{(1-\psi_n)\delta^s}{1-\delta} + (1-\psi_m)(1-\psi_n)\psi_m^{*s}\kappa(\psi_n^*)\Big) \\
= &O\Big(\delta^s\Big) + O\Big((1-\psi_m)\psi_m^{*s}\Big) + O\Big((1-\psi_n)\psi_n^{*s}\Big)
\end{aligned}
\tag{1.45}
$$

and for the third term we have

$$
\begin{aligned}
&\sum_l \Big[(\delta^{l+j} + (1-\psi_m)\psi_m^{*l+j})(\delta^l + (1-\psi_n)\psi_n^{*l}) \\
&\qquad (\delta^{l+j+s} + (1-\psi_m)\psi_m^{*l+j+s})(\delta^{l+s} + (1-\psi_n)\psi_n^{*l+s})\Big] \\
= &\sum_l \Bigg[\Big(\delta^{2l+j} + (1-\psi_m)\psi_m^{*l+j}\delta^l + (1-\psi_n)\psi_n^{*l}\delta^{l+j} + (1-\psi_m)(1-\psi_n)\psi_m^{*l+j}\psi_n^{*l}\Big) \\
&\qquad \Big(\delta^{2l+j+s} + (1-\psi_m)\psi_m^{*l+j+s}\delta^{l+s} + \\
&\qquad (1-\psi_n)\psi_n^{*l+s}\delta^{l+j+s} + (1-\psi_m)(1-\psi_n)\psi_m^{*l+j+s}\psi_n^{*l+s}\Big)\Bigg] \\
\leq &\sum_l \Big[(3\delta^l + (1-\psi_m)(1-\psi_n)\psi_m^{*l}\psi_n^{*l})(3\delta^{l+s} + (1-\psi_m)(1-\psi_n)\psi_m^{*l+s}\psi_n^{*l+s})\Big] \\
= &O\Big(\delta^s\Big) + O\Big((1-\psi_m)\psi_m^{*s}\Big)
\end{aligned}
\tag{1.46}
$$

Therefore

$$
\begin{aligned}
&\left| Var\left(\frac{1}{T}\sum_{t=1}^{T} e_m \Delta y_{t-1}^{\dagger} \Delta y_{t-j}^{\dagger\prime} e_n^{\prime}\right)\right| \\
\leq &\frac{1}{T}\max_t \sum_{s=-t}^{T-t} \left\| Cov\left(e_m \Delta y_{t-1}^{\dagger}\Delta y_{t-j}^{\dagger\prime} e_n^{\prime}, e_m \Delta y_{t+s-1}^{\dagger}\Delta y_{t+s-j}^{\dagger\prime} e_n^{\prime}\right)\right\| \\
= &O\Big(T^{-1}\Big)\cdot\Big(O\Big(1\Big) + O\Big((1-\psi_m)\kappa(\psi_m^*)\Big) + O\Big((1-\psi_n)\kappa(\psi_n^*)\Big)\Big) \\
= &o(1)
\end{aligned}
\tag{1.47}
$$

for all m,n. Thus, by the multivariate Chebyshev inequality,

$$\sup_{\theta\in\Theta_{k,T}} P\left(\left\|vec\left(\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-1}^{\dagger}\Delta y_{t-j}^{\dagger'}\right)-E\left[vec\left(\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-1}^{\dagger}\Delta y_{t-j}^{\dagger'}\right)\right]\right\|>x\right)=o(1) \tag{1.48}$$

which implies that

$$\sup_{\theta\in\Theta_{k,T}} P\left(\left\|vec\left(\frac{1}{T}\sum_{t=1}^{T}x_{3,t}x_{3,t}'\right)-E\left[vec\left(\frac{1}{T}\sum_{t=1}^{T}x_{3,t}x_{3,t}'\right)\right]\right\|>x\right)=o(1) \tag{1.49}$$

(vi) Consider the $(1,3)$ and $(3,1)$ blocks. Since $T(1-\psi_m^*)\to+\infty$ for $1\le m\le N-k$, we have that $\psi_{m,T}^*<1$ for large enough T. Notice that

$$\begin{aligned}\left|E\left[e_m y_{1,t}^{\dagger}y_{1,t+s}^{\dagger'}e_m'\right]\right| \le & \left|\sum_{j_1=0}^{t}\sum_{j_2=0}^{t+s}E\left[e_mF_{j_1}\varepsilon_{t-j_1}\varepsilon_{t+s-j_2}'F_{j_2}'e_m'\right]\right| \\ =&O\left(\sum_{l}\psi_m^{*2l+s}\right)=O\left(\frac{\psi_m^{*s}}{1-\psi_m^*}\right)\end{aligned} \tag{1.50}$$

Then

$$\begin{aligned}&\left|E\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}e_m\Lambda_1 y_{1,t-p-1}^{\dagger}\right)^2\right]\right| \\ \le&\frac{1}{T}(1-\psi_m^{*2})\max_t\sum_{s=-t}^{T-t}\left|E\left[e_m y_{1,t}^{\dagger}y_{2t+s}^{\dagger'}e_m'\right]\right| \\ =&O\Big(T^{-1}\Big)\cdot O\Big((1-\psi_m^*)\Big)\cdot O\left(\frac{1}{(1-\psi_m^*)^2}\right)=o(1)\end{aligned} \tag{1.51}$$

since $T(1-\psi_m^*)\to+\infty$. By the multivariate Chebyshev inequality, this implies that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Lambda_1 y_{1,t-p-1}^{\dagger}\xrightarrow{p}0 \tag{1.52}$$

(vii) Finally, consider the $(1,2)$ and $(2,1)$ blocks. Notice that for any $1 \leq j \leq p$

$$\begin{aligned}
&\left| E\left[e_m \Delta y^{\dagger}_{t-j} \Delta y^{\dagger'}_{t-j+s} e'_m\right] \right| \leq \left| \sum_{j_1=0}^{t-j} \sum_{j_2=0}^{t-j+s} E\left[e_m G_{j_1} \varepsilon_{t-j-j_1} \varepsilon'_{t-j+s-j_2} G'_{j_2} e'_m\right] \right| \\
=&O\left(\sum_l \delta^{2l+s}\right) + O\left((1-\psi_m)\sum_l \psi_m^{*l} \delta^{l+s}\right) \\
&+O\left((1-\psi_m)^2 \sum_l \psi_m^{*2l+s}\right) + O\left((1-\psi_m)\sum_l \psi_m^{*l+s}\delta^l\right) \\
=&O\left(\delta^s\right) + O\left((1-\psi_m)\psi_m^{*s}\right)
\end{aligned} \tag{1.53}$$

Then

$$\begin{aligned}
&\left| E\left[\left(\frac{1}{T}\sum_{t=1}^{T} e_m \Delta y^{\dagger}_{t-j}\right)^2\right] \right| \leq \frac{1}{T} \max_t \sum_{s=-t}^{T-t} \left| E\left[e_m \Delta y^{\dagger}_{t-j} \Delta y^{\dagger'}_{t-j+s} e'_m\right] \right| \\
=&O\Big(T^{-1}\Big) \cdot \Big(O\Big(1\Big) + O\Big((1-\psi_m)\kappa(\psi_m^*)\Big)\Big) = o(1)
\end{aligned} \tag{1.54}$$

which, by the multivariate Chebyshev inequality, gives that

$$\frac{1}{T}\sum_{t=1}^{T} \Delta y^{\dagger}_{t-j} \xrightarrow{p} 0 \tag{1.55}$$

This implies that

$$\frac{1}{T}\sum_{t=1}^{T} x_{3,t} \xrightarrow{p} 0 \tag{1.56}$$

□

Lemma 1.2 establishes the convergence of $\sum_{t=1}^{T} \varepsilon_t x'_t$ at the appropriate rate. Lemmas 1.1 and 1.2, once combined, give the asymptotic distribution of the slope estimator.

Lemma 1.2. *For sequences $\{\theta_T\} \subset \Theta$ satisfying* (1.8)*, suppose there exists $k \in \{1,\ldots,N\}$ such that $c_j \in [0,+\infty)$ for all $j > N-k$, and $c_j = +\infty$ for all $j \leq N-k$. Let $S_\Sigma = \Sigma^{1/2}$, $\zeta_t^* = [x'_{3,t}, y^{\dagger'}_{1,t-p-1}]'$,*

$S_{\zeta^*} = (E[\sum_{t=1}^{T} \zeta_t^* \zeta_t^{*'}])^{1/2}$, *and let* $\tilde{\Lambda} = \begin{bmatrix} \frac{1}{\sqrt{T}} & 0 & 0 \\ 0 & S_{\zeta^*}^{-1} & 0 \\ 0 & 0 & \Lambda_2 \end{bmatrix}$, *then*

$$\sum_{t=1}^{T} vec\left(S_{\Sigma}^{-1} \varepsilon_t x_t' \tilde{\Lambda}'\right) \xrightarrow{d} \begin{bmatrix} Z_1 \\ Z_2 \\ vec\left(S_{\Sigma}^{-1}\left(\int_0^1 J_{C_2}(x) dB(x)'\right)'\right) \end{bmatrix}$$

where $[Z_1', Z_2']' \sim N(0, I_{N(Np+N-k+1)})$.

Proof. Again, we take a block by block approach. The third block follows from Lemma 1(iv) in Elliott (1998).

Now consider the first and second block. Let $S_\Sigma = \Sigma^{1/2}$. In addition, let $\zeta_t = [1, \zeta_t^{*'}]'$ and $S_\zeta = \left(E\left[\sum_{t=1}^{T} \zeta_t \zeta_t'\right]\right)^{1/2}$, $X_{Tt} = vec\left(S_\Sigma^{-1}(\theta_T)\varepsilon_t(\theta_T)\zeta_t(\theta_T)' S_\zeta(\theta_T)^{-1'}\right)$ and let $\mathscr{F}_{Tt}$ be a filtration

$$\mathscr{F}_{Tt} = \sigma(\varepsilon_s(\theta_T) : 0 \leq s \leq t)$$

i.e. the smallest σ-algebra such that all ε_s are measurable for $0 \leq s \leq t$. The desired result can be shown by verifying the following (Davidson (1994)):

(1) $E[X_{Tt}] = 0$;

(2) $\sum_{t=1}^{T} E[X_{Tt} X_{Tt}' | \mathscr{F}_{Tt-1}] \to I$ in probability;

(3) $\sum_{t=1}^{T} E\left[\|X_{Tt}\|^2 \mathbb{1}\{\|X_{Tt}\| > \varepsilon\} \middle| \mathscr{F}_{Tt-1}\right] \to 0$ in probability for all $\varepsilon > 0$;

(1) is straightforward. To verify (2), let

$$\tilde{\zeta}_t = S_\zeta^{-1} \zeta_t$$
$$\tilde{\varepsilon}_t = S_\Sigma^{-1} \varepsilon_t$$

and

$$\Lambda^* = \begin{bmatrix} \frac{1}{\sqrt{T}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{T}} I_{Np} & 0 \\ 0 & 0 & \Lambda_1 \end{bmatrix} \tag{1.57}$$

Then a typical element in $\sum_{t=1}^{T} E\left[X_{Tt} X_{Tt}' | \mathscr{F}_{Tt-1}\right]$ is

$$\sum_{t=1}^{T} E\left[e_i \tilde{\varepsilon}_t \tilde{\zeta}_t' e_j' e_m \tilde{\zeta}_t \tilde{\varepsilon}_t' e_n' \middle| \mathscr{F}_{Tt-1}\right] = \sum_{t=1}^{T} e_j \tilde{\zeta}_t \tilde{\zeta}_t' e_m' \mathbb{1}\{i = n\} \tag{1.58}$$

Meanwhile, in Lemma 1.1 we showed that

$$\Lambda^* \sum_{t=1}^{T} \zeta_t \zeta_t' \Lambda^{*'} - E\left[\Lambda^* \sum_{t=1}^{T} \zeta_t \zeta_t' \Lambda^{*'}\right] \xrightarrow{p} 0 \tag{1.59}$$

Therefore by Assumption 1.2, we have

$$S_\zeta^{-1} \sum_{t=1}^{T} \zeta_t \zeta_t' S_\zeta^{-1'} \xrightarrow{p} I \tag{1.60}$$

and thus (2) holds.

To verify (3), it is sufficient to show that

$$\sum_{t=1}^{T} E\left[\left\|\tilde{\varepsilon}_t \tilde{\zeta}_t'\right\|^4 \middle| \mathscr{F}_{Tt-1}\right] \xrightarrow{p} 0 \tag{1.61}$$

Notice that

$$\begin{aligned} &\sum_{t=1}^{T} E\left[\left\|\tilde{\varepsilon}_t \tilde{\zeta}_t'\right\|^4 \middle| \mathscr{F}_{Tt-1}\right] \\ \leq &\sum_{t=1}^{T} E\left[\|\tilde{\varepsilon}_t\|^4 \left\|\tilde{\zeta}_t\right\|^4 \middle| \mathscr{F}_{Tt-1}\right] = E\left[\|\tilde{\varepsilon}_t\|^4\right] \sum_{t=1}^{T} \|\tilde{\zeta}_t\|^4 \\ \leq &E\left[\|\tilde{\varepsilon}_t\|^4\right] \left\|S_\zeta^{-1} \Lambda^{*-1}\right\|^4 \sum_{t=1}^{T} \|\Lambda^* \zeta_t\|^4 \\ = &E\left[\|\tilde{\varepsilon}_t\|^4\right] \left[tr\left((\Lambda^* S_\zeta S_\zeta' \Lambda^{*'})^{-1}\right)\right]^2 \sum_{t=1}^{T} \left\|\Lambda^* \zeta_t\right\|^4 \\ \leq &Const \cdot \left[\lambda_{\min}\left(\Lambda^* S_\zeta S_\zeta' \Lambda^{*'}\right)\right]^{-2} \sum_{t=1}^{T} \left\|\Lambda^* \zeta_t\right\|^4 \end{aligned} \tag{1.62}$$

Then by Assumption 1.2, it is sufficient to show that

$$\sum_{t=1}^{T} E\left[\|\Lambda^* \zeta_t\|^4\right] \to 0 \tag{1.63}$$

To show this, notice that

$$\|\Lambda^* \zeta_t\|^4 = \left(\frac{1}{T} + \sum_{n=1}^{N-k}\left[\frac{1-\psi_n^{*2}}{T}(e_n y_{t-p-1}^{\dagger})^2\right] + \sum_{l=1}^{p}\sum_{n=1}^{N}\left[\frac{1}{T}(e_n \Delta y_{t-l}^{\dagger})^2\right]\right)^2 \tag{1.64}$$

Therefore we may expand the right-hand side, and it is sufficient to show the convergence for each element of the expansion. As a demonstration, here we show that

$$\frac{1-\psi_m^{*2}}{T^2}\sum_{t=1}^{T} E\left[(e_m y_{t-p-1}^{\dagger})^2 (e_n \Delta y_{t-j}^{\dagger})^2\right] \to 0 \tag{1.65}$$

for any $1 \leq j \leq p$ and any $1 \leq m \leq N-k$. To see this, notice that

$$\begin{aligned}
&\left\|E\left[e_m y_{t-p-1}^{\dagger} \Delta y_{t-j}^{\dagger\prime} e_n' e_m y_{t-p-1}^{\dagger} \Delta y_{t-j}^{\dagger\prime} e_n'\right]\right\| \\
=&\left\|\sum_{j_1=0}^{t-p-1}\sum_{j_2=0}^{t-j}\sum_{j_3=0}^{t-p-1}\sum_{j_4=0}^{t-j} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t-j_3}\varepsilon_{t-j-j_4}' G_{j_4}' e_n'\right]\right\| \\
\leq&\left\|\sum_{j_1=j_2-(p-j+1)\neq j_3=j_4-(p-j+1)} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t-j_3}\varepsilon_{t-j-j_4}' G_{j_4}' e_n'\right]\right\| \\
&+\left\|\sum_{j_1=j_3\neq j_2-(p-j+1)=j_4-(p-j+1)} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t-j_3}\varepsilon_{t-j-j_4}' G_{j_4}' e_n'\right]\right\| \\
&+\left\|\sum_{j_1=j_4-(p-j+1)\neq j_2-(p-j+1)=j_3} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t-j_3}\varepsilon_{t-j-j_4}' G_{j_4}' e_n'\right]\right\| \\
&+\left\|\sum_{j_1=j_2-(p-j+1)=j_3=j_4-(p-j+1)} E\left[e_m F_{j_1}\varepsilon_{t-j_1}\varepsilon_{t-j-j_2}' G_{j_2}' e_n' e_m F_{j_3}\varepsilon_{t-j_3}\varepsilon_{t-j-j_4}' G_{j_4}' e_n'\right]\right\| \\
=&O\Big(1\Big) + O\Big(\frac{1}{1-\psi_m^*}\Big) + O\Big(1\Big) + O\Big(1\Big)
\end{aligned} \tag{1.66}$$

Therefore

$$\left\|E\left[(1-\psi_m^{*2}) e_m y_{t-p-1}^{\dagger} \Delta y_{t-j}^{\dagger\prime} e_n' e_m y_{t-p-1}^{\dagger} \Delta y_{t-j}^{\dagger\prime} e_n'\right]\right\| = O(1) \tag{1.67}$$

and thus

$$\frac{1-\psi_m^{*2}}{T^2}\sum_{t=1}^{T} E\left[(e_m y_{t-p-1}^{\dagger})^2 (e_n \Delta y_{t-j}^{\dagger})^2\right] \to 0 \tag{1.68}$$

Other elements from the right-hand side of (1.64) can be bounded similarly. □

Lemma 1.2 extends the argument in Lemma 1.2, and is useful to show the joint convergence of the slope and variance estimators.

Lemma 1.3. *For sequences* $\{\theta_T\} \subset \Theta$ *satisfying* (1.8)*, suppose there exists* $k \in \{1,\ldots,N\}$ *such that* $c_j \in [0,+\infty)$ *for all* $j > N-k$*, and* $c_j = +\infty$ *for all* $j \leq N-k$*. Let* $\zeta_t = [1, x_{3,t}', y_{1,t-1}^{\dagger'}]'$*, and* $S_\zeta = \left(E\left[\sum_{t=1}^T \zeta_t\zeta_t'\right]\right)^{1/2}$*. Let* $S_\Sigma = \Sigma^{*1/2}$ *and* $S_4 = \Sigma_4^{*1/2}$*. Then*

$$\begin{bmatrix} vec\left(S_\Sigma^{-1}\sum_{t=1}^T \varepsilon_t\zeta_t' S_\zeta^{-1'}\right) \\ \sqrt{T}S_4^{-1}vech\left(\hat{\Sigma}^* - \Sigma^*\right) \end{bmatrix} \xrightarrow{d} N\left(0, I_{N(Np-k+1)+N(N+1)/2}\right) \tag{1.69}$$

where $\hat{\Sigma}^* = \frac{1}{T}\sum_{t=1}^T \hat{\varepsilon}_t^*\hat{\varepsilon}_t^{*'}$*.*

Proof. First, following the proof of Proposition 1 in Inoue and Kilian (2020), we have that

$$\begin{aligned} &\sqrt{T}S_4^{-1}vech\left(\hat{\Sigma}^* - \Sigma^*\right) \\ =&\frac{1}{\sqrt{T}}S_4^{-1}\sum_{t=1}^T vech\left(\hat{\varepsilon}_t^*\hat{\varepsilon}_t^{*'} - \Sigma^*\right) \\ =&\frac{1}{\sqrt{T}}S_4^{-1}\sum_{t=1}^T vech\Big(\varepsilon_t^*\varepsilon_t^{*'} - \Sigma^* + (\hat{\Pi}-\Pi)x_t^*x_t^{*'}(\hat{\Pi}-\Pi)' \\ &-(\hat{\Pi}-\Pi)x_t^*\varepsilon_t^{*'} - \varepsilon_t^*x_t^{*'}(\hat{\Pi}-\Pi)'\Big) \\ =&\frac{1}{\sqrt{T}}S_4^{-1}\left[\sum_{t=1}^T vech\Big(\varepsilon_t^*\varepsilon_t^{*'} - \Sigma^*\Big) + vech\Big((\hat{\Pi}-\Pi)P\Lambda^{-1}\Lambda\sum_{t=1}^T x_tx_t'\Lambda'\Lambda^{-1'}P'(\hat{\Pi}-\Pi)'\Big)\right. \\ &\left.-vech\Big((\hat{\Pi}-\Pi)P\Lambda^{-1}\Lambda\sum_{t=1}^T x_t\varepsilon_t^{*'} + \sum_{t=1}^T \varepsilon_t^*x_t'\Lambda'\Lambda^{-1'}P(\hat{\Pi}-\Pi)'\Big)\right] \end{aligned} \tag{1.70}$$

By Lemmas 1.1 and 1.2, we have

$$\begin{aligned} &\frac{1}{\sqrt{T}}\left[vech\Big((\hat{\Pi}-\Pi)P\Lambda^{-1}\Lambda\sum_{t=1}^T x_tx_t'\Lambda'\Lambda^{-1'}P'(\hat{\Pi}-\Pi)'\Big)\right. \\ &\left.-vech\Big((\hat{\Pi}-\Pi)P\Lambda^{-1}\Lambda\sum_{t=1}^T x_t\varepsilon_t^{*'} + \sum_{t=1}^T \varepsilon_t^*x_t'\Lambda'\Lambda^{-1'}P(\hat{\Pi}-\Pi)'\Big)\right] \xrightarrow{p} 0 \end{aligned} \tag{1.71}$$

Now let

$$X_{1t} = vec\left(S_\Sigma^{-1}\varepsilon_t\zeta_t'S_\zeta^{-1'}\right), \quad X_{2t} = \frac{1}{\sqrt{T}}S_4^{-1}vech\left(\varepsilon_t^*\varepsilon_t^{*'} - \Sigma^*\right) \tag{1.72}$$

and let $X_{Tt} = [X_{1t}(\theta_T)', X_{2t}(\theta_T)']'$ and $S_\Sigma = \Sigma^{1/2}$. Additionally, let $\mathscr{F}_{Tt}$ be a filtration

$$\mathscr{F}_{Tt} = \sigma(\varepsilon_s(\theta_T) : 0 \le s \le t)$$

i.e. the smallest σ-algebra such that all ε_s are measurable for $0 \le s \le t$.

(1.71) suggests that it is sufficient to show that

$$\sum_{t=1}^{T} X_{Tt} \xrightarrow{d} N(0, I) \tag{1.73}$$

The desired result can be shown by verifying the following (Davidson (1994)):

(1) $E[X_{Tt}] = 0$;

(2) $\sum_{t=1}^{T} E[X_{Tt} X_{Tt}' | \mathscr{F}_{Tt-1}] \to I$ in probability;

(3) $\sum_{t=1}^{T} E\left[\|X_{Tt}\|^2 \mathbb{1}\{\|X_{Tt}\| > \varepsilon\} \middle| \mathscr{F}_{Tt-1} \right] = 0$ in probability for all $\varepsilon > 0$;

Again, (1) is straightforward. To show (2), Notice that we already showed that

$$\sum_{t=1}^{T} E\left[X_{1t} X_{1t}' | \mathscr{F}_{Tt-1}\right] \xrightarrow{p} I_{N^2 p - Nk}$$

in Lemma 1.2. Meanwhile, we also have

$$\begin{aligned}
&\sum_{t=1}^{T} E\left[X_{2t} X_{2t}' | \mathscr{F}_{Tt-1}\right] \\
=&\sum_{t=1}^{T} E\left[\frac{1}{T} \left(S_4^{-1} vech\left(\varepsilon_t^* \varepsilon_t^{*'} - \Sigma \right) \right) \left(S_4^{-1} vech\left(\varepsilon_t^* \varepsilon_t^{*'} - \Sigma \right) \right)' \middle| \mathscr{F}_{Tt-1} \right] \\
=&S_4^{-1} E\left[vech\left(\varepsilon_t^* \varepsilon_t^{*'} - \Sigma^* \right) vech\left(\varepsilon_t^* \varepsilon_t^{*'} - \Sigma^* \right)' \right] S_4^{-1'} = I_{N(N+1)/2}
\end{aligned} \tag{1.74}$$

and for a typical element of X_{1t}, $e_m S_\Sigma^{-1} \varepsilon_t \zeta_t' S_\zeta^{-1'} e_n'$,

$$\begin{aligned}
&\sum_{t=1}^{T} E\left[\left(e_m S_\Sigma^{-1} \varepsilon_t \zeta_t' S_\zeta^{-1'} e_n' \right) X_{2t}' \middle| \mathscr{F}_{Tt-1} \right] \\
=&\frac{1}{\sqrt{T}} \sum_{t=1}^{T} E\left[\left(e_m S_\Sigma^{-1} \varepsilon_t \zeta_t' S_\zeta^{-1'} e_n' \right) \left(S_4^{-1} vech\left(\varepsilon_t \varepsilon_t' - \Sigma \right) \right)' \middle| \mathscr{F}_{Tt-1} \right] \\
=&\frac{1}{\sqrt{T}} \sum_{t=1}^{T} (e_n S_\zeta^{-1} \zeta_t) E\left[\left(e_m S_\Sigma^{-1} \varepsilon_t \right) \left(S_4^{-1} vech\left(\varepsilon_t \varepsilon_t' - \Sigma \right) \right)' \right] = 0
\end{aligned} \tag{1.75}$$

Therefore (2) holds. To verify (3), it is sufficient to show that

$$\sum_{t=1}^{T} E\left[\|X_{Tt}\|^4 \,\middle|\, \mathscr{F}_{Tt-1}\right] \xrightarrow{p} 0 \tag{1.76}$$

To show this, notice that

$$\|X_{Tt}\|^4 = \left(\|X_{1t}\|^2 + \|X_{2t}\|^2\right)^2 \tag{1.77}$$

We already showed that

$$\sum_{t=1}^{T} E\left[\|X_{1t}\|^4 \,\middle|\, \mathscr{F}_{Tt-1}\right] \xrightarrow{p} 0 \tag{1.78}$$

in Lemma 1.2, and by construction, it is not difficult to show that

$$\sum_{t=1}^{T} E\left[\|X_{2t}\|^4 \,\middle|\, \mathscr{F}_{Tt-1}\right] \xrightarrow{p} 0 \tag{1.79}$$

Meanwhile, we have

$$\|X_{1t}\|^2 = \sum_{m=1}^{N} \sum_{n=1}^{Np-k} (e_m S_\Sigma^{-1} \varepsilon_t \zeta_t' S_\zeta^{-1'} e_n')^2 \tag{1.80}$$

and

$$\begin{aligned} &\sum_{t=1}^{T} E\left[\left(e_m S_\Sigma^{-1} \varepsilon_t \zeta_t' S_\zeta^{-1'} e_n'\right)^2 \|X_{2t}\|^2 \,\middle|\, \mathscr{F}_{Tt-1}\right] \\ =&E\left[\left(e_m S_\Sigma^{-1} \varepsilon_t\right)^2 \left\|S_4^{-1} vech\left(\varepsilon_t \varepsilon_t' - \Sigma\right)\right\|^2\right] \frac{1}{T} \sum_{t=1}^{T} (e_n S_\zeta^{-1} \zeta_t)^2 \end{aligned} \tag{1.81}$$

Now notice that

$$\begin{aligned} &\frac{1}{T} \sum_{t=1}^{T} (e_n S_\zeta^{-1} \zeta_t)^2 \\ \leq&\frac{1}{T} \sum_{t=1}^{T} \left\|S_\zeta^{-1} \zeta_t\right\|^2 \leq \frac{1}{T} \left\|S_\zeta^{-1} \Lambda^{*-1}\right\|^2 \sum_{t=1}^{T} \|\Lambda^* \zeta_t\|^2 \\ =&\frac{1}{T} tr\left((\Lambda^* S_\zeta S_\zeta' \Lambda^{*'})^{-1}\right) \sum_{t=1}^{T} tr(\Lambda^* \zeta_t \zeta_t' \Lambda^{*'}) \\ \leq&Const \cdot \left[\lambda_{\min}\left(\Lambda^* S_\zeta S_\zeta' \Lambda^{*'}\right)\right]^{-1} tr\left(\frac{1}{T} \sum_{t=1}^{T} \Lambda^* \zeta_t \zeta_t' \Lambda^{*'}\right) \xrightarrow{p} 0 \end{aligned} \tag{1.82}$$

Combining the above, we have

$$\sum_{t=1}^{T} E\left[\|X_{1t}\|^2 \|X_{2t}\|^2 \,\middle|\, \mathscr{F}_{Tt-1}\right] \xrightarrow{p} 0 \tag{1.83}$$

Therefore (1.76) holds which implies that (3) holds. □

Lemma 1.4 establishes the consistency of the estimator for the fourth moment of the errors.

Lemma 1.4. *For sequences $\{\theta_T\} \subset \Theta$ satisfying* (1.8)*, suppose there exists $k \in \{1,\ldots,N\}$ such that $c_j \in [0,+\infty)$ for all $j > N-k$, and $c_j = +\infty$ for all $j \leq N-k$. Let*

$$\hat{\Sigma}_4^* = \frac{1}{T}\sum_{t=1}^{T} vech(\hat{\varepsilon}_t^* \hat{\varepsilon}_t^{*'} - \hat{\Sigma}^*) vech(\hat{\varepsilon}_t^* \hat{\varepsilon}_t^{*'} - \hat{\Sigma}^*)' \tag{1.84}$$

then

$$\hat{\Sigma}_4^* \xrightarrow{p} \Sigma_4^* \tag{1.85}$$

Proof. The proof is similar to that of Lemma B.1 in Inoue and Kilian (2020). Let $\xi_t = (\hat{\Pi} - \Pi)x_t^*$, then we have

$$\begin{aligned}
&\hat{\Sigma}_4 - \Sigma_4 \\
=&\frac{1}{T}\sum_{t=1}^{T}\left[vech(\hat{\varepsilon}_t^* \hat{\varepsilon}_t^{*'} - \hat{\Sigma}^*) vech(\hat{\varepsilon}_t^* \hat{\varepsilon}_t^{*'} - \hat{\Sigma}^*)' - \Sigma_4^*\right] \\
=&\frac{1}{T}\sum_{t=1}^{T}\Bigg[vech\Big((\varepsilon_t^* - \xi_t)(\varepsilon_t^* - \xi_t)' - \Sigma\Big) vech\Big((\varepsilon_t^* - \xi_t)(\varepsilon_t^* - \xi_t)' - \Sigma^*\Big)' - \Sigma_4^* \\
&+ vech\Big((\varepsilon_t^* - \xi_t)(\varepsilon_t^* - \xi_t)' - \Sigma^*\Big) vech\Big(\hat{\Sigma}^* - \Sigma^*\Big)' \\
&+ vech\Big(\hat{\Sigma}^* - \Sigma^*\Big) vech\Big((\varepsilon_t^* - \xi_t)(\varepsilon_t^* - \xi_t)' - \Sigma^*\Big)' \\
&+ vech\Big(\hat{\Sigma}^* - \Sigma^*\Big) vech\Big(\hat{\Sigma}^* - \Sigma^*\Big)'\Bigg]
\end{aligned} \tag{1.86}$$

In addition, for the first term,

$$
\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\Bigg[vech\bigg(\Big(\varepsilon_t^* - (\hat{\Pi}-\Pi)x_t^*\Big)\Big(\varepsilon_t^* - (\hat{\Pi}-\Pi)x_t^*\Big)' - \Sigma^*\bigg) \\
&vech\bigg(\Big(\varepsilon_t^* - (\hat{\Pi}-\Pi)x_t^*\Big)\Big(\varepsilon_t^* - (\hat{\Pi}-\Pi)x_t^*\Big)' - \Sigma^*\bigg)' - \Sigma_4^*\Bigg] \\
=&\frac{1}{T}\sum_{t=1}^{T}\Bigg[vech\Big(\varepsilon_t^*\varepsilon_t^{*'} - \Sigma^*\Big)vech\Big(\varepsilon_t^*\varepsilon_t^{*'} - \Sigma^*\Big)' - \Sigma_4^* \\
&+vech\Big(\varepsilon_t^*\varepsilon_t^{*'} - \Sigma^*\Big)vech\Big(\xi_t\xi_t' - \xi_t\varepsilon_t^{*'} - \varepsilon_t^*\xi_t'\Big)' \\
&+vech\Big(\xi_t\xi_t' - \xi_t\varepsilon_t^{*'} - \varepsilon_t^*\xi_t'\Big)vech\Big(\varepsilon_t^*\varepsilon_t^{*'} - \Sigma^*\Big)' \\
&+vech\Big(\xi_t\xi_t' - \xi_t\varepsilon_t^{*'} - \varepsilon_t^*\xi_t'\Big)vech\Big(\xi_t\xi_t' - \xi_t\varepsilon_t^{*'} - \varepsilon_t^*\xi_t'\Big)'\Bigg]
\end{aligned} \tag{1.87}
$$

Now notice that

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_t' = \frac{1}{T}\sum_{t=1}^{T}(\hat{\Pi}-\Pi)P\Lambda^{-1}\Lambda x_t x_t'\Lambda'\Lambda^{-1'}P'(\hat{\Pi}-\Pi)' \xrightarrow{p} 0 \tag{1.88}$$

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^*\xi_t' = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^* x_t'\Lambda'\Lambda^{-1'}P'(\hat{\Pi}-\Pi) \xrightarrow{p} 0 \tag{1.89}$$

$$vech(\hat{\Sigma}^* - \Sigma^*) \xrightarrow{p} 0 \tag{1.90}$$

By Lemmas 1.1, 1.2, and 1.3. Then by applying the arguments in Inoue and Kilian (2020) for each entry, we obtain

$$\hat{\Sigma}_4^* - \Sigma_4^* \xrightarrow{p} 0 \tag{1.91}$$

□

1.3.3 Uniform validity of a Wald test

After obtaining the asymptotics of the statistics in the previous section, we can now construct a Wald test that is uniformly valid across the parameter space. First, in Lemma 1.5, we examine the asymptotic distribution of the Wald statistic under a drifting parameter sequence (1.8).

Lemma 1.5. *For sequences $\{\theta_T\} \subset \Theta$ satisfying* (1.8)*, suppose there exists $k \in \{1,\ldots,N\}$ such that*

$c_j \in [0, +\infty)$ for all $j > N-k$, and $c_j = +\infty$ for all $j \leq N-k$.

$$\gamma = \begin{bmatrix} vec(\Pi^*) \\ vech(\Sigma^*) \end{bmatrix}, \quad \hat{\gamma} = \begin{bmatrix} vec(\hat{\Pi}^*) \\ vech(\hat{\Sigma}^*) \end{bmatrix} \tag{1.92}$$

and

$$\begin{aligned} \Sigma_\gamma &= \begin{bmatrix} M_x^* \otimes \Sigma^* & 0 \\ 0 & T^{-1}\Sigma_4^* \end{bmatrix} \\ \hat{\Sigma}_\gamma &= \begin{bmatrix} M_x^* \otimes \hat{\Sigma}^* & 0 \\ 0 & T^{-1}\hat{\Sigma}_4^* \end{bmatrix} \end{aligned} \tag{1.93}$$

Then the Wald statistic

$$(\hat{\gamma} - \gamma)'\hat{\Sigma}_\gamma^{-1}(\hat{\gamma} - \gamma) \xrightarrow{d} \chi^2_{N^2p + N(N+1)/2} \tag{1.94}$$

Proof. First let

$$S_M = M_x^{*1/2},\ S_{\Sigma^*} = QS_\Sigma,\ S_\Sigma = \Sigma^{1/2},\ S_4 = \Sigma_4^{*1/2} \tag{1.95}$$

and

$$S_\gamma = \begin{bmatrix} S_M' \otimes S_{\Sigma^*} & 0 \\ 0 & T^{-1/2}S_4 \end{bmatrix},\ \hat{S}_\gamma = \begin{bmatrix} S_M' \otimes \hat{S}_{\Sigma^*} & 0 \\ 0 & T^{-1/2}\hat{S}_4 \end{bmatrix} \tag{1.96}$$

Then

$$\hat{S}_\gamma^{-1}(\hat{\gamma} - \gamma) = \begin{bmatrix} \left(S_M' \otimes \hat{S}_\Sigma^{-1}Q^{-1}\right) vec(\hat{\Pi} - \Pi) \\ \sqrt{T}\hat{S}_4^{-1} vech(\hat{\Sigma}^* - \Sigma^*) \end{bmatrix} \tag{1.97}$$

The first block can be rewritten as

$$\begin{aligned}
&\left(S_M' \otimes \hat{S}_\Sigma^{-1} Q^{-1}\right) vec\left(\hat{\Pi} - \Pi\right) \\
=& vec\left(\hat{S}_\Sigma^{-1} Q^{-1}\left(\hat{\Pi} - \Pi\right) S_M\right) \\
=& vec\left(\hat{S}_\Sigma^{-1}\left(\sum_{t=1}^{T} \varepsilon_t x_{2,t}^{*'}\right) M_x^{*-1} S_M\right) \\
=& vec\Bigg\{\hat{S}_\Sigma^{-1}\left(\sum_{t=1}^{T} \varepsilon_t x_t' \Lambda\Lambda^{-1} P_2\right)\Bigg[\Bigg\{P_2\Lambda^{-1}\Lambda\sum_{t=1}^{T} x_t x_t' \Lambda\Lambda^{-1} P_2' \\
&-\left(P_2\Lambda^{-1}\Lambda\sum_{t=1}^{T} x_t x_t' \Lambda\Lambda^{-1} P_1'\right)\left(P_1\Lambda^{-1}\Lambda\sum_{t=1}^{T} x_t x_t' P_1' \Lambda\Lambda^{-1}\right)^{-1} \\
&\left(P_1\Lambda^{-1}\Lambda\sum_{t=1}^{T} x_t x_t' \Lambda\Lambda^{-1} P_2\right)\Bigg\}^{1/2}\Bigg]^{-1'}\Bigg\}
\end{aligned} \tag{1.98}$$

Now we inspect P_1 and P_2 more closely. For P_1, we notice that

$$P_1 = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix} \begin{bmatrix} 1 & 0_{1\times Np} \dots & 0_{1\times N} \\ Q^{-1}d^\dagger & 0_{N\times Np} & I_N \end{bmatrix} \equiv Q^* P_1^* \tag{1.99}$$

For P_2, we notice that

$$P_2 = (I_N \otimes Q) \begin{bmatrix} Q^{-1}d^\dagger & I_N & I_N & \dots & I_N & I_N \\ Q^{-1}d^\dagger & 0 & I_N & \dots & I_N & I_N \\ \ddots & \ddots & \ddots & \vdots & \ddots & \ddots \\ Q^{-1}d^\dagger & 0 & 0 & \dots & 0 & I_N \end{bmatrix} \equiv (I_N \otimes Q) P_2^* \tag{1.100}$$

Now we defined two diagonal matrices: $(N+1)\times(N+1)$ matrix Λ_1^* and $Np \times Np$ matrix Λ_2^*:

$$\Lambda_1^* = \begin{bmatrix} \frac{1}{\sqrt{T}} & 0_{1\times N} \\ 0_{N\times 1} & \Lambda^\dagger \end{bmatrix}, \qquad \Lambda_2^* = \frac{1}{T} I_{Np} \tag{1.101}$$

Using the above characterization, we can rewrite (1.98) as

$$
\begin{aligned}
& vec\left\{\hat{S}_\Sigma^{-1}\left(\sum_{t=1}^T \varepsilon_t x_t'\Lambda\Lambda^{-1}P_2'\right)\left[\left\{P_2\Lambda^{-1}\Lambda\sum_{t=1}^T x_t x_t'\Lambda\Lambda^{-1}P_2'\right.\right.\right.\\
&-\left(P_2\Lambda^{-1}\Lambda\sum_{t=1}^T x_t x_t'\Lambda\Lambda^{-1}P_1'\right)\left(P_1\Lambda^{-1}\Lambda\sum_{t=1}^T x_t x_t'\Lambda\Lambda^{-1}P_1'\right)^{-1}\\
&\qquad\left.\left.\left.\left(P_1\Lambda^{-1}\Lambda\sum_{t=1}^T x_t x_t'\Lambda\Lambda^{-1}P_2\right)\right\}^{1/2}\right]^{-1'}\right\}\\
=& vec\left\{\hat{S}_\Sigma^{-1}\left(\sum_{t=1}^T \varepsilon_t x_t'\Lambda\Lambda^{-1}P_2^{*'}(I_N\otimes Q)\right)\left[\left\{(I_N\otimes Q)P_2^*\Lambda^{-1}\Lambda\sum_{t=1}^T x_t x_t'\Lambda^{-1}\Lambda P_2^{*'}(I_N\otimes Q')\right.\right.\right.\\
&-\left((I_N\otimes Q)P_2^{*'}\Lambda^{-1}\Lambda\sum_{t=1}^T x_t x_t'\Lambda\Lambda^{-1}P_1^{*'}Q^{*'}\right)\left(Q^*P_1^*\Lambda^{-1}\Lambda\sum_{t=1}^T x_t x_t'\Lambda\Lambda^{-1}P_1^{*'}Q^{*'}\right)^{-1}\\
&\qquad\left.\left.\left.\left(Q^*P_1^*\Lambda^{-1}\Lambda\sum_{t=1}^T x_t x_t'\Lambda\Lambda^{-1}P_2^{*'}(I_N\otimes Q')\right)\right\}^{1/2}\right]^{-1'}\right\}\\
=& vec\left\{\hat{S}_\Sigma^{-1}\left(\sum_{t=1}^T \varepsilon_t x_t'\Lambda(\Lambda^{-1}P_2^{*'}\Lambda_2^*)\right)\left[\left\{(\Lambda_2^*P_2^*\Lambda^{-1})\Lambda\sum_{t=1}^T x_t x_t'(\Lambda^{-1}\Lambda P_2^{*'}\Lambda_2^*)\right.\right.\right.\\
&-\left((\Lambda_2^*P_2^*\Lambda^{-1})\Lambda\sum_{t=1}^T x_t x_t'\Lambda(\Lambda^{-1}P_1^{*'}\Lambda_1^*)\right)\left((\Lambda_1^*P_1^*\Lambda^{-1})\Lambda\sum_{t=1}^T x_t x_t'\Lambda(\Lambda^{-1}P_1^{*'}\Lambda_1^*)\right)^{-1}\\
&\qquad\left.\left.\left.\left((\Lambda_1^*P_1^*\Lambda^{-1})\Lambda\sum_{t=1}^T x_t x_t'\Lambda(\Lambda^{-1}P_2^{*'}\Lambda_2^*)\right)\right\}^{1/2}\right]^{-1'}\right\}
\end{aligned}
\tag{1.102}
$$

The last equation holds because Q^*, $(I_N\otimes Q)$, Λ_1^* and Λ_2^* are invertible.

Meanwhile, for P_1^* we have

$$
\Lambda_1^*P_1^*\Lambda^{-1}=\begin{bmatrix} 1 & 0_{1\times Np} & 0_{1\times N}\\ \sqrt{T}\Lambda^\dagger Q^{-1}d^\dagger & 0 & I_N\end{bmatrix} \tag{1.103}
$$

Notice that $\sqrt{T}\Lambda^\dagger$ is a diagonal matrix with elements being of the form

$$
\max\left\{\frac{1}{\sqrt{T}},(1-\psi_j^{*2})^{1/2}\right\}\le 1 \tag{1.104}
$$

Similarly for P_2^*, we have

$$\Lambda_2^* P_2^* \Lambda^{-1} = \begin{bmatrix} Q^{-1}d^\dagger & I_N & I_N & \dots & I_N & \frac{1}{T}\Lambda^{\dagger-1} \\ Q^{-1}d^\dagger & 0 & I_N & \dots & I_N & \frac{1}{T}\Lambda^{\dagger-1} \\ \ddots & \ddots & \ddots & \vdots & \ddots & \ddots \\ Q^{-1}d^\dagger & 0 & 0 & \dots & 0 & \frac{1}{T}\Lambda^{\dagger-1} \end{bmatrix} \tag{1.105}$$

and $\frac{1}{T}\Lambda^{\dagger-1}$ is a diagonal matrix with elements being of the form

$$\min\left\{1, (T(1-\psi_j^{*2}))^{-1/2}\right\} \leq 1 \tag{1.106}$$

These characterizations guarantee that the elements of $\Lambda_1^* P_1^* \Lambda^{-1}$ and $\Lambda_2^* P_2^* \Lambda^{-1}$ are bounded. Then by Lemmas 1.1 and 1.2, we have

$$\left(S_M' \otimes \hat{S}_\Sigma^{-1} Q^{-1}\right) vec\left(\hat{\Pi} - \Pi\right) \xrightarrow{d} N(0, I_{N^2 p}) \tag{1.107}$$

This combined with Lemma 1.3 and 1.4 gives

$$\hat{S}_\gamma^{-1}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, I_{N^2 p + N(N+1)/2}) \tag{1.108}$$

which implies that

$$(\hat{\gamma} - \gamma)' \hat{\Sigma}_\gamma^{-1} (\hat{\gamma} - \gamma) \xrightarrow{d} \chi^2_{N^2 p + N(N+1)/2} \tag{1.109}$$

□

Notice that the limiting distribution does not depend on c_T, so naturally we have the following theorem giving the uniform validity of the Wald statistic:

Theorem 1.6. *Let*

$$CP_T(\theta) = P\Big((\hat{\gamma}_T - \gamma)' \hat{\Sigma}_{\gamma,T}^{-1} (\hat{\gamma}_T - \gamma) \leq q_{1-\alpha}\left(\chi^2_{N^2 p + N(N+1)/2}\right)\Big) \tag{1.110}$$

where $q_{1-\alpha}(\cdot)$ denotes the $1-\alpha$ quantile. Then

$$\limsup_{T\to\infty}\sup_{\theta\in\Theta} CP_T(\theta) = \liminf_{T\to\infty}\inf_{\theta\in\Theta} CP_T(\theta) = 1-\alpha \tag{1.111}$$

Proof. By Corollary 2.1 in Andrews et al. (2020), we only need to verify their Assumption $B1^*$ and Assumption $B2^*$. It is straightforward to verify Assumption $B2^*$ for the parameter space Θ and the way we set up the parameter sequence (1.8). To verify Assumption $B1^*$, consider a sequence θ_T satisfying (1.8) with the corresponding c_T. By Lemma 1.5, we have

$$(\hat{\gamma}_T-\gamma)'\hat{\Sigma}_{\gamma,T}^{-1}(\hat{\gamma}_T-\gamma) \xrightarrow{d} \chi^2_{N^2p+N(N+1)/2} \tag{1.112}$$

Since this limiting distribution does not depend on c_T, we have $CP_T(\theta) \to 1-\alpha$ and hence Assumption $B1^*$ holds. □

Based on Theorem 1.6, we obtain the following result for any function of the VAR parameters:

Corollary 1.7. *Let $f(\gamma)$ be a function of γ and*

$$\mathbb{S}_{f,T}^{1-\alpha} = \left\{ f(\gamma) \middle| (\hat{\gamma}_T-\gamma)'\hat{\Sigma}_{\gamma,T}^{-1}(\hat{\gamma}_T-\gamma) \le q_{1-\alpha}\left(\chi^2_{N^2p+N(N+1)/2}\right) \right\} \tag{1.113}$$

Then

$$\liminf_{T\to\infty}\inf_{\theta\in\Theta} P\Big(f(\gamma(\theta)) \in \mathbb{S}_{f,T}^{1-\alpha}\Big) \ge 1-\alpha \tag{1.114}$$

In particular, if $f(\gamma)$ is one-to-one, then

$$\limsup_{T\to\infty}\sup_{\theta\in\Theta} P\Big(f(\gamma(\theta)) \in \mathbb{S}_{f,T}^{1-\alpha}\Big) = \liminf_{T\to\infty}\inf_{\theta\in\Theta} P\Big(f(\gamma(\theta)) \in \mathbb{S}_{f,T}^{1-\alpha}\Big) = 1-\alpha \tag{1.115}$$

Regarding impulse response functions, consider the confidence intervals obtained through inverting the Wald test for candidate parameters and calculating corresponding impulse responses. By Corollary 1.7, these intervals should have uniformly valid (potentially conservative) coverage rate.

1.4 Simulation

In this section, we examine the finite sample performance of our method. We consider bivariate Gaussian VAR(1) models of the following form:

$$y_t = \begin{bmatrix} \pi_1 & 0 \\ 0.5 & \pi_2 \end{bmatrix} y_{t-1} + u_t \tag{1.116}$$

$$u_t \overset{\text{iid}}{\sim} N\left(0, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 3 \end{bmatrix}\right) \tag{1.117}$$

with $y_0 = 0$. This DGP is similar to the bivariate process used in Kilian (1998a) and Lütkepohl et al. (2015). Due to the setup of our model, we assume that $\pi_1 \neq \pi_2$ so that the slope coefficient matrix can be diagonalized. For structural identification, we assume a recursive structure such that the first variable does not respond to the second structural shock.

1.4.1 Coverage rate

To demonstrate the finite-sample coverage accuracy of our method, we conduct two simulation studies for reduced-form and structural impulse responses respectively. We consider the reduced-form response since the uniform local projection method in Montiel Olea and Plagborg-Møller (2021) and Xu (2023) are proposed for reduced-form models, although the former could be extended to accommodate recursive identification as discussed in Plagborg-Møller and Wolf (2021)

Now for reduced-form impulse responses, we consider the following five methods of constructing confidence interval for the impulse responses, in order to compare the performance of our method with other uniformly valid methods, especially local projection based procedures.

(1) The first method, referred to as LAVAR, is our method based on the Wald statistic constructed after lag-augmented regression on the VAR model. For this method, we report the size of the Wald test based on the VAR slope parameters, which does not depend on the horizon.

(2) The second method is based on the test proposed in Montiel Olea and Plagborg-Møller (2021). They recommend running lag-augmented local projections, and using the Eicker-Huber-White heteroskedasticity robust standard errors. We refer to this method as LALPHC.

(3) The third method is similar to the second except it uses the HAR approach recommended in Lazarus et al. (2018). More specifically, we use an equally weighted cosine (EWC) long run variance estimator, following their suggestions on selecting the critical value and tuning parameter. This method is referred to as LALPHAR.

(4) The fourth method, referred to as LALPMG, again runs lag-augmented local projections, but uses the martingale variance estimator proposed in Xu (2023). In contrast to Xu (2023) which assumes infinite order DGP, we maintain the assumption that the lag order is known in order to obtain consistent comparison with the other methods.

(5) The final method, referred to as LALPB, follows the bootstrap procedure in Montiel Olea and Plagborg-Møller (2021). Although the uniform validity of the method is not established in their work, their simulation study has shown that the bootstrap procedure significantly improves coverage rates when roots are close to 1.

The nominal size is set to 0.1, and for the local projection methods, we consider five horizons $h \in 1, 6, 12, 36, 60$. The sample size is set to $T = 240$ and the number of Monte Carlo simulations is $M = 1000$. For LALPB, we set the number of bootstrap repetitions to be 2000. The results are summarized in Table 1.1.

From the results, the Wald test in our method yields valid size over various degrees of persistence. This implies that if the researcher first invert the test to obtain candidate parameters, then calculate the impulse responses using these parameters, the obtained confidence interval should have valid size across different degrees of persistence and across all horizons. For the local projection methods, all of them lead to under coverage when at least one of the roots is close to 1 and horizon is not short, which is consistent with the theoretical and simulation results in Montiel Olea and Plagborg-Møller (2021) and Xu (2023). Among these methods, while the martingale variance estimator proposed in Xu (2023) is not meant to improve coverage rates at near unit root, it does seem to give slightly better coverage rates than HC/HAR variance estimators, which is also demonstrated in Xu (2023). The bootstrap method in Montiel Olea and Plagborg-Møller (2021) mitigates the coverage rate problem significantly, but still fails to reach the nominal coverage rate at longer horizons. Compared to the univariate results in Montiel Olea and Plagborg-Møller (2021), the coverage distortion of the local projection methods start to be significant at much shorter horizons,

Table 1.1: Coverage rate comparison for reduced-form impulse responses

	h	LAVAR	LALPHC	LALPHAR	LALPMG	LALPB
$\pi_1 = 0.50, \pi_2 = 0.49$	1	0.90	0.90	0.90	0.90	0.89
	6	0.90	0.88	0.88	0.89	0.88
	12	0.90	0.89	0.88	0.89	0.89
	36	0.90	0.90	0.89	0.92	0.89
	60	0.90	0.88	0.88	0.92	0.88
$\pi_1 = 0.90, \pi_2 = 0.89$	1	0.89	0.87	0.87	0.89	0.88
	6	0.89	0.85	0.84	0.86	0.86
	12	0.89	0.84	0.83	0.84	0.86
	36	0.89	0.82	0.82	0.87	0.85
	60	0.89	0.86	0.85	0.93	0.90
$\pi_1 = 0.99, \pi_2 = 0.98$	1	0.89	0.89	0.90	0.90	0.89
	6	0.89	0.83	0.81	0.82	0.89
	12	0.89	0.73	0.70	0.72	0.88
	36	0.89	0.49	0.45	0.50	0.85
	60	0.89	0.28	0.27	0.34	0.71
$\pi_1 = 1.00, \pi_2 = 0.99$	1	0.88	0.89	0.89	0.89	0.90
	6	0.88	0.80	0.79	0.80	0.87
	12	0.88	0.65	0.63	0.64	0.87
	36	0.88	0.31	0.32	0.34	0.80
	60	0.88	0.15	0.15	0.20	0.69
$\pi_1 = 1.00, \pi_2 = 0.50$	1	0.89	0.90	0.90	0.90	0.90
	6	0.89	0.84	0.84	0.85	0.87
	12	0.89	0.74	0.74	0.75	0.83
	36	0.89	0.50	0.50	0.56	0.75
	60	0.89	0.33	0.37	0.42	0.60
The results are for the response of the second variable to the first reduced form shock						

when there are two roots close to 1.

For the second study, we consider a similar comparison for structural impulse responses. Based on Plagborg-Møller and Wolf (2021), the recursive identification assumption in VAR models can be equivalently implemented in local projection by regressing $y_{2,t+h}$ on $y_{1,t}$ (and y_{t-1} when implementing lag augmentation). In this case, the martingale variance estimator in Xu (2023) could not be constructed, but the other LALP methods could be accommodated in a straightforward fashion. Therefore, we consider LAVAR, LALPHC, LALPHAR and LALPB. For our result, LAVAR, we report the size of the Wald test based on both the slope and variance parameters. For LALPB, again we set the number of bootstrap repetitions to be 2000. We summarize the results in Table 1.2.

Results for structural impulse responses are overall similar to those for reduced-form impulse

Table 1.2: Coverage rate comparison for structural impulse responses

	h	LAVAR	LALPHC	LALPHAR	LALPB
$\pi_1 = 0.50$, $\pi_2 = 0.49$	1	0.87	0.90	0.88	0.91
	6	0.87	0.89	0.88	0.90
	12	0.87	0.90	0.89	0.90
	36	0.87	0.92	0.90	0.92
	60	0.87	0.90	0.89	0.90
$\pi_1 = 0.90$, $\pi_2 = 0.89$	1	0.87	0.91	0.88	0.94
	6	0.87	0.87	0.86	0.90
	12	0.87	0.82	0.80	0.86
	36	0.87	0.82	0.79	0.86
	60	0.87	0.86	0.85	0.90
$\pi_1 = 0.99$, $\pi_2 = 0.98$	1	0.85	0.91	0.88	0.94
	6	0.85	0.84	0.81	0.90
	12	0.85	0.72	0.65	0.89
	36	0.85	0.43	0.40	0.82
	60	0.85	0.23	0.23	0.70
$\pi_1 = 1.00$, $\pi_2 = 0.99$	1	0.85	0.92	0.89	0.94
	6	0.85	0.80	0.76	0.87
	12	0.85	0.67	0.62	0.87
	36	0.85	0.30	0.30	0.83
	60	0.85	0.16	0.15	0.74
$\pi_1 = 1.00$, $\pi_2 = 0.50$	1	0.85	0.90	0.89	0.93
	6	0.85	0.84	0.81	0.88
	12	0.85	0.75	0.71	0.87
	36	0.85	0.49	0.48	0.70
	60	0.85	0.35	0.35	0.60
The results are for the response of the second variable to the first structural shock					

responses. Our method now has a slightly larger size than the nominal size, mostly due to the relatively small sample size, as it is well known that the estimation of higher moments of the error tend to be less accurate compared to the other components of the model. The size distortion of the local projection methods is similar to that for reduced-form impulse responses. In addition, in contrast to the reduced-form results, the bootstrap method gives larger coverage rates at short horizons for near unit root/unit root processes.

1.4.2 Median length

In this section, we provide a method of constructing confidence intervals using our Wald test in practice. In order to provide better guidance for empirical application of uniform inference methods, we also compare the median length of the confidence intervals for structural impulse responses using the four methods discussed in the previous section: LAVAR, LALPHC, LALPHAR and LALPB.

To construct confidence interval using our method (LAVAR), we need to invert our test statistic. Ideally we could evaluate the Wald statistic across a grid over the entire parameter space to construct the confidence set, but this might be computationally burdensome when the dimension of the parameter space is large. To attenuate such burden, we recommend searching over a Bayesian VAR posterior obtained with an uninformative prior in practice. More specifically, we implement the following procedure:

1. Estimate the model to obtain $\hat{\Pi}$, $\hat{\Sigma}$ and $\hat{\Sigma}_4$. Construct $\hat{\gamma}$ and $\hat{\Sigma}_\gamma$.

2. Obtain the posterior distribution of (Π, Σ) from a Bayesian VAR model with diffuse prior.

3. Draw a (Π, Σ) from the posterior. Construct γ.

4. Discard the draw if it does not fit the parameterization (1.1). Otherwise, construct the Wald statistic

$$(\hat{\gamma} - \gamma)' \hat{\Sigma}_\gamma^{-1} (\hat{\gamma} - \gamma)$$

   and conduct the Wald test using the critical value $q_{1-\alpha}(\chi^2_{N^2 p + N(N+1)/2})$.

5. Repeat (3) & (4) until we have enough accepted draws of Π and Σ. We set this number to 5000.

6. Calculate the upper and lower bound of the structural impulse responses corresponding to these draws to construct the confidence interval.

For each Monte Carlo simulation, we generate $T = 240$ samples, and the number of bootstrap repetitions is set to 2000 for LALPB. Similar to previous simulation studies, we consider five horizons: $h \in 1, 6, 12, 36, 60$. Total number of simulations is $M = 1000$. The results are summarized in Table 1.3.

Overall, consistent with the observations in Li et al. (2021), our simulation results corroborate that neither VAR methods nor LP methods necessarily produces shorter confidence intervals across all horizons. In fact, the relative performance highly depends on the horizons and persistence.

We make the following observations on the results:

Table 1.3: Median lengths and coverage rates of structural impulse response confidence intervals

	h	LAVAR	LALPHC	LALPHAR	LALPB
$\pi_1 = 0.50, \pi_2 = 0.49$	1	0.72 (0.99)	0.43 (0.91)	0.42 (0.89)	0.46 (0.92)
	6	0.36 (0.99)	0.49 (0.90)	0.49 (0.89)	0.50 (0.90)
	12	0.08 (0.99)	0.50 (0.90)	0.50 (0.88)	0.50 (0.90)
	36	1.5e-4 (0.99)	0.53 (0.91)	0.53 (0.88)	0.53 (0.90)
	60	3.1e-7 (0.99)	0.56 (0.90)	0.57 (0.88)	0.57 (0.90)
$\pi_1 = 0.90, \pi_2 = 0.89$	1	0.75 (0.98)	0.52 (0.92)	0.50 (0.90)	0.60 (0.94)
	6	2.01 (0.99)	1.02 (0.88)	0.99 (0.87)	1.09 (0.90)
	12	3.37 (0.99)	1.46 (0.83)	1.39 (0.80)	1.65 (0.88)
	36	5.16 (0.99)	1.80 (0.80)	1.73 (0.79)	1.89 (0.85)
	60	5.23 (0.99)	1.87 (0.86)	1.82 (0.83)	1.94 (0.90)
$\pi_1 = 0.99, \pi_2 = 0.98$	1	0.73 (0.98)	0.54 (0.91)	0.52 (0.90)	0.65 (0.94)
	6	2.22 (0.98)	1.31 (0.85)	1.25 (0.80)	1.41 (0.88)
	12	4.30 (0.97)	2.55 (0.71)	2.29 (0.65)	3.02 (0.87)
	36	13.63 (0.97)	7.18 (0.39)	6.36 (0.35)	12.72 (0.83)
	60	22.57 (0.97)	8.83 (0.23)	8.10 (0.22)	17.68 (0.67)
$\pi_1 = 1.00, \pi_2 = 0.99$	1	0.74 (0.98)	0.55 (0.91)	0.53 (0.90)	0.66 (0.95)
	6	2.27 (0.95)	1.35 (0.81)	1.32 (0.77)	1.49 (0.86)
	12	4.30 (0.91)	2.70 (0.64)	2.52 (0.59)	3.30 (0.85)
	36	13.60 (0.86)	8.61 (0.29)	7.80 (0.26)	15.72 (0.81)
	60	22.75 (0.85)	11.76 (0.14)	15.72 (0.14)	28.66 (0.72)
$\pi_1 = 1.00, \pi_2 = 0.50$	1	0.44 (0.99)	0.43 (0.91)	0.43 (0.88)	0.46 (0.92)
	6	0.61 (0.97)	0.61 (0.85)	0.59 (0.83)	0.65 (0.90)
	12	0.72 (0.97)	0.78 (0.73)	0.75 (0.71)	0.89 (0.86)
	36	1.06 (0.97)	1.05 (0.44)	1.00 (0.44)	1.32 (0.72)
	60	1.13 (0.97)	1.09 (0.31)	1.08 (0.31)	1.31 (0.57)
The results are for the response of the second variable to the first structural shock and are reported as "median length (coverage rate)"					

First, for strictly stationary model where $\pi_1 = 0.5$ and $\pi_2 = 0.49$, the length of our intervals shrink to 0 as horizons increase, although the exact horizon at which our method start to have lower length depends on the persistence of the DGP. Meanwhile, the length of LP confidence intervals does not shrink to zero as horizon increases, which is also noted in Li et al. (2021).

In contrast, when $\pi_1 = 0.9$ and $\pi_2 = 0.89$, the intervals of our methods are larger than those from local projection methods. However, consistent with the earlier size studies and Corollary 1.7, our method provides conservative intervals, while local projection methods have coverage rate distortions, especially at longer horizons. In this scenario, LALPB manages to largely, although not completely, mitigate the distortions while having comparable lengths to non-bootstrap methods.

For models with roots very close to 1 or exactly 1, the confidence intervals from our method still manage to reach nominal coverage rate, except for $\pi_1 = 1$ and $\pi_2 = 0.99$ at larger horizons, in which case the coverage rates are still fairly close to the nominal rate. In comparison, the local projection methods all suffer from severe size distortions. Similar to previous results, The bootstrap method LALPB mitigates the distortions, but still has a significantly lower coverage rate than the nominal rate at larger horizons. Moreover, when exact unit root exists, LALPB actually yields larger confidence intervals that has worse coverage rates, compared to our method.

Finally, for our method, as briefly mentioned in the above discussion, the actual coverage rates of the confidence intervals are higher than the nominal rate (and the coverage rates from size studies) for stationary models and slightly lower for highly persistent models. This could be attributed to multiple factors. First, for a specific component of the impulse response at a fixed horizon, the mapping between it and the VAR parameters is not one-to-one. By Corollary 1.7, our method will give conservative coverage rates when we are focusing on a specific component at a fixed horizon. Second, we construct the confidence interval using a projection approach. The true confidence set might not be a connected interval, which implies that our confidence intervals can be slightly conservative. Lastly, when the DGP is highly persistent, the magnitude of the impulse responses becomes highly sensitive to the model parameters. This means that a finer grid or more candidate draws are needed for our method to approximate the true confidence set, and the coverage rates might suffer if the amount of draws is not enough.

1.5 Conclusions

While uniform inference methods eliminate pretesting issues and provide reliable results for empirical applications, the discussion on uniformly valid inference methods for impulse responses in autoregressive models in the literature has been largely focusing on univariate models. In this chapter, we propose a method to construct uniformly valid confidence intervals for impulse responses in a large class of vector autoregressive models, via a uniformly valid Wald test. Our method not only allows for uniform inference on the structural impulse responses, but also for any function of the VAR parameters.

We also provide a simulation study that not only establishes the finite sample validity of our method, but also provides a comparison between inference under VAR and LP method of constructing impulse responses. We find that there is no universally better performing method, but at very high persistence level, our method leads to significantly less coverage distortions than local projection methods, while our confidence intervals have comparable or lower lengths. In addition, when the roots are strictly stationary, confidence intervals from VAR methods gradually shrinks towards 0 length as horizon increases while those from local projection methods do not. However, we do note that when the roots are mildly large, local projection with bootstrap tend to give shorter confidence intervals, without significant size distortions.

As to future directions of research, our results could be extended from both a theoretical and an applied perspective. Theoretically, our model setup does not include all non-stationary models, and relaxing our modelling assumptions could lead to a more general uniform inference theory. It is also worth investigating the possibility of a uniformly valid test constructed without lag augmentation, which might involve further characterization of the limiting behavior of multivariate stochastic integrals. From an application perspective, the general question of numerically inverting a statistical test could be further examined, and development in this aspect could lead to more efficient and wider empirical implementation of our proposed method.

CHAPTER 2

Identification through non-Gaussianity in factor-augmented vector autoregression models

2.1 Introduction

A common theme of criticisms of VAR models is the relative small amount of information used in them. The number of VAR parameters increases rapidly when the number of variables increases, which severely limits the amount of information a researcher is able to include in the model.

On one hand, this could lead to omitted variable bias and as a result, puzzling empirical results. For example, Christiano et al. (1999) find the existence of the puzzle when using a three variable structural VAR to examine the effects of monetary policy. From a theoretical perspective, a small VAR system also does not incorporate enough information compared to the actual decision-making process of agents or central banks. Stock and Watson (2005), Ramey (2016) and Stock and Watson (2016) provide more discussion on the impact of not including enough variables. On the other hand, the small amount of variables included also limits the amount and scale of economic interaction we could estimate and investigate, as impulse responses can only be calculated for the included variables in VARs.

To address such degree-of-freedom problem, Bernanke et al. (2005) proposes the factor augmented vector autoregression model (FAVAR) which, instead of a priori selecting a few economic variables, includes a low dimensional vector of observed and unobserved factors that contains the bulk of information about the economy. Such approach allows researchers to include a rich set of information while keeping the amount of parameters, or the dimensionality of the problem, tractable. This provides researcher the benefits of both avoiding selection of specific variables from a variety of conceptually similar noisy time series, and avoiding the problem that a small set of variables might not properly span the entire space of shocks which lead to inaccurate structural shocks. Due to these benefits, FAVAR has been used in the empirical literature, to study a wide range of domestic and international macroeconomic problems, such as in Mumtaz and Surico (2009), Wu and Xia (2016), Ho et al. (2018) and Dahlhaus et al. (2018).

Structural FAVAR includes structural VAR as a component, and similar to the latter, requires additional identification assumptions. For example, Bai et al. (2016) proposes three sets of statistical identification restrictions that assume no contemporaneous correlation between observed and unobserved factors, although the economic interpretation of these restrictions might not always be of economic interest. Meanwhile, as Stock and Watson (2016) has pointed out, the identification problem in FAVAR is similar to that of structural VAR, with the additional complication of estimating the unobserved factors. Therefore, in the literature, various economic assumptions similar to that of structural VAR are imposed in order to identify the structural FAVAR model. For example, Bernanke et al. (2005) focuses on the identification of impulse responses, and assumes a recursive structure on the factors. To accommodate such structure, the researcher needs to a priori separates the variables into slow-moving and fast-moving (compared to the structural shock of interest) categories. In contrast to the above assumptions directly on the elements of covariance matrices or the structural matrix, Yamamoto and Hara (2022) assumes the existence of structural breaks that leads to identification through heteroskedasticity.

However, conventional identification restrictions for structural VAR might face additional shortcomings for FAVAR models. For example, in a small scale structural VAR, a researcher utilizing short run recursive identification scheme could conduct additional robustness checks through changing the order of the variables. However, in typical recursive FAVAR identification schemes, such as the "slow-R-fast" scheme as seen in Bernanke et al. (2005) or Ho et al. (2018), a researcher has to a priori separate a large amount of economic variables into slow moving and fast moving categories, and then extract the slow moving and fast moving factors from variables in the two categories respectively. This implies that the robustness check is not as simple as reordering the variables as in structural VAR models. In other words, the need for additional identification schemes exist, especially in a FAVAR context, so that the conventional identification restrictions could be tested as over-identifying restrictions.

This chapter aims to complement the aforementioned research and provide a method to identify, estimate and perform inference on structural impulse responses in FAVAR models using alternative statistical assumptions, following the literature of identification through non-Gaussianity in structural VAR, such as Lanne et al. (2017), Gouriéroux et al. (2017), Lanne and Luoto (2021), and Fiorentini and Sentana (2023). For a literature review and a more detailed overview of such identification scheme, we refer the readers to Kilian and Lütkepohl (2017) and Montiel Olea et al. (2022). Originating from the engineering problem of independent component analysis (ICA), the assumption of independent and non-Gaussian structural shocks implies that we can decompose the reduced form shocks without additional economic assumptions. In other words, such assumption reduces the identification problem of the structural shocks to normalization that does not affect structural impulse responses to each shock. From a practitioner's perspective, this method can be used to supplement conventional economically oriented identification restrictions, and provide additional testing tools and robustness checks.

Independence and non-Gaussian distribution might sound like strong assumptions, but they are often well justified in empirical economic applications. While a large body of empirical and theoretical macroeconomic research assumes Gaussian shocks, there has been an increasing number of studies that question such assumptions. For example, Christiano (2007) presents evidence against assuming Gaussian likelihood in DSGE modelling, while Mishkin (2011) also suggested that many shocks hitting the economies could display non-Gaussian features such as fat tail. Furthermore, Cúrdia et al. (2014) shows strong evidence against Gaussian shocks, and Müller (2013) also discusses the potential problems when shocks are misspecified. All of these studies suggest not only the possibility, but also the necessity of investigating non-Gaussian shocks in statistical modelling. As to independence, the structural shocks, by its definition and construction, are meant to be orthogonal and be able to span the space of the economy. This means that in traditional Gaussian settings, while independence is not explicitly required, the commonly used uncorrelatedness assumption implies independence between the shocks. In fact, as discussed in Lanne et al. (2017) and Gouriéroux et al. (2017), independence might be the more appropriate concept of orthogonality for structural shocks.

FAVAR models are traditionally estimated in two steps, with a principal component first step followed by a least square estimation of the VAR model using the estimated factors. To incorporate non-Gaussian assumptions, it is natural to consider a similar two-step estimator with the second step substituted with a proper estimator proposed in the structural VAR literature. In this chapter, we choose to implement maximum likelihood estimation in Lanne et al. (2017), which assumes correct specification of the distribution of the structural shocks. Alternatively, there are also methods that relax such assumption. For example, Gouriéroux et al. (2017) estimates all the reduced form parameters by quasi-MLE using Gaussian density, and the structural matrix by quasi-MLE using non-Gaussian densities, although their assumption on the normalization of the structural matrix is somewhat uncommon in empirical research. Alternatively, Lanne and Luoto (2021) provides a GMM estimator that exploits fourth moment of the structural shocks, and Fiorentini and Sentana (2023) recommends a quasi-MLE using discrete location scale mixture of normal distributions. We leave the possible implementation of these procedures to future research, and hope that our theoretical results provide some insights on alternative estimators of FAVAR identified through non-Gaussianity.

Our contribution to the literature is fourfold:

First, as mentioned above, we propose a two-step estimator and the relevant inference results, under the assumption of independence and non-Gaussianity of structural shocks, for structural impulse responses in FAVAR models. The first step is a principal component estimation of the unobserved factors, and the second step, assuming that the distribution of the structural shocks are known, implements a maximum likelihood estimation (MLE). While in practice additional economic interpretation is needed to label the structural shocks, this method could accommodate a variety of economic intuitions and is a useful complement to the usual identification assumptions. We also briefly discuss testing conventional identification restrictions as over-identifying restrictions under our setup.

Other than proving the consistency and asymptotic normality of our estimator, our theoretical results also provide insights in two-step estimators with general M-estimation second step in FAVAR models, which could potentially accommodate various other estimation methods under non-Gaussianity assumption, such as the quasi-MLE in Gouriéroux et al. (2017) and Fiorentini and Sentana (2023), or GMM in Lanne and Luoto (2021).

In addition, we provide simulation results that not only validates our method in finite sample, but also complement the existing literature by examining the effect of assuming $\frac{\sqrt{T}}{N} \to 0$. The results show that Delta method intervals based on such assumption suffers significantly when T is much larger than N, but bootstrap intervals based on this seem to be less impacted.

Finally, we revisit the results in Bernanke et al. (2005) as a demonstration of our method, and show that their identification assumption is rejected under our setting. We also produce impulse responses of the same macroeconomic variables to monetary policy shock, and show that while the general implications are not affected by this difference in identification strategy, there are some key differences in the results. Most importantly, our results do not support the existence of price puzzle as strongly as the original results, and our results indicate a stronger and more consistent impact of monetary policy on output. This suggests that while their identification scheme is convenient and standard in the literature, empirical researchers should be wary of the economic implications and possibly refer to alternative identification schemes as robustness check.

The chapter is organized as follows. In section 2, we set up the model, define the notations and present the identification result on structural impulse response functions. Section 3 provides the relevant consistency and inference results for a two-step estimator. Section 4 presents simulation results that examine the finite sample validity of the estimation method. In section 5, we demonstrate the empirical application of our method by re-examining the results of Bernanke et al. (2005). Section 6 summarizes and concludes the chapter and provides discussion on potential future research directions.

2.2 Model and identification

2.2.1 Assumptions

Throughout this chapter, we consider a standard structural FAVAR model,

$$x_t = \begin{bmatrix} \Lambda & \Gamma \end{bmatrix} \begin{bmatrix} f_t \\ y_t \end{bmatrix} + u_t \tag{2.1}$$

$$g_t = \sum_{j=1}^{p} A_j g_{t-j} + e_t \tag{2.2}$$

where $g_t = (f_t', y_t')'$ is an $r \times 1$ vector of factors and $e_t = B\varepsilon_t$ is an $r \times 1$ error vector. f_t is an $r_1 \times 1$ vector of unobserved factors, y_t is an $r_2 \times 1$ vector of observed factors and $r_1 + r_2 = r$. $A_1, \ldots, A_p$ and B are $r \times r$ matrices. Λ and Γ are $N \times r_1$ and $N \times r_2$ matrices of factor loading respectively. Let $\Xi = (\Lambda, \Gamma)$, $X = (x_1, \ldots, x_T)'$, $F = (f_1, \ldots, f_T)'$, $Y = (y_1, \ldots, y_T)'$ and $G = (g_1, \ldots, g_T)' = (F, Y)$. For identification of the structural impulse responses, we consider the following assumption on the structural shocks ε_t:

Assumption 2.1. *(Identification through non-Gaussianity) Assume that the structural shocks ε_t are i.i.d. and zero mean. Assume that the components $\varepsilon_{1,t}, \ldots, \varepsilon_{r,t}$ have variances σ_i^2 respectively, are mutually independent, and at most one of them is Gaussian.*

This assumption is similar to its counterparts in Lanne et al. (2017) and Gouriéroux et al. (2017), and is in contrary to traditional identification schemes which normally imposes restrictions on A_j and B while requiring only cross-sectional uncorrelatedness and no distributional assumption on ε_t. In principle, non-Gaussian of the structural errors implies non-Gaussian reduced form errors, which can be tested in data, such as in Kilian and Demiroglu (2000). Testing of independence is significantly more difficult. While there exists several statistical tests for independence (see Herwartz and Maxand (2020) for a review), the structural shocks identified under different schemes might have different economic interpretation, and simply testing independence for shocks identified under an alternative identification scheme does not necessarily validate or invalidate independence assumption presented above. Kilian and Lütkepohl (2017), Montiel Olea et al. (2022) and Lanne and Luoto (2021) also pointed out some potential problems of independence and non-Gaussian assumptions, but we leave addressing these issues to future research.

In addition to Assumption 2.1, we need assumptions to guarantee extraction of factors and stationarity of the VAR.

Assumption 2.2. *(Factor model)*

(1) The factors F satisfy $E\left[\|F_t\|^4\right] \leq M$ and $T^{-1}F'M_YF \xrightarrow{p} \Sigma_F$ as $T \to \infty$, where Σ_F is positive definite and $M_Y = I_T - Y(Y'Y)^{-1}Y'$.

(2) The factor loading of the unobserved factors are deterministic and $N^{-1}\Lambda'\Lambda \to \Sigma_\Lambda$, a fixed positive definite matrix as $N \to \infty$.

(3) The eigenvalues of the $r_1 \times r_1$ matrix $\Sigma_\Lambda\Sigma_F$ are distinct

Assumption 2.3. *(Idiosyncratic errors)*

(1) $\{f_t\}$, $\{\varepsilon_t\}$ and $\{u_t\}$ are three mutually independent groups.

(2) $E(u_{it}) = 0$, $E(u_{it}^8) \leq M$.

(3) $E(u_{is}u_{jt}) = \tau_{ijst}$, $|\tau_{ijst}| < \tau_{st}^M$ for all i, j and $|\tau_{ijst}| < \tau_{ij}^M$ for all s,t such that $T^{-1}\sum_{s,t=1}^T \tau_{st}^M \leq M$, $N^{-1}\sum_{i,j=1}^N \tau_{ij}^M \leq M$ and $(NT)^{-1}\sum_{i,j,s,t}|\tau_{ijst}| \leq M$.

(4) For all s,t, $E(|N^{-1/2}\sum_{i=1}^N(u_{is}u_{it} - E(u_{is}u_{it}))|^4) \leq M$

Assumption 2.4. *(VAR model)*

(1) The structural matrix B is positive, i.e. $\det(B) > 0$.

(2) The roots of $\det(I_n - zA_1 - \cdots - z^p A_p) = 0$ *are all outside the unit circle.*

Assumptions 2.2, 2.3 and 2.4 are standard assumptions in the FAVAR literature. They are similar to Assumption A(iii), A(iv), B and C in Yamamoto and Hara (2022). Assumption 2.2 is a standard regularity condition to guarantee extraction of r_1 unobserved factors. In particular, Assumption 2.2(1) guarantees that factors are non-degenerate and (2) guarantees that all elements of the factors contribute to the variance of x_t, whereas (3) is a technical assumption to ensure the uniqueness of the principal component procedure. Assumption 2.3 imposes restrictions on the dependence structure of the idiosyncratic errors in the factor model (2.1). Assumption 2.3(1) is standard in the literature, see Bai and Ng (2006). Although Gonçalves and Perron (2014) provides a weaker set of alternative assumptions, we choose not to adopt them for simplicity. Assumption 2.3(2) and (3) permits weak serial and cross-sectional dependence, and leads to the approximate factor model as in Chamberlain and Rothschild (1983). Assumption 2.4 guarantees that the structural VAR is well-defined and stationary.

2.2.2 Identification through non-Gaussianity

Under the assumptions in the previous section, which both allows for identification of the VAR through non-Gaussianity and guarantees the existence of a proper principal component first step estimator, we have the following identification result on the structural impulse responses:

Proposition 2.1. *If Assumptions 2.1, 2.2, 2.3 and 2.4 hold, then the impulse response function of the observed factors y_t to the structural shocks $\Phi_y(h)$ is point identified. In addition, the approximate impulse response function of the variables x_t to the structural shocks $\Phi_x(h)$ is also point identified.*

Proof. Following the literature, (see Bai and Ng (2006) and Gonçalves and Perron (2014)), a random rotation of the unobserved factors f_t, Hf_t can be consistently estimated by a principal component estimator. Then the factors g_t are extracted as a rotation of itself, Qg_t, where

$$Q = \begin{bmatrix} H & 0 \\ 0 & I_{r_2} \end{bmatrix}$$

Now left multiply equation (2.2) by Q, we can write it as

$$g_t^* = \sum_{j=1}^{p} A_j^* g_{t-j}^* + B^* \varepsilon_t \tag{2.3}$$

where $g_t^* = Qg_t$, $A_j^* = QA_jQ^{-1}$ and $B^* = QB$. Under Assumption 2.1, by Proposition 1 in Lanne et al. (2017), if there exists another set of $A_1^\dagger, \ldots, A_p^\dagger, B^\dagger$ and $\varepsilon_T^\dagger$ such that

$$g_t^* = \sum_{j=1}^{p} A_j^\dagger g_{t-j}^* + B^\dagger \varepsilon_t^\dagger \tag{2.4}$$

then

$$(A_1^\dagger, \ldots, A_p^\dagger) = (A_1^*, \ldots, A_p^*), \quad B^\dagger = B^* DP, \quad \varepsilon_t^\dagger = P' D^{-1} \varepsilon_t \tag{2.5}$$

where D is a diagonal matrix with nonzero diagonal elements, and P is a permutation matrix. In other words, the structural impulse responses

$$\Phi^\dagger(h) = J\boldsymbol{A}^{\dagger h} J' B^{\dagger'} P' D^{-1} \Sigma_\varepsilon^{1/2} = J\boldsymbol{A}^{*h} J' B^* \Sigma_\varepsilon^{1/2} = \Phi^*(h) \tag{2.6}$$

where $J = [I_{r\times r}, 0_{r\times(n-1)r}]$. And more importantly,

$$\Phi^*(h) = J\boldsymbol{A}^{*h} J' B^* \Sigma_\varepsilon^{1/2} = JM\boldsymbol{A}^h M^{-1} J' QB \Sigma_\varepsilon^{1/2} = QJ\boldsymbol{A}^h J' B \Sigma_\varepsilon^{1/2} = Q\Phi(h) \tag{2.7}$$

where $M = diag\left(Q, I_{(N-1)r\times(N-1)r}\right)$. Now we partition $\Phi^*(h)$ (and similarly $\Phi(h)$) into

$$\Phi^*(h) = \begin{bmatrix} \Phi_{f1}^*(h) & \Phi_{f2}^*(h) \\ \Phi_{y1}^*(h) & \Phi_{y2}^*(h) \end{bmatrix} \tag{2.8}$$

such that $\Phi_{f1}^*(h)$ is the response of f_t to the first r_1 structural shocks, $\Phi_{f2}^*(h)$ is the response of f_t to the rest r_2 structural shocks, and $\Phi_{g1}^*(h), \Phi_{g2}^*(h)$ are similarly defined. Then considering the partition of Q, we have

$$\begin{aligned} &\Phi_{f1}^*(h) = H\Phi_{f1}(h),\ \Phi_{f2}^*(h) = H\Phi_{f2}(h), \\ &\Phi_{g1}^*(h) = \Phi_{g1}(h),\ \Phi_{g2}^*(h) = \Phi_{g2}(h) \end{aligned} \tag{2.9}$$

Now the response of y_t to the structural shocks under the alternative set of parameters is

$$\Phi_y^*(h) = \begin{bmatrix} \Phi_{g1}^*(h) & \Phi_{g2}^*(h) \end{bmatrix} = \begin{bmatrix} \Phi_{g1}(h) & \Phi_{g2}(h) \end{bmatrix} = \Phi_y(h) \tag{2.10}$$

So that $\Phi_y(h)$ is point identified. In addition, we have

$$\begin{aligned} \Phi_x^*(h) &= \Lambda H^{-1} \begin{bmatrix} \Phi_{f1}^*(h) & \Phi_{f2}^*(h) \end{bmatrix} + \Gamma \begin{bmatrix} \Phi_{g1}^*(h) & \Phi_{g2}^*(h) \end{bmatrix} \\ &= \Lambda \begin{bmatrix} \Phi_{f1}(h) & \Phi_{f2}(h) \end{bmatrix} + \Gamma \begin{bmatrix} \Phi_{g1}(h) & \Phi_{g2}(h) \end{bmatrix} = \Phi_x(h) \end{aligned} \tag{2.11}$$

So that the approximate response $\Phi_x(h)$ is point identified. □

The above result shows that the structural impulse response of the observed factor y_t, along with the approximate responses of the information variables X_t, are impacted by neither the rotation introduced in the principal component first step, nor the potential permutation and scaling from identification through non-Gaussianity. However, we do stress that this is a statistical identification result, rather than an economical one, as the structural shocks do not necessarily carry any specific economic meaning. To properly label the structural shocks, as suggested in Lanne et al. (2017), one could inspect the shape and signs of the estimated impulse responses, and refer to a variety of economic theory and intuition. This process will be further discussed and demonstrated in section 2.5

2.3 Estimation, inference and testing

In this section, we propose a two-step estimator of the FAVAR model and provide the relevant consistency and inference results. For the first step, we consider a principal component estimation for the factors. Notice that (2.1) can be written as

$$M_Y X = M_Y F \Lambda' + M_Y U \tag{2.12}$$

Let $\tilde{F}$ consist of $\sqrt{T}$ times the eigenvectors corresponding to the r_1 largest eigenvalues of

$$T^{-1} N^{-1} M_Y X X' M_Y'$$

arranged in descending order, where the normalization $T^{-1}\tilde{F}'\tilde{F} = I_{r_1}$ is used. It is commonly known in the factor model literature (see Bai and Ng (2006) and Gonçalves and Perron (2014)) that $\tilde{f}_t$ actually estimates a rotation of the true factors, i.e. Hf_t, where

$$H = \tilde{V}^{-1}\frac{\tilde{F}'M_Y F}{T}\frac{\Lambda'\Lambda}{N} \tag{2.13}$$

where $\tilde{V}$ is the $r_1 \times r_1$ diagonal matrix with the diagonal elements being the r_1 the largest eigenvalues of the matrix $T^{-1}N^{-1}M_Y XX'M_Y'$, in descending order. Correspondingly, let $\tilde{g}_t = (\tilde{f}_t', y_t')'$, $\tilde{G} = (\tilde{g}_1, \ldots, \tilde{g}_T)' = (\tilde{F}, Y)$. Then $\tilde{g}_t = Qg_t$ where

$$Q = \begin{bmatrix} H & 0 \\ 0 & I_{r_2} \end{bmatrix} \tag{2.14}$$

with H defined above.

After obtaining an estimate of the unobserved factors, following Lanne et al. (2017), we consider a maximum likelihood estimation of the VAR model. To this end, assume that $\varepsilon_{i,t}$ has a density $\sigma_i^{-1} f_i(\sigma_i^{-1}x; \lambda_i)$ where λ_i is a vector of parameters.

Let $\theta = (\sigma', \lambda', \pi', \beta')'$ where $\sigma = (\sigma_1, \ldots, \sigma_r)'$, $\lambda = (\lambda_1', \ldots, \lambda_r')'$, $\pi = vec(A_1, \ldots, A_p)$ and $\beta = vecd^{\circ}(B)$ where $vecd^{\circ}$ denotes the vector obtained by removing the diagonal elements of B from $vec(B)$. We consider the VAR parameter space $\Theta = \Theta_\sigma \times \Theta_\lambda \times \Theta_\pi \times \Theta_\beta$ where

1. $\Theta_\sigma = \mathbb{R}_+^r$.
2. $\Theta_\lambda = \Theta_{\lambda_1} \times \ldots \Theta_{\lambda_r}$ with each Θ_{λ_i} being an open subset of $\mathbb{R}^{d_i}$. Let $d = \sum_{i=1}^r d_i$.
3. $\Theta_\pi \subset \mathbb{R}^{n^2 p}$ an open set such that Assumption 2.4 is satisfied. In addition, for any $\pi \in \Theta_\pi$ we have that the corresponding π^* where $\pi^* = vec(A_1^*, \ldots, A_p^*)$ and $A_i^* = QA_iQ^{-1}$ satisfies $\pi^* \in \Theta_\pi$ almost surely.
4. $\Theta_\beta = vecd^{\circ}(\mathscr{B})$ with $\mathscr{B}$ is an open set of structural matrices satisfying a normalization scheme with diagonal elements being 1. For an example of such set (and the corresponding normalization scheme), see Lanne et al. (2017).

Suppose the true parameter is $\theta_0 = (\sigma_0, \lambda_0, \pi_0, \beta_0)$, with $\pi_0 = vec(A_{01}, \ldots, A_{0p})$ and $\beta_0 = vecd^{\circ}(B_0)$.

The parameter space assumptions guarantee that θ_0 is an interior point. To guarantee the validity of maximum likelihood estimation, we need to be more specific with the distribution of the structural shocks, on which we impose the following assumption:

Assumption 2.5. *(Density) The following conditions hold for $f_1,\dots,f_r$:*

(1) $f_i(x;\lambda_i)>0$ *and is twice continuously differentiable with respect to* $(x;\lambda_i)$.

(2) $f_{i,x}(x;\lambda_{i,0})>0$, $f_{i,xx}(x;\lambda_{i,0})$ *and* $f_{i,x\lambda}(x;\lambda_{i,0})$ *are integrable with respect to x.*

(3) $\int\sup_{\lambda_i\in\Theta_{\lambda_i}}\|f_{i,\lambda_i}(x;\lambda_i)\|dx<\infty$ *and* $\int\sup_{\lambda_i\in\Theta_{\lambda_i}}\|f_{i,\lambda\lambda}(x;\lambda_i)\|dx<\infty$

(4) The matrix $E[l_{\theta,t}(\theta_0,G)l'_{\theta,t}(\theta_0,G)]$ *is positive definite.*

(5) For all $x\in\mathbb{R}$,

$$x^2\frac{f_{i,x}^2(x;\lambda_{i,0})}{f_i^2(x;\lambda_{i,0})} \text{ and } \frac{\|f_{i,\lambda}(x;\lambda_{i,0})\|^2}{f_i^2(x;\lambda_{i,0})}$$

are dominated by $c_0(1+|x|^{c_1})$ *with* $c_0\geq 0, 0\leq c_1\leq 2$ *and*

$$\int|x|^{c_2}f_i(x;\lambda_{i,0})dx<\infty$$

(6) For all $x\in\mathbb{R}$ *and* $\lambda_i\in\Theta_{\lambda_i}$,

$$\frac{f_{i,x}^2(x;\lambda_i)}{f_i^2(x;\lambda_i)} \text{ and } \left|\frac{f_{i,xx}(x;\lambda_i)}{f_i(x;\lambda_i)}\right|$$

are dominated by $a_0(1+|x|^{a_1})$,

$$\left\|\frac{f_{i,x\lambda}(x;\lambda_i)}{f_i(x;\lambda_i)}\right\| \text{ and } \left\|\frac{f_{i,x}(x;\lambda_i)}{f_i(x;\lambda_i)}\frac{f_{i,\lambda}(x;\lambda_i)}{f_i(x;\lambda_i)}\right\|$$

are dominated by $a_0(1+|x|^{a_2})$,

$$\left\|\frac{f_{i,\lambda}(x;\lambda_i)}{f_i^2(x;\lambda_i)}\right\|^2 \text{ and } \left\|\frac{f_{i,\lambda\lambda}(x;\lambda_i)}{f_i(x;\lambda_i)}\right\|$$

are dominated by $a_0(1+|x|^{a_3})$, *with* $a_0\geq 0$, $0\leq a_1,a_2,a_3\leq 2$, *and*

$$\int(|x|^{2+a_1}+|x|^{1+a_2}+|x|^{a_3})f_i(x;\lambda_{i,0})dx<\infty$$

(7) For all $x \in \mathbb{R}$ and $\lambda_i \in \Theta_{\lambda_i}$,

$$\left|\frac{f_{i,xxx}(x;\lambda_i)}{f_i(x;\lambda_i)}\right|, \left|\frac{f_{i,xx}(x;\lambda_i)f_{i,x}(x;\lambda_i)}{f_i^2(x;\lambda_i)}\right| \text{ and } \left|\frac{f_{i,x}(x;\lambda_i)}{f_i(x;\lambda_i)}\right|^3$$

are dominated by $b_0(1+|x|^{b_1})$,

$$\left\|\frac{f_{i,\lambda\lambda x}(x;\lambda_i)}{f_i(x;\lambda_i)}\right\|, \left\|\frac{f_{i,x}(x;\lambda_i)f_{i,\lambda\lambda}(x;\lambda_i)}{f_i^2(x;\lambda_i)}\right\|,$$
$$\left\|\frac{f_{i,\lambda}(x;\lambda_i)f'_{i,x\lambda}(x;\lambda_i)}{f_i^2(x;\lambda_i)}\right\| \text{ and } \left\|\frac{f_{i,x}(x;\lambda_i)f_{i,\lambda}(x;\lambda_i)f_{i,\lambda}(x;\lambda_i)'}{f_i^3(x;\lambda_i)}\right\|$$

are dominated by $b_0(1+|x|^{b_2})$, with $b_0 \geq 0$, $0 \leq b_1, b_2 \leq 2$, and

$$\int (|x|^{b_1} + |x|^{b_2}) f_i(x;\lambda_{i,0}) dx < \infty$$

These assumptions impose restrictions on the smoothness and tail behavior of the density function, and are similar to Assumptions 4 and 5 in Lanne et al. (2017), with the additional requirement of $a_1, a_2, a_3, b_1, b_2, c_1 \leq 2$. More specifically, (1), (2), (3) and (4) guarantees the differentiability and the information matrix equality. (5), (6) and (7) ensures the proper asymptotic behavior of the score vector and hessian matrix, and also allows for bounding the estimation error introduced in the principal component estimation of the factors. While the additional requirement on the magnitude of the parameters might seem restrictive, notice that these parameters represents a trade-off between smoothness and heavy tail. Larger parameters allow for the density and derivatives to have large fluctuations, but at the price of more restrictive tail behavior. In macroeconomic applications, most non-Gaussian distributional assumptions are used to accommodate heavy tails of the variables and shocks, which implies that the magnitude of the aforementioned parameters should not be too large. In fact, Assumption 2.5 is satisfied by many commonly used non-Gaussian density function, such as the Student's t-distribution (with degrees of freedom larger than 4) and the logistic density.

Now we introduce some additional notations due to the principal component estimation. Taking into account the rotation introduced in factor estimation, let

$$
\begin{aligned}
&V = p\lim \tilde{V}, \quad \Sigma_{\tilde{F}F} = p\lim\left(\frac{\tilde{F}' M_Y F}{T}\right), \\
&H_0 = p\lim H = V^{-1}\Sigma_{\tilde{F}F}\Sigma_{\Lambda}, \\
&Q_0 = p\lim Q = \begin{bmatrix} H_0 & 0 \\ 0 & I_{r_2} \end{bmatrix}
\end{aligned}
\tag{2.15}
$$

Let $A^*_{i,0} = Q_0 A_{i,0} Q_0^{-1}$. While $Q_0 B_0$ may not be an element of $\mathscr{B}$, by Proposition 1 and 2 in Lanne et al. (2017), there must exist matrices diagonal matrices D_1, D_2 and permutation matrix P such that $B^*_0 = Q_0 B_0 D_1 P D_2 \in \mathscr{B}$. Correspondingly, let $\sigma^*_0 = D_2^{-1} P' D_1^{-1} \sigma_0$ denote the standard error of $D_2^{-1} P' D_1^{-1} \varepsilon_t$ and λ^*_0 denote the reassembled parameter vector after permutation P'. Now let $\pi^*_0 = vec(A^*_{1,0}, \ldots, A^*_{p,0})$, $\beta^*_0 = vecd^\circ(B^*_0)$ and $\theta^*_0 = (\sigma^{*'}_0, \lambda^{*'}_0, \pi^{*'}_0, \beta^{*'}_0)'$.

Now we show that the MLE exists and consistently estimates such transformation of the true VAR parameters. The likelihood function is

$$
\begin{aligned}
L_T(\theta, G) =& \frac{1}{T}\sum_{t=p+1}^{T} l_t(\theta, \boldsymbol{g}_t) \\
=& \frac{1}{T}\sum_{t=p+1}^{T}\left[\sum_{i=1}^{r}\log f_i(\sigma_i^{-1}\iota_i B^{-1} e_t(\theta, \boldsymbol{g}_t); \lambda_i) - \log(\det(B)) - \sum_{i=1}^{r}\log\sigma_i\right]
\end{aligned}
\tag{2.16}
$$

where $\boldsymbol{g}_t = [g'_t, \ldots, g'_{t-p}]'$, ι_i is a $1 \times r$ vector with its i-th element being 1 and other elements being 0 and

$$
e_t(\theta, \boldsymbol{g}_t) = g_t - A_1 g_{t-1} - \cdots - A_p g_{t-p} \tag{2.17}
$$

For simplicity of notation, here we ignore the fact that e_t does not depend on the elements of θ other than π.

Theorem 2.2. *(Consistency) Under Assumptions 2.1, 2.2, 2.3, 2.4 and 2.5, there exists a double indexed sequence of solutions $\hat{\theta}_{N,T}$ to the FOCs $L_{\theta,T}(\theta, \tilde{G}) = 0$ such that $\hat{\theta}_{N,T} - \theta^*_0 \xrightarrow{p} 0$ as $N, T \to \infty$.*

Proof. Consider a sphere $\Theta^*_\varepsilon = \{\theta^* : \|\theta^* - \theta^*_0\| = \varepsilon\}$ for some sufficiently small ε. We have

$$
\begin{aligned}
L_T(\theta^*, \tilde{G}) - L_T(\theta^*_0, \tilde{G}) = & \Big[L_T(\theta^*, \tilde{G}) - L_T(\theta^*, GQ_0')\Big] \\
& + \Big[L_T(\theta^*_0, GQ_0') - L_T(\theta^*_0, \tilde{G})\Big] \\
& + \Big[L_T(\theta^*, GQ_0') - L_T(\theta^*_0, GQ_0')\Big]
\end{aligned} \tag{2.18}
$$

Let $\tilde{\boldsymbol{g}}_t = [\tilde{g}_t', \ldots, \tilde{g}_{t-p}']'$ and $\boldsymbol{g}^*_t = [Q_0 g_t', \ldots, Q_0 g_{t-p}']'$.

(i) For the first term, a mean value expansion gives

$$
\begin{aligned}
& L_T(\theta^*, \tilde{G}) - L_T(\theta^*, GQ_0') \\
= & \frac{1}{T}\sum_{t=p+1}^{T}\Big[vec(\tilde{\boldsymbol{g}}_t - \boldsymbol{g}^*_t)' l_{g,t}(\theta^*, \boldsymbol{g}^\dagger_t)\Big]
\end{aligned} \tag{2.19}
$$

where $\boldsymbol{g}^\dagger_t$ is between $\tilde{\boldsymbol{g}}_t$ and $\boldsymbol{g}^*_t$. Let

$$
x_i(\theta, \boldsymbol{g}_t) = \sigma_i(\theta)^{-1}\iota_i B(\theta)^{-1} e_t(\pi(\theta), \boldsymbol{g}_t) \tag{2.20}
$$

Then

$$
\begin{aligned}
& vec(\tilde{\boldsymbol{g}}_t - \boldsymbol{g}^*_t)' l_{g,t}(\theta^*, \boldsymbol{g}^\dagger_t) \\
= & (\tilde{\boldsymbol{g}}_t - \boldsymbol{g}^*_t)' \begin{bmatrix} I_r \\ -A_1^{*'} \\ \vdots \\ -A_p^{*'} \end{bmatrix} B^{*-1'} \begin{bmatrix} \sigma_1^{*-1}\frac{f_{1,x}(x_1(\theta^*, \boldsymbol{g}^\dagger_t); \lambda_1^*)}{f_1(x_1(\theta^*, \boldsymbol{g}^\dagger_t); \lambda_1^*)} \\ \vdots \\ \sigma_r^{*-1}\frac{f_{r,x}(x_r(\theta^*, \boldsymbol{g}^\dagger_t); \lambda_r^*)}{f_r(x_r(\theta^*, \boldsymbol{g}^\dagger_t); \lambda_r^*)} \end{bmatrix}
\end{aligned} \tag{2.21}
$$

We have

$$
\begin{aligned}
& |L_T(\theta^*, \tilde{G}) - L_T(\theta^*, GQ_0')| \\
\leq & \left(\frac{1}{T}\sum_{t=p+1}^{T}\left(\sum_{s=t-p}^{t}\|\tilde{g}_s - Q_0 g_s\|^2\right)\right)^{1/2}\left(\frac{1}{T}\sum_{t=p+1}^{T}\|l_{g,t}(\theta^*, \boldsymbol{g}^\dagger_t)\|^2\right)^{1/2}
\end{aligned} \tag{2.22}
$$

By Lemma A.1 in Bai (2003), we have

$$\frac{1}{T}\sum_{t=1}^{T}\|\tilde{g}_t - Q_0 g_t\|^2 = O_p(\delta_{NT}^{-2}) \tag{2.23}$$

where $\delta_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$. Meanwhile, by Assumption 2.5, for any t, we have

$$\sup_{\theta^* \in \Theta_\varepsilon^*} \|l_{g,t}(\theta^*, \boldsymbol{g}_t^\dagger)\|^2 \le Const \cdot \sup_{\theta^* \in \Theta_\varepsilon^*} \max_{1 \le i \le r} (1 + |x_i(\theta^*, \boldsymbol{g}_t^\dagger)|^{a_1}) \tag{2.24}$$

Notice that

$$\begin{aligned} & x_i(\theta^*, \boldsymbol{g}_t^\dagger) \\ =& x_i(\theta^*, \boldsymbol{g}_t^*) + \sigma^{*-1}\iota_i B^{*-1}(g_t^\dagger - Q_0 g_t - A_1^*(g_{t-1}^\dagger - Q_0 g_{t-1}) - \cdots - A_p^*(g_{t-p}^\dagger - Q_0 g_{t-p})) \end{aligned} \tag{2.25}$$

Then by the Minkowski inequality we have

$$\begin{aligned} & \sup_{\theta^* \in \Theta_\varepsilon^*} \frac{1}{T}\sum_{t=1}^{T}|x_i(\theta^*, \boldsymbol{g}_t^\dagger)|^2 \\ \le & \left(\left(\sup_{\theta^* \in \Theta_\varepsilon^*} \frac{1}{T}\sum_{t=1}^{T}|x_i(\theta^*, \boldsymbol{g}_t^*)|^2 \right)^{1/2} + Const \cdot \left(\frac{1}{T}\sum_{t=1}^{T}\|g_t^\dagger - Q_0 g_t\|^2 \right)^{1/2} \right)^2 \\ \le & \left(\left(\sup_{\theta^* \in \Theta_\varepsilon^*} \frac{1}{T}\sum_{t=1}^{T}|x_i(\theta^*, \boldsymbol{g}_t^*)|^2 \right)^{1/2} + Const \cdot \left(\frac{1}{T}\sum_{t=1}^{T}\|\tilde{g}_t - Q_0 g_t\|^2 \right)^{1/2} \right)^2 \end{aligned} \tag{2.26}$$

From the expression $x_i(\theta^*, \boldsymbol{g}_t^*)$, due to boundedness of the coefficients for $\boldsymbol{g}_t^*$ when $\theta^* \in \Theta_\varepsilon^*$, it is straightforward to verify that $\{x_i(\theta^*, \boldsymbol{g}_t^*) | \theta^* \in \Theta_\varepsilon^*\}$ is equicontinuous (as functions of $\boldsymbol{g}_t^*$). Then by Assumption 5 and Theorem 3.2 in Rao (1962), we can establish that

$$\sup_{\theta^* \in \Theta_\varepsilon^*} \frac{1}{T}\sum_{t=1}^{T}|x_i(\theta^*, \boldsymbol{g}_t^*)|^2 = O_p(1) \tag{2.27}$$

Combining the above, we have

$$\sup_{\theta^* \in \Theta_\varepsilon^*} \frac{1}{T}\sum_{t=p+1}^{T}\|l_{g,t}(\theta^*, \boldsymbol{g}_t^\dagger)\|^2 = O_p(1) \tag{2.28}$$

Therefore

$$\sup_{\theta^* \in \Theta^*_\varepsilon} |L_T(\theta^*, \tilde{G}) - L_T(\theta^*, GQ_0')| = O_p(\delta_{NT}^{-2}) = o_p(1) \tag{2.29}$$

as $N, T \to \infty$.

(ii) Similarly, for the second term we have

$$\sup_{\theta^* \in \Theta^*_\varepsilon} |L_T(\theta_0^*, GQ_0') - L_T(\theta_0^*, \tilde{G})| = o_p(1) \tag{2.30}$$

as $N, T \to \infty$.

(iii) For the third term, notice that this term is not affected by the estimation of factors and does not depend on N. Then following the proof of Theorem 1 in Lanne et al. (2017), we have a Taylor expansion

$$\begin{aligned}
& L_T(\theta^*, GQ_0') - L_T(\theta_0^*, GQ_0') \\
=& (\theta^* - \theta_0^*)' L_{\theta,T}(\theta_0^*, GQ_0') \\
& + \frac{1}{2}(\theta^* - \theta_0^*)' \left[L_{\theta,T}(\theta^\dagger, \boldsymbol{g}_t^*) - E\left(l_{\theta\theta,t}(\theta^\dagger, \boldsymbol{g}_t^*) \right) \right] (\theta^* - \theta_0^*) \\
& + \frac{1}{2}(\theta^* - \theta_0^*)' \left[E(l_{\theta\theta,t}(\theta^\dagger, \boldsymbol{g}_t^*)) - E\left(l_{\theta\theta,t}(\theta_0^*, \boldsymbol{g}_t^*) \right) \right] (\theta^* - \theta_0^*) \\
& + \frac{1}{2}(\theta^* - \theta_0^*)' E\left(l_{\theta\theta,t}(\theta_0^*, \boldsymbol{g}_t^*) \right) (\theta^* - \theta_0^*) \\
=& S_1 + S_2 + S_3 + S_4
\end{aligned} \tag{2.31}$$

For S_1, notice that

$$e_t(\pi_0^*, \boldsymbol{g}_t^*) = Q_0 g_t - A_{1,0}^* Q_0 g_{t-1} - \cdots - A_{p,0}^* Q_0 g_{t-p} = Q_0 e_t(\pi_0, \boldsymbol{g}_t) \tag{2.32}$$

we have

$$
\begin{aligned}
L_T(\theta_0^*, GQ_0') =& \frac{1}{T}\sum_{t=p+1}^{T} l_t(\theta_0^*, \boldsymbol{g}_t^*) \\
=& \sum_{t=p+1}^{T}\left[\sum_{i=1}^{r}\log f_i(\sigma_i^{*-1}\iota_i D_2^{-1}P'D_1^{-1}B_0^{-1}Q_0^{-1}Q_0 e_t(\pi_0, G);\lambda_i^*)\right. \\
& \left. -\log|\det(Q_0B_0D_1PD_2)| - \sum_{i=1}^{r}\log\sigma_i^*\right] \\
=& \frac{1}{T}\sum_{t=p+1}^{T}(l_t(\theta_0, \boldsymbol{g}_t) - Const) \\
=& L_T(\theta_0, G) - \frac{(T-p)}{T}\cdot Const
\end{aligned}
\tag{2.33}
$$

which implies that $l_{\theta,t}(\theta_0^*, \boldsymbol{g}_t^*) = l_{\theta,t}(\theta_0, \boldsymbol{g}_t)$ and $L_{\theta,T}(\theta_0^*, GQ_0') = L_{\theta,T}(\theta_0, G)$. Hence, we have

$$
L_{\theta,T}(\theta_0^*, GQ_0') = L_{\theta,T}(\theta_0, G) \xrightarrow{p} 0 \tag{2.34}
$$

which gives $S_1 \xrightarrow{p} 0$ and thus

$$
\sup_{\theta^*\in\Theta_\varepsilon^*} S_1 \xrightarrow{p} 0 \tag{2.35}
$$

For S_2, for a compact and convex set $\Theta_0^* \subset \Theta$ that contains θ_0^* as an interior point, similar to Lemma 2 in Lanne et al. (2017), we can establish that:

$$
\sup_{\theta^*\in\Theta_0^*}\left\|\frac{1}{T}\sum_{t=p\ 1}^{T} l_{\theta\theta,t}(\theta^*, \boldsymbol{g}_t^*) - E[l_{\theta\theta,t}(\theta^*, \boldsymbol{g}_t^*)]\right\| \xrightarrow{p} 0 \tag{2.36}
$$

and that $E[l_{\theta\theta,t}(\theta^*, \boldsymbol{g}_t^*)]$ is continuous in θ^* at θ_0^*. This implies that for small enough ε such that $\Theta_\varepsilon^* \subset \Theta_0^*$, we have

$$
\sup_{\theta^*\in\Theta_\varepsilon^*} S_2 \xrightarrow{p} 0 \tag{2.37}
$$

For S_3 and S_4, we use the continuity of $E[l_{\theta\theta,t}(\theta^*, \boldsymbol{g}_t^*)]$ established above, and the negative definiteness of $E[l_{\theta\theta,t}(\theta_0^*, \boldsymbol{g}_t^*)]$. Then we have

$$
\sup_{\theta^*\in\Theta_\varepsilon^*} S_3 + S_4 < -Const\cdot\varepsilon^2 \tag{2.38}
$$

Combining (i), (ii) and (iii), we have

$$P\left(\sup_{\theta^*\in\Theta^*_\varepsilon} L_T(\theta^*,\tilde{G}) < L_T(\theta^*_0,\tilde{G})\right) \to 1 \tag{2.39}$$

as $N,T\to\infty$, which implies the existence and the consistency of a solution sequence, see Serfling (1980), pp. 147-148 and Shao (2003), pp. 290. □

To perform inference, we now derive the asymptotic distribution of the parameter estimates.

Theorem 2.3. *(Asymptotic Normality) Under Assumptions 2.1, 2.2, 2.3, 2.4 and 2.5, we have*

$$\sqrt{T}(\hat{\theta}^* - \theta^*_0) \xrightarrow{d} N(0,(-E[l_{\theta\theta,t}(\theta_0,\boldsymbol{g}_t)])^{-1}) \tag{2.40}$$

as $N,T\to\infty$ *and* $\frac{\sqrt{T}}{N}\to 0$.

In addition, $\left(-L_{\theta\theta}(\tilde{\theta}^*,\tilde{G})\right)^{-1}$ *is a consistent estimator of* $(-E[l_{\theta\theta,t}(\theta_0,\boldsymbol{g}_t)])^{-1}$.

Proof. Let $\hat{\theta}^*$ be the MLE. We have

$$0 = \sqrt{T}L_{\theta,T}(\hat{\theta}^*,\tilde{G}) = \sqrt{T}L_{\theta,T}(\theta^*_0,\tilde{G}) + \sqrt{T}L_{\theta\theta,T}(\tilde{\theta}^*,\tilde{G})(\hat{\theta}^*-\theta^*_0) \tag{2.41}$$

So

$$\sqrt{T}(\hat{\theta}^*-\theta^*_0) = (L_{\theta\theta,T}(\tilde{\theta}^*,\tilde{G}))^{-1}\sqrt{T}L_{\theta,T}(\theta^*_0,\tilde{G}) \tag{2.42}$$

For $L_{\theta\theta,T}(\tilde{\theta}^*,\tilde{G})$, we have

$$\begin{aligned} L_{\theta\theta,T}(\tilde{\theta}^*,\tilde{G}) &= \frac{1}{T}\sum_{t=p+1}^{T} l_{\theta\theta g,t}(\theta^*,\tilde{\boldsymbol{g}}_t) \\ &= \frac{1}{T}\sum_{t=p+1}^{T}\begin{bmatrix} l_{\sigma\sigma,t}(\theta^*,\tilde{\boldsymbol{g}}_t) & l_{\sigma\lambda,t}(\theta^*,\tilde{\boldsymbol{g}}_t) & l_{\sigma\pi,t}(\theta^*,\tilde{\boldsymbol{g}}_t) & l_{\sigma\beta,t}(\theta^*,\tilde{\boldsymbol{g}}_t) \\ l_{\sigma\lambda,t}(\theta^*,\tilde{\boldsymbol{g}}_t)' & l_{\lambda\lambda,t}(\theta^*,\tilde{\boldsymbol{g}}_t) & l_{\lambda\pi,t}(\theta^*,\tilde{\boldsymbol{g}}_t) & l_{\lambda\beta,t}(\theta^*,\tilde{\boldsymbol{g}}_t) \\ l_{\sigma\pi,t}(\theta^*,\tilde{\boldsymbol{g}}_t)' & l_{\lambda\pi,t}(\theta^*,\tilde{\boldsymbol{g}}_t)' & l_{\pi\pi,t}(\theta^*,\tilde{\boldsymbol{g}}_t) & l_{\pi\beta,t}(\theta^*,\tilde{\boldsymbol{g}}_t) \\ l_{\sigma\beta,t}(\theta^*,\tilde{\boldsymbol{g}}_t)' & l_{\lambda\beta,t}(\theta^*,\tilde{\boldsymbol{g}}_t)' & l_{\pi\beta,t}(\theta^*,\tilde{\boldsymbol{g}}_t)' & l_{\beta\beta,t}(\theta^*,\tilde{\boldsymbol{g}}_t) \end{bmatrix} \end{aligned} \tag{2.43}$$

Following Lanne et al. (2017), let

$$\Sigma = diag(\sigma_1, \ldots, \sigma_r) \tag{2.44}$$

$$\mathbb{g}_t = (1, g'_{t-1}, \ldots, g'_{t-p})' \tag{2.45}$$

and

$$\mathbb{e}_t(\theta, \boldsymbol{g}_t) = diag(\mathbb{e}_{1,t}(\theta, \boldsymbol{g}_t), \ldots, \mathbb{e}_{r,t}(\theta, \boldsymbol{g}_t)) \tag{2.46}$$

$$\mathbb{e}_{x,t}(\theta, \boldsymbol{g}_t) = diag(\mathbb{e}_{1,x,t}(\theta, \boldsymbol{g}_t), \ldots, \mathbb{e}_{r,x,t}(\theta, \boldsymbol{g}_t)) \tag{2.47}$$

where

$$\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t) = \iota_i' B^{-1} e_t(\theta, \boldsymbol{g}_t) \tag{2.48}$$

$$\mathbb{e}_{i,x,t}(\theta, \boldsymbol{g}_t) = \frac{f_{i,x}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} \tag{2.49}$$

In addition, define block diagonal matrices

$$\mathbb{e}_{xx,t}(\theta, \boldsymbol{g}_t) = diag(\mathbb{e}_{1,xx,t}(\theta, \boldsymbol{g}_t), \ldots, \mathbb{e}_{r,xx,t}(\theta, \boldsymbol{g}_t)) \tag{2.50}$$

$$\mathbb{e}_{\lambda\lambda,t}(\theta, \boldsymbol{g}_t) = diag(\mathbb{e}_{1,\lambda\lambda,t}(\theta, \boldsymbol{g}_t), \ldots, \mathbb{e}_{r,\lambda\lambda,t}(\theta, \boldsymbol{g}_t)) \tag{2.51}$$

$$\mathbb{e}_{x\lambda,t}(\theta, \boldsymbol{g}_t) = diag(\mathbb{e}_{1,x\lambda,t}(\theta, \boldsymbol{g}_t), \ldots, \mathbb{e}_{r,x\lambda,t}(\theta, \boldsymbol{g}_t)) \tag{2.52}$$

where

$$\mathbb{e}_{i,xx,t}(\theta, \boldsymbol{g}_t) = \frac{f_{i,xx}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} - \left(\frac{f_{i,x}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}\right)^2 \tag{2.53}$$

$$\begin{aligned} \mathbb{e}_{i,x\lambda,t}(\theta, \boldsymbol{g}_t) = {} & \frac{f'_{i,x\lambda}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} \\ & - \frac{f_{i,x}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} \frac{f'_{i,\lambda}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} \end{aligned} \tag{2.54}$$

$$\begin{aligned} \mathbb{e}_{i,\lambda\lambda,t}(\theta, \boldsymbol{g}_t) = {} & \frac{f_{i,\lambda_i\lambda_i}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} \\ & - \frac{f_{i,\lambda}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} \frac{f'_{i,\lambda}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} \end{aligned} \tag{2.55}$$

Then the blocks of the Hessian are

$$l_{\sigma\sigma,t} = \Sigma^{-2} + 2\Sigma^{-3}\mathbb{e}_t(\theta,\boldsymbol{g}_t)\mathbb{e}_{x,t}(\theta,\boldsymbol{g}_t) + \Sigma^{-4}\mathbb{e}_t^2(\theta,\boldsymbol{g}_t)\mathbb{e}_{xx,t}(\theta,\boldsymbol{g}_t) \tag{2.56}$$

$$l_{\sigma\lambda,t} = -\Sigma^{-2}\mathbb{e}_t(\theta,\boldsymbol{g}_t)\mathbb{e}_{x\lambda,t}(\theta,\boldsymbol{g}_t) \tag{2.57}$$

$$l_{\sigma\pi,t} = \mathbb{g}_t \otimes B^{-1'}(\Sigma^{-2}\mathbb{e}_{x,t}(\theta,\boldsymbol{g}_t) + \Sigma^{-3}\mathbb{e}_t(\theta,\boldsymbol{g}_t)\mathbb{e}_{xx,t}(\theta,\boldsymbol{g}_t)) \tag{2.58}$$

$$l_{\sigma\beta,t} = \mathbb{H}'(B^{-1}\otimes B^{-1'})[e_t(\theta,\boldsymbol{g}_t)\otimes(\Sigma^{-2}\mathbb{e}_{x,t}(\theta,\boldsymbol{g}_t) + \Sigma^{-3}\mathbb{e}_t(\theta,\boldsymbol{g}_t)\mathbb{e}_{xx,t}(\theta,\boldsymbol{g}_t))] \tag{2.59}$$

$$l_{\lambda\lambda,t} = \mathbb{e}_{\lambda\lambda,t} \tag{2.60}$$

$$l_{\lambda\pi,t} = -(I_{rp+1}\otimes B^{-1'}\Sigma^{-1})(\mathbb{g}_t\otimes\mathbb{e}_{x\lambda,t}(\theta,\boldsymbol{g}_t)) \tag{2.61}$$

$$l_{\lambda\beta,t} = -\mathbb{H}'(B^{-1}\otimes B^{-1'}\Sigma^{-1})(e_t(\theta,\boldsymbol{g}_t)\otimes\mathbb{e}_{x\lambda,t}(\theta,\boldsymbol{g}_t)) \tag{2.62}$$

$$l_{\pi\pi,t} = (I_r\otimes B^{-1'}\Sigma^{-1})(\mathbb{g}_t\mathbb{g}_t'\otimes\mathbb{e}_{xx,t}(\theta,\boldsymbol{g}_t))(I_r\otimes B^{-1'}\Sigma^{-1})' \tag{2.63}$$

$$\begin{aligned} l_{\pi\beta,t} &= \mathbb{g}_t\otimes\left[\left(I_r\otimes diag(\mathbb{e}_{x,t}(\theta,\boldsymbol{g}_t))'\right)(B^{-1'}\otimes\Sigma^{-1}B^{-1})\mathbb{H}\right] \\ &\quad + \mathbb{g}_t\otimes\left[B^{-1'}\Sigma^{-1}\left(e_t(\theta,\boldsymbol{g}_t)'\otimes\mathbb{e}_{xx,t}(\theta,\boldsymbol{g}_t)\right)(B^{-1'}\otimes\Sigma^{-1}B^{-1})\mathbb{H}\right] \end{aligned} \tag{2.64}$$

$$\begin{aligned} l_{\beta\beta,t} &= \mathbb{H}'(B^{-1}\otimes B^{-1'})(e_t(\theta,\boldsymbol{g}_t)e_t(\theta,\boldsymbol{g}_t)'\otimes\mathbb{e}_{xx,t}(\theta,\boldsymbol{g}_t))(B^{-1'}\otimes\Sigma^{-1}B^{-1})\mathbb{H} \\ &\quad + \mathbb{H}'(B^{-1}\otimes I_r)\left((e_t(\theta,\boldsymbol{g}_t)diag(\mathbb{e}_{x,t}(\theta,\boldsymbol{g}_t))')\otimes I_r\right)(\Sigma^{-1}B^{-1}\otimes B^{-1'}K_{rr}\mathbb{H}) \\ &\quad + \mathbb{H}'K_{rr}(B^{-1'}\Sigma^{-1}\otimes B^{-1})\left((diag(\mathbb{e}_{x,t}(\theta,\boldsymbol{g}_t))e_t(\theta,\boldsymbol{g}_t))\otimes I_r\right)(B^{-1'}\otimes I_r)\mathbb{H} \\ &\quad + \mathbb{H}'(B^{-1}\otimes B^{-1'})K_{rr}\mathbb{H} \end{aligned} \tag{2.65}$$

where $\mathbb{H}$ is a special elimination matrix defined following footnote 8 of Lanne et al. (2017), and K_{rr} is the commutation matrix such that $K_{rr}vec(A) = vec(A')$ for any $r\times r$ matrix A.

Now notice that

$$\begin{aligned} &vech(L_{\theta\theta,T}(\tilde{\theta}^*,\tilde{G}) - L_{\theta\theta,T}(\tilde{\theta}^*,GQ_0')) \\ =&\frac{1}{T}\sum_{t=p+1}^{T}\left[vec(\tilde{\boldsymbol{g}}_t - Q_0\boldsymbol{g}_t)'\frac{\partial vech(l_{\theta\theta,t}(\theta^*,\boldsymbol{g}_t^\dagger))}{\partial\boldsymbol{g}}\right] \end{aligned} \tag{2.66}$$

To characterize $\frac{\partial vech(l_{\theta\theta,t}(\theta^*,\boldsymbol{g}_t^\dagger))}{\partial\boldsymbol{g}}$, consider the derivative of an arbitrary element from each block.

First for $l_{\sigma\sigma,t}$, let

$$
\begin{aligned}
\mathbb{e}_{i,xxx,t}(\theta,\boldsymbol{g}_t) =& \frac{f_{i,xxx}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta,\boldsymbol{g}_t);\lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta,\boldsymbol{g}_t);\lambda_i)} + 2\left(\frac{f_{i,x}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta,\boldsymbol{g}_t);\lambda_i)}{f_i(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta,\boldsymbol{g}_t);\lambda_i)}\right)^3 \\
&-3\frac{f_{i,xx}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta,\boldsymbol{g}_t);\lambda_i)f_{i,x}(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta,\boldsymbol{g}_t);\lambda_i)}{f_i^2(\sigma_i^{-1}\mathbb{e}_{i,t}(\theta,\boldsymbol{g}_t);\lambda_i)}
\end{aligned} \tag{2.67}
$$

Then consider the i-th diagonal element of $l_{\sigma\sigma,t}(\theta^*,\boldsymbol{g}_t^\dagger)$, we have

$$
\begin{aligned}
&\frac{\partial}{\partial g_{t-s}}\left(\sigma_i^{*-2} + 2\sigma_i^{*-3}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{i,x,t}(\theta^*,\boldsymbol{g}_t^\dagger) + \sigma_i^{*-4}\mathbb{e}_{i,t}^2(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{i,xx,t}(\theta^*,\boldsymbol{g}_t^\dagger)\right) \\
=&\Big(2\sigma_i^{*-3}\mathbb{e}_{i,x,t}(\theta^*,\boldsymbol{g}_t^\dagger) + 2\sigma_i^{*-4}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{i,xx,t}(\theta^*,\boldsymbol{g}_t^\dagger) + 2\sigma_i^{*-4}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{i,xx,t}(\theta^*,\boldsymbol{g}_t^\dagger) \\
&+\sigma_i^{*-5}\mathbb{e}_{i,t}^2(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{i,xxx,t}(\theta^*,\boldsymbol{g}_t^\dagger)\Big)\iota_i'B^{*-1}(-A_s^*)
\end{aligned} \tag{2.68}
$$

For $l_{\sigma\lambda,t}$, consider the j-th element of its i-th diagonal block, we have

$$
\begin{aligned}
&\frac{\partial}{\partial g_{t-s}}\left(-\sigma_i^{*-2}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{i,x\lambda_i,t}^{(j)}(\theta^*,\boldsymbol{g}_t^\dagger)\right) \\
=&\Bigg[\mathbb{e}_{i,x\lambda_i,t}^{(j)}(\theta^*,\boldsymbol{g}_t^\dagger) - \mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger)\Bigg(\frac{f_{i,x\lambda}^{(j)}(\sigma_i^{*-1}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger);\lambda_i)}{f_i(\sigma_i^{*-1}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger);\lambda_i)} \\
&-\frac{f_{i,xx}(\sigma_i^{*-1}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger);\lambda_i)f_{i,\lambda}^{(j)}(\sigma_i^{*-1}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger);\lambda_i)}{f_i^2(\sigma_i^{*-1}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger);\lambda_i)} \\
&+\frac{2f_{i,x}^2(\sigma_i^{*-1}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger);\lambda_i)f_{i,\lambda}^{(j)}(\sigma_i^{*-1}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger);\lambda_i)}{f_i^3(\sigma_i^{*-1}\mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger);\lambda_i)}\Bigg)\Bigg]\sigma^{*-3}\iota_i'B^*A_s^* \\
\equiv&\left(\mathbb{e}_{i,x\lambda_i,t}^{(j)}(\theta^*,\boldsymbol{g}_t^\dagger) - \mathbb{e}_{i,t}(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{i,xx\lambda_i}^{(j)}\right)\sigma^{*-3}\iota_i'B^*A_s^*
\end{aligned} \tag{2.69}
$$

Now $l_{\sigma\pi,t}$ is a Kronecker product of two matrices, so we consider the product of arbitrary elements of each matrix. We have

$$
\begin{aligned}
&\frac{\partial}{\partial g_{t-s}}\left[g_{t-j}^{\dagger(m)}(B^{*-1})^{(n,m)}\left(\sigma_n^{*-2}\mathbb{e}_{n,x,t}(\theta^*,\boldsymbol{g}_t^\dagger) + \sigma_n^{*-3}\mathbb{e}_{n,t}(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{n,xx,t}(\theta^*,\boldsymbol{g}_t^\dagger)\right)\right] \\
=&(B^{*-1})^{(n,m)}\Bigg[\left(\sigma_n^{*-2}\mathbb{e}_{n,x,t}(\theta^*,\boldsymbol{g}_t^\dagger) + \sigma_n^{*-3}\mathbb{e}_{n,t}(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{n,xx,t}(\theta^*,\boldsymbol{g}_t^\dagger)\right)\iota_m \\
&+g_{t-j}^{\dagger(m)}\left(2\sigma_n^{*-3}\mathbb{e}_{n,xx,t}(\theta^*,\boldsymbol{g}_t^\dagger) + \sigma_n^{*-4}\mathbb{e}_{n,t}(\theta^*,\boldsymbol{g}_t^\dagger)\mathbb{e}_{n,xxx,t}(\theta^*,\boldsymbol{g}_t^\dagger)\right)\iota_n'B^{*-1}(-A_s^*)\Bigg]
\end{aligned} \tag{2.70}
$$

For $l_{\sigma\beta,t}$, notice that $\mathbb{H}'(B^{*-1}\otimes B^{*-1'})$ does not depend on G, so we consider an element from

$[e_t(\theta^*, \boldsymbol{g}_t^\dagger) \otimes (\Sigma^{-2} \mathbb{e}_{x,t}(\theta^*, \boldsymbol{g}_t^\dagger) + \Sigma^{*-3} \mathbb{e}_t(\theta^*, \boldsymbol{g}_t^\dagger) \mathbb{e}_{xx,t}(\theta^*, \boldsymbol{g}_t^\dagger))]$.

Similar as above, we obtain

$$\begin{aligned}
&\frac{\partial}{\partial g_{t-s}} \left[e_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger) \left(\sigma_j^{*-2} \mathbb{e}_{j,x,t}(\theta^*, \boldsymbol{g}_t^\dagger) + \sigma_j^{*-3} \mathbb{e}_{j,t}(\theta^*, \boldsymbol{g}_t^\dagger) \mathbb{e}_{j,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger) \right) \right] \\
=& \left(\sigma_j^{*-2} \mathbb{e}_{j,x,t}(\theta^*, \boldsymbol{g}_t^\dagger) + \sigma_j^{*-3} \mathbb{e}_{j,t}(\theta^*, \boldsymbol{g}_t^\dagger) \mathbb{e}_{j,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger) \right) \iota_i'(-A_s^*) \\
&+ e_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger) \left(2\sigma_j^{*-3} \mathbb{e}_{j,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger) + \sigma_j^{*-4} \mathbb{e}_{j,t}(\theta^*, \boldsymbol{g}_t^\dagger) \mathbb{e}_{j,xxx,t}(\theta^*, \boldsymbol{g}_t^\dagger) \right) \iota_j' B^{*-1}(-A_s^*)
\end{aligned} \tag{2.71}$$

For $l_{\lambda\lambda,t}$, consider the (j,k)-th element of its i-th diagonal block, we have

$$\begin{aligned}
&\frac{\partial}{\partial g_{t-s}} \left(\mathbb{e}_{i,\lambda_i\lambda_i,t}^{(j,k)}(\theta^*, \boldsymbol{g}_t^\dagger) \right) \\
=& \Bigg\{ \frac{f_{i,x\lambda\lambda}^{(j,k)}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i)}{f_i(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t); \lambda_i)} - \frac{f_{i,\lambda\lambda}^{(j,k)}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i) f_{i,x}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)}{f_i^2(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta, \boldsymbol{g}_t); \lambda_i)} \\
&- \frac{1}{f_i^2(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i)} \Big[f_{i,\lambda}^{(j)}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i) f_{i,x\lambda}^{(k)}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i) \\
&+ f_{i,x\lambda}^{(j)}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i) f_{i,\lambda}^{(k)}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i) \Big] \\
&- \frac{f_{i,\lambda}^{(j)}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i) f_{i,\lambda}^{(k)}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i) f_{i,x}(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i)}{f_i^3(\sigma_i^{*-1} \mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger); \lambda_i)} \Bigg\} \\
&\cdot \sigma^{*-1} \iota_i' B^{*-1}(-A_s^*) \\
\equiv& \mathbb{e}_{i,\lambda_i\lambda_i x}^{(j,k)} \sigma^{*-1} \iota_i' B^{*-1}(-A_s^*)
\end{aligned} \tag{2.72}$$

For $l_{\lambda\pi,t}$, the first term $(I_{rp+1} \otimes B^{*-1}\Sigma^{*-1})$ does not depend on G, thus we consider an element of the second term $\mathbb{g}_t \otimes \mathbb{e}_{x\lambda,t}(\theta^*, \boldsymbol{g}_t^\dagger)$.

$$\begin{aligned}
&\frac{\partial}{\partial g_{t-s}} \left(g_{t-i}^{\dagger(m)} \mathbb{e}_{j,x\lambda_j,t}^{(n)}(\theta^*, \boldsymbol{g}_t^\dagger) \right) \\
=& \mathbb{1}\{i = s\} \mathbb{e}_{j,x\lambda_j,t}^{(n)}(\theta^*, \boldsymbol{g}_t^\dagger) \iota_m + \sigma_j^{*-1} g_{t-i}^{\dagger(m)} \mathbb{e}_{j,xx\lambda_j,t}^{(n)}(\theta^*, \boldsymbol{g}_t^\dagger) \iota_j' B^{*-1}(-A_s^*)
\end{aligned} \tag{2.73}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function.

For $l_{\lambda\beta,t}$, again $-\mathbb{H}'(B^{-1} \otimes B^{*-1'}\Sigma^{*-1})$ does not depend on G, so we consider an element of $e_t(\theta^*, \boldsymbol{g}_t^\dagger) \otimes \mathbb{e}_{x\lambda,t}(\theta^*, \boldsymbol{g}_t^\dagger)$.

$$\frac{\partial}{\partial g_{t-s}} \left(e_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger) \mathbb{e}_{j,x\lambda_j,t}^{(m)}(\theta^*, \boldsymbol{g}_t^\dagger) \right) = \iota_i'(-A_s^*) + \mathbb{e}_{j,xx\lambda_j,t}^{(m)}(\theta^*, \boldsymbol{g}_t^\dagger) \iota_j' B^{*-1}(-A_s^*) \tag{2.74}$$

For $l_{\pi\pi,t}$, notice that $(I_r \otimes B^{*-1'}\Sigma^{*-1})$ does not depend on G, we consider an element from the middle component $(\mathbb{g}_t\mathbb{g}_t' \otimes \mathbb{e}_{xx,t}(\theta, \boldsymbol{g}_t))$:

$$\begin{aligned}\frac{\partial}{\partial g_{t-s}}\left(g_{t-i}^{\dagger(m)} g_{t-j}^{\dagger(n)} \mathbb{e}_{i,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger)\right) =& \mathbb{e}_{i,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger)\left(\mathbb{1}\{i=s\}\iota_m + \mathbb{1}\{j=s\}\iota_n\right) \\ &+ g_{t-i}^{\dagger(m)} g_{t-j}^{\dagger(n)} \mathbb{e}_{i,xxx,t}(\theta^*, \boldsymbol{g}_t^\dagger)\sigma^{*-1}\iota_i' B^{*-1}(-A_s^*)\end{aligned} \tag{2.75}$$

For $l_{\pi\beta,t}$, again we have $B^{*-1'} \otimes \Sigma^{*-1}B^{*-1}\mathbb{H}$ and $B^{*-1'}\Sigma^{*}-1'$ does not depend on G. Then for the first term, we have

$$\frac{\partial}{\partial g_{t-s}} g_{t-i}^{\dagger(m)} \mathbb{e}_{j,x,t}(\theta^*, \boldsymbol{g}_t^\dagger) = \mathbb{e}_{j,x,t}(\theta^*, \boldsymbol{g}_t^\dagger)\mathbb{1}\{i=s\}\iota_m + g_{t-i}^{(m)} \mathbb{e}_{j,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger)\sigma^{*-1}\iota_j' B^{*-1}(-A_s^*) \tag{2.76}$$

For the second term, we have

$$\begin{aligned}&\frac{\partial}{\partial g_{t-s}} g_{t-i}^{\dagger(m)} e_{j,t}(\theta^*, \boldsymbol{g}_t^\dagger)\mathbb{e}_{k,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger) \\ =& e_{j,t}(\theta^*, \boldsymbol{g}_t^\dagger)\mathbb{e}_{k,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger)\mathbb{1}\{i=s\}\iota_m \\ &+ g_{t-i}^{\dagger(m)}\left(e_{k,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger)\iota_j' + e_{j,t}(\theta^*, \boldsymbol{g}_t^\dagger)\mathbb{e}_{k,xxx,t}\sigma^{*-1}\iota_k'\right)B^{*-1}(-A_s^*)\end{aligned} \tag{2.77}$$

Now that we have the above characterization of $\frac{\partial vech(l_{\theta\theta,t}(\theta^*, \boldsymbol{g}_t^\dagger))}{\partial \boldsymbol{g}}$, we notice that Assumption 2.5 implies bounds on

$$\begin{aligned}&\|\mathbb{e}_{i,t}(\theta^*, \boldsymbol{g}_t^\dagger)\|,\ \|\mathbb{e}_{i,x,t}(\theta^*, \boldsymbol{g}_t^\dagger)\|, \\ &\|\mathbb{e}_{i,xx,t}(\theta^*, \boldsymbol{g}_t^\dagger)\|,\ \|\mathbb{e}_{i,x\lambda_i,t}(\theta^*, \boldsymbol{g}_t^\dagger)\|, \\ &\|\mathbb{e}_{i,xxx,t}(\theta^*, \boldsymbol{g}_t^\dagger)\|,\ \|\mathbb{e}_{i,xx\lambda_i,t}(\theta^*, \boldsymbol{g}_t^\dagger)\|, \text{ and } \|\mathbb{e}_{i,x\lambda_i\lambda_i}(\theta^*, \boldsymbol{g}_t^\dagger)\|\end{aligned} \tag{2.78}$$

by

$$Const \cdot \sup_{\theta^* \in \Theta_\varepsilon^*}\left(1 + |x_i(\theta^*, \boldsymbol{g}_t^\dagger)|^c\right) \tag{2.79}$$

where $c \in \{c_1, a_1, a_2, a_3, b_1, b_2\}$ is the corresponding power.

Then following the same arguments as in eq(2.22) to eq(2.29), we obtain

$$L_{\theta\theta,T}(\tilde{\theta}^*, \tilde{G}) - L_{\theta\theta,T}(\tilde{\theta}^*, GQ_0') \xrightarrow{p} 0 \tag{2.80}$$

Combined with eq(2.36), we have

$$L_{\theta\theta,T}(\tilde{\theta}^*,\tilde{G}) \xrightarrow{p} E(l_{\theta\theta,t}(\theta_0^*,\boldsymbol{g}_t^*)) = E(l_{\theta\theta,t}(\theta_0,\boldsymbol{g}_t)) \tag{2.81}$$

Now notice that

$$\sqrt{T}L_{\theta,T}(\theta_0^*,\tilde{G}) = (\sqrt{T}L_{\theta,T}(\theta_0^*,\tilde{G}) - \sqrt{T}L_{\theta,T}(\theta_0^*,GQ_0')) + \sqrt{T}L_{\theta,T}(\theta_0^*,GQ_0') \tag{2.82}$$

For the second term, by Theorem 1 in Lanne et al. (2017), we have

$$\sqrt{T}L_{\theta,T}(\theta_0^*,GQ_0') \xrightarrow{d} N(0,(-E[l_{\theta\theta,t}(\theta_0,\boldsymbol{g}_t)])^{-1}) \tag{2.83}$$

Meanwhile, for the first term, recall that we previously showed that

$$\sup_{\theta^*\in\Theta_\varepsilon^*} |L_T(\theta^*,\tilde{G}) - L_T(\theta^*,GQ_0')| = O_p(\delta_{NT}^{-2}) \tag{2.84}$$

which implies that

$$\sup_{\theta^*\in\Theta_\varepsilon^*} |\sqrt{T}\left(L_T(\theta^*,\tilde{G}) - L_T(\theta^*,GQ_0')\right)| = O_p(\sqrt{T}\delta_{NT}^{-2}) = o_p(1) \tag{2.85}$$

as $N,T \to \infty$ when $\frac{\sqrt{T}}{N} \to 0$. Combining all of the above, we obtain that

$$\sqrt{T}(\hat{\theta}^* - \theta_0^*) \xrightarrow{d} N(0,(-E[l_{\theta\theta,t}(\theta_0,\boldsymbol{g}_t)])^{-1}) \tag{2.86}$$

as $N,T \to \infty$ and $\frac{\sqrt{T}}{N} \to 0$. To estimate the asymptotic variance, notice that eq(2.81) implies that

$$\left(-L_{\theta\theta}(\tilde{\theta}^*,\tilde{G})\right)^{-1} \xrightarrow{p} (-E(l_{\theta\theta,t}(\theta_0,g_t)))^{-1} \tag{2.87}$$

Therefore $\left(-L_{\theta\theta}(\tilde{\theta}^*,\tilde{G})\right)^{-1}$ is a consistent estimator of the asymptotic variance. □

The above result, Theorem 2.3, implies that the first step principal component estimation does not affect the asymptotic variance of the second step estimates. This is consistent with the result in Bai and Ng (2006) where they consider a least square second step. Gonçalves and Perron (2014) shows that under the assumption $\frac{\sqrt{T}}{N} \to c \neq 0$, such result no longer holds and an asymptotic bias will be introduced in the least square second step. For our maximum likelihood second step, a similar result might also hold, but we leave this issue for future work. In our simulation practice, we will evaluate the impact of this assumption in finite samples.

Other than inference, the above asymptotic normality result also implies that standard testing procedures could be used to test many conventional identification assumptions, as discussed in Lanne et al. (2017). For example, short-run restrictions, which normally involves assuming some elements of B to be zero, could be tested via standard Wald or likelihood ratio tests. Although already stressed in the literature, we reiterate the following remarks due to their importance in empirical research. First, not all restrictions could be tested, since we restrict B to have unit diagonal elements. Second, the test results should be interpreted as under a specific ordering, rather than under all ordering. To be more specific, for example, when testing $B(1,3) = B(2,3) = 0$ when there are three factors in total, this should be interpreted as the first two factors do not respond to the third shock contemporaneously, rather than that there exists a shock that do not affect two of the factors. These problems are no different from those of testing identification restrictions in structural VAR, which are further discussed in section 4.6 in Lanne et al. (2017). In a FAVAR context, the identification assumptions, while are often mathematically identical or similar to the structural VAR identification restrictions, sometimes involve a slightly different economic interpretation due to the factor estimation, which the researcher should be careful with. We will discuss this point further in Section 2.5.

2.4 Simulation

In this section, we examine the finite sample performance of the proposed method via Monte Carlo simulations. Following Bai et al. (2016) and Yamamoto and Hara (2022), we consider a data generating process with a VAR(1) for the autoregressive part, as following:

$$x_t = \begin{bmatrix} \Lambda & \Gamma \end{bmatrix} \begin{bmatrix} f_t \\ y_t \end{bmatrix} + u_t \tag{2.88}$$

$$g_t = Ag_{t-1} + B\varepsilon_t \tag{2.89}$$

For the VAR model(2.88), we set $r_1 = 2$ and $r_2 = 1$ so that $r = 3$.

$$A = \begin{bmatrix} 0.6 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}, \quad B = \begin{bmatrix} 1.0 & 0.4 & 0.8 \\ 0.2 & 1.0 & -0.3 \\ -0.6 & 0.4 & 1.0 \end{bmatrix} \tag{2.90}$$

For the factor model(2.89), the idiosyncratic errors are i.i.d. across both time and cross-sections, generated from the standard normal distribution. The factor loading are generated as an $r \times 1$ random vector where each element is independently drawn from the standard normal distribution. Since we assume the factor loading are deterministic, we keep the realization of the factor loading the same through different Monte Carlo repetitions. We set the total number of simulations to $M = 1000$. The structural errors ε_{it} are generated using independent t location scale distributions, with the location parameter being 0, scale parameter $\sigma_i = 1$ and degree of freedom $\lambda_i = 6$. We focus on the estimation and inference of the structural impulse response of the third factor, i.e. the observed factor, to the third shock. We assume that the number of unobserved factor r_1 is known.

To construct confidence intervals for impulse responses, the theoretical results in previous sections implies a Delta method approach. However, in practice, the derivatives of the impulse responses with respect to parameters, especially the shape parameters, can be difficult to evaluate depending on the choice of distribution. To address this, we also consider a residual-based bootstrap method. The detailed steps are:

1. Estimate the model and obtaining the parameter estimates and the residuals $\hat{u}_t$ and $\hat{\varepsilon}_t$.
2. Demean $\hat{\varepsilon}_t$ over T and i.i.d. resample to obtain $e_t^\dagger$. Resample $\hat{u}_t$ i.i.d. as well to obtain $u_t^\dagger$.
3. Generate the bootstrap sample $G^\dagger$ and $X^\dagger$ through the parameter estimates.
4. Regress $X^\dagger$ on $G^\dagger$ to get bootstrap estimates of factor loading. Use $G^\dagger$ to obtain maximum likelihood estimates of the VAR parameters. Calculate the impulse responses of interest using the bootstrap estimates.
5. Repeat step 2-4 500 times.
6. Construct the confidence intervals using the bootstrap impulse responses.

This takes into the account of our theoretical results which suggest that the first step estimation has no effect on the second step variances asymptotically. For a more detailed discussion on bootstrap methods in two-step estimation of FAVAR, see Gonçalves and Perron (2014) and Yamamoto (2019). The simulation results are presented in Table 2.1

Table 2.1: Coverage rates of 90% confidence intervals

	coverage rates					
T=50, N=50	$h=0$	1	2	3	4	5
Delta method	0.79	0.84	0.82	0.82	0.82	0.83
Bootstrap	0.87	0.81	0.82	0.82	0.85	0.87
T=50, N=150	$h=0$	1	2	3	4	5
Delta method	0.81	0.87	0.83	0.84	0.84	0.85
Bootstrap	0.88	0.85	0.84	0.88	0.89	0.90
T=150, N=50	$h=0$	1	2	3	4	5
Delta method	0.77	0.82	0.82	0.81	0.78	0.81
Bootstrap	0.88	0.84	0.83	0.84	0.87	0.88
T=150, N=150	$h=0$	1	2	3	4	5
Delta method	0.81	0.88	0.86	0.84	0.86	0.85
Bootstrap	0.93	0.87	0.84	0.86	0.90	0.91
T=300, N=300	$h=0$	1	2	3	4	5
Delta method	0.84	0.87	0.87	0.87	0.89	0.89
Bootstrap	0.92	0.86	0.88	0.91	0.90	0.91

We can see that both the Delta method and bootstrap confidence intervals performed reasonably well, especially when the sample sizes increase.

Based on the simulation, we make the following observation:

1. While the bootstrap confidence intervals have good coverage rates across the board, the coverage rate distortion does not seem to completely disappear even when $T = N = 300$. A possible explanation for this could be our assumption of $\frac{\sqrt{T}}{N} \to 0$. As demonstrated in Gonçalves and Perron (2014), an asymptotic bias arise when $\frac{\sqrt{T}}{N} \to c$ where $c \neq 0$, for a two-step estimator of FAVAR with principal component first step and linear regression second step, in contrast to the result in Bai and Ng (2006). Although we have a maximum likelihood second step, it is likely that under when $\frac{\sqrt{T}}{N} \to c \neq 0$, our method also results in a bias. To remedy this, we attempt an alternative bootstrap procedure that add re-estimation of the factors in step

4, although results (not reported) show no significant difference. This is consistent with the simulation results in Yamamoto (2019) that such correction might not significantly improve the coverage rates unless T is significantly larger than N.

2. The Delta method intervals seem to suffer significantly when T is significantly larger than N, which is a sign of the potential violation of our assumption of $\frac{\sqrt{T}}{N} \to 0$. This is possibly explained by the potential bias we discussed in the earlier point. However, it is interesting that the bootstrap intervals do not seem to suffer as much. It might also relate to the computational problem in the next point.

3. The Delta method intervals of impact period response seem to show signs of under coverage, although the coverage rates gradually increase as T increases. This might be a computational issue, as we use the numerical hessian as the estimator for variance. Alternatively, we use the numerical score to construct the estimator, but the result (not reported) is largely similar.

2.5 Empirical application

In this section, we replicate the results of Bernanke et al. (2005) as a demonstration of our method.

To investigate the role of monetary policy shocks in explaining economic fluctuations, Bernanke et al. (2005) proposed and implemented FAVAR to incorporate more information compared to a small scale baseline structural VAR. They examine monthly data of 120 macroeconomic variables from January 1959 through August 2001, and found that by incorporating these additional variables, the price puzzle is significantly attenuated compared to a baseline 3 variable structural VAR model.

To identify the structural matrix, Bernanke et al. (2005) chooses the federal funds rate as the observed factor and assumed that the unobserved factors are all slow moving, i.e. do not respond to monetary policy shocks contemporaneously. To properly justify this, part of the macroeconomic variables are assumed to have no contemporaneous response to a monetary policy shock.

On the contrary, we do not need to impose contemporaneous response restrictions on any variable. Our method, however, as in Lanne et al. (2017), does not naturally imply economic interpretation of the structural shocks. To achieve identification, we inspect impulse responses for the federal funds rate, and choose the shock that incurs a positive response at impact which decreases gradually as horizon increases. While not reported here, the responses of most observed macroeconomic variables to the four statistically identified structural shocks are quite distinct, such that the selection of variables used to economically label the monetary shock is not crucial.

To better compare the results, we use the same data set as in Bernanke et al. (2005). A detailed description of the variables and the corresponding transformation (first difference and logarithms, etc.) can be found in their Appendix 1. The baseline specification of Bernanke et al. (2005) includes three unobserved factors and thirteen lags, while five unobserved factors and seven lags are also investigated respectfully and yield similar results as the baseline specification. Considering the computational burden of maximum likelihood estimation, we include three unobserved factors and seven lags. We assume the structural shocks all follow t-distribution, which is widely used in empirical literature as a heavier tailed replacement of Gaussian distribution. Due to our theoretical assumptions, we restrict the degrees of freedoms to be larger than 4. The impulse responses are reported in standard deviation units, and are normalized such that the federal funds rate has a 25-basis point increase at impact from the monetary policy shock.

We first estimate the model and conduct a Wald test on the assumption that all the unobserved factors do not respond to the monetary shock contemporaneously. The test yield a p-value of 0.0355, which rejects the null hypothesis. Notice that this is not a test of the economical identification assumption in Bernanke et al. (2005). They a priori categorize the macroeconomic variables in slow and fast categories, and by assuming that the unobserved factors do not respond to the monetary shock contemporaneously, they also impose additional assumptions on the factor loading (see Stock and Watson (2016) for a detailed discussion and Han (2015) for an over-identification test on these type of restrictions). Our test, to be more specific, shows that if we use information from all variables without imposing these assumptions, then the factors indeed respond to monetary shock contemporaneously. Alternatively, we extract the factors assuming slow and fast moving variables as in Bernanke et al. (2005). Now a Wald test on the no contemporaneous response assumption yields a p-value of 0.4357, which no longer rejects the null. Again, this does not directly verify their identification assumption. However, it does seem to indicate that their extraction method, under their assumption, correctly produces factors that have no contemporaneous response to the monetary shock.

To further investigate the empirical implication of the aforementioned problem, we also examine the responses to a monetary policy shock of the 20 macroeconomic variables as in figure II-V in Bernanke et al. (2005). The impulse responses and the 90% confidence intervals, constructed using the bootstrap procedure given in Section 2.4, are presented in Figure 2.1.

To better demonstrate the differences between our results and figure II in Bernanke et al. (2005), we also plot their point estimates along with our results in Figure 2.2. Their point estimates are labeled as IRF-BBE, while our point estimates and confidence intervals are labeled as IRF-nonG and CI-nonG. We also provide a comparison of the 90% confidence intervals in Figure 2.3, where their confidence intervals are labeled CI-BBE and our confidence intervals are labeled CI-nonG.

The results show that, while our testing procedure indicates problems in the identification strategy in Bernanke et al. (2005), the impulse responses, especially the point estimates, are not significantly different from theirs. The point estimate of the response of CPI to monetary shocks still suggests the existence of price puzzle. However, the lower bounds of the confidence intervals are now negative across all horizons, which allows for non-existence of price puzzle. This result is in stark contrast to the results in Bernanke et al. (2005). Their two-step estimator produces confidence intervals that are non-negative at least in the very short run, and while their Bayesian estimator allows for no price puzzle, the confidence intervals are too wide to be properly interpreted.

In addition, our method produces a significantly larger trough effect on industrial production, and the negative impact seem to continue well past the 48 months mark. Similarly, the effect on exchange rate also seems to be more persistent compared to Bernanke et al. (2005). One final major difference is that several real variables, such as total employment and unemployment rate, seem to respond to the monetary shock contemporaneously, which is consistent with the testing result we discussed earlier.

2.6 Conclusions

This chapter proposes a method of incorporating the identification strategy using non-Gaussianity assumptions, recently developed in the structural VAR literature, into FAVAR settings. We first show that the structural impulse responses of the observed variables are point identified, regardless of the potential rotation, permutation and scaling introduced by factor estimation and VAR estimation. Then following such idea, we propose a two-step estimator with a principal component first stage to estimate the factors and a maximum likelihood second stage to estimate the VAR parameters, and provide the relevant asymptotic results for the estimator. The asymptotic results not only provide the foundation for inference on the structural impulse responses, but also implies testing methods for many conventional identification restrictions. We also conduct a simulation study that demonstrates the construction of confidence intervals using our proposed estimator, and the results from the study confirms the finite sample validity our method.

To further demonstrate the application of our method in an empirical setting, we revisit the data set used in Bernanke et al. (2005) and compare the results. Despite obtaining impulse responses of relatively similar shape for the key macroeconomic variables, we do note two key differences in the results: First, our results, while not completely eliminating price puzzle, also do not support its existence as strongly as Bernanke et al. (2005). Furthermore, our results also suggest a stronger and more persistent effect of monetary policy on output and exchange rate.

As to potential future work, there remains several important unresolved issues on identification through non-Gaussianity in FAVAR models. From a technical viewpoint, the estimator proposed in our study requires correct specification of the likelihood. Several recent studies in structural VAR with non-Gaussian errors suggest that such issue could be avoided through Quasi-MLE or GMM estimators, and future research could further investigate whether these methods could be incorporated in the estimation and inference in FAVAR. From an empirical perspective, the general issue of labeling shocks under a statistical identification scheme could use more work. In addition, allowing usage of categorical variables and variables of different frequencies could significantly expand the possibility of empirical application. We hope that this study stimulates future research to address these directions.

Figure 2.1: Impulse responses to monetary policy shock

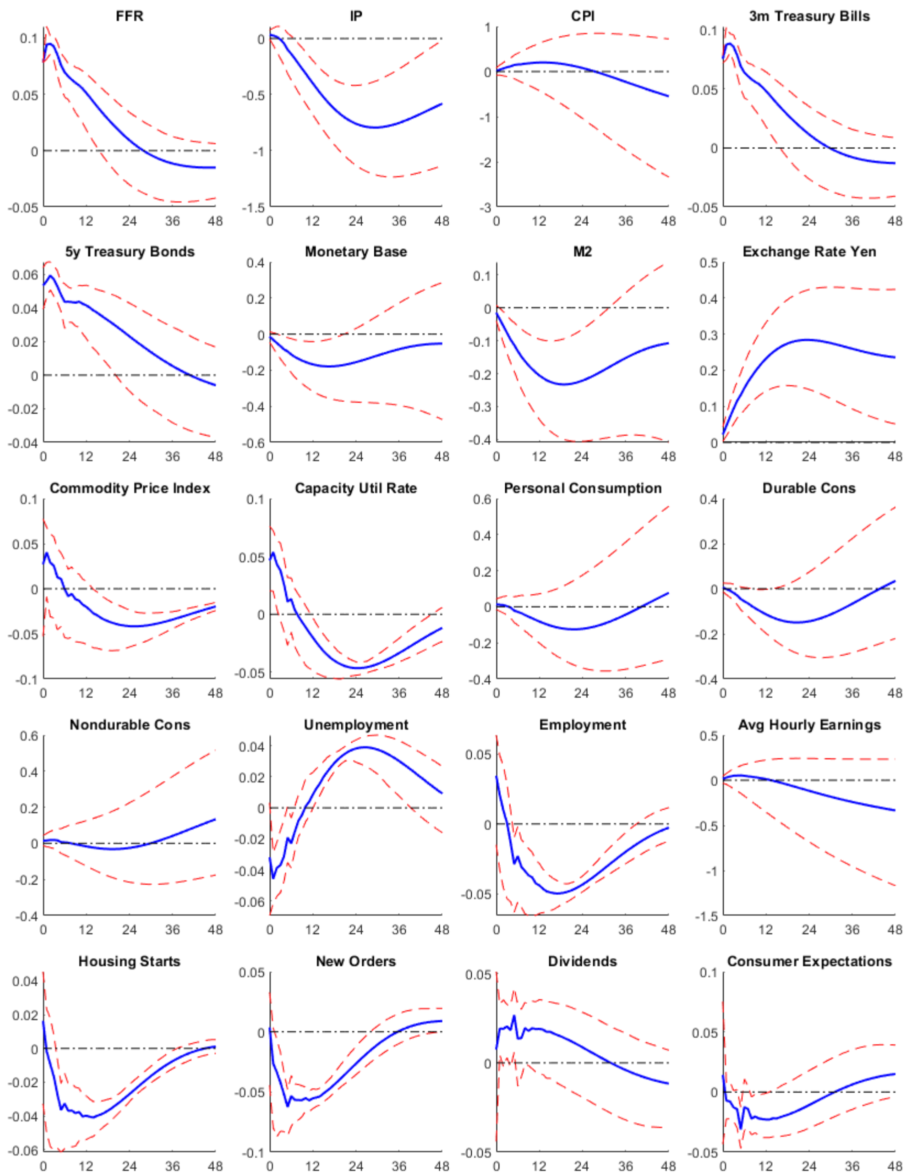


Figure 2.2: Point estimates comparison

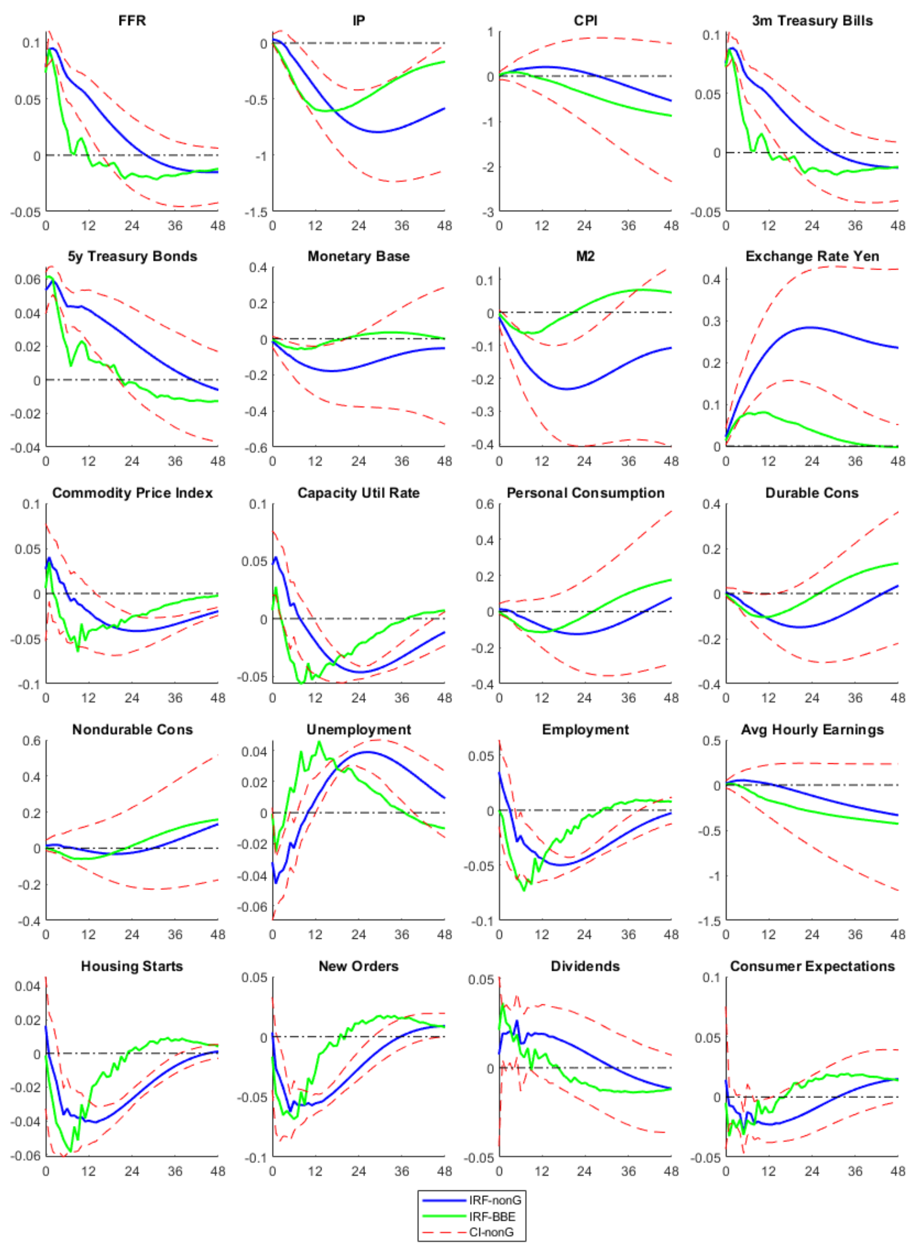
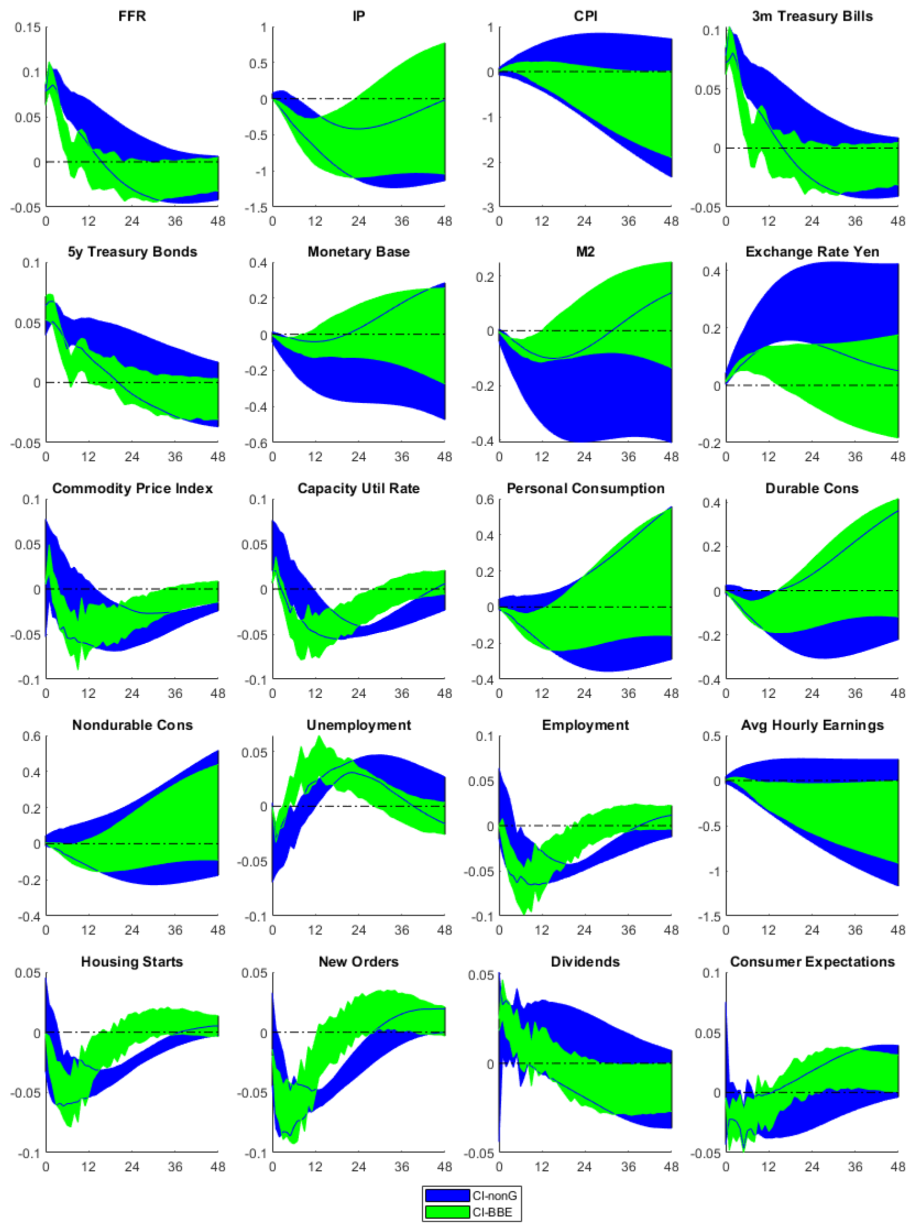


Figure 2.3: Confidence intervals comparison



Appendix A

Appendix for Chapter 1

A.1 Notations

For an $m \times n$ matrix A, $\|A\|$ denote the Frobenius norm of A. Let $vec(A)$ denote the vectorization of A:

$$vec(A) = [A_{1,1}, \ldots, A_{m,1}, A_{1,2}, \ldots, A_{1,n}, \ldots, A_{m,n}]'$$

where $A_{i,j}$ denotes the i, j-th element of A.

If A is an $n \times n$ square matrix, let $\lambda_{\min}(A)$ denote the smallest eigenvalue of A. Furthermore, if A is positive definite, let $A^{1/2}$ denote the Cholesky factor of A.

If A is an $n \times n$ symmetric matrix, let $vech(A)$ denote the half vectorization of A:

$$vech(A) = [A_{1,1}, \ldots, A_{n,1}, A_{2,2}, \ldots, A_{n,2}, \ldots, A_{n-1,n-1}, A_{n,n-1}, A_{n,n}]$$

The duplication matrix D_n for $n \times n$ matrices is the unique $n^2 \times \frac{n(n+1)}{2}$ matrix such that for any $n \times n$ symmetric matrix A,

$$D_n vech(A) = vec(A)$$

The elimination matrix L_n for $n \times n$ matrices is the unique $\frac{n(n+1)}{2} \times n^2$ matrix such that for any $n \times n$ symmetric matrix A,

$$L_n vec(A) = vech(A)$$

Throughout chapter 1, all $o_p(\cdot)$, $O_p(\cdot)$, $o(\cdot)$ and $O(\cdot)$ are uniform over the relevant parameter spaces unless stated otherwise.

References

Abbate, A., Eickmeier, S., Lemke, W., and Marcellino, M. (2016). The changing international transmission of financial shocks: evidence from a classical time-varying favar. *Journal of Money, Credit and Banking*, 48(4):573–601.

Anderson, T. and Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry*, pages 111–150. University of California Press.

Andrews, D. W., Cheng, X., and Guggenberger, P. (2020). Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics*, 218(2):496–531.

Andrews, D. W. and Guggenberger, P. (2014). A conditional-heteroskedasticity-robust confidence interval for the autoregressive parameter. *Review of Economics and Statistics*, 96(2):376–381.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171.

Bai, J., Li, K., and Lu, L. (2016). Estimation and inference of favar models. *Journal of Business & Economic Statistics*, 34(4):620–641.

Bai, J. and Ng, S. (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica*, 74(4):1133–1150.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of econometrics*, 176(1):18–29.

Barakchian, S. M. and Crowe, C. (2013). Monetary policy matters: Evidence from new shocks data. *Journal of Monetary Economics*, 60(8):950–966.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.

Billingsley, P. (2012). *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley.

Boivin, J. and Giannoni, M. P. (2007). Global Forces and Monetary Policy Effectiveness. In *International Dimensions of Monetary Policy*, NBER Chapters, pages 429–478. National Bureau of Economic Research, Inc.

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

Christiano, L. J. (2007). On the fit of new keynesian models: Comment. *Journal of Business & Economic Statistics*, 25(2):143–151.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? *Handbook of macroeconomics*, 1:65–148.

Culver, W. J. (1966). On the existence and uniqueness of the real logarithm of a matrix. *Proceedings of the American Mathematical Society*, 17(5):1146–1146.

Cúrdia, V., Del Negro, M., and Greenwald, D. L. (2014). Rare shocks, great recessions. *Journal of Applied Econometrics*, 29(7):1031–1052.

Dahlhaus, T., Hess, K., and Reza, A. (2018). International transmission channels of us quantitative easing: Evidence from canada. *Journal of Money, Credit and Banking*, 50(2-3):545–563.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford University Press, Oxford.

Eichenbaum, M. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36(5):1001–1011.

Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of Multivariate Analysis*, 28(1):20–68.

Elliott, G. (1998). On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica*, 66(1):149–158.

Engle, R. F. and Yoo, B. S. (1991). Cointegrated economic time series: An overview with new results. In *Long-Run Economic Relations: Readings in Cointegration*, chapter 12, pages 237–266. Oxford University Press, Oxford.

Ferguson, T. S. (1996). *A Course in Large Sample Theory*. Chapman & Hall Texts in Statistical Science Series. Springer US.

Fiorentini, G. and Sentana, E. (2019). Consistent non-gaussian pseudo maximum likelihood estimators. *Journal of Econometrics*, 213(2):321–358.

Fiorentini, G. and Sentana, E. (2023). Discrete mixtures of normals pseudo maximum likelihood estimators of structural vector autoregressions. *Journal of Econometrics*, 235(2):643–665.

Franchi, M. and Paruolo, P. (2019). A general inversion theorem for cointegration. *Econometric Reviews*, 38(10):1176–1201.

Georgiadis, G. (2017). To bi, or not to bi? differences between spillover estimates from bilateral and multilateral multi-country models. *Journal of International Economics*, 107:1–18.

Gonçalves, S. and Perron, B. (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics*, 182(1):156–173.

Gouriéroux, C., Monfort, A., and Renne, J.-P. (2017). Statistical inference for independent component analysis: Application to structural var models. *Journal of Econometrics*, 196(1):111–126.

Haldrup, N. and Salmon, M. (1998). Representations of I(2) cointegrated systems using the Smith-McMillan form. *Journal of Econometrics*, 84(2):303–325.

Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Application*. Academic Press.

Han, X. (2015). Tests for overidentifying restrictions in factor-augmented var models. *Journal of Econometrics*, 184(2):394–419.

Herwartz, H. and Maxand, S. (2020). Nonparametric tests for independence: a review and comparative simulation study with an application to malnutrition data in india. *Statistical papers*, 61:2175–2201.

Ho, S. W., Zhang, J., and Zhou, H. (2018). Hot money and quantitative easing: The spillover effects of us monetary policy on the chinese economy. *Journal of Money, Credit and Banking*, 50(7):1543–1569.

Inoue, A. and Kilian, L. (2020). The uniform validity of impulse response inference in autoregressions. *Journal of Econometrics*, 215(2):450–472.

Kilian, L. (1998a). Confidence intervals for impulse responses under departures from normality. *Econometric Reviews*, 17(1):1–29.

Kilian, L. (1998b). Small-sample Confidence Intervals for Impulse Response Functions. *The Review of Economics and Statistics*, 80(2):218–230.

Kilian, L. and Demiroglu, U. (2000). Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence. *Journal of Business & Economic Statistics*, 18(1):40–50.

Kilian, L. and Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge University Press.

Kilian, L., Plante, M. D., and Richter, A. W. (2022). Macroeconomic responses to uncertainty shocks: The perils of recursive orderings. Technical report, Federal Reserve Bank of Dallas.

Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 34(1):33–58.

Kuo, H. (2006). *Introduction to Stochastic Integration*. Universitext. Springer New York.

Lanne, M. and Luoto, J. (2021). Gmm estimation of non-gaussian structural vector autoregression. *Journal of Business & Economic Statistics*, 39(1):69–81.

Lanne, M., Meitz, M., and Saikkonen, P. (2017). Identification and estimation of non-gaussian structural vector autoregressions. *Journal of Econometrics*, 196(2):288–304.

Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018). Har inference: Recommendations for practice. *Journal of Business & Economic Statistics*, 36(4):541–559.

Li, D., Plagborg-Møller, M., and Wolf, C. K. (2021). Local projections vs. vars: Lessons from thousands of dgps. *arXiv preprint arXiv:2104.00655*.

Liu, Z., Spiegel, M. M., and Tai, A. (2017). Measuring the effects of dollar appreciation on asia: A favar approach. *Journal of International Money and Finance*, 74:353–370.

Lütkepohl, H., Staszewska-Bystrova, A., and Winker, P. (2015). Confidence bands for impulse responses: Bonferroni vs. wald. *Oxford Bulletin of Economics and Statistics*, 77(6):800–821.

Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, 75(5):1411–1452.

Mikusheva, A. (2012). One-Dimensional Inference in Autoregressive Models With the Potential Presence of a Unit Root. *Econometrica*, 80(1):173–212.

Mishkin, F. S. (2011). Monetary policy strategy: lessons from the crisis. Technical report, National Bureau of Economic Research.

Montiel Olea, J. L. and Plagborg-Møller, M. (2019). Simultaneous confidence bands: Theory, implementation, and an application to svars. *Journal of Applied Econometrics*, 34(1):1–17.

Montiel Olea, J. L. and Plagborg-Møller, M. (2021). Local projection inference is simpler and more robust than you think. *Econometrica*, 89(4):1789–1823.

Montiel Olea, J. L., Plagborg-Møller, M., and Qian, E. (2022). Svar identification from higher moments: Has the simultaneous causality problem been solved? In *AEA Papers and Proceedings*, volume 112, pages 481–85.

Müller, U. K. (2013). Risk of bayesian inference in misspecified models, and the sandwich covariance matrix. *Econometrica*, 81(5):1805–1849.

Mumtaz, H. and Surico, P. (2009). The transmission of international shocks: a factor-augmented var approach. *Journal of Money, Credit and Banking*, 41:71–100.

Mumtaz, H. and Theodoridis, K. (2018). The changing transmission of uncertainty shocks in the us. *Journal of Business & Economic Statistics*, 36(2):239–252.

Nakamura, E. and Steinsson, J. (2018). High-frequency identification of monetary non-neutrality: the information effect. *The Quarterly Journal of Economics*, 133(3):1283–1330.

Park, J. Y. and Phillips, P. C. (1988). Statistical inference in regressions with integrated processes: Part 1. *Econometric Theory*, 4(3):468–497.

Park, J. Y. and Phillips, P. C. (1989). Statistical inference in regressions with integrated processes: Part 2. *Econometric Theory*, 5(1):95–131.

Phillips, P. C. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, 74(3):535–547.

Phillips, P. C. (1988). Regression Theory for Near-Integrated Time Series. *Econometrica*, 56(5):1021.

Phillips, P. C. (2007). Unit root log periodogram regression. *Journal of Econometrics*, 138(1):104–124.

Phillips, P. C. and Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.

Phillips, P. C. and Solo, V. (1992). Asymptotics for Linear Processes. *The Annals of Statistics*, 20(2):971–1001.

Phillips, P. C. B. (2014). On confidence intervals for autoregressive roots and predictive regression. *Econometrica*, 82(3):1177–1195.

Plagborg-Møller, M. and Wolf, C. K. (2021). Local projections and vars estimate the same impulse responses. *Econometrica*, 89(2):955–980.

Potjagailo, G. (2017). Spillover effects from euro area monetary policy across europe: A factor-augmented var approach. *Journal of International Money and Finance*, 72:127–147.

Ramey, V. A. (2016). Macroeconomic shocks and their propagation. *Handbook of macroeconomics*, 2:71–162.

Rao, R. R. (1962). Relations between weak and uniform convergence of measures with applications. *The Annals of Mathematical Statistics*, pages 659–680.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Wiley series in probability and mathematical statistics. Wiley New York.

Shao, J. (2003). *Mathematical Statistics*. Springer Texts in Statistics. Springer.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European economic review*, 36(5):975–1000.

Skorokhod, A. V. (1982). *Studies in the theory of random processes*, volume 7021. Courier Dover Publications.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for var analysis. Working Paper 11467, National Bureau of Economic Research.

Stock, J. H. and Watson, M. W. (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, volume 2, pages 415–525. Elsevier.

Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics*, 66(1-2):225–250.

Uhlig, H. (2005). What are the effects of monetary policy on output? results from an agnostic identification procedure. *Journal of Monetary Economics*, 52(2):381–419.

Watson, M. W. (1994). Vector autoregressions and cointegration. *Handbook of econometrics*, 4:2843–2915.

Wu, J. C. and Xia, F. D. (2016). Measuring the macroeconomic impact of monetary policy at the zero lower bound. *Journal of Money, Credit and Banking*, 48(2-3):253–291.

Xu, K.-L. (2023). Local Projection Based Inference under General Conditions. CAEPR Working Papers 2023-001 Classification-C, Center for Applied Economics and Policy Research, Department of Economics, Indiana University Bloomington.

Yamamoto, Y. (2019). Bootstrap inference for impulse response functions in factor-augmented vector autoregressions. *Journal of Applied Econometrics*, 34(2):247–267.

Yamamoto, Y. and Hara, N. (2022). Identifying factor-augmented vector autoregression models via changes in shock variances. *Journal of Applied Econometrics*.