Supplemental Appendices for Asymptotically Valid Bootstrap Inference for Proxy SVARs*

Carsten Jentsch†
TU Dortmund University

Kurt G. Lunsford‡
Federal Reserve Bank of Cleveland

May 2021

*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Cleveland or the Federal Reserve System.

†TU Dortmund University, Faculty of Statistics, D-44221 Dortmund, Germany; jentsch@statistik.tu-dortmund.de

‡Federal Reserve Bank of Cleveland, P.O. Box 6387, Cleveland, OH 44101; kurt.lunsford@clev.frb.org

Contents

A Additional Theoretical Detail **2**

A.1 Detail for Theorem 2.1 . . . 2
A.2 CLT for IRFs and FEVDs . . . 3
A.3 CLT with u_t and m_t both iid . . . 4
A.4 Detail for Theorem 3.1 . . . 5

B Proofs **6**

B.1 Proof of Theorem 2.1 . . . 6
B.2 Proof of Corollary A.1 . . . 13
B.3 Proof of Theorem 3.1 . . . 17
B.4 Proof of Theorem 3.2 . . . 17
B.5 Proof of Theorem 4.1 . . . 19
B.6 Proof of Theorem 4.2 . . . 22
B.7 Proof of Theorem 4.3 . . . 22

C Residual-Based Wild Bootstrap **24**

C.1 Residual-based Wild Bootstrap Algorithm . . . 24
C.2 Inconsistency of the Wild Bootstrap . . . 25
C.3 Proof of Theorem C.1 . . . 27

D Montiel Olea, Stock and Watson Analytical Confidence Sets **28**

E Tests of Proxy Strength **31**

F Computation Details for Grid MBB AR Confidence Sets **32**

G Additional Monte Carlo Results **34**

G.1 Additional Confidence Interval Constructions . . . 34
G.2 One Standard Deviation IRFs . . . 40
G.3 95% Confidence Intervals . . . 42
G.4 Results with $T = 2000$ . . . 42
G.5 Less Persistent VAR Data . . . 48
G.6 More Persistent Stochastic Volatility . . . 52
G.7 A DGP with Censoring . . . 56
G.8 Results with Lag Augmentation . . . 56

A Additional Theoretical Detail

This appendix provides additional detail for our theoretical results to supplement what we show in the paper. We also provide some complementing additional results.

A.1 Detail for Theorem 2.1

In this section, we provide additional mathematical details for Theorem 2.1. To do so, we make use of some additional notation. Define the $(Kp \times K)$ matrices $C_j = (\Phi'_{j-1}, \ldots, \Phi'_{j-p})'$ and the $((Kp+1) \times (Kp+1))$ matrix $\Gamma_{ZZ} = \mathbb{E}(T^{-1}ZZ')$, which we can write as

$$\Gamma_{ZZ} = \begin{pmatrix} 1 & (\mathbf{1}_p \otimes \mu)' \\ (\mathbf{1}_p \otimes \mu) & (\mathbf{1}_p \otimes \mu)(\mathbf{1}_p \otimes \mu)' + \sum_{j=1}^{\infty} C_j \Sigma_u C_j' \end{pmatrix}, \tag{A.1}$$

in which $\mathbf{1}_p$ is a $(p \times 1)$ vector of ones. Further, for $a, b, c \in \mathbb{Z}$, we define the $(K^2 \times K)$ and $(Kr \times K)$ covariance matrices

$$\begin{aligned} \kappa_{a,b} &= Cov\left(\text{vec}(u_t u'_{t-a}), u_{t-b}\right) = \mathbb{E}\left(\text{vec}(u_t u'_{t-a}) u'_{t-b}\right), \\ \lambda_{a,b} &= Cov\left(\text{vec}(u_{t-a} m'_t), u_{t-b}\right) = \mathbb{E}\left(\text{vec}(u_{t-a} m'_t) u'_{t-b}\right), \end{aligned}$$

and the $(K^2 \times K^2)$, $(Kr \times K^2)$ and $(Kr \times Kr)$ covariance matrices

$$\begin{aligned} \tau_{a,b,c} &= Cov\left(\text{vec}(u_t u'_{t-a}), \text{vec}(u_{t-b} u'_{t-c})\right) \\ &= \begin{cases} \mathbb{E}\left(\text{vec}(u_t u'_{t-a}) \text{vec}(u_{t-b} u'_{t-c})'\right) - \text{vec}(\Sigma_u)\,\text{vec}(\Sigma_u)', & a = 0, b = c \\ \mathbb{E}\left(\text{vec}(u_t u'_{t-a}) \text{vec}(u_{t-b} u'_{t-c})'\right), & \text{otherwise} \end{cases}, \\ \nu_{a,b,c} &= Cov\left(\text{vec}(u_{t-a} m'_t), \text{vec}(u_{t-b} u'_{t-c})\right) \\ &= \begin{cases} \mathbb{E}\left(\text{vec}(u_{t-a} m'_t) \text{vec}(u_{t-b} u'_{t-c})'\right) - \text{vec}(H^{(1)}\Psi')\,\text{vec}(\Sigma_u)', & a = 0, b = c \\ \mathbb{E}\left(\text{vec}(u_{t-a} m'_t) \text{vec}(u_{t-b} u'_{t-c})'\right) - \mathbb{E}\left(\text{vec}(u_{t-a} m'_t)\right) \text{vec}(\Sigma_u)', & b = c \\ \mathbb{E}\left(\text{vec}(u_{t-a} m'_t) \text{vec}(u_{t-b} u'_{t-c})'\right), & b \neq c \end{cases}, \\ \zeta_{a,b,c} &= Cov\left(\text{vec}(u_{t-a} m'_t), \text{vec}(u_{t-b} m'_{t-c})\right) \\ &= \begin{cases} \mathbb{E}\left(\text{vec}(u_{t-a} m'_t) \text{vec}(u_{t-b} m'_{t-c})'\right) - \text{vec}(H^{(1)}\Psi')\,\text{vec}(H^{(1)}\Psi')', & a = 0, b = c \\ \mathbb{E}\left(\text{vec}(u_{t-a} m'_t) \text{vec}(u_{t-b} m'_{t-c})'\right) - \mathbb{E}\left(\text{vec}(u_{t-a} m'_t)\right) E\left(\text{vec}(u_{t-b} m'_{t-c})'\right), & \text{otherwise.} \end{cases}. \end{aligned}$$

Then, the formulas for the sub-matrices of V can be obtained from $V = Q\Omega^A Q'$, where Q is defined in (B.15) and Ω^A is a symmetric block matrix consisting of 3×3 blocks $[\Omega^A]^{(i,j)}$, $i, j = 1, 2, 3$. Precisely, we have $[\Omega^A]^{(i,j)} = \Omega^{(i,j)}$, $i, j \in \{2, 3\}$ as defined in Lemma B.1 and

$$
\begin{aligned}
[\Omega^A]^{(1,1)} &= \begin{pmatrix} \Sigma_u & \sum_{j=1}^{\infty} \left(\sum_{h=-\infty}^{\infty} \check{\kappa}'_{j,h}\right) (C_j \otimes I_K)' \\ \sum_{i=1}^{\infty} (C_i \otimes I_K) \left(\sum_{h=-\infty}^{\infty} \check{\kappa}_{i,h}\right) & \sum_{i,j=1}^{\infty} \sum_{h=-\infty}^{\infty} (C_i \otimes I_K) \check{\tau}_{i,h,h+j} (C_j \otimes I_K)' \end{pmatrix}, \\
[\Omega^A]^{(2,1)} &= L_K \left(\sum_{h=-\infty}^{\infty} \kappa_{0,h}, \quad \sum_{j=1}^{\infty} \sum_{h=-\infty}^{\infty} (\tau_{0,h,h+j} + (\nu' \otimes \kappa_{0,h}))(C_j \otimes I_K)' \right), \\
[\Omega^A]^{(3,1)} &= \left(\sum_{h=-\infty}^{\infty} \lambda_{0,h}, \quad \sum_{j=1}^{\infty} \sum_{h=-\infty}^{\infty} (\nu_{0,h,h+j} + (\nu' \otimes \lambda_{0,h}))(C_j \otimes I_K)' \right),
\end{aligned}
$$

where

$$
\begin{aligned}
\check{\kappa}_{i,h} &:= \kappa_{i,h} + (\nu \otimes \Sigma_u)\mathbf{1}(h = 0), \\
\check{\tau}_{i,h,h+j} &:= \tau_{i,h,h+j} + (\nu' \otimes \kappa_{i,h}) + (\nu \otimes \kappa'_{j,h}) + (\nu\nu' \otimes \Sigma_u)\mathbf{1}(h = 0)
\end{aligned}
$$

and L_K is the $(K(K+1)/2 \times K^2)$ elimination matrix such that $\text{vech}(A) = L_K \text{vec}(A)$ for any $(K \times K)$ matrix A. More precisely, we have $V^{(i,j)} = V^{(j,i)\prime}$, $i, j = 1, 2, 3$ and

$$
\begin{aligned}
V^{(1,1)} &= (\Gamma_{ZZ}^{-1} \otimes I_K)[\Omega^A]^{(1,1)}(\Gamma_{ZZ}^{-1} \otimes I_K)', \\
V^{(2,1)} &= [\Omega^A]^{(2,1)}(\Gamma_{ZZ}^{-1} \otimes I_K)', \\
V^{(2,2)} &= [\Omega^A]^{(2,2)} = \Omega^{(2,2)}, \\
V^{(3,1)} &= -\left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right) [\Omega^A]^{(1,1)}(\Gamma_{ZZ}^{-1} \otimes I_K)' + [\Omega^A]^{(3,1)}(\Gamma_{ZZ}^{-1} \otimes I_K)' \\
V^{(3,2)} &= -\left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right) [\Omega^A]^{(1,2)} + [\Omega^A]^{(3,2)} \\
&= -\left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right) [\Omega^A]^{(1,2)} + \Omega^{(3,2)} \\
V^{(3,3)} &= \left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right) [\Omega^A]^{(1,1)} \left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right)' - [\Omega^A]^{(3,1)} \left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right)' \\
&\quad - \left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right) [\Omega^A]^{(1,3)} + [\Omega^A]^{(3,3)} \\
&= \left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right) [\Omega^A]^{(1,1)} \left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right)' - [\Omega^A]^{(3,1)} \left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right)' \\
&\quad - \left(\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K\right) [\Omega^A]^{(1,3)} + \Omega^{(3,3)}.
\end{aligned}
$$

A.2 CLT for IRFs and FEVDs

Corollary A.1 (CLTs for IRFs and FEVDs) *Let $r = g = 1$. Under the the assumptions of Theorem 2.1, i.e. Assumptions 2.1, 2.2 and 2.4, for any $s \in \mathbb{R}$, $j, m \in \{1, \ldots, K\}$*

and $i, h \in \{0, 1, 2, \ldots\}$*, we have*

$$
\begin{aligned}
&(i) \quad \sqrt{T}(\widehat{\Theta}_{\bullet 1,i} - \Theta_{\bullet 1,i}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{\widehat{\Theta}_{\bullet 1,i}}\right), \\
&(ii) \quad \sqrt{T}(\widehat{\Xi}_{\bullet 1,i}(s;m,1) - \Xi_{\bullet 1,i}(s;m,1)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}\right), \\
&(iii) \quad \sqrt{T}(\widehat{\omega}_{j1,h} - \omega_{j1,h}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{\widehat{\omega}_{j1,h}}\right),
\end{aligned}
$$

where the exact formulas for the limiting variances can be found in Equations (B.22)*,* (B.23)*, and* (B.24)*, respectively.*

Remark A.1 *Equations* (B.22) *and* (B.23) *that give* $\Sigma_{\widehat{\Theta}_{\bullet 1,i}}$ *and* $\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$ *are generally similar. However, we make three observations to highlight the differences between the asymptotic variances of one standard deviation IRFs and normalized IRFs. First, compared to* $\Sigma_{\widehat{\Theta}_{\bullet 1,i}}$*, all components of* $\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$ *are scaled by* $(s/(e_m' H^{(1)} \sigma_{\epsilon^{(1)}}))^2$*. This is a natural scaling as* $s/(e_m' H^{(1)} \sigma_{\epsilon^{(1)}})$ *is the normalization that researchers get to choose when converting a one standard deviation IRF to a normalized IRF. Second, how the VMA coefficients,* Φ_i*, interact with* $V^{(2,1)}$*,* $V^{(2,2)}$*,* $V^{(3,1)}$*,* $V^{(3,2)}$*, and* $V^{(3,3)}$ *changes when computing* $\Sigma_{\widehat{\Theta}_{\bullet 1,i}}$ *versus* $\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$*. Specifically,* Φ_i *is post-multiplied by the matrix* $(I_K - H^{(1)} \sigma_{\epsilon^{(1)}} e_m' / (e_m' H^{(1)} \sigma_{\epsilon^{(1)}}))$ *when switching from* $\Sigma_{\widehat{\Theta}_{\bullet 1,i}}$ *to* $\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$*. Hence, changing from a one standard deviation IRF to a normalized IRF changes how the asymptotic variances and covariances of* $\boldsymbol{\sigma}$ *and* $\boldsymbol{\varphi}$ *impact the variance of the IRF at horizon* i*. Third, in comparison to* $\Sigma_{\widehat{\Theta}_{\bullet 1,i}}$ *which depends on both* $\boldsymbol{\sigma}$ *and* $\boldsymbol{\varphi}$*,* $\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$ *does not depend on* $\boldsymbol{\sigma}$ *or its asymptotic variances and covariances,* $V^{(2,1)}$*,* $V^{(2,2)}$*, and* $V^{(3,2)}$*. This is natural as Equations (5), (6), and (16) in the paper imply* $\Xi_{j1,i}(s;m,1) = s e_j' \Phi_i \boldsymbol{\varphi} / (e_m' \boldsymbol{\varphi})$ *when* $r = 1$*, which does not depend on* $\boldsymbol{\sigma}$*.*

A.3 CLT with u_t and m_t both iid

Corollary A.2 (V for iid innovations and iid proxies) *Suppose that in addition to Assumptions 2.1, 2.2 and 2.4, the process* (x_t) *with* $x_t = (u_t', m_t')'$ *is iid. Then, we have*

$$
\sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \\ \widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, V_{iid}\right), \quad V_{iid} = \begin{pmatrix} V_{iid}^{(1,1)} & V_{iid}^{(2,1)\prime} & V_{iid}^{(3,1)\prime} \\ V_{iid}^{(2,1)} & V_{iid}^{(2,2)} & V_{iid}^{(3,2)\prime} \\ V_{iid}^{(3,1)} & V_{iid}^{(3,2)} & V_{iid}^{(3,3)} \end{pmatrix},
$$

where

$$
\begin{aligned}
V_{iid}^{(1,1)} &= \Gamma_{ZZ}^{-1} \otimes \Sigma_u \\
V_{iid}^{(2,1)} &= L_K \left(\mathbb{E}(vec(u_t u_t') u_t'), \quad O_{K^2 \times K^2 p} \right), \\
V_{iid}^{(2,2)} &= L_K \left(\mathbb{E}\left(vec(u_t u_t') vec(u_t u_t')' \right) - \mathrm{vec}(\Sigma_u) \, \mathrm{vec}(\Sigma_u)' \right) L_K', \\
V_{iid}^{(3,1)} &= \left(\mathbb{E}(\mathrm{vec}(u_t (m_t - \mu_m)) u_t'), \quad O_{Kr \times K^2 p} \right), \\
V_{iid}^{(3,2)} &= [\mathbb{E}(\mathrm{vec}(u_t (m_t - \mu_m)) \, \mathrm{vec}(u_t u_t)') - \mathrm{vec}(H^{(1)} \Psi') \, \mathrm{vec}(\Sigma_u)] L_K', \\
V_{iid}^{(3,3)} &= \mathbb{E}(\mathrm{vec}(u_t (m_t - \mu_m)) \, \mathrm{vec}(u_t (m_t - \mu_m))') - \mathrm{vec}(H^{(1)} \Psi') \, \mathrm{vec}(H^{(1)} \Psi')',
\end{aligned}
$$

in which $\mu_m = \mathbb{E}(m_t)$.

We make two remarks about the form of V_{iid}. First, $V_{iid}^{(2,1)}$ and $V_{iid}^{(3,1)}$ include $O_{K^2 \times K^2 p}$ and $O_{Kr \times K^2 p}$, respectively. The position of these zeros indicates that the estimates of $\boldsymbol{\alpha} = \mathrm{vec}(A_1, \ldots, A_p)$ (excluding the intercept, ν) are asymptotically uncorrelated with the estimates of $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, when u_t and m_t are *both* iid. Second, the joint third and fourth moments of u_t and m_t in $V_{iid}^{(3,1)}$, $V_{iid}^{(3,2)}$, and $V_{iid}^{(3,3)}$ can all be written in terms of $m_t - \mu_m$. That is, the joint third and fourth moments of u_t and m_t in V_{iid} only depend on the centered proxy variable.

A.4 Detail for Theorem 3.1

The sub-matrices $V_{weak}^{(i,j)}$ of V_{weak} are defined as those of $V^{(i,j)}$ as given in Section A.1, but with $\kappa_{a,b}$, $\lambda_{a,b}$, $\tau_{a,b,c}$, $\nu_{a,b,c}$ and $\zeta_{a,b,c}$ replaced by $\kappa_{a,b;\infty}$, $\lambda_{a,b;\infty}$, $\tau_{a,b,c;\infty}$, $\nu_{a,b,c;\infty}$ and $\zeta_{a,b,c;\infty}$. In particular, $\nu_{a,b,c;\infty} = plim_{T\to\infty} \mathbb{E}\left(\mathrm{vec}(m_t u_t') \mathrm{vec}(u_{t-h} u_{t-h}')'\right)$ and $\zeta_{a,b,c;\infty} = plim_{T\to\infty} \mathbb{E}\left(\mathrm{vec}(m_t u_t') \mathrm{vec}(m_{t-h} u_{t-h}')'\right)$, because $\psi_T = \frac{C}{\sqrt{T}} \to 0$ as $T \to \infty$ by Assumption 3.2(iii).

B Proofs

B.1 Proof of Theorem 2.1

We define $\widetilde{\boldsymbol{\sigma}} = \text{vech}(\widetilde{\Sigma}_u)$, where $\widetilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} u_t u_t'$ and $\widetilde{\boldsymbol{\varphi}} = \text{vec}(\widetilde{H^{(1)}\Psi'})$, where $\widetilde{H^{(1)}\Psi'} = \frac{1}{T}\sum_{t=1}^{T} u_t m_t'$ to get

$$\sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \\ \widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \end{pmatrix} = \sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ (\widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}) + (\widehat{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}) \\ (\widetilde{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) + (\widehat{\boldsymbol{\varphi}} - \widetilde{\boldsymbol{\varphi}}) \end{pmatrix} \tag{B.1}$$

By standard arguments, we can show that $\sqrt{T}(\widehat{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}) = o_P(1)$. In contrast, it does not hold $\sqrt{T}(\widehat{\boldsymbol{\varphi}} - \widetilde{\boldsymbol{\varphi}}) = o_P(1)$ as we do not impose $\mathbb{E}(m_t) = 0$ and $\mathbb{E}(m_t y_{t-j}') = 0$, $j = 1, \ldots, p$ (as it was e.g. done in Mertens and Ravn (2013)) in our framework. Using

$$\widehat{u}_t = u_t + \left[(\nu, A_1, \ldots, A_p) - (\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_p)\right] Z_{t-1} \tag{B.2}$$

we can show

$$\begin{aligned}
\widehat{\boldsymbol{\varphi}} - \widetilde{\boldsymbol{\varphi}} &= -\frac{1}{T}\sum_{t=1}^{T} \text{vec}\left(\left[(\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_p) - (\nu, A_1, \ldots, A_p)\right] Z_{t-1} m_t'\right) && \text{(B.3)} \\
&= -\frac{1}{T}\sum_{t=1}^{T} \left((m_t Z_{t-1}') \otimes I_K\right) \text{vec}\left(\left[(\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_p) - (\nu, A_1, \ldots, A_p)\right]\right) && \text{(B.4)} \\
&= -\left(\left(\frac{1}{T}\sum_{t=1}^{T} m_t Z_{t-1}'\right) \otimes I_K\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) && \text{(B.5)} \\
&= -\left(\left(\frac{1}{T} M Z'\right) \otimes I_K\right)\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right), && \text{(B.6)}
\end{aligned}$$

where $M = [m_1, \ldots, m_T]$ is an $(r \times T)$ matrix. Using $\widetilde{K} = K(K+1)/2$, (B.1) is asymptotically equivalent to the following expression

$$\sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \\ \widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \end{pmatrix} = \begin{pmatrix} I_{K^2p+K} & O_{K^2p+K\times\widetilde{K}} & O_{K^2p+K\times Kr} \\ O_{\widetilde{K}\times K^2p+K} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr} \\ -\left(\frac{1}{T}MZ'\right)\otimes I_K & O_{Kr\times\widetilde{K}} & I_{Kr} \end{pmatrix} \sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \\ \widetilde{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \end{pmatrix}. \tag{B.7}$$

Furthermore, by using

$$Z_{t-1} = \begin{pmatrix} 1 \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix} = \begin{pmatrix} 1 \\ \mu + \sum_{j=0}^{\infty} \Phi_j u_{t-1-j} \\ \vdots \\ \mu + \sum_{j=0}^{\infty} \Phi_j u_{t-p-j} \end{pmatrix} = \begin{pmatrix} 1 \\ \sum_{j=0}^{\infty} \begin{pmatrix} \Phi_j(u_{t-1-j} + \nu) \\ \vdots \\ \Phi_j(u_{t-p-j} + \nu) \end{pmatrix} \end{pmatrix} \tag{B.8}$$

$$= \begin{pmatrix} 1 \\ \sum_{j=1}^{\infty} \begin{pmatrix} \Phi_{j-1}(u_{t-j} + \nu) \\ \vdots \\ \Phi_{j-p}(u_{t-j} + \nu) \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 1 \\ \sum_{j=1}^{\infty} C_j(u_{t-j} + \nu) \end{pmatrix}, \tag{B.9}$$

where we used $\mu = (I_K - A_1 - \cdots - A_p)^{-1}\nu = \sum_{j=0}^{\infty} \Phi_j \nu$, it can be shown that

$$\sqrt{T} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \\ \widetilde{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \end{pmatrix} \tag{B.10}$$

$$= \begin{pmatrix} \left\{ \left(\frac{1}{T} ZZ'\right)^{-1} \otimes I_K \right\} \begin{pmatrix} I_K & O_{K \times K^2} \\ O_{K^2 p \times K} & \sum_{j=1}^{\infty} (C_j \otimes I_K) \end{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} u_t \\ \mathrm{vec}(u_t(u'_{t-j} + \nu')) \end{pmatrix} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} L_K \left\{ \mathrm{vec}(u_t u'_t) - \mathrm{vec}(\Sigma_u) \right\} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ \mathrm{vec}(u_t m'_t) - \mathrm{vec}(H^{(1)} \Psi') \right\} \end{pmatrix}$$

By combining (B.10) and (B.7), we get

$$\sqrt{T} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \\ \widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \end{pmatrix} = \begin{pmatrix} \left(\frac{1}{T} ZZ'\right)^{-1} \otimes I_K & O_{K^2 p + K \times \widetilde{K}} & O_{K^2 p + K \times Kr} \\ O_{\widetilde{K} \times K^2 p + K} & I_{\widetilde{K}} & O_{\widetilde{K} \times Kr} \\ -\left[\left(\frac{1}{T} MZ'\right) \left(\frac{1}{T} ZZ'\right)^{-1} \right] \otimes I_K & O_{Kr \times \widetilde{K}} & I_{Kr} \end{pmatrix} \tag{B.11}$$

$$\times \begin{pmatrix} \begin{pmatrix} I_K & O_{K \times K^2} \\ O_{K^2 p \times K} & \sum_{j=1}^{\infty} (C_j \otimes I_K) \end{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} u_t \\ \mathrm{vec}(u_t(u'_{t-j} + \nu')) \end{pmatrix} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} L_K \left\{ \mathrm{vec}(u_t u'_t) - \mathrm{vec}(\Sigma_u) \right\} \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left\{ \mathrm{vec}(u_t m'_t) - \mathrm{vec}(H^{(1)} \Psi') \right\} \end{pmatrix} \tag{B.12}$$

$$=: \widehat{Q}_T \left(A_q + (A - A_q) \right), \tag{B.13}$$

with an obvious notation for the $((K^2 p + K + \widetilde{K} + Kr) \times (K^2 p + K + \widetilde{K} + Kr))$ matrix $\widehat{Q}_T$ in (B.11) and where A denotes the term in brackets in equation (B.12) and A_q is the same

expression, but with $\sum_{j=1}^{\infty}$ replaced by $\sum_{j=1}^{q}$ for some $q \in \mathbb{N}$. In the following, we make use of Proposition 6.3.9 of Brockwell and Davis (1991) and, to show asymptotic normality of A, it suffices to show

$$\begin{aligned}
&(a) \quad A_q \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega_q^A) \text{ as } T \to \infty \\
&(b) \quad \Omega_q^A \to \Omega^A \text{ as } q \to \infty \\
&(c) \quad \forall\ \delta > 0 : \lim_{q\to\infty} \limsup_{T\to\infty} P(|A - A_q|_1 > \delta) = 0.
\end{aligned}$$

To prove (a), we can write

$$\begin{aligned}
A_q &= \begin{pmatrix}
I_K & O_{K\times K^2} & \cdots & O_{K\times K^2} & O_{K\times\widetilde{K}} & O_{K\times Kr} \\
O_{K^2p\times K} & C_1 \otimes I_K & \cdots & C_q \otimes I_K & O_{K^2p\times\widetilde{K}} & O_{K^2p\times Kr} \\
O_{\widetilde{K}\times K} & O_{\widetilde{K}\times K^2} & \cdots & O_{\widetilde{K}\times K^2} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr} \\
O_{Kr\times K} & O_{Kr\times K^2} & \cdots & O_{Kr\times K^2} & O_{Kr\times\widetilde{K}} & I_{Kr}
\end{pmatrix} \\
&\quad \times \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix}
u_t \\
\mathrm{vec}(u_t(u_{t-1}' + \nu')) \\
\vdots \\
\mathrm{vec}(u_t(u_{t-q}' + \nu')) \\
L_K \left\{\mathrm{vec}(u_t u_t') - \mathrm{vec}(\Sigma_u)\right\} \\
\mathrm{vec}(u_t m_t') - \mathrm{vec}(H^{(1)}\Psi')
\end{pmatrix} \\
&= R_q \frac{1}{\sqrt{T}} \sum_{t=1}^{T} W_{t,q}
\end{aligned}$$

with an obvious notation for the $((K + K^2p + \widetilde{K} + Kr) \times (K + K^2q + \widetilde{K} + Kr))$ matrix R_q, and the $(K + K^2q + \widetilde{K} + Kr)$-dimensional vector $W_{t,q}$. By using Lemma B.1, this leads to

$$A_q \xrightarrow{D} \mathcal{N}(0, \Omega_q^A), \quad \text{where} \quad \Omega_q^A = R_q \Omega_q R_q' = \begin{pmatrix} [\Omega_q^A]^{(i,j)}, \\ i, j = 1, 2, 3 \end{pmatrix} \tag{B.14}$$

for $T \to \infty$. Here, Ω_q^A is a symmetric block matrix consisting of 3×3 blocks $[\Omega_q^A]^{(i,j)}$,

$i,j=1,2,3$. Precisely, we have $[\Omega_q^A]^{(i,j)}=\Omega_q^{(i,j)}$, $i,j\in\{2,3\}$ and

$$[\Omega_q^A]^{(1,1)}=\begin{pmatrix}\Sigma_u & \sum_{j=1}^q\left[\left(\sum_{h=-\infty}^{\infty}\kappa_{j,h}\right)+(\nu\otimes\Sigma_u)\right]'(C_j\otimes I_K)' \\ \sum_{i=1}^q(C_i\otimes I_K)\left[\left(\sum_{h=-\infty}^{\infty}\kappa_{i,h}\right)+(\nu\otimes\Sigma_u)\right] & \sum_{i,j=1}^q\sum_{h=-\infty}^{\infty}(C_i\otimes I_K)\check{\tau}_{i,h,h+j}(C_j\otimes I_K)'\end{pmatrix},$$

$$[\Omega_q^A]^{(2,1)}=L_K\left(\sum_{h=-\infty}^{\infty}\kappa_{0,h},\quad\sum_{j=1}^q\sum_{h=-\infty}^{\infty}(\tau_{0,h,h+j}+(\nu'\otimes\kappa_{0,h}))(C_j\otimes I_K)'\right),$$

$$[\Omega_q^A]^{(3,1)}=\left(\sum_{h=-\infty}^{\infty}\lambda_{0,h},\quad\sum_{j=1}^q\sum_{h=-\infty}^{\infty}(\nu_{0,h,h+j}+(\nu'\otimes\lambda_{0,h}))(C_j\otimes I_K)'\right),$$

where

$$\check{\tau}_{i,h,h+j}:=\tau_{i,h,h+j}+(\nu'\otimes\kappa_{i,h})+(\nu\otimes\kappa'_{j,h})+(\nu\nu'\otimes\Sigma_u)\mathbf{1}(h=0).$$

Letting $q\to\infty$, part (b) follows from the summability of fourth order cumulants imposed in Assumption 2.4(iii) and the exponential decay of the sequence $(\Phi_j,j\in\mathbb{N})$ and, consequently, also of $(C_j,j\in\mathbb{N})$. It remains to show part (c). Noting that the first part of the first sub-vector and the second and third sub-vectors of $A-A_q$ in Equation (B.12) are zero, it suffices to show (c) for the second part of the first sub-vector only. Let $d\in\mathbb{R}^{K^2p}$ and $\delta>0$. Then, using Markov inequality, finiteness of (all entries of) Ω^A and exponential decay of $(C_j,j\in\mathbb{N})$, we get

$$\begin{aligned}
&P\left(\left|\sum_{j=q+1}^{\infty}d'(C_j\otimes I_K)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\mathrm{vec}(u_t(u'_{t-j}+\nu'))\right|>\delta\right)\\
\leq\;&\frac{1}{\delta^2T}\mathbb{E}\left(\left|\sum_{j=q+1}^{\infty}d'(C_j\otimes I_K)\sum_{t=1}^{T}\mathrm{vec}(u_t(u'_{t-j}+\nu'))\right|^2\right)\\
=\;&\frac{1}{\delta^2}\sum_{i,j=q+1}^{\infty}d'(C_i\otimes I_K)\left\{\frac{1}{T}\sum_{t_1,t_2=1}^{T}\mathbb{E}\left(\mathrm{vec}(u_{t_1}(u'_{t_1-i}+\nu'))\,\mathrm{vec}(u_{t_2}(u'_{t_2-j}+\nu'))'\right)\right\}(C_j\otimes I_K)'d\\
=\;&\frac{1}{\delta^2}\sum_{i,j=q+1}^{\infty}d'(C_i\otimes I_K)\left(\sum_{h=-(T-1)}^{T-1}\left(1-\frac{|h|}{T}\right)\check{\tau}_{i,h,h+j}\right)(C_j\otimes I_K)'d\\
\underset{T\to\infty}{\to}\;&\frac{1}{\delta^2}\sum_{i,j=q+1}^{\infty}d'(C_i\otimes I_K)\sum_{h=-\infty}^{\infty}\check{\tau}_{i,h,h+j}(C_j\otimes I_K)'d\\
\underset{q\to\infty}{\to}\;&0.
\end{aligned}$$

When dealing with $\widehat{Q}_T$, using similar arguments as in (Brüggemann, Jentsch, and Trenkler, 2016, Lemma A.2), it is possible to show that

$$\widehat{Q}_T \xrightarrow{P} Q = \begin{pmatrix} \Gamma_{ZZ}^{-1} \otimes I_K & O_{K^2p+K\times\widetilde{K}} & O_{K^2p+K\times Kr} \\ O_{\widetilde{K}\times K^2p+K} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr} \\ -\left[\Gamma_{MZ}\Gamma_{ZZ}^{-1}\right] \otimes I_K & O_{Kr\times\widetilde{K}} & I_{Kr} \end{pmatrix}, \tag{B.15}$$

where

$$\Gamma_{ZZ} := \mathbb{E}\left(\frac{1}{T}ZZ'\right) = \mathbb{E}(Z_iZ_i') = \begin{pmatrix} 1 & (\mathbf{1}_p \otimes \mu)' \\ (\mathbf{1}_p \otimes \mu) & \begin{pmatrix} E(y_{t-i}y_{t-j}'), \\ i,j = 1,\ldots,p \end{pmatrix} \end{pmatrix}, \tag{B.16}$$

$$\Gamma_{MZ} := \mathbb{E}\left(\frac{1}{T}MZ'\right) = \mathbb{E}(m_iZ_i') = \left[\mu_m, \sum_{j=1}^{\infty}\left(\mathbb{E}(m_tu_{t-j}') + \mu_m\nu'\right)C_j'\right]. \tag{B.17}$$

Finally, an application of Slutsky's lemma completes the proof of the CLT leading to a limiting normal distribution with mean zero and covariance matrix $V = Q\Omega^A Q'$. □

Lemma B.1 (CLT for innovations) *Let $q \in \mathbb{N}$ and $W_{t,q} = (W_t^{(0)\prime}, W_{t,q}^{(1)\prime}, W_t^{(2)\prime}, W_t^{(3)\prime})'$, where*

$$\begin{aligned} W_t^{(0)} &= u_t \\ W_{t,q}^{(1)} &= (\mathrm{vec}(u_t(u_{t-1}' + \nu'))', \ldots, \mathrm{vec}(u_t(u_{t-q}' + \nu'))')' \\ W_t^{(2)} &= L_K\{\mathrm{vec}(u_tu_t') - \mathrm{vec}(\Sigma_u)\} = \mathrm{vech}(u_tu_t') - \mathrm{vech}(\Sigma_u) \\ W_t^{(3)} &= \mathrm{vec}(u_tm_t') - \mathrm{vec}(H^{(1)}\Psi'). \end{aligned}$$

Under Assumptions 2.1, 2.2 and 2.4, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} W_{t,q} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega_q),$$

where Ω_q is a $((K + K^2q + \widetilde{K} + Kr) \times (K + K^2q + \widetilde{K} + Kr))$ block matrix

$$\Omega_q = \begin{pmatrix} \Omega^{(0,0)} & \Omega_q^{(0,1)} & \Omega^{(0,2)} & \Omega^{(0,3)} \\ \Omega_q^{(1,0)} & \Omega_q^{(1,1)} & \Omega_q^{(1,2)} & \Omega_q^{(1,3)} \\ \Omega^{(2,0)} & \Omega_q^{(2,1)} & \Omega^{(2,2)} & \Omega^{(2,3)} \\ \Omega^{(3,0)} & \Omega_q^{(3,1)} & \Omega^{(3,2)} & \Omega^{(3,3)} \end{pmatrix}. \tag{B.18}$$

Here, $\Omega^{(i,j)} = \Omega^{(j,i)\prime}$ *for* $i, j \in \{0,1,2,3\}$ *and*

$$
\begin{aligned}
\Omega^{(0,0)} &= \Sigma_u \\
\Omega_q^{(1,0)} &= \sum_{h=-\infty}^{\infty} \begin{bmatrix} Cov\left(\text{vec}(u_t(u_{t-1}' + \nu')), u_{t-h}\right) \\ \vdots \\ Cov\left(\text{vec}(u_t(u_{t-q}' + \nu')), u_{t-h}\right) \end{bmatrix} = \sum_{h=-\infty}^{\infty} \begin{bmatrix} \kappa_{1,h} + (\nu \otimes I_K)\Sigma_u \mathbf{1}(h=0) \\ \vdots \\ \kappa_{q,h} + (\nu \otimes I_K)\Sigma_u \mathbf{1}(h=0) \end{bmatrix} \\
\Omega_q^{(1,1)} &= \sum_{h=-\infty}^{\infty} \begin{pmatrix} \tau_{i,h,h+j} + \kappa_{i,h}(\nu \otimes I_K)' + (\nu \otimes I_K)\kappa_{j,h}' + (\nu \otimes I_K)\Sigma_u(\nu \otimes I_K)'\mathbf{1}(h=0) \\ i,j = 1,\ldots,q \end{pmatrix} \\
\Omega^{(2,0)} &= L_K \left(\sum_{h=-\infty}^{\infty} Cov\left(\text{vec}(u_t u_t'), u_{t-h}\right) \right) = L_K \left(\sum_{h=-\infty}^{\infty} \kappa_{0,h} \right) \\
\Omega_q^{(2,1)} &= L_K \left(\sum_{h=-\infty}^{\infty} [\tau_{0,h,h+1} + \kappa_{0,h}(\nu \otimes I_K)', \ldots, \tau_{0,h,h+q} + \kappa_{0,h}(\nu \otimes I_K)'] \right) \\
\Omega^{(2,2)} &= L_K \left(\sum_{h=-\infty}^{\infty} Cov\left(\text{vec}(u_t u_t'), \text{vec}(u_{t-h} u_{t-h}')\right) \right) L_K' = L_K \left(\sum_{h=-\infty}^{\infty} \tau_{0,h,h} \right) L_K' \\
\Omega^{(3,0)} &= \sum_{h=-\infty}^{\infty} Cov\left(\text{vec}(u_t m_t'), u_{t-h}\right) = \sum_{h=-\infty}^{\infty} \lambda_{0,h} \\
\Omega_q^{(3,1)} &= \sum_{h=-\infty}^{\infty} [\nu_{0,h,h+1} + \lambda_{0,h}(\nu \otimes I_K)', \ldots, \nu_{0,h,h+q} + \lambda_{0,h}(\nu \otimes I_K)'] \\
\Omega^{(3,2)} &= \left(\sum_{h=-\infty}^{\infty} Cov\left(\text{vec}(u_t m_t'), \text{vec}(u_{t-h} u_{t-h}')\right) \right) L_K' = \left(\sum_{h=-\infty}^{\infty} \nu_{0,h,h} \right) L_K' \\
\Omega^{(3,3)} &= \sum_{h=-\infty}^{\infty} Cov\left(\text{vec}(u_t m_t'), \text{vec}(m_{t-h} u_{t-h}')\right) = \sum_{h=-\infty}^{\infty} \zeta_{0,h,h}
\end{aligned}
$$

Proof.

The result follows analogously to the proof of Lemma A.1 (ii) in Brüggemann, Jentsch, and Trenkler (2014) when allowing for an intercept and extending it to the proxy SVAR setup. In particular, for $\Omega_q^{(1,0)}$, we have

$$
\Omega_q^{(1,0)} = \sum_{h=-\infty}^{\infty} \begin{bmatrix} Cov\left(\text{vec}(u_t u_{t-1}'), u_{t-h}\right) \\ \vdots \\ Cov\left(\text{vec}(u_t u_{t-q}'), u_{t-h}\right) \end{bmatrix} + \sum_{h=-\infty}^{\infty} \begin{bmatrix} Cov\left(\text{vec}(u_t \nu'), u_{t-h}\right) \\ \vdots \\ Cov\left(\text{vec}(u_t \nu'), u_{t-h}\right) \end{bmatrix},
$$

and the summands of the second term become

$$\mathbf{1}_q \otimes Cov\left(\text{vec}(u_t\nu'), u_{t-h}\right) = \mathbf{1}_q \otimes (\nu \otimes I_K) Cov\left(u_t, u_{t-h}\right) = \mathbf{1}_q \otimes (\nu \otimes I_K)\Sigma_u \mathbf{1}(h=0)$$

leading to

$$\Omega_q^{(1,0)} = \sum_{h=-\infty}^{\infty} \begin{bmatrix} \kappa_{1,h} \\ \vdots \\ \kappa_{q,h} \end{bmatrix} + \mathbf{1}_q \otimes (\nu \otimes I_K)\Sigma_u$$

For $\Omega^{(1,1)}$, we get

$$\Omega^{(1,1)} = \sum_{h=-\infty}^{\infty} \begin{pmatrix} Cov\left(\text{vec}(u_t(u'_{t-i} + \nu')), \text{vec}(u_{t-h}(u'_{t-h-j} + \nu'))\right) \\ i, j = 1, \ldots, q \end{pmatrix}$$

and for the summands, using $\text{vec}(AB) = (B' \otimes I)\,\text{vec}(A) = (B' \otimes I)A$ if A is a (column) vector and $E(\text{vec}(u_t(u'_{t-i} + \nu'))) = 0$ for all i , we get

$$\begin{aligned}
&Cov\left(\text{vec}(u_t(u'_{t-i} + \nu')), \text{vec}(u_{t-h}(u'_{t-h-j} + \nu'))\right) \\
=&E\left(\text{vec}(u_t u'_{t-i})\,\text{vec}(u_{t-h}u'_{t-h-j})'\right) + E\left(\text{vec}(u_t u'_{t-i})\,\text{vec}(u_{t-h}\nu')'\right) \\
&+ E\left(\text{vec}(u_t\nu')\,\text{vec}(u_{t-h}u'_{t-h-j})'\right) + E\left(\text{vec}(u_t\nu')\,\text{vec}(u_{t-h}\nu')'\right) \\
=&\tau_{i,h,h+j} + \kappa_{i,h}(\nu \otimes I_K)' + (\nu \otimes I_K)\kappa'_{j,h} + (\nu \otimes I_K)\Sigma_u(\nu \otimes I_K)'\mathbf{1}(h=0).
\end{aligned}$$

For $\Omega_q^{(2,1)}$, we get

$$\Omega_q^{(2,1)} = \quad L_K \left(\sum_{h=-\infty}^{\infty} \left[Cov\left(\text{vec}(u_t u'_t) - \text{vec}(\Sigma_u), \text{vec}(u_{t-h}(u'_{t-h-j} + \nu'))\right), j = 1, \ldots, q \right] \right)$$

with summands

$$\begin{aligned}
&Cov\left(\text{vec}(u_t u'_t) - \text{vec}(\Sigma_u), \text{vec}(u_{t-h}(u'_{t-h-j} + \nu'))\right) \\
=&E\left(\text{vec}(u_t u'_t)(\text{vec}(u_{t-h}u'_{t-h-j}))'\right) + E\left(\text{vec}(u_t u'_t)(\text{vec}(u_{t-h}\nu'))'\right) \\
=&\tau_{0,h,h+j} + E\left(\text{vec}(u_t u'_t)u'_{t-h}\right)(\nu \otimes I_K)' \\
=&\tau_{0,h,h+j} + \kappa_{0,h}(\nu \otimes I_K)'.
\end{aligned}$$

For $\Omega_q^{(3,1)}$, we get

$$\Omega_q^{(3,1)} = \sum_{h=-\infty}^{\infty} \left[Cov\left(\text{vec}(u_t m_t') - \text{vec}(\Psi H^{(1)\prime}), \text{vec}(u_{t-h}(u_{t-h-j}' + \nu')) \right), j = 1, \ldots, q \right]$$

with summands

$$\begin{aligned}
& Cov\left(\text{vec}(u_t m_t') - \text{vec}(H^{(1)}\Psi'), \text{vec}(u_{t-h}(u_{t-h-j}' + \nu')) \right) \\
=& E\left(\text{vec}(u_t m_t')(\text{vec}(u_{t-h}u_{t-h-j}'))' \right) + E\left(\text{vec}(u_t m_t')(\text{vec}(u_{t-h}\nu'))' \right) \\
=& \nu_{0,h,h+j} + E\left(\text{vec}(u_t m_t')u_{t-h}' \right)(\nu \otimes I_K)' \\
=& \nu_{0,h,h+j} + \lambda_{0,h}(\nu \otimes I_K)'.
\end{aligned}$$

□

B.2 Proof of Corollary A.1

As $r = g = 1$, we can make use of the identification and estimation schemes (up to sign restriction) given in Equations (14) and (15) in the paper. It becomes apparent that $H^{(1)}\sigma_{\epsilon^{(1)}}$ is a smooth function of Σ_u and $\boldsymbol{\varphi}$ as, by assumption, Σ_u is positive definite and $\boldsymbol{\varphi}$ is not the zero vector. Further, the VMA coefficients, Φ_i's, are smooth functions of $\boldsymbol{\alpha} = \text{vec}(A_1, \ldots, A_p) = \widetilde{J}\boldsymbol{\beta}$, where $\widetilde{J} = [O_{K^2p \times K} : I_{K^2p}]$. Hence, estimates of $\boldsymbol{\varphi}$ are smooth functions of $\widehat{\boldsymbol{\alpha}} = \text{vec}(\widehat{A}_1, \ldots, \widehat{A}_p) = \widetilde{J}\widehat{\boldsymbol{\beta}}$. In this section, we use $q \in \{0, 1, \ldots\}$ to denote the IRF or FEVD horizon, with $q = 0$ being the horizon of the immediate impact of a shock. This is a different use of q than what we use in Section B.1. Using the Delta method similar to (and borrowing some of the notation from) Lütkepohl (2005, Proposition 3.6), we get that

$$\sqrt{T}\begin{pmatrix} \text{vec}(\widehat{\Phi}_0) - \text{vec}(\Phi_0) \\ \text{vec}(\widehat{\Phi}_1) - \text{vec}(\Phi_1) \\ \vdots \\ \text{vec}(\widehat{\Phi}_q) - \text{vec}(\Phi_q) \\ \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \\ \widehat{H^{(1)}\sigma_{\epsilon^{(1)}}} - H^{(1)}\sigma_{\epsilon^{(1)}} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}(0, W), \quad W = \begin{pmatrix} W^{(1,1)} & W^{(2,1)\prime} & W^{(3,1)\prime} \\ W^{(2,1)} & W^{(2,2)} & W^{(3,2)\prime} \\ W^{(3,1)} & W^{(3,2)} & W^{(3,3)} \end{pmatrix} \quad \text{(B.19)}$$

holds, where

$$
\begin{aligned}
W^{(1,1)} &= G_{0,q}\widetilde{J}V^{(1,1)}\widetilde{J}'G_{0,q}' \\
W^{(2,1)} &= V^{(2,1)}\widetilde{J}'G_{0,q}' \\
W^{(2,2)} &= V^{(2,2)} \\
W^{(3,1)} &= M_{\boldsymbol{\sigma}}V^{(2,1)}\widetilde{J}'G_{0,q}' + M_{\boldsymbol{\varphi}}V^{(3,1)}\widetilde{J}'G_{0,q}' \\
W^{(3,2)} &= M_{\boldsymbol{\sigma}}V^{(2,2)} + M_{\boldsymbol{\varphi}}V^{(3,2)} \\
W^{(3,3)} &= M_{\boldsymbol{\sigma}}V^{(2,2)}M_{\boldsymbol{\sigma}}' + M_{\boldsymbol{\varphi}}V^{(3,2)}M_{\boldsymbol{\sigma}}' + M_{\boldsymbol{\sigma}}V^{(3,2)\prime}M_{\boldsymbol{\varphi}}' + M_{\boldsymbol{\varphi}}V^{(3,3)}M_{\boldsymbol{\varphi}}'
\end{aligned}
\tag{B.20}
$$

with $V^{(i,j)}$ defined in Theorem 2.1, $G_{0,q} = [G_0' : G_1' : \cdots : G_q']'$ is a $(K^2(q+1) \times K^2p)$ matrix, G_i is a $(K^2 \times K^2p)$ matrix given by

$$
G_i = \frac{\partial \mathrm{vec}(\Phi_i)}{\partial \boldsymbol{\alpha}'} = \sum_{s=0}^{i-1} J(\boldsymbol{A}')^{i-1-s} \otimes \Phi_s,
$$

$J = [I_K : 0_K : \cdots : 0_K]$ is a $(K \times Kp)$ matrix,

$$
\boldsymbol{A} = \begin{pmatrix}
A_1 & A_2 & \cdots & \cdots & A_p \\
I_K & 0_K & \cdots & \cdots & 0_K \\
0_K & I_K & 0_K & \cdots & 0_K \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0_K & \cdots & 0_K & I_K & 0_K
\end{pmatrix},
$$

is the companion matrix, and following standard rules for matrix differentiation from, for example, Sections A.12 and A.13 in Lütkepohl (2005),

$$
\begin{aligned}
M_{\boldsymbol{\sigma}} &= \partial \mathrm{vec}(H^{(1)}\sigma_{\epsilon^{(1)}})/\partial \boldsymbol{\sigma}' = \boldsymbol{\varphi}\left((1/2)(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-3/2}\right)\left(\boldsymbol{\varphi}' \otimes \boldsymbol{\varphi}'\right)\left((\Sigma_u^{-1})' \otimes (\Sigma_u^{-1})\right) D_K \\
M_{\boldsymbol{\varphi}} &= \partial \mathrm{vec}(H^{(1)}\sigma_{\epsilon^{(1)}})/\partial \boldsymbol{\varphi}' = -\boldsymbol{\varphi}\left((1/2)(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-3/2}\right)\left(\boldsymbol{\varphi}'((\Sigma_u^{-1})' + \Sigma_u^{-1})\right) + \left(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi}\right)^{-1/2} I_K.
\end{aligned}
$$

This is sufficient for using the Delta method and deriving the limiting distributions of one standard deviation IRFs $\widehat{\Theta}_{j1,i}$, normalized IRFs $\widehat{\Xi}_{j1,i}(s;m,1)$ and FEVDs $\widehat{\omega}_{j1,h}$ as all of them are smooth functions of $\widehat{\Phi}_i$, $i = 0, \ldots, q$, $\widehat{\Sigma}_u$ and $\widehat{H^{(1)}\sigma_{\epsilon^{(1)}}}$.

(i) For one standard deviation IRFs $\widehat{\Theta}_{\bullet 1,i} = \widehat{\Phi}_i\widehat{H^{(1)}\sigma_{\epsilon^{(1)}}}$, we can make use of the relevant parts of (B.19) as the basis result for another application of the Delta method. That

is, we shall use

$$\sqrt{T}\begin{pmatrix} \text{vec}(\widehat{\Phi}_i) - \text{vec}(\Phi_i) \\ \widehat{H^{(1)}\sigma_{\epsilon^{(1)}}} - H^{(1)}\sigma_{\epsilon^{(1)}} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \begin{pmatrix} G_i\widetilde{J}V^{(1,1)}\widetilde{J}'G_i' & \left(M_{\boldsymbol{\sigma}}V^{(2,1)}\widetilde{J}'G_i' + M_{\boldsymbol{\varphi}}V^{(3,1)}\widetilde{J}'G_i'\right)' \\ M_{\boldsymbol{\sigma}}V^{(2,1)}\widetilde{J}'G_i' + M_{\boldsymbol{\varphi}}V^{(3,1)}\widetilde{J}'G_i' & W^{(3,3)} \end{pmatrix}\right) \tag{B.21}$$

Together with

$$\partial\text{vec}(\Phi_i H^{(1)}\sigma_{\epsilon^{(1)}})/\partial\text{vec}(\Phi_i)' = (H^{(1)}\sigma_{\epsilon^{(1)}})' \otimes I_K \quad \text{and} \quad \partial\text{vec}(\Phi_i H^{(1)}\sigma_{\epsilon^{(1)}})/\partial\text{vec}(H^{(1)}\sigma_{\epsilon^{(1)}})' = \Phi_i,$$

Equation (B.21) implies

$$\sqrt{T}(\widehat{\Theta}_{\bullet 1,i} - \Theta_{\bullet 1,i}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{\widehat{\Theta}_{\bullet 1,i}}\right),$$

where

$$\begin{aligned} \Sigma_{\widehat{\Theta}_{\bullet 1,i}} &= \left(\sigma_{\epsilon^{(1)}}H^{(1)\prime} \otimes I_K\right) G_i\widetilde{J}V^{(1,1)}\widetilde{J}'G_i' \left(H^{(1)}\sigma_{\epsilon^{(1)}} \otimes I_K\right) \\ &\quad + \Phi_i\left(M_{\boldsymbol{\sigma}}V^{(2,1)}\widetilde{J}'G_i' + M_{\boldsymbol{\varphi}}V^{(3,1)}\widetilde{J}'G_i'\right)\left(H^{(1)}\sigma_{\epsilon^{(1)}} \otimes I_K\right) \\ &\quad + \left(\sigma_{\epsilon^{(1)}}H^{(1)\prime} \otimes I_K\right)\left(M_{\boldsymbol{\sigma}}V^{(2,1)}\widetilde{J}'G_i' + M_{\boldsymbol{\varphi}}V^{(3,1)}\widetilde{J}'G_i'\right)'\Phi_i' + \Phi_i W^{(3,3)}\Phi_i'. \end{aligned} \tag{B.22}$$

(ii) For normalized IRFs $\widehat{\Xi}_{\bullet 1,i}(s;m,1) = s\widehat{\Theta}_{\bullet 1,i}/(e_m'\widehat{H^{(1)}\sigma_{\epsilon^{(1)}}}) = s\widehat{\Phi}_i\widehat{H^{(1)}\sigma_{\epsilon^{(1)}}}/(e_m'\widehat{H^{(1)}\sigma_{\epsilon^{(1)}}})$, together with

$$\begin{aligned} \frac{\partial\text{vec}(s\Phi_i H^{(1)}\sigma_{\epsilon^{(1)}}/(e_m'H^{(1)}\sigma_{\epsilon^{(1)}}))}{\partial\text{vec}(\Phi_i)'} &= s\left((H^{(1)}\sigma_{\epsilon^{(1)}}/(e_m'H^{(1)}\sigma_{\epsilon^{(1)}}))' \otimes I_K\right), \\ \frac{\partial\text{vec}(s\Phi_i H^{(1)}\sigma_{\epsilon^{(1)}}/(e_m'H^{(1)}\sigma_{\epsilon^{(1)}}))}{\partial\text{vec}(H^{(1)}\sigma_{\epsilon^{(1)}})'} &= s\Phi_i\left(H^{(1)}\sigma_{\epsilon^{(1)}}\left(-\frac{1}{(e_m'H^{(1)}\sigma_{\epsilon^{(1)}})^2}\right)e_m' + \frac{1}{(e_m'H^{(1)}\sigma_{\epsilon^{(1)}})}I_K\right) \end{aligned}$$

Equation (B.21) gives

$$\sqrt{T}(\widehat{\Xi}_{\bullet 1,i}(s;m,1) - \Xi_{\bullet 1,i}(s;m,1)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}\right),$$

where

$$\begin{aligned}\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)} &= \left(\frac{s}{e_m'H^{(1)}\sigma_{\epsilon^{(1)}}}\right)^2\Bigg[\left(\sigma_{\epsilon^{(1)}}H^{(1)\prime}\otimes I_K\right)G_i\widetilde{J}V^{(1,1)}\widetilde{J}'G_i'\left(H^{(1)}\sigma_{\epsilon^{(1)}}\otimes I_K\right) \\ &+ \left(\Phi_i\left(I_K - H^{(1)}\sigma_{\epsilon^{(1)}}e_m'\left(\frac{1}{e_m'H^{(1)}\sigma_{\epsilon^{(1)}}}\right)\right)\right)\left(M_\sigma V^{(2,1)}\widetilde{J}'G_i' + M_\varphi V^{(3,1)}\widetilde{J}'G_i'\right)\left(H^{(1)}\sigma_{\epsilon^{(1)}}\otimes I_K\right) \\ &+ \left(\sigma_{\epsilon^{(1)}}H^{(1)\prime}\otimes I_K\right)\left(M_\sigma V^{(2,1)}\widetilde{J}'G_i' + M_\varphi V^{(3,1)}\widetilde{J}'G_i'\right)'\left(\Phi_i\left(I_K - H^{(1)}\sigma_{\epsilon^{(1)}}e_m'\left(\frac{1}{e_m'H^{(1)}\sigma_{\epsilon^{(1)}}}\right)\right)\right)' \\ &+ \left(\Phi_i\left(I_K - H^{(1)}\sigma_{\epsilon^{(1)}}e_m'\left(\frac{1}{e_m'H^{(1)}\sigma_{\epsilon^{(1)}}}\right)\right)\right)W^{(3,3)}\left(\Phi_i\left(I_K - H^{(1)}\sigma_{\epsilon^{(1)}}e_m'\left(\frac{1}{e_m'H^{(1)}}\sigma_{\epsilon^{(1)}}\right)\right)\right)'\Bigg].\end{aligned} \tag{B.23}$$

Further, it is the case that $\left(I_K - H^{(1)}\sigma_{\epsilon^{(1)}}e_m'/(e_m'H^{(1)}\sigma_{\epsilon^{(1)}})\right)H^{(1)}\sigma_{\epsilon^{(1)}} = 0$. Then, $H^{(1)}\sigma_{\epsilon^{(1)}} = \boldsymbol{\varphi}(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-1/2}$ from (14) in the paper and after imposing a sign restriction implies $\left(I_K - H^{(1)}\sigma_{\epsilon^{(1)}}e_m'/(e_m'H^{(1)}\sigma_{\epsilon^{(1)}})\right)M_\sigma = 0$ and $\left(I_K - H^{(1)}\sigma_{\epsilon^{(1)}}e_m'/(e_m'H^{(1)}\sigma_{\epsilon^{(1)}})\right)M_\varphi = (I_K - \boldsymbol{\varphi}e_m'/(e_m'\boldsymbol{\varphi}))(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-1/2}$. It follows that

$$\begin{aligned}\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)} &= \left(\frac{s}{e_m'\boldsymbol{\varphi}}\right)^2\Bigg[(\boldsymbol{\varphi}'\otimes I_K)G_i\widetilde{J}V^{(1,1)}\widetilde{J}'G_i'(\boldsymbol{\varphi}\otimes I_K) \\ &+ \left(\Phi_i\left(I_K - \boldsymbol{\varphi}e_m'\left(\frac{1}{e_m'\boldsymbol{\varphi}}\right)\right)\right)V^{(3,1)}\widetilde{J}'G_i'(\boldsymbol{\varphi}\otimes I_K) \\ &+ (\boldsymbol{\varphi}'\otimes I_K)G_i\widetilde{J}V^{(1,3)}\left(\Phi_i\left(I_K - \boldsymbol{\varphi}e_m'\left(\frac{1}{e_m'\boldsymbol{\varphi}}\right)\right)\right)' \\ &+ \left(\Phi_i\left(I_K - \boldsymbol{\varphi}e_m'\left(\frac{1}{e_m'\boldsymbol{\varphi}}\right)\right)\right)V^{(3,3)}\left(\Phi_i\left(I_K - \boldsymbol{\varphi}e_m'\left(\frac{1}{e_m'\boldsymbol{\varphi}}\right)\right)\right)'\Bigg],\end{aligned}$$

so that $\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$ does not depend on $\boldsymbol{\sigma}$, $V^{(2,1)}$, $V^{(2,2)}$ or $V^{(3,2)}$.

(iii) For FEVDs

$$\widehat{\omega}_{j1,h} = \frac{\sum_{i=0}^{h-1}\widehat{\Theta}_{j1,i}^2}{\sum_{i=0}^{h-1}e_j'\widehat{\Phi}_i\widehat{\Sigma}_u\widehat{\Phi}_i'e_j} = \frac{\sum_{i=0}^{i-1}(e_j'\widehat{\Phi}_i\widehat{H}^{(1)}\sigma_{\epsilon^{(1)}})^2}{\widehat{MSE}_j(h)},$$

together with

$$\begin{aligned}F_{j1,h}(r) &:= \frac{\partial \mathrm{vec}(\omega_{j1,h})}{\partial \mathrm{vec}(\Phi_r)'} \\ &= -2\frac{\sum_{i=0}^{h-1}(e_j'\Phi_iH^{(1)}\sigma_{\epsilon^{(1)}})^2}{(MSE_j(h))^2}(e_j'\Phi_r\Sigma_u\otimes e_j') + 2\frac{e_j'\Phi_rH^{(1)}\sigma_{\epsilon^{(1)}}}{MSE_j(h)}(\sigma_{\epsilon^{(1)}}H^{(1)\prime}\otimes e_j')\end{aligned}$$

for $r = 0, \ldots, h-1$ and

$$\begin{aligned} Q_{j1,h} &:= \frac{\partial \text{vec}(\omega_{j1,h})}{\partial \boldsymbol{\sigma}'} = -\frac{\sum_{i=0}^{h-1}(e_j'\Phi_i H^{(1)}\sigma_{\epsilon^{(1)}})^2}{(MSE_j(h))^2}\sum_{i=0}^{h-1}(e_j'\Phi_i \otimes e_j'\Phi_i)D_K \\ N_{j1,h} &:= \frac{\partial \text{vec}(\omega_{j1,h})}{\partial \text{vec}(H^{(1)}\sigma_{\epsilon^{(1)}})'} = \frac{2}{MSE_j(h)}\sum_{i=0}^{h-1}(e_j'\Phi_i H^{(1)}\sigma_{\epsilon^{(1)}})(e_j'\Phi_i) \end{aligned}$$

Equation (B.19) with $q = h-1$ gives

$$\sqrt{T}(\widehat{\omega}_{j1,h} - \omega_{j1,h}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{\widehat{\omega}_{j1,h}}\right),$$

where

$$\Sigma_{\widehat{\omega}_{j1,h}} = [\boldsymbol{F}_{j1,h} : Q_{j1,h} : N_{j1,h}] \begin{pmatrix} W^{(1,1)} & W^{(2,1)\prime} & W^{(3,1)\prime} \\ W^{(2,1)} & W^{(2,2)} & W^{(3,2)\prime} \\ W^{(3,1)} & W^{(3,2)} & W^{(3,3)} \end{pmatrix} \begin{bmatrix} \boldsymbol{F}_{j1,h}' \\ Q_{j1,h}' \\ N_{j1,h}' \end{bmatrix} \tag{B.24}$$

and $\boldsymbol{F}_{j1,h} = [F_{j1,h}(0) : F_{j1,h}(1) : \cdots : F_{j1,h}(h-1)]$. □

B.3 Proof of Theorem 3.1

By replacing the α-mixing condition in Assumption 2.4 imposed for the strong proxy case by the uniform mixing condition in Assumption 3.2 used for a weak proxy, the proof follows by the same arguments as used in Theorem 2.1.

B.4 Proof of Theorem 3.2

Throughout this proof, we make use of $\widehat{\boldsymbol{\varphi}}_T = \widehat{H^{(1)}\psi_T}$ in the $r = g = 1$ case.
(i) To prove (22) in the paper, we write

$$\begin{aligned} \sqrt{T}\begin{pmatrix} se_j'\widehat{\Phi}_i\widehat{\boldsymbol{\varphi}}_T \\ e_m'\widehat{\boldsymbol{\varphi}}_T \end{pmatrix} &= \sqrt{T}\begin{pmatrix} se_j'\widehat{\Phi}_i\widehat{H^{(1)}\psi_T} \\ e_m'\widehat{H^{(1)}\psi_T} \end{pmatrix} \\ &= \begin{pmatrix} (H^{(1)}\psi_T)' \otimes (se_j') & se_j'\widehat{\Phi}_i \\ O_{1\times K^2} & e_m' \end{pmatrix}\sqrt{T}\begin{pmatrix} \text{vec}(\widehat{\Phi}_i) - \text{vec}(\Phi_i) \\ \widehat{H^{(1)}\psi_T} - H^{(1)}\psi_T \end{pmatrix} + \begin{pmatrix} se_j'\Phi_i H^{(1)}\sqrt{T}\psi_T \\ e_m' H^{(1)}\sqrt{T}\psi_T \end{pmatrix}. \end{aligned}$$

By making use of $\psi_T = C/\sqrt{T} \to 0$, $\widehat{\Phi}_i \xrightarrow{P} \Phi_i$, Slutsky's Lemma and of the asymptotic normality in (21) in the paper, this leads to

$$\begin{aligned}
\sqrt{T}\begin{pmatrix} se_j'\widehat{\Phi}_i\widehat{H^{(1)}\psi_T} \\ e_m'\widehat{H^{(1)}\psi_T} \end{pmatrix} \xrightarrow{D} &\mathcal{N}\left(\begin{pmatrix} se_j'\Phi_i H^{(1)}C \\ e_m'H^{(1)}C \end{pmatrix}, \begin{pmatrix} O_{1\times K^2} & se_j'\Phi_i \\ O_{1\times K^2} & e_m' \end{pmatrix} R_{weak} \begin{pmatrix} O_{1\times K^2} & se_j'\Phi_i \\ O_{1\times K^2} & e_m' \end{pmatrix}'\right) \\
=&\mathcal{N}\left(\begin{pmatrix} se_j'\Phi_i H^{(1)}C \\ e_m'H^{(1)}C \end{pmatrix}, \begin{pmatrix} se_j'\Phi_i \\ e_m' \end{pmatrix} V_{weak}^{(3,3)} \begin{pmatrix} s\Phi_i'e_j, & e_m \end{pmatrix}\right) \\
=&\begin{pmatrix} se_j'\Phi_i \\ e_m' \end{pmatrix} \mathcal{N}\left(H^{(1)}C, V_{weak}^{(3,3)}\right).
\end{aligned}$$

The proof of (23) in part (i)(a) in the paper, follows immediately from (22) in the paper using

$$\widehat{\Xi}_{j1,i}(s;m,1) = \frac{se_j'\widehat{\Phi}_i\widehat{H^{(1)}\psi_T}}{e_m'\widehat{H^{(1)}\psi_T}} = \frac{\sqrt{T}se_j'\widehat{\Phi}_i\widehat{H^{(1)}\psi_T}}{\sqrt{T}e_m'\widehat{H^{(1)}\psi_T}}.$$

By multiplying (22) in the paper from the left with $(1, -\xi_0)$, we get

$$\sqrt{T}\left(se_j'\widehat{\Phi}_i\widehat{\boldsymbol{\varphi}} - \xi_0 e_m'\widehat{\boldsymbol{\varphi}}\right) \xrightarrow{D} (se_j'\Phi_i - \xi_0 e_m')\mathcal{N}\left(H^{(1)}C, V_{weak}^{(3,3)}\right).$$

In part (i)(b) in the paper under the null, then

$$(1, -\xi_0)\begin{pmatrix} se_j'\Phi_i \\ e_m' \end{pmatrix} H^{(1)}C = 0C = 0,$$

yielding mean zero in Equation (24) in the paper. Under the alternative, $se_j'\Phi_i\boldsymbol{\varphi}_T - \xi_A e_m'\boldsymbol{\varphi}_T = 0$ with $\xi_A \neq \xi_0$, we have

$$\begin{aligned}
se_j'\Phi_i H^{(1)}C - \xi_0 e_m'H^{(1)}C &= se_j'\Phi_i H^{(1)}C - \xi_0 e_m'H^{(1)}C - (se_j'\Phi_i H^{(1)}C - \xi_A e_m'H^{(1)}C) \\
&= \xi_A e_m'H^{(1)}C - \xi_0 e_m'H^{(1)}C \\
&= (\xi_A - \xi_0)e_m'H^{(1)}C,
\end{aligned}$$

yielding a mean of $(\xi_A - \xi_0)e_m'H^{(1)}C$ in Equation (25) in the paper.

(ii) From Equation (18) in the paper, under the null of $se_j'\Phi_i H^{(1)}\psi - \xi_0 e_m'H^{(1)}\psi = 0$, we

get

$$\begin{aligned}
&\sqrt{T}\left(se_j'\widehat{\Phi}_i\widehat{H^{(1)}\psi} - \xi_0 e_m'\widehat{H^{(1)}\psi}\right) \\
&= \left((H^{(1)}\psi)' \otimes (se_j'), se_j'\widehat{\Phi}_i - \xi_0 e_m'\right)\sqrt{T}\begin{pmatrix} \mathrm{vec}(\widehat{\Phi}_i) - \mathrm{vec}(\Phi_i) \\ \widehat{H^{(1)}\psi} - H^{(1)}\psi \end{pmatrix} \\
&\xrightarrow{D} \left((H^{(1)}\psi)' \otimes (se_j'), se_j'\Phi_i - \xi_0 e_m'\right)\mathcal{N}(0, R),
\end{aligned}$$

analogously to (20) in the paper, but based on Theorem 2.1, where R is defined as R_{weak}, but with V_{weak} replaced by V.

Under the alternative, $se_j'\Phi_i\boldsymbol{\varphi} - \xi_A e_m'\boldsymbol{\varphi} = 0$ with $\xi_A \neq \xi_0$, Equation (18) in the paper can be rewritten as

$$\begin{aligned}
&\sqrt{T}\left(se_j'\widehat{\Phi}_i\widehat{H^{(1)}\psi} - \xi_0 e_m'\widehat{H^{(1)}\psi}\right) \\
&= \left((H^{(1)}\psi)' \otimes (se_j'), se_j'\widehat{\Phi}_i - \xi_0 e_m'\right)\sqrt{T}\begin{pmatrix} \mathrm{vec}(\widehat{\Phi}_i) - \mathrm{vec}(\Phi_i) \\ \widehat{H^{(1)}\psi} - H^{(1)}\psi \end{pmatrix} + \sqrt{T}(\xi_A - \xi_0)e_m'H^{(1)}\psi,
\end{aligned}$$

with the $\sqrt{T}(\xi_A - \xi_0)e_m'H^{(1)}\psi$ term going to ∞ as $T \to \infty$ a.s.. □

B.5 Proof of Theorem 4.1

To disentangle the autoregressive part and the MBB part of the bootstrap proposal, we will proceed in two steps. First, using Lemma B.2, we have that it is asymptotically equivalent if the estimator $\widehat{\boldsymbol{\beta}}$ is replaced by the true VAR coefficients $\boldsymbol{\beta}$ when conducting the residual-based MBB scheme as described in Section 4.1 of the paper. Second, we will resemble the proof of Theorem 2.1 in the bootstrap world and make use of Lemma B.3 to prove an asymptotic normality results that corresponds to what is obtained in Theorem 2.1.

Hence, using Polya's Theorem and (the notation introduced in) Lemma B.2, it remains to show that $\sqrt{T}((\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}})', (\widetilde{\boldsymbol{\sigma}}^* - \widehat{\boldsymbol{\sigma}})', (\widehat{\boldsymbol{\varphi}}^* - \widetilde{\boldsymbol{\varphi}})')' \xrightarrow{\mathcal{D}} \mathcal{N}(0, V)$ in probability with limiting covariance matrix V as obtained in Theorem 2.1. Following the lines of the proof of Theorem

2.1 from (B.1) - (B.13), we get

$$\sqrt{T}\begin{pmatrix}\widetilde{\boldsymbol{\beta}}^*-\widetilde{\boldsymbol{\beta}}\\ \widetilde{\boldsymbol{\sigma}}^*-\widetilde{\boldsymbol{\sigma}}\\ \widetilde{\boldsymbol{\varphi}}^*-\widetilde{\boldsymbol{\varphi}}\end{pmatrix}=\begin{pmatrix}\left\{(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\otimes I_K\right\} & O_{K^2p+K\times\widetilde{K}} & O_{K^2p+K\times Kr}\\ O_{\widetilde{K}\times K^2p+K} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr}\\ -\left[\left(\frac{1}{T}M^*\widetilde{Z}^{*\prime}\right)\left(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime}\right)^{-1}\right]\otimes I_K & O_{Kr\times\widetilde{K}} & I_{Kr}\end{pmatrix} \tag{B.25}$$

$$\times\begin{pmatrix}\begin{pmatrix}I_K & O_{K\times K^2}\\ O_{K^2p\times K} & \sum_{j=1}^{T-1}(C_j\otimes I_K)\end{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\widetilde{u}_t^*\\ \mathrm{vec}(\widetilde{u}_t^*(\widetilde{u}_{t-j}^{*\prime}+\nu'))\end{pmatrix}\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}L_K\left\{\mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime})-\mathbb{E}^*(\mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}))\right\}\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\mathrm{vec}(\widetilde{u}_t^*m_t^{*\prime})-\mathbb{E}^*(\mathrm{vec}(\widetilde{u}_t^*m_t^{*\prime}))\right\}\end{pmatrix} \tag{B.26}$$

$$=:\widetilde{Q}_T^*\left(A_q^*+(A^*-A_q^*)\right), \tag{B.27}$$

where $M^*=[m_1^*,\ldots,m_T^*]$ and A^* denotes the term in brackets in equation (B.26) and A_q^* is the same expression, but with $\sum_{j=1}^{T-1}$ replaced by $\sum_{j=1}^{q}$ for some $q\in\mathbb{N}$. In the following, we make use of Proposition 6.3.9 of Brockwell and Davis (1991) and, to show asymptotic normality of A, it suffices to show

$$\begin{aligned}&(a)\quad A_q^*\xrightarrow{\mathcal{D}}\mathcal{N}(0,\Omega_q^A)\quad\text{in probability as } T\to\infty\\ &(b)\quad \Omega_q^A\to\Omega^A\text{ as } q\to\infty\\ &(c)\quad \forall\,\delta>0:\ \lim_{q\to\infty}\limsup_{T\to\infty}P^*(|A^*-A_q^*|_1>\delta)=0\quad\text{in probability.}\end{aligned}$$

To prove (a), we can write

$$A_q^*=R_q\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\widetilde{u}_t^*\\ \mathrm{vec}(\widetilde{u}_t^*(\widetilde{u}_{t-1}^{*\prime}+\nu'))\\ \vdots\\ \mathrm{vec}(\widetilde{u}_t^*(\widetilde{u}_{t-q}^{*\prime}+\nu'))\\ L_K\left\{\mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime})-\mathbb{E}^*(\mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}))\right\}\\ \mathrm{vec}(\widetilde{u}_t^*m_t^{*\prime})-\mathbb{E}^*(\mathrm{vec}(\widetilde{u}_t^*m_t^{*\prime}))\end{pmatrix}=R_q\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{W}_{t,q}^*$$

with R_q as defined in the proof of Theorem 2.1 and with an obvious notation for the $(K+$

$K^2q + \widetilde{K} + Kr$)-dimensional vector $\widetilde{W}^*_{t,q}$. Using Lemma B.3, we get

$$A^*_q \xrightarrow{D} \mathcal{N}(0, \Omega^A_q), \quad \text{where} \quad \Omega^A_q = R_q \Omega_q R_q' = \begin{pmatrix} [\Omega^A_q]^{(i,j)}, \\ i,j = 1,2,3 \end{pmatrix} \tag{B.28}$$

for $T \to \infty$ and with Ω^A_q as defined in the proof of Theorem 2.1. Part (b) follows from the cumulant summability condition imposed in Assumption 4.1 in the paper and the exponential decay of $(C_j, j \in \mathbb{N})$. As it remains to consider the second part of the first sub-vector of $A^* - A^*_q$ in Equation (B.26) only, part (c) follows as in Theorem 4.1 in Brüggemann, Jentsch, and Trenkler (2016). Finally, using similar arguments as in the proof of Lemma A.2 in Brüggemann, Jentsch, and Trenkler (2016), it is possible to show that $\widetilde{Q}^*_T \to Q$ with respect to P^*, where Q is defined in (B.15), which concludes the proof. □

Lemma B.2 (Equivalence of bootstrap estimators) *Under the assumptions of Theorem 4.1 in the paper, we have*

$$\sqrt{T}\left(\left(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\right) - \left(\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}}\right)\right) = o_{P^*}, \tag{B.29}$$

$$\sqrt{T}\left((\widehat{\boldsymbol{\sigma}}^* - \widehat{\boldsymbol{\sigma}}) - (\widetilde{\boldsymbol{\sigma}}^* - \widetilde{\boldsymbol{\sigma}})\right) = o_{P^*}, \tag{B.30}$$

$$\sqrt{T}\left((\widehat{\boldsymbol{\varphi}}^* - \widehat{\boldsymbol{\varphi}}) - (\widetilde{\boldsymbol{\varphi}}^* - \widetilde{\boldsymbol{\varphi}})\right) = o_{P^*}, \tag{B.31}$$

where $\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}} := ((\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\widetilde{Z}^* \otimes \underbrace{I_K}) \widetilde{\mathbf{u}}^*$, $\widetilde{\boldsymbol{\sigma}}^* - \widetilde{\boldsymbol{\sigma}} = \underbrace{\text{vech}}(\widetilde{\Sigma}^*_u) - \mathbb{E}^*(\underbrace{\text{vech}}(\widetilde{\Sigma}^*_u))$ *with* $\widetilde{\Sigma}^*_u = \frac{1}{T}\sum_{t=1}^T \widehat{\widetilde{u}}^*_t \widehat{\widetilde{u}}^{*\prime}_t$, *and* $\widetilde{\boldsymbol{\varphi}}^* - \widetilde{\boldsymbol{\varphi}} = \text{vec}(\underbrace{H^{(1)*}\Psi^{*\prime}}) - \mathbb{E}^*(\text{vec}(\underbrace{H^{(1)*}\Psi^{*\prime}}))$ *with* $\underbrace{H^{(1)*}\Psi^{*\prime}} = \frac{1}{T}\sum_{t=1}^T \widehat{\widetilde{u}}^*_t m^{*\prime}_t$. *Here, pre-sample values* $\widetilde{y}^*_{-p+1}, \ldots, \widetilde{y}^*_0$ *are set to zero and* $\widetilde{y}^*_1, \ldots, \widetilde{y}^*_T$ *is generated according to* $\widetilde{y}^*_t = \nu + A_1\widetilde{y}^*_{t-1} + \cdots + A_p\widetilde{y}^*_{t-p} + \widetilde{u}^*_t$, *where* $\widetilde{u}^*_1, \ldots, \widetilde{u}^*_T$ *is an analogously drawn version of* $u^*_1, \ldots, u^*_T$ *as described in Steps 2 and 3 of the MBB procedure introduced in Section 4 of the paper, but from* $u_1, \ldots, u_T$ *instead of* $\widehat{u}_1, \ldots, \widehat{u}_T$. *Further, we use the notation* $\widetilde{Z}^*_t = (1, \widetilde{y}^{*\prime}_t, \ldots, \widetilde{y}^{*\prime}_{t-p+1})'$ $((Kp+1)\times 1)$, $\widetilde{Z}^* = (\widetilde{Z}^*_0, \ldots, \widetilde{Z}^*_{T-1})$ $((Kp+1)\times T)$, *and* $\widetilde{\mathbf{u}}^* = \text{vec}(\widetilde{u}^*_1, \ldots, \widetilde{u}^*_T)$ $(KT\times 1)$. *Finally,* $\widehat{\widetilde{u}}^*_t = \widetilde{y}^*_t - \widetilde{\nu}^* - \widetilde{A}^*_1\widetilde{y}^*_{t-1} - \cdots - \widetilde{A}^*_p\widetilde{y}^*_{t-p}$ *are the residuals from a VAR(p) model fitted to the bootstrap sample* $\widetilde{y}^*_{-p+1}, \ldots, \widetilde{y}^*_0, \widetilde{y}^*_1, \ldots, \widetilde{y}^*_T$ *leading to coefficients* $\widetilde{\nu}^*, \widetilde{A}^*_1, \ldots, \widetilde{A}^*_p$.

Proof.

Using similar arguments as in the proof of Lemma A.1 in Brüggemann, Jentsch, and Trenkler (2016), the claim follows. However, note that in contrast to that Lemma A.1, we allow for an intercept in (B.29), include proxy variables to get (B.31) and we do not (yet) replace the bootstrap residuals $\widehat{\widetilde{u}}^*_t$ by the bootstrap errors $\widetilde{u}^*_t$ to get (B.30) and (B.31). □

Lemma B.3 (CLT for bootstrap innovations) *Suppose the assumptions of Theorem 4.1 hold and let* $\widetilde{W}^*_{t,q} = (\widetilde{W}^{*(0)\prime}_t, \widetilde{W}^{*(1)\prime}_{t,q}, \widetilde{W}^{*(2)\prime}_t, \widetilde{W}^{*(3)\prime}_t)'$, *where*

$$\begin{aligned}
\widetilde{W}^{*(0)}_t &= \widetilde{u}^*_t \\
\widetilde{W}^{*(1)}_{t,q} &= (\mathrm{vec}(\widetilde{u}^*_t(\widetilde{u}^{*\prime}_{t-1} + \nu'))', \ldots, \mathrm{vec}(\widetilde{u}^*_t(\widetilde{u}^{*\prime}_{t-q} + \nu'))')' \\
\widetilde{W}^{*(2)}_t &= L_K\{\mathrm{vec}(\widetilde{u}^*_t\widetilde{u}^{*\prime}_t) - \mathbb{E}^*(\mathrm{vec}(\widetilde{u}^*_t\widetilde{u}^{*\prime}_t))\} = \mathrm{vech}(\widetilde{u}^*_t\widetilde{u}^{*\prime}_t) - \mathbb{E}^*(\mathrm{vech}(\widetilde{u}^*_t\widetilde{u}^{*\prime}_t)) \\
\widetilde{W}^{*(3)}_t &= \mathrm{vec}(\widetilde{u}^*_t m^{*\prime}_t) - \mathbb{E}^*(\mathrm{vec}(\widetilde{u}^*_t m^{*\prime}_t)).
\end{aligned}$$

Then, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{W}^*_{t,q} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega_q),$$

in probability, where Ω_q *as defined in Lemma B.1.*

Proof.
Using similar arguments as in the proof of Lemma A.3 in Brüggemann, Jentsch, and Trenkler (2016), asymptotic normality can be deduced for the properly centered quantity $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{W}^*_{t,q} - \mathbb{E}^*(\widetilde{W}^*_{t,q})$ (note that $\mathbb{E}^*(\widetilde{W}^{*(1)}_{t,q})$ might be not exactly zero) making use of the α-mixing and cumulant conditions imposed in Assumptions 2.4 and 4.1 in the paper. Finally, because $\sqrt{T}\mathbb{E}^*(\widetilde{W}^*_{t,q}) = o_P(1)$, we get the claimed result. □

B.6 Proof of Theorem 4.2

By replacing the α-mixing condition in Assumption 2.4 imposed for the strong proxy case by the uniform mixing condition in Assumptions 3.2 and 4.2 used for a weak proxy, the proof follows by the same arguments as used in Theorem 4.1.

B.7 Proof of Theorem 4.3

(i) To prove (30) in the paper, we can write

$$\begin{aligned}
\sqrt{T}\begin{pmatrix} se_j'\widehat{\Phi}^*_i\widehat{\boldsymbol{\varphi}}^* \\ e_m'\widehat{\boldsymbol{\varphi}}^* \end{pmatrix} &= \begin{pmatrix} se_j'\widehat{\Phi}^*_i\left(\widehat{\boldsymbol{\varphi}}^* - \widehat{\boldsymbol{\varphi}}\right) + se_j'\left(\widehat{\Phi}^*_i - \widehat{\Phi}_i\right)\widehat{\boldsymbol{\varphi}} + se_j'\widehat{\Phi}_i\widehat{\boldsymbol{\varphi}} \\ e_m'\left(\widehat{\boldsymbol{\varphi}}^* - \widehat{\boldsymbol{\varphi}}\right) + e_m'\widehat{\boldsymbol{\varphi}} \end{pmatrix} \\
&= \begin{pmatrix} se_j'\widehat{\Phi}^*_i \\ e_m' \end{pmatrix}\sqrt{T}\left(\widehat{\boldsymbol{\varphi}}^* - \widehat{\boldsymbol{\varphi}}\right) + \begin{pmatrix} se_j'\left(\widehat{\Phi}^*_i - \widehat{\Phi}_i\right)\widehat{\boldsymbol{\varphi}} \\ 0 \end{pmatrix} + \begin{pmatrix} se_j'\widehat{\Phi}_i \\ e_m' \end{pmatrix}\sqrt{T}\widehat{\boldsymbol{\varphi}}
\end{aligned}$$

By Theorem 4.2, we have

$$\begin{pmatrix} se_j'\widehat{\Phi}_i^* \\ e_m' \end{pmatrix} \sqrt{T}\left(\widehat{\boldsymbol{\varphi}}^* - \widehat{\boldsymbol{\varphi}}\right) \xrightarrow{D} \begin{pmatrix} se_j'\Phi_i \\ e_m' \end{pmatrix} \mathcal{N}\left(0, V_{weak}^{(3,3)}\right),$$

because $\sqrt{T}\widehat{\boldsymbol{\varphi}} = \sqrt{T}(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_T) + \sqrt{T}\widehat{\boldsymbol{\varphi}} = \sqrt{T}(\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_T) + C = O_{P^*}(1)$, the second summand $se_j'\left(\widehat{\Phi}_i^* - \widehat{\Phi}_i\right)\widehat{\boldsymbol{\varphi}}$ vanishes asymptotically and the third term converges by Theorem 3.2(i) such that $\sqrt{T}(se_j'\widehat{\Phi}_i^*\widehat{\boldsymbol{\varphi}}^*, e_m'\widehat{\boldsymbol{\varphi}}^*)$ converges (conditional on the data) to

$$\begin{pmatrix} X_{1,weak} \\ X_{2,weak} \end{pmatrix} + \begin{pmatrix} Z_{1,weak} \\ Z_{2,weak} \end{pmatrix},$$

where $(Z_{1,weak}, Z_{2,weak})'$ as in Theorem 3.2(i.a) in the paper. Hence, (31) follows immediately and to prove (32) in the paper, similarly to (24) in the paper, we can write

$$\begin{aligned}
&\sqrt{T}[(se_j'\widehat{\Phi}_i^* - \xi_0 e_m')\widehat{\boldsymbol{\varphi}}^* - (se_j'\widehat{\Phi}_i - \xi_0 e_m')\widehat{\boldsymbol{\varphi}}] \\
=&\left(se_j'\widehat{\Phi}_i^* - \xi_0 e_m'\right)\sqrt{T}\left(\widehat{\boldsymbol{\varphi}}^* - \widehat{\boldsymbol{\varphi}}\right) + se_j'\left(\sqrt{T}(\widehat{\Phi}_i^* - \widehat{\Phi}_i)\right)\widehat{\boldsymbol{\varphi}} \\
=&\left(se_j'\widehat{\Phi}_i^* - \xi_0 e_m'\right)\left(\sqrt{T}(\widehat{\boldsymbol{\varphi}}^* - \widehat{\boldsymbol{\varphi}})\right) + \left(\widehat{\boldsymbol{\varphi}}' \otimes (se_j')\right)\left(\sqrt{T}(\mathrm{vec}(\widehat{\Phi}_i^*) - \mathrm{vec}(\widehat{\Phi}_i))\right) \\
=&\left(\widehat{\boldsymbol{\varphi}}' \otimes (se_j'), se_j'\widehat{\Phi}_i^* - \xi_0 e_m'\right)\sqrt{T}\begin{pmatrix} \mathrm{vec}(\widehat{\Phi}_i^*) - \mathrm{vec}(\widehat{\Phi}_i) \\ \widehat{\boldsymbol{\varphi}}^* - \widehat{\boldsymbol{\varphi}} \end{pmatrix} \\
\xrightarrow{D}&\left(O_{1\times K} \otimes (se_j'), se_j'\Phi_i - \xi_0 e_m'\right)\mathcal{N}(0, R_{weak}) \quad \text{in probability} \\
=&\mathcal{N}(0, (se_j'\Phi_i - \xi_0 e_m')V_{weak}^{(3,3)}(se_j'\Phi_i - \xi_0 e_m')'),
\end{aligned}$$

which uses Theorems 3.1 and 4.2, $\widehat{\boldsymbol{\varphi}} \xrightarrow{P} O_{K\times 1}$ with a weak proxy, and $\widehat{\Phi}_i^* \xrightarrow{P} \Phi_i$ in probability.
(ii) Similarly, for a strong proxy, we can show that

$$\begin{aligned}
&\sqrt{T}[(se_j'\widehat{\Phi}_i^* - \xi_0 e_m')\widehat{\boldsymbol{\varphi}}^* - (se_j'\widehat{\Phi}_i - \xi_0 e_m')\widehat{\boldsymbol{\varphi}}] \\
\xrightarrow{D}&\left(\boldsymbol{\varphi}' \otimes (se_j'), se_j'\Phi_i - \xi_0 e_m'\right)\mathcal{N}(0, R)
\end{aligned}$$

□

C Residual-Based Wild Bootstrap

This appendix is about the residual-based wild bootstrap and has three sections. Section C.1 gives the algorithm for the residual-based wild bootstrap. Section C.2 gives the theorem showing the asymptotic inconsistency of the residual-based wild bootstrap and provides some discussion. Section C.3 gives the proof of inconsistency of the residual-based wild bootstrap.

The wild bootstrap approach was proposed in Mertens and Ravn (2013) in a framework that imposed centered proxy variables, i.e. $\mathbb{E}(m_t) = 0$, and that the proxy is uncorrelated with lags of the time series process, i.e. $\mathbb{E}(m_t y_{t-j}) = 0$, $j = 1, \ldots, p$. As the wild bootstrap requires generally centeredness of the random variables and it may capture only rather specific types of dependence, we stick to such conditions in this section. Nevertheless, we are able to prove that even under these additional conditions, the wild bootstrap will generally fail to correctly mimic the limiting distributions in proxy SVAR setups. Note also that without imposing the assumption $\mathbb{E}(m_t) = 0$, an additional (initial) centering step of the proxy, i.e. using $\widetilde{m}_t = m_t - \frac{1}{T}\sum_{t=1}^{T} m_t$ as the proxy to be fed in the wild bootstrap scheme will also not cure this wild bootstrap inconsistently.

For this appendix, we also do not include the intercept, ν, in VAR. Because of this, we use $\boldsymbol{\alpha} = \text{vec}(A_1, \ldots, A_p)$ instead of $\boldsymbol{\beta} = \text{vec}(\nu, A_1, \ldots, A_p)$, which are linked with $\widetilde{J} = [O_{K^2p \times K} : I_{K^2p}]$ via $\boldsymbol{\alpha} = \widetilde{J}\boldsymbol{\beta}$.

C.1 Residual-based Wild Bootstrap Algorithm

The algorithm for the recursive-design residual-based wild bootstrap is as follows:

1. Independently draw T observations of the scalar random sequence $(\eta_t, t \in \mathbb{Z})$ from a distribution with $\mathbb{E}(\eta_t) = 0$, $\mathbb{E}(\eta_t^2) = 1$, and $\mathbb{E}(\eta_t^4) < \infty$.

2. Use $u_t^+ = \widehat{u}_t \eta_t$ to produce $(u_1^+, \ldots, u_T^+)$ and $m_t^+ = m_t \eta_t$ to produce $(m_1^+, \ldots, m_T^+)$. Here, we use " $^+$ " to denote bootstrap quantities obtained from the wild bootstrap.

3. Set the initial condition $(y_{-p+1}^+, \ldots, y_0^+) = (y_{-p+1}, \ldots, y_0)$. Use the initial condition along with $\widehat{A}_1, \ldots, \widehat{A}_p$ and u_t^+ to recursively produce $(y_1^+, \ldots, y_T^+)$ with

$$y_t^+ = \widehat{A}_1 y_{t-1}^+ + \cdots + \widehat{A}_p y_{t-p}^+ + u_t^+.$$

4. Estimate $\widehat{A}_1^+, \ldots, \widehat{A}_p^+$ by least squares from the bootstrap sample $(y_{-p+1}^+, \ldots, y_T^+)$ and set $\widehat{u}_t^+ = y_t^+ - \widehat{A}_1^+ y_{t-1}^+ - \cdots - \widehat{A}_p^+ y_{t-p}^+$.

5. Use $\widehat{u}_t^+$ and m_t^+ for $t = 1, \ldots, T$ to estimate $\widehat{\Sigma}_u^+ = T^{-1}\sum_{t=1}^T \widehat{u}_t^+ \widehat{u}_t^{+\prime}$ and $\widehat{H^{(1)}\Psi'}^+ = T^{-1}\sum_{t=1}^T \widehat{u}_t^+ m_t^{+\prime}$.

6. Use $\widehat{A}_1^+, \ldots, \widehat{A}_p^+$, $\widehat{\Sigma}_u^+$ and $\widehat{H^{(1)}\Psi'}^+$ to produce the bootstrapped IRFs and FEVDs. In this step, use same the identification scheme as when computing the point estimates of the IRFs and FEVDs. This includes using the same sign and scale normalizations.

Repeat the algorithm a large number of times and collect the bootstrapped IRFs and FEVDs. Confidence intervals are commonly produced with a standard percentile interval by sorting the bootstrapped IRFs and FEVDs and keeping the $\alpha/2$- and $1-\alpha/2$-percentiles as the confidence interval, where α is the level of significance. For example, see Mertens and Ravn (2013) and Gertler and Karadi (2015). Using Hall's percentile intervals (Hall (1992) and Lütkepohl (2005, Appendix D)) does not change the asymptotic results below.

Many choices of distribution are available when drawing the bootstrap multiplier η_t. For example, drawing η_t from the standard normal distribution satisfies the conditions in step 1 of the algorithm. A second choice is the Rademacher distribution, which sets $\eta_t = 1$ with probability 0.5 and $\eta_t = -1$ with probability 0.5. This distribution is used in, for example, Mertens and Ravn (2013) and Gertler and Karadi (2015). Because of its use in the proxy SVAR literature, we provide additional results and discussion for the Rademacher distribution in Section C.2. We also use the Rademacer distribution in the simulations in Section G.

C.2 Inconsistency of the Wild Bootstrap

In this section, we show that the wild bootstrap is generally not consistent for inference on $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$ and, consequently, also for statistics that are functions of these estimators. We only study the strong proxy case, $\mathbb{E}(m_t \epsilon_t^{(1)\prime}) = \Psi$ with $\Psi(r \times g)$ and rank g, and we do not discuss the weak proxy case.

Define $\widehat{\boldsymbol{\alpha}}^+ = \text{vec}(\widehat{A}_1^+, \ldots, \widehat{A}_p^+)$, $\widehat{\boldsymbol{\sigma}}^+ = \text{vech}(\widehat{\Sigma}_u^+)$, and $\widehat{\boldsymbol{\varphi}}^+ = \text{vec}(\widehat{H^{(1)}\Psi'}^+)$ to be the estimators from the wild bootstrap that correspond to $\boldsymbol{\alpha}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, respectively. We derive the joint limiting variance of $\sqrt{T}((\widehat{\boldsymbol{\alpha}}^+ - \widehat{\boldsymbol{\alpha}})', (\widehat{\boldsymbol{\sigma}}^+ - \widehat{\boldsymbol{\sigma}})', (\widehat{\boldsymbol{\varphi}}^+ - \widehat{\boldsymbol{\varphi}})')'$ in the following theorem.

Theorem C.1 (Residual-based Wild Bootstrap Limiting Variance) *Suppose Assumptions 2.1, 2.2, 2.4, and 4.1 in the paper all hold as in Theorem 4.1 in the paper. In addition, we assume that $x_t = (u_t', m_t')'$ is a martingale difference sequence (mds), i.e., in particular, $\mathbb{E}(m_t) = 0$. If the residual-based wild bootstrap from Section C.1 is used to*

compute bootstrap statistics $\widehat{\boldsymbol{\alpha}}^+, \widehat{\boldsymbol{\sigma}}^+$ *and* $\widehat{\boldsymbol{\varphi}}^+$*, then we have*

$$T\,Var^+\begin{pmatrix}\widehat{\boldsymbol{\alpha}}^+-\widehat{\boldsymbol{\alpha}}\\ \widehat{\boldsymbol{\sigma}}^+-\widehat{\boldsymbol{\sigma}}\\ \widehat{\boldsymbol{\varphi}}^+-\widehat{\boldsymbol{\varphi}}\end{pmatrix}\to\begin{pmatrix}V_{mds}^{(1,1)} & O_{K^2p\times\widetilde{K}} & O_{K^2p\times Kr}\\ O_{\widetilde{K}\times K^2p} & \tau_{0,0,0}\{\mathbb{E}(\eta_t^4)-1\} & \nu'_{0,0,0}\{\mathbb{E}(\eta_t^4)-1\}\\ O_{Kr\times K^2p} & \nu_{0,0,0}\{\mathbb{E}(\eta_t^4)-1\} & \zeta_{0,0,0}\{\mathbb{E}(\eta_t^4)-1\}\end{pmatrix}=:V_{WB},$$

in which $V_{mds}^{(1,1)}$ denotes $\widetilde{J}V^{(1,1)}\widetilde{J}'$ with $V^{(1,1)}$ as given in Section A.1 but with an additional mds assumption imposed on $x_t=(u_t',m_t')'$.

As $V_{WB}\neq V$ for V as defined in Theorem 2.1 of the paper, a consequence of Theorem C.1 is that the residual-based wild bootstrap is generally inconsistent for statistics that are functions of $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$. However, $V_{WB}^{(1,1)}$ does equal $\widetilde{J}V^{(1,1)}\widetilde{J}'$ after imposing an additional mds assumption on $x_t=(u_t',m_t')'$. Hence, the only exclusion to the invalidity of the wild bootstrap is the case where the statistic of interest is a (smooth) function of $\widehat{\boldsymbol{\alpha}}$ under an additional mds assumption.[1] The validity of the wild bootstrap for smooth functions of $\widehat{\boldsymbol{\alpha}}$ was already addressed for the univariate case by Gonçalves and Kilian (2004) and for the multivariate case by Brüggemann, Jentsch, and Trenkler (2014). The general asymptotic inconsistency of the residual-based wild bootstrap for functions of $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$, such as, for example, structural IRFs, without adding proxy variables to the VAR setup has already been discussed in Brüggemann, Jentsch, and Trenkler (2016), who show that the wild bootstrap cannot replicate the fourth moments of the VAR innovations. Note also that imposing iid-ness for the process $(x_t, t\in\mathbb{Z})$ does not lead to wild bootstrap consistency either. Compare Corollary A.2 in their paper.

If the i.i.d. bootstrap multipliers $(\eta_t, t\in\mathbb{Z})$ follow a Rademacher distribution, we have $\mathbb{E}^+(\eta_t^4)=\mathbb{E}(\eta_t^4)=1$, which immediately leads to the following corollary.

Corollary C.1 (Residual-based Rademacher Wild Bootstrap Limiting Variance) *Under the assumptions of Theorem C.1 and if the (iid) bootstrap multipliers* $(\eta_t, t\in\mathbb{Z})$ *follow a Rademacher distribution, that is* $P(\eta_t=-1)=P(\eta_t=1)=0.5$*, we get*

$$V_{WB}=\begin{pmatrix}V_{mds}^{(1,1)} & O_{K^2p\times\widetilde{K}+Kr}\\ O_{\widetilde{K}+Kr\times K^2p} & O_{\widetilde{K}+Kr\times\widetilde{K}+Kr}\end{pmatrix}. \tag{C.1}$$

A comparison of V_{WB} in Equation (C.1) with V from Theorem 2.1 of the paper leads to the conclusion that a considerable amount of estimation uncertainty caused by estimating

[1] The wild bootstrap would also remain valid under mds assumptions in a very special and unrealistic scenario where $V^{(2,1)}$ and $V^{(3,1)}$ vanish and $\mathbb{E}(\eta_t^4)$ accidently yields $V_{WB}^{(i,j)}=V^{(i,j)}$ for $i,j=1,2$.

Σ_u and $H^{(1)}\Psi'$ with $\widehat{\Sigma}_u$ and $\widehat{\Psi H^{(1)\prime}}$, respectively, is simply ignored by the wild bootstrap using a Rademacher distribution for the bootstrap multipliers. Consequently, as can also be seen in the Monte Carlo simulations conducted in Section G of the paper, the wild bootstrap clearly leads to considerable undercoverage of corresponding bootstrap confidence intervals for IRFs and FEVDs.

To see why the wild bootstrap asymptotically ignores, for example, the variance of $H^{(1)}\Psi'$, we temporarily consider a simpler specification than the VAR and assume that u_t can be directly observed. Then, the Rademacher wild bootstrap estimate of $H^{(1)}\Psi'$ is given by

$$\widehat{H^{(1)}\Psi'}^{+} = T^{-1}\sum_{t=1}^{T} u_t^{+} m_t^{+\prime}.$$

Because $u_t^{+} = u_t\eta_t$ and $m_t^{+} = m_t\eta_t$ and η_t equals 1 or -1, it is the case that

$$\widehat{H^{(1)}\Psi'}^{+} = T^{-1}\sum_{t=1}^{T}(\eta_t)^2 u_t m_t' = T^{-1}\sum_{t=1}^{T} u_t m_t' = \widehat{H^{(1)}\Psi'}.$$

That is, when u_t is directly observable, the Rademacher wild bootstrap estimator is simply the non-bootstrapped sample estimate and $\widehat{H^{(1)}\Psi'}^{+} = \widehat{H^{(1)}\Psi'}$ holds for every bootstrap replication. This implies that the uncertainty in the estimation of the covariance $H^{(1)}\Psi'$ is completely ignored and hence not captured by the Rademacher wild bootstrap.

Going back to the VAR, it is not the case that u_t is directly observable. Thus, in the bootstrap, we use $\widehat{u}_t^{+}$ rather than u_t^{+} to estimate the covariances. Because $\widehat{u}_t^{+}$ is different for each bootstrap replication, it will not be the case that $\widehat{H^{(1)}\Psi'}^{+} = \widehat{H^{(1)}\Psi'}$ holds exactly, but $\widehat{H^{(1)}\Psi'}^{+} = \widehat{H^{(1)}\Psi'} + o_{P^+}(1)$ as $T \to \infty$, where P^+ denotes the probability measure induced by the Rademacher wild bootstrap. Hence, although the bootstrapped variance of $H^{(1)}\Psi'$ will generally not be zero in finite samples with the Rademacher wild bootstrap, it will converge to zero as the sample size increases.

C.3 Proof of Theorem C.1

As $u_t^{+} = \widehat{u}_t\eta_t$ and $m_t^{+} = m_t\eta_t$, by taking conditional expectations, we get

$$\mathbb{E}^{+}\left(\text{vec}(u_t^{+}u_{t-a}^{+\prime})\text{vec}(u_{t-b}^{+}u_{t-c}^{+\prime})'\right) = \text{vec}(\widehat{u}_t\widehat{u}_{t-a}')\text{vec}(\widehat{u}_{t-b}\widehat{u}_{t-c}')'\mathbb{E}\left(\eta_t\eta_{t-a}\eta_{t-b}\eta_{t-c}\right), \qquad \text{(C.2)}$$

where

$$\mathbb{E}\left(\eta_t\eta_{t-a}\eta_{t-b}\eta_{t-c}\right) = \begin{cases} \mathbb{E}(\eta_t^4), & a=b=c=0 \\ 1, & a=0\neq b=c \text{ or } b=0\neq a=c \text{ or } c=0\neq a=b\,. \\ 0, & \text{otherwise} \end{cases} \quad \text{(C.3)}$$

Note that analogous representations also hold for $\mathbb{E}^+\left(\text{vec}(u^+_{t-a}m^{+\prime}_t)\text{vec}(u_{t-b}u^{+\prime}_{t-c})'\right)$ as well as $\mathbb{E}^+\left(\text{vec}(u^+_{t-a}m^{+\prime}_t)\text{vec}(u^+_{t-b}m^{+\prime}_{t-c})'\right)$. Now, by using arguments similar to those used in the proof of Theorem 4.1 in Section B.5, we can show that the variance of $\sqrt{T}((\widehat{\boldsymbol{\alpha}}^+ - \widehat{\boldsymbol{\alpha}})', (\widehat{\boldsymbol{\sigma}}^+ - \widehat{\boldsymbol{\sigma}})', (\widehat{\boldsymbol{\varphi}}^+ - \widehat{\boldsymbol{\varphi}})')'$ converges to a quantity corresponding to V as defined in Theorem 2.1 in the paper, where all $\tau_{a,b,c}, \nu_{a,b,c}$ and $\zeta_{a,b,c}$ terms have to be replaced by $\tau_{a,b,c}\mathbb{E}\left(\eta_t\eta_{t-a}\eta_{t-b}\eta_{t-c}\right), \nu_{a,b,c}\mathbb{E}\left(\eta_t\eta_{t-a}\eta_{t-b}\eta_{t-c}\right)$ and $\zeta_{a,b,c}\mathbb{E}\left(\eta_t\eta_{t-a}\eta_{t-b}\eta_{t-c}\right)$, respectively, leading to the claimed result. □

D Montiel Olea, Stock and Watson Analytical Confidence Sets

In this section, we provide additional detail about the Montiel Olea, Stock, and Watson (forthcoming) (MSW) confidence sets that we use in our Monte Carlo simulations.

MSW construct confidence sets by collecting all values of ξ_0 such that $\sqrt{T}(se_j'\widehat{\Phi}_i - \xi_0 e_m')\widehat{H^{(1)}\psi} = \sqrt{T}(se_j'\widehat{\Phi}_i - \xi_0 e_m')\widehat{\boldsymbol{\varphi}}$ is not statistically different from zero. We are in the $r=g=1$ case so that $H^{(1)}\psi = \boldsymbol{\varphi}$ is $K\times 1$. Then, building off of Equation (18) in the paper, we can write $\sqrt{T}(se_j'\widehat{\Phi}_i - \xi_0 e_m')\widehat{\boldsymbol{\varphi}}$ as

$$\begin{pmatrix} 1 & -\xi_0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\varphi}'\otimes(se_j') & se_j'\widehat{\Phi}_i \\ 0 & e_m' \end{pmatrix} \sqrt{T} \begin{pmatrix} \text{vec}(\widehat{\Phi}_i) - \text{vec}(\Phi_i) \\ \widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi} \end{pmatrix}.$$

We can write this as

$$\begin{pmatrix} 1 & -\xi_0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\varphi}'\otimes(se_j') & se_j'\widehat{\Phi}_i \\ 0 & e_m' \end{pmatrix} \sqrt{T} \begin{pmatrix} \text{vec}(\widehat{\Phi}_i) - \text{vec}(\Phi_i) \\ \widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_T \end{pmatrix}$$

for the weak proxy case, in which $\boldsymbol{\varphi}_T = H^{(1)}\psi_T = H^{(1)}\psi_T/\sqrt{T}$.

In the strong proxy case, we have

$$\sqrt{T}\begin{pmatrix}\text{vec}(\widehat{\Phi}_i)-\text{vec}(\Phi_i)\\ \widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi}\end{pmatrix}\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,R\right),\quad R=\begin{pmatrix}G_i\widetilde{J}V^{(1,1)}\widetilde{J}'G_i' & G_i\widetilde{J}V^{(3,1)\prime}\\ V^{(3,1)}\widetilde{J}'G_i' & V^{(3,3)}\end{pmatrix}$$

and in the weak proxy case, we have

$$\sqrt{T}\begin{pmatrix}\text{vec}(\widehat{\Phi}_i)-\text{vec}(\Phi_i)\\ \widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi}\end{pmatrix}\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,R_{weak}\right),\quad R=\begin{pmatrix}G_i\widetilde{J}V_{weak}^{(1,1)}\widetilde{J}'G_i' & G_i\widetilde{J}V_{weak}^{(3,1)\prime}\\ V_{weak}^{(3,1)}\widetilde{J}'G_i' & V_{weak}^{(3,3)}\end{pmatrix}.$$

In both the strong and weak case, we use $\widetilde{J}=[O_{K^2p\times K}:I_{K^2p}]$ and $G_i=\partial\text{vec}(\Phi_i)/\partial\boldsymbol{\alpha}'$ with $\boldsymbol{\alpha}=\text{vec}(A_1,\ldots,A_p)$. We use the formula in Lütkepohl (2005, Proposition 3.6) to compute estimates of G_i, denoted with $\widehat{G}_i$, by plugging $\widehat{A}_1,\ldots,\widehat{A}_p$ in for $A_1,\ldots,A_p$.

To proceed in computing the MSW sets, we need estimates of $V^{(1,1)}$, $V^{(3,1)}$, and $V^{(3,3)}$. We also need estimates of $V_{weak}^{(1,1)}$, $V_{weak}^{(3,1)}$, and $V_{weak}^{(3,3)}$. We use the same estimator in both the strong and weak case. For the estimator, we use the $[\Omega^A]^{(i,j)}$ notation from Section A.1 and add hats to denote estimators. We first compute

$$\begin{pmatrix}[\widehat{\Omega}^A]^{(1,1)} & [\widehat{\Omega}^A]^{(1,3)}\\ [\widehat{\Omega}^A]^{(3,1)} & [\widehat{\Omega}^A]^{(3,3)}\end{pmatrix}=T^{-1}\sum_{t=1}^{T}\begin{pmatrix}\text{vec}(\widehat{u}_tZ_{t-1}')\\ \widehat{u}_tm_t\end{pmatrix}\begin{pmatrix}\text{vec}(\widehat{u}_tZ_{t-1}')\\ \widehat{u}_tm_t\end{pmatrix}'$$

where Z_t is defined in Equation (13) of the paper. Then, we compute

$$\begin{aligned}\begin{pmatrix}\widehat{V}^{(1,1)} & \widehat{V}^{(3,1)\prime}\\ \widehat{V}^{(3,1)} & \widehat{V}^{(3,3)}\end{pmatrix}&=\begin{pmatrix}\widehat{V}_{weak}^{(1,1)} & \widehat{V}_{weak}^{(3,1)\prime}\\ \widehat{V}_{weak}^{(3,1)} & \widehat{V}_{weak}^{(3,3)}\end{pmatrix}\\&=\begin{pmatrix}\widehat{\Gamma}_{ZZ}^{-1}\otimes I_K & O_{K^2p+K\times K}\\ \widehat{\Gamma}_{MZ}\widehat{\Gamma}_{ZZ}^{-1}\otimes I_K & I_K\end{pmatrix}\begin{pmatrix}[\widehat{\Omega}^A]^{(1,1)} & [\widehat{\Omega}^A]^{(1,3)}\\ [\widehat{\Omega}^A]^{(3,1)} & [\widehat{\Omega}^A]^{(3,3)}\end{pmatrix}\begin{pmatrix}\widehat{\Gamma}_{ZZ}^{-1}\otimes I_K & O_{K^2p+K\times K}\\ \widehat{\Gamma}_{MZ}\widehat{\Gamma}_{ZZ}^{-1}\otimes I_K & I_K\end{pmatrix}',\end{aligned}$$

in which $\widehat{\Gamma}_{ZZ}=T^{-1}\sum_{t=1}^{T}Z_{t-1}Z_{t-1}'$ and $\widehat{\Gamma}_{MZ}=T^{-1}\sum_{t=1}^{T}m_tZ_{t-1}'$. For the rest of this section, we only use the $\widehat{V}^{(1,1)}$, $\widehat{V}^{(3,1)}$, and $\widehat{V}^{(3,3)}$ notation with the understanding that is covers both the strong and weak cases.

Next, under the null of $\Xi_{j1,i}(s;m,1)=\xi_0$ we estimate the asymptotic variance of

$\sqrt{T}(se_j'\Phi_i - \xi_0 e_m')\boldsymbol{\varphi}$ to be

$$\widehat{V}_{(se_j'\Phi_i-\xi_0 e_m')\boldsymbol{\varphi}}$$
$$= \begin{pmatrix} 1 & -\xi_0 \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\varphi}}' \otimes (se_j') & se_j'\widehat{\Phi}_i \\ 0 & e_m' \end{pmatrix} \begin{pmatrix} \widehat{G}_i \widetilde{J}\widehat{V}^{(1,1)}\widetilde{J}'\widehat{G}_i' & \widehat{G}_i\widetilde{J}\widehat{V}^{(3,1)\prime} \\ \widehat{V}^{(3,1)}\widetilde{J}'\widehat{G}_i' & \widehat{V}^{(3,3)} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\varphi}}' \otimes (se_j') & se_j'\widehat{\Phi}_i \\ 0 & e_m' \end{pmatrix}' \begin{pmatrix} 1 \\ -\xi_0 \end{pmatrix},$$

which is quadratic in ξ_0.

To produce the $1-\alpha$ MSW confidence set, we follow MSW and compute a Wald statistic with $T(se_j'\widehat{\Phi}_i\widehat{\boldsymbol{\varphi}} - \xi_0 e_m'\widehat{\boldsymbol{\varphi}})^2$ in the numerator and $\widehat{V}_{(se_j'\Phi_i-\xi_0 e_m')\boldsymbol{\varphi}}$ in the denominator. Then, we keep all values ξ_0 that yield a Wald statistic less than or equal to $\chi^2_{1,1-\alpha}$. We write out the numerator as

$$T(se_j'\widehat{\Phi}_i\widehat{\boldsymbol{\varphi}})^2 - \xi_0 2T(se_j'\widehat{\Phi}_i\widehat{\boldsymbol{\varphi}})(e_m'\widehat{\boldsymbol{\varphi}}) + \xi_0^2 T(e_m'\widehat{\boldsymbol{\varphi}})^2 \tag{D.1}$$

and the denominator as

$$\begin{aligned} s^2(\widehat{\boldsymbol{\varphi}}\otimes e_j)'\widehat{G}_i\widetilde{J}\widehat{V}^{(1,1)}\widetilde{J}'\widehat{G}_i'(\widehat{\boldsymbol{\varphi}}\otimes e_j) + 2s^2(\widehat{\boldsymbol{\varphi}}\otimes e_j)'\widehat{G}_i\widetilde{J}\widehat{V}^{(3,1)\prime}\widehat{\Phi}_i' e_j + s^2 e_j'\widehat{\Phi}_i\widehat{V}^{(3,3)}\widehat{\Phi}_i' e_j \\ -\xi_0[2s(\widehat{\boldsymbol{\varphi}}\otimes e_j)'\widehat{G}_i\widetilde{J}\widehat{V}^{(3,1)\prime}e_m + 2se_j'\widehat{\Phi}_i\widehat{V}^{(3,3)}e_m] + \xi_0^2 e_m'\widehat{V}^{(3,3)}e_m. \end{aligned} \tag{D.2}$$

Written out in this way, we can see that the Wald statistic being less than or equal to $\chi^2_{1,1-\alpha}$ can be written as a quadratic inequality in ξ_0. That is, it can take the form

$$a\xi_0^2 + b\xi_0 + c \leq 0. \tag{D.3}$$

Notice that if $a > 0$, then a plot of $a\xi_0^2 + b\xi_0 + c$ as a function of ξ_0 opens up. In this case, the set of ξ_0 that satisfies Equation (D.3) is the interval $(-b \pm \sqrt{b^2 - 4ac})/2a$. If $a < 0$, then a plot of $a\xi_0^2 + b\xi_0 + c$ as a function of ξ_0 opens down. In this case, the set of ξ_0 that satisfies Equation (D.3) is either the entire real line or the union of $(-\infty, -(b + \sqrt{b^2 - 4ac})/2a)$ and $(-(b - \sqrt{b^2 - 4ac})/2a, \infty)$.

Equations (D.1), (D.2) and (D.3) imply that

$$a = T(e_m'\widehat{\boldsymbol{\varphi}})^2 - \chi^2_{1,1-\alpha} e_m'\widehat{V}^{(3,3)}e_m. \tag{D.4}$$

Then, the condition that $a > 0$ to ensure that the MSW set is one bounded interval is equivalent to

$$\frac{T(e_m'\widehat{\boldsymbol{\varphi}})^2}{e_m'\widehat{V}^{(3,3)}e_m} > \chi^2_{1,1-\alpha}, \tag{D.5}$$

which is a Wald test that would reject the null hypothesis of $e_m'\boldsymbol{\varphi} = 0$ for the alternative

$e_m'\boldsymbol{\varphi} \neq 0$. That is, if the covariance of $\widehat{u}_t^{(m)}$ and m_t is statistically different from zero, then the MSW confidence set is one bounded interval. We use this test in our Monte Carlo simulations to verify if the MSW sets are bounded or not.

E Tests of Proxy Strength

In Section 5 of the paper, we discuss tests of proxy strength. We use this section to lay out the details of these tests. For all of these tests, we only consider the $r = g = 1$ case. In this section, we use the notation $M = [m_1, \ldots, m_T]'$, $\widehat{U} = [\widehat{u}_1, \ldots, \widehat{u}_T]'$ in which $\widehat{u}_t$ is a VAR residual, $\widehat{U}_1 = [\widehat{u}_{1,1}, \ldots, \widehat{u}_{1,T}]'$ where $\widehat{u}_{1,t}$ is the first element of $\widehat{u}_t$, and $\mathbf{1}_T$ is a $(T \times 1)$ vector of 1s.

We consider two different tests, both using F statistics. For the first test, consider the least squares regression of m_t on a constant and $\widehat{u}_t$. We denote the vector of regression errors as

$$\hat{E} = M - [\mathbf{1}_T, \widehat{U}]([\mathbf{1}_T, \widehat{U}]'[\mathbf{1}_T, \widehat{U}])^{-1}([\mathbf{1}_T, \widehat{U}]'M).$$

Then, the F statistic is

$$F = \left(\frac{T-1}{K}\right)\left(\frac{\sum_{t=1}^T (m_t - \bar{m})^2}{\widehat{E}'\widehat{E}} - 1\right), \tag{E.1}$$

in which $\bar{m} = T^{-1}\sum_{t=1}^T m_t$. Stock and Watson (2012) use this F statistic and Lunsford (2015) provides further analysis.

For the second test, we partition u_1 and H so that

$$\begin{bmatrix} u_t^{(1)} \\ (1 \times 1) \\ \\ u_t^{(2)} \\ (K-1 \times 1) \end{bmatrix} = \begin{bmatrix} H^{(1,1)} & H^{(1,2)} \\ (1 \times 1) & (1 \times K-1) \\ \\ H^{(2,1)} & H^{(2,2)} \\ (K-1 \times 1) & (K-1 \times K-1) \end{bmatrix} \begin{bmatrix} \epsilon_t^{(1)} \\ (1 \times 1) \\ \\ \epsilon_t^{(2)} \\ (K-1 \times 1) \end{bmatrix}.$$

In this partition, $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ are defined as in the paper. In practice, the partition of the VAR innovations often follows an economic logic. For example, if $\epsilon_t^{(1)}$ is a monetary policy shock, then $u_t^{(1)}$ is the VAR innovation that corresponds to the monetary policy indicator – typically a short-term interest rate (Gertler and Karadi, 2015). We assume that $H^{(1,1)} \neq 0$ and that $H^{(2,2)}$ is invertible.

With this partition, we can rearrange terms to get

$$u_t^{(2)} = H^{(2,1)}H^{(1,1)-1}u_t^{(1)} + \left(H^{(2,2)} - H^{(2,1)}H^{(1,1)-1}H^{(1,2)}\right)\epsilon_t^{(2)},$$

which shows that $u_t^{(2)}$ can be written as a function of $u_t^{(1)}$ and $\epsilon_t^{(2)}$. We now make two remarks. First, because this is the $r = g = 1$ case, the $j$the element of $H^{(2,1)}H^{(1,1)-1}$ can be written as $(e'_{j+1}H^{(1)}\sigma^2_{\epsilon^{(1)}})/(e'_1H^{(1)}\sigma^2_{\epsilon^{(1)}}) = \Xi_{j+1,1,0}(1;1,1)$. Hence, $H^{(2,1)}H^{(1,1)-1}$ yields the normalized IRF on impact with $s = 1$. Second, $u_t^{(1)}$ is correlated with $\left(H^{(2,2)} - H^{(2,1)}H^{(1,1)-1}H^{(1,2)}\right)\epsilon_t^{(2)}$. Because of this correlation, we cannot consistently estimate $H^{(2,1)}H^{(1,1)-1}$ by regressing $u_t^{(2)}$ on $u_t^{(1)}$. Rather, we can use m_t as an instrument and compute

$$\widehat{H^{(2,1)}H^{(1,1)-1}} = \left(\sum_{t=1}^{T}\widehat{u}_t^{(2)}m_t\right)\left(\sum_{t=1}^{T}\widehat{u}_t^{(1)}m_t\right)^{-1}.$$

This is the IV estimator of the normalized IRF. The corresponding first-stage is the regression of $\widehat{u}_t^{(1)}$ on m_t. Hence, we test the strength of m_t by computing the F statistic of the regression of $\widehat{u}_t^{(1)}$ on m_t, which we denote with F_{IV} to avoid confusion with F statistic in (E.1). First, we compute

$$\widetilde{E} = \widehat{U}_1 - [\mathbf{1}_T, M]([\mathbf{1}_T, M]'[\mathbf{1}_T, M])^{-1}([\mathbf{1}_T, M]'\widehat{U}_1).$$

Then, we have

$$F_{IV} = (T-1)\left(\frac{\widehat{U}'\widehat{U}}{\widetilde{E}'\widetilde{E}} - 1\right) \tag{E.2}$$

The F statistics in Equations (E.1) and (E.2) are the ones that we make use of in the paper. In addition, a third statistic that we compute is the left-hand side of Equation (D.5). As discussed in Section D, this is the Wald statistic for testing the null hypothesis of $e'_m\boldsymbol{\varphi} = 0$ against the alternative $e'_m\boldsymbol{\varphi} \neq 0$, and this test indicates when the MSW analytic sets will be one bounded interval.

F Computation Details for Grid MBB AR Confidence Sets

The grid bootstrap AR confidence sets constructed with the MBB that we provide an algorithm for in Section 4.3 of the paper is grid-based and can be computationally intensive. Further, in Section G.1 below, we provide an algorithm for a similar AR confidence set

in which asymptotic simulation replaces the MBB. In this section, we briefly discuss some computational details.

Based on the algorithm in the paper, we can bootstrap a large number of $\sqrt{T}[(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^* - (se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}}]$ for each ξ_g in a grid. Then, for each grid point, we can sort the bootstrapped values $\sqrt{T}[(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^* - (se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}}]$ and compute the quantiles, $\widehat{q}_{q,\alpha/2}$ and $\widehat{q}_{q,1-\alpha/2}$, of $\sqrt{T}[(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^* - (se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}}]$. Finally, we check $-\widehat{q}_{q,1-\alpha/2} \leq \sqrt{T}(se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}} \leq -\widehat{q}_{q,\alpha/2}$. We now make two comments about this algorithm.

First, when computing $\sqrt{T}[(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^* - (se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}}]$, $(se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}}$ is the same in every bootstrap loop. Based on this, we can sort $\sqrt{T}(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$ and compute the quantiles, $\widetilde{q}_{q,\alpha/2}$ and $\widetilde{q}_{q,1-\alpha/2}$, of $\sqrt{T}(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$. Then, we can check $-\widetilde{q}_{q,1-\alpha/2} + \sqrt{T}(se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}} \leq \sqrt{T}(se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}} \leq -\widetilde{q}_{q,\alpha/2} + \sqrt{T}(se_j\widehat{\phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}}$. This simplifies to checking $-\widetilde{q}_{q,1-\alpha/2} \leq 0 \leq -\widetilde{q}_{q,\alpha/2}$ or, equivalently, $\widetilde{q}_{q,\alpha/2} \leq 0 \leq \widetilde{q}_{q,1-\alpha/2}$.

Second, we find that sorting $\sqrt{T}(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$ to compute $\widetilde{q}_{q,\alpha/2}$ and $\widetilde{q}_{q,1-\alpha/2}$ is computationally expensive. Because of this, we use a different approach to check the condition $\widetilde{q}_{q,\alpha/2} \leq 0 \leq \widetilde{q}_{q,1-\alpha/2}$. Let F_g^* denote the bootstrap analog of an empirical distribution function. That is, F_g^* is the distribution of the bootstrapped values of $\sqrt{T}(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$ so that $F_g^*(\widetilde{q}_{q,\alpha/2}) = \alpha/2$ and $F_g^*(\widetilde{q}_{q,1-\alpha/2}) = 1-\alpha/2$. Because cumulative distributions are non-decreasing, it is the case that $\widetilde{q}_{q,\alpha/2} \leq 0 \leq \widetilde{q}_{q,1-\alpha/2}$ implies $F_g^*(\widetilde{q}_{q,\alpha/2}) \leq F_g^*(0) \leq F_g^*(\widetilde{q}_{q,1-\alpha/2})$ or, equivalently, $\alpha/2 \leq F_g^*(0) \leq 1-\alpha/2$. Hence, we can check either $\widetilde{q}_{q,\alpha/2} \leq 0 \leq \widetilde{q}_{q,1-\alpha/2}$ or $\alpha/2 \leq F_g^*(0) \leq 1-\alpha/2$ when constructing our confidence intervals. We use the latter and compute $F_g^*(0)$ as the fraction of the bootstrapped values of $\sqrt{T}(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$ that are less than or equal to zero. We find that computing this fraction is computationally faster than sorting the bootstrapped values.

Based on these two comments, we compute the grid MBB AR confidence sets with the following algorithm.

1. Construct a grid $\boldsymbol{\xi} = \{\xi_1, \ldots \xi_G\}$, with ξ_g denoting one grid point, of null hypotheses $\Xi_{j1,i}(s; m, 1) = \xi_g$.

2. Follow steps 1 through 7 of the residual-based MBB algorithm from Section 4.1 of the paper to compute a large number of the bootstrapped estimates $\widehat{A}_1^*, \ldots, \widehat{A}_p^*$ and $\widehat{\boldsymbol{\varphi}}^* = \widehat{H^{(1)}\psi}^*$.

3. For each grid point, use the bootstrapped estimates to compute $\sqrt{T}(se_j'\widehat{\Phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$ for each bootstrap loop. Compute the fraction of the bootstrapped values $\sqrt{T}(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$ that are less than or equal to zero.

4. Construct the confidence interval for $\Xi_{j1,i}(s;m,1)$ by including any grid point, $\xi_g \in \boldsymbol{\xi}$, with the property that the fraction in the previous step is between $\alpha/2$ and $1-\alpha/2$.

For the simulation approach in Section G.1 below, we use an analogous algorithm.

G Additional Monte Carlo Results

In this section, we provide additional Monte Carlo simulation results. In Subsection G.1, we show the simulation results from the paper but include three additional confidence interval constructions: the Rademacher wild bootstrap, the MBB with Hall's percentile intervals, and a grid AR confidence set constructed with asymptotic simulation. In Subsection G.2, we show results for one standard deviation IRFs. In Subsection G.3, we show results for 95% confidence intervals. In Subsection G.4, we show results based on simulations with an effective sample size of $T = 2000$. In Subsection G.5, we show results for a data generating process (DGP) with less persistence in the VAR slope coefficients. In Subsection G.6, we show results for a DGP with more persistent stochastic volatility. In Subsection G.7, we show results for a DGP in which the proxy variable can be censored to zero. Finally, in section G.8, we show results when the VAR is estimated with lag augmentation in the spirit of Inoue and Kilian (2020).

For all of the DGPs in this section, except for in Subsection G.5, the IRFs and FEVD are the same and shown in Figure G.1. These are also the IRFs and FEVD in the simulations in the paper. The IRFs have hump shapes that are common in macroeconomic studies. We show the IRFs and FEVDs for the less persistent DGP in Subsection G.5.

G.1 Additional Confidence Interval Constructions

We begin by using the data generating process (DGP) in the paper and including results for three additional confidence interval constructions.

The first additional confidence interval is constructed using the residual-based wild bootstrap, used in Mertens and Ravn (2013), Gertler and Karadi (2015) and other proxy SVAR papers. The algorithm is described above in Section C.1. Following Mertens and Ravn (2013) and Gertler and Karadi (2015), we draw η_t from the Rademacher distribution: $\eta_t = -1$ with probability 0.5 and $\eta_t = 1$ with probability 0.5.

The second additional confidence interval uses the MBB but constructs Hall's percentile interval (Hall (1992) and Lütkepohl (2005, Appendix D)) instead of the standard percentile

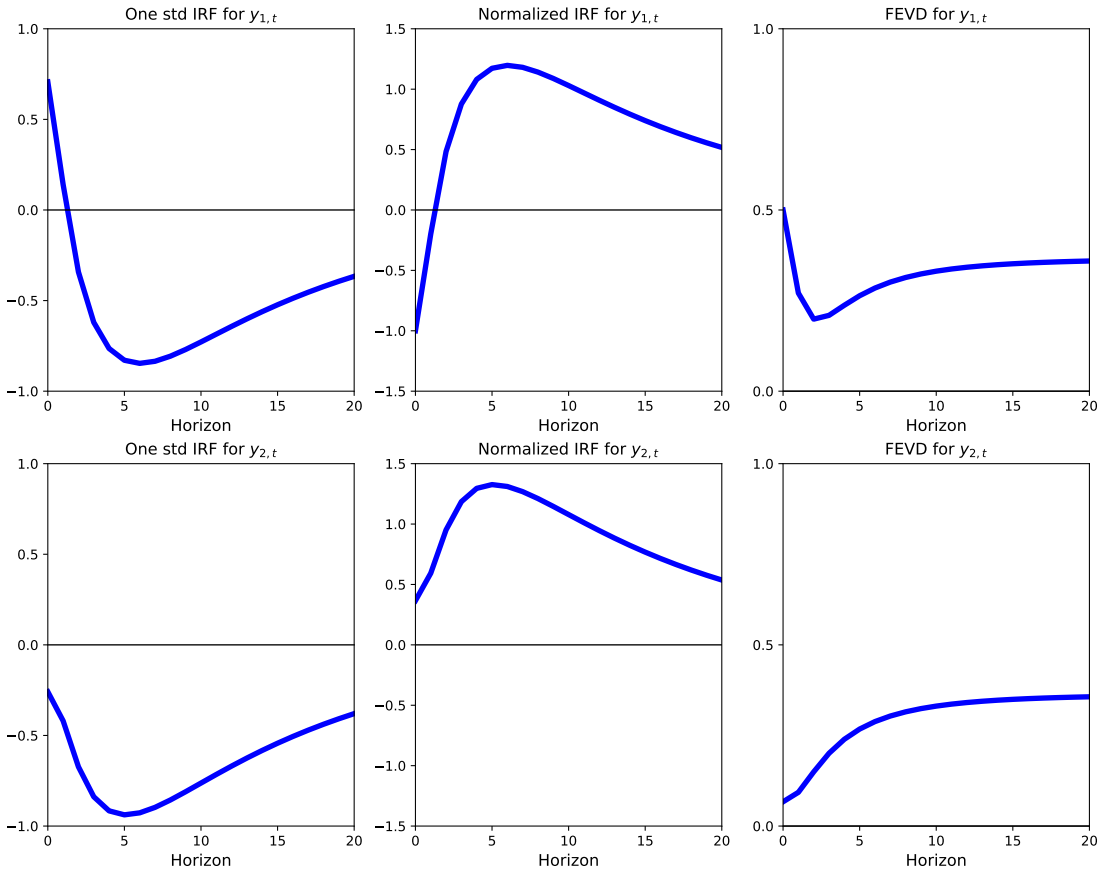

Figure G.1: The one standard deviation IRF, normalized IRF and FEVD from the structural VAR used in the paper and in every simulation in this supplemental appendix, except for in Subsection G.5.

interval. Intuitively, this interval flips the standard percentile interval around the respective IRF or FEVD in order to adjust for bias. Asymptotically, Hall's percentile interval is equivalent to the standard percentile interval because the limiting distribution of the IRFs and FEVDs are normal and, hence, symmetric.

The third confidence interval is constructed similarly to the grid MBB AR confidence sets; however, the MBB is replaced with an asymptotic simulation design. The algorithm is initialized by estimating $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\varphi}}$ and their corresponding asymptotic variance. We use

$$\begin{pmatrix} \widehat{V}^{(1,1)} & \widehat{V}^{(3,1)\prime} \\ \widehat{V}^{(3,1)} & \widehat{V}^{(3,3)} \end{pmatrix}$$

as estimated in Section D to be the asymptotic variance. That is, the estimated asymptotic variance used for simulation is the same as the estimated asymptotic variance used for the

MSW analytical sets. The algorithm proceeds as follows

1. Construct a grid $\boldsymbol{\xi} = \{\xi_1, \dots \xi_G\}$, with ξ_g denoting one grid point, of null hypotheses $\Xi_{j1,i}(s; m, 1) = \xi_g$.

2. Simulate a large number of $\boldsymbol{\beta}^{sim}$, yielding $A_1^{sim}, \dots, A_p^{sim}$, and $\boldsymbol{\varphi}^{sim}$ from

$$\mathcal{N}\left(\begin{pmatrix}\widehat{\boldsymbol{\beta}}\\ \widehat{\boldsymbol{\varphi}}\end{pmatrix}, \begin{pmatrix}\widehat{V}^{(1,1)} & \widehat{V}^{(3,1)\prime}\\ \widehat{V}^{(3,1)} & \widehat{V}^{(3,3)}\end{pmatrix}\right)$$

3. For each grid point, use the simulated values $A_1^{sim}, \dots, A_p^{sim}$ and $\boldsymbol{\varphi}^{sim}$ to compute $\sqrt{T}[(se_j'\Phi_i^{sim} - \xi_g e_m')\boldsymbol{\varphi}^{sim} - (se_j'\widehat{\Phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}}]$ for each bootstrap loop. Define $q_{g,\alpha/2}^{sim}$ and $q_{g,1-\alpha/2}^{sim}$ to be the $\alpha/2$ and $1-\alpha/2$ quantiles of this simulated statistic for grid point ξ_g.

4. Construct the confidence interval for $\Xi_{j1,i}(s; m, 1)$ by including any grid point, $\xi_g \in \boldsymbol{\xi}$, with the property $-q_{g,1-\alpha/2}^{sim} \leq \sqrt{T}(se_j'\widehat{\Phi}_i - \xi_g e_m')\widehat{\boldsymbol{\varphi}} \leq -q_{g,\alpha/2}^{sim}$.

This algorithm is conceptually the same as for the grid MBB AR confidence sets used in the paper, but step 2 now uses asymptotic simulation instead of the MBB to draw values of $A_1, \dots, A_p$, and $\boldsymbol{\varphi}$. Further, we use the algorithmic approach in Section F to speed up computation. We call these confidence sets "grid simulation AR" confidence sets.

Before discussing the results, we note that both the Rademacher wild bootstrap and the MBB with Hall's percentile intervals can be used for both IRFs and FEVDs. However, the grid simulation AR confidence sets can only be used for normalized IRFs.

Coverage rates for normalized IRFs: Figure G.2 shows the coverage rates of 68% MBB percentile intervals, grid MBB AR confidence sets, and MSW analytical sets for normalized IRFs. These confidence intervals and sets are as described in the paper. Figure G.2 also shows the coverage rates of 68% Rademacher wild bootstrap percentile intervals, MBB Hall's percentile intervals, and grid simulation AR confidence sets for normalized IRFs.

Throughout this section and the next several sections, the legend of the figures are as follows. "MBB" is the standard percentile interval. "grid MBB AR" is the grid MBB Anderson and Rubin (1949) confidence set. "MSW" is the Montiel Olea, Stock, and Watson (forthcoming) analytical set. "WB-Rademacher" is the percentile interval from the Rademacher wild bootstrap. "MBB Halls" is Hall's percentile interval with the MBB. "grid Sim AR" is the grid simulation AR confidence set.

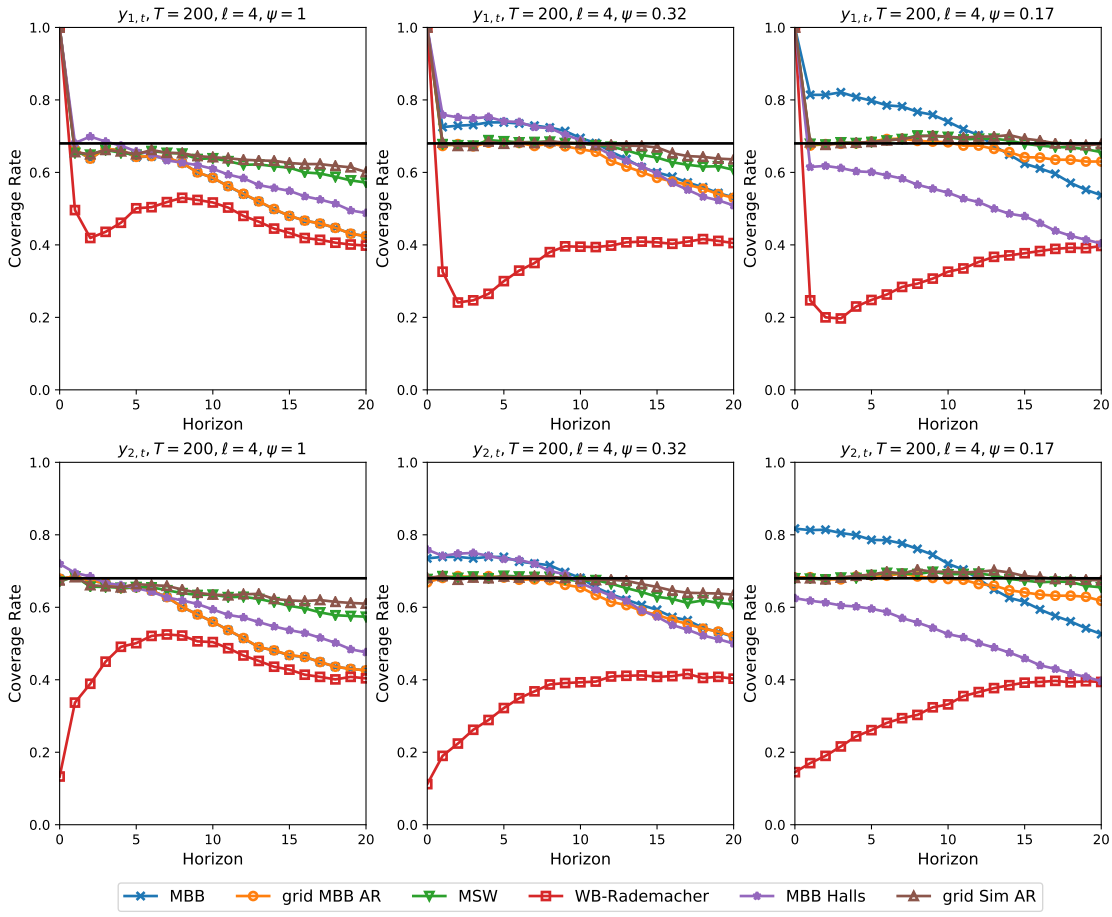

Figure G.2: Coverage rates of 68% confidence intervals for normalized IRFs. The solid horizontal line shows the 0.68 target level.

The top and bottom panels of G.2 show the coverage rates for $y_{1,t}$ and $y_{2,t}$, respectively. The columns show coverage rates for different values of ψ. We normalize $y_{1,t}$ to fall by 1 on impact and do this within every bootstrap loop. Hence, the coverage rates for $y_{1,t}$ are always 1 at horizon 0.

The Rademacher wild bootstrap produces coverage rates that are too low for both $y_{1,t}$ and $y_{2,t}$ at every horizon and for every value of ψ. As discussed in Jentsch and Lunsford (2016, 2019a,b) and in Section C.2, the Rademacher wild bootstrap asymptotically ignores the uncertainty surrounding the estimates of Σ_u and $H^{(1)}\Psi'$. Hence, it produces confidence intervals that are too small and coverage rates that are too low.

For $\psi = 1$, the coverage rates of the MBB with Hall's percentile intervals behave similarly to the coverage rates of the MBB with the standard percentile intervals. The coverage rates from Hall's intervals are close to the target level at short horizons but become too low at

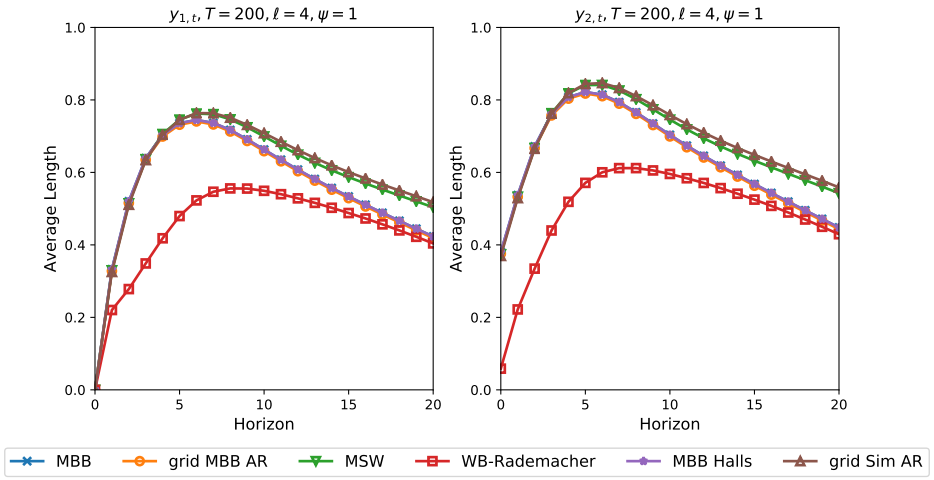

Figure G.3: Average lengths of 68% confidence intervals for normalized IRFs.

longer horizons. For $\psi = 0.32$, the coverage rates from Hall's intervals are very close to the standard percentile intervals: above the target level at short horizons and falling to below the target level at longer horizons. Unlike with the standard percentile intervals, the coverage rates from Hall's intervals are below the target level for every horizon with $\psi = 0.17$.

Finally, the coverage rates of the simulation AR grid confidence sets are very close to the coverage rates of the MSW analytical confidence sets for both $y_{1,t}$ and $y_{2,t}$ at every horizon and for every value of ψ.

Average lengths for normalized IRFs: Figure G.3 shows the average lengths of 68% MBB percentile intervals, grid MBB AR confidence sets, and MSW analytical sets for normalized IRFs. These lengths are as in the paper. Figure G.3 also shows the average lengths of 68% Rademacher wild bootstrap percentile intervals, MBB Hall's percentile intervals, and grid simulation AR confidence sets for normalized IRFs. The left and right panels show average lengths for $y_{1,t}$ and $y_{2,t}$, respectively. Because we normalize $y_{1,t}$ to fall by 1 on impact, the average length for $y_{1,t}$ is always 0 at horizon 0.

Consistent with its low coverage rates in Figure G.2, the Rademacher wild bootstrap has average lengths that are smaller than every other confidence interval or set.

As noted above, Hall's percentile intervals essentially flip the standard percentile intervals around the respective IRF or FEVD. Because of this, the MBB with Hall's percentile interval has the same lengths as the MBB with the standard percentile interval by construction.

Finally, the average lengths of the grid simulation AR confidence set are very similar to those of the MSW analytical sets.

Coverage rates for FEVDs: Figure G.4 shows the coverage rates of the 68% MBB

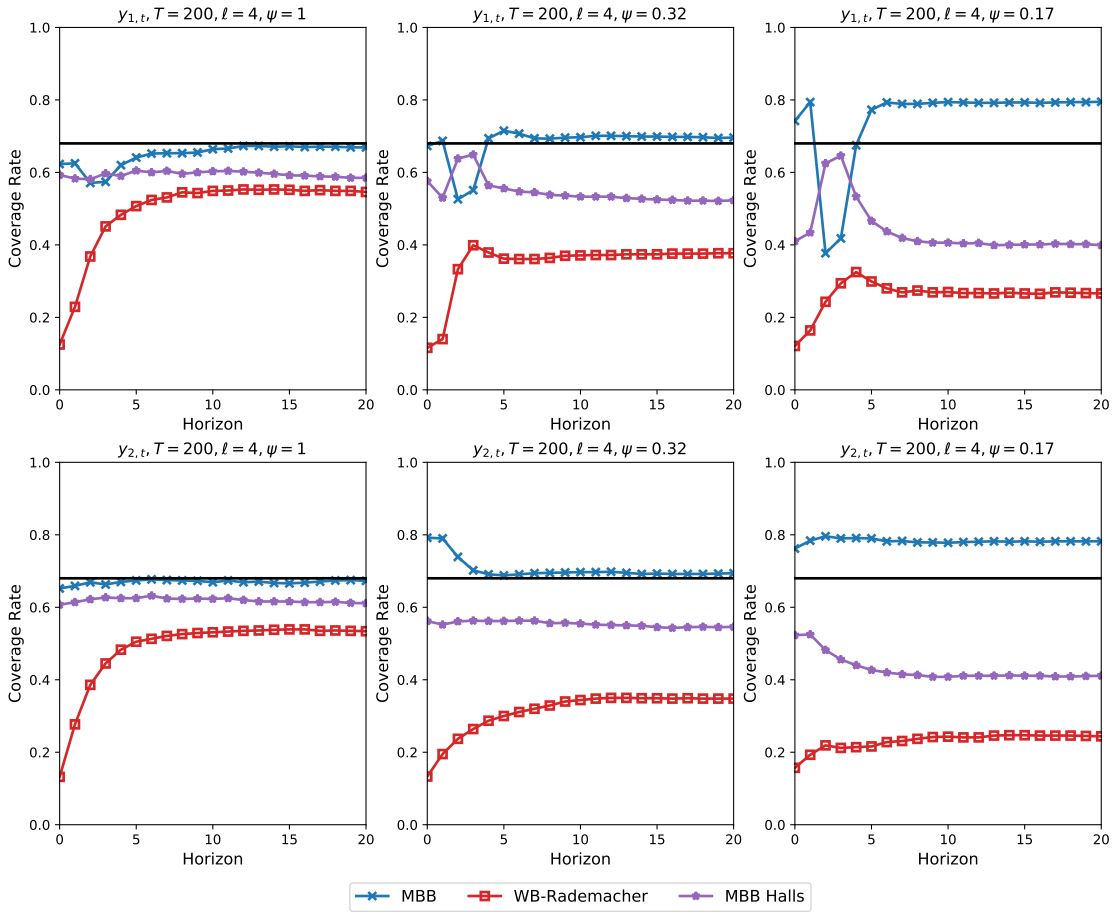

Figure G.4: Coverage rates of 68% confidence intervals for FEVDs. The solid horizontal line shows the 0.68 target level.

percentile intervals for FEVDs as in the paper. Figure G.4 also shows the coverage rates of the 68% Rademacher wild bootstrap percentile intervals and the MBB with Hall's percentile intervals for FEVDs. The top and bottom panels show the coverage rates for $y_{1,t}$ and $y_{2,t}$, respectively. The columns show coverage rates for different values of ψ.

As with the normalized IRFs, Figure G.4 shows that the Rademacher wild bootstrap produces coverage rates for FEVDs that are too low for both $y_{1,t}$ and $y_{2,t}$ at every horizon and for every value of ψ.

For $\psi = 1$, the MBB with Hall's percentile interval has coverage rates that are generally lower and farther from the target level than the MBB with the standard percentile interval. This result contrasts with the coverage rates for normalized IRFs for $\psi = 1$. In the left panels of Figure G.2, Hall's percentile intervals have coverage rates that are generally closer to target than the standard percentile intervals. Comparing the left panels of Figures G.2 and

G.4, we see that Hall's percentile intervals do not strictly dominate the standard percentile intervals in terms of coverage rates nor vice versa. Instead, the ranking of the interval types may change depending on which statistic is of interest.

For $\psi = 0.32$ and $\psi = 0.17$, Figure G.4 shows that the Hall's percentile intervals have coverage rates for FEVDs that are too low at every horizon. In general, the coverage rates for Hall's percentile intervals are also farther from the target level than the standard percentile intervals. The exception is for $y_{1,t}$ for horizons 2 and 3. As shown in Figure G.1, there is a clear trough in the FEVD for $y_{1,t}$ at horizons 2 and 3. Overall, Figure G.4 show that Hall's percentile intervals can have coverage rates that are closer to target than standard percentile intervals, but only for some variables and some horizons.

G.2 One Standard Deviation IRFs

In this section, we show coverage rates of confidence intervals for one standard deviation IRFs. We do not show these in the body of the paper. However, they remain relevant as researchers, e.g. Gertler and Karadi (2015), occasionally show one standard deviation IRFs as their main results.

Figure G.5 shows the coverage rates of 68% MBB percentile intervals, Rademacher wild bootstrap percentile intervals, and MBB Hall's percentile intervals for one standard deviation IRFs. The top and bottom panels show the coverage rates for $y_{1,t}$ and $y_{2,t}$, respectively. The columns show coverage rates for different values of ψ.

For both variables and all proxy strengths, the MBB percentile intervals have coverage rates that are near the target level at short horizons but fall below the target level at longer horizons. Overall, the coverage rates of the MBB percentile intervals do not appear to change much with different values of ψ.

As with the normalized IRFs and FEVDs, Figure G.5 shows that the Rademacher wild bootstrap produces coverage rates for one standard deviation IRFs that are too low for both $y_{1,t}$ and $y_{2,t}$ at every horizon and for every value of ψ. For $\psi = 1$, the Rademacher wild bootstrap's coverage rates are similar to those from the MBB percentile interval at longer horizons. However, with weaker proxy variables, the Rademacher wild bootstrap's coverage rates stay more persistently below those of the MBB percentile interval.

For $\psi = 1$, the MBB with Hall's percentile intervals has coverage rates that are generally below target. At short horizons, the coverage rates from Hall's percentile intervals are slightly below those of the standard percentile interval. However, at longer horizons, the coverage rates from Hall's percentile intervals are generally closer to target.

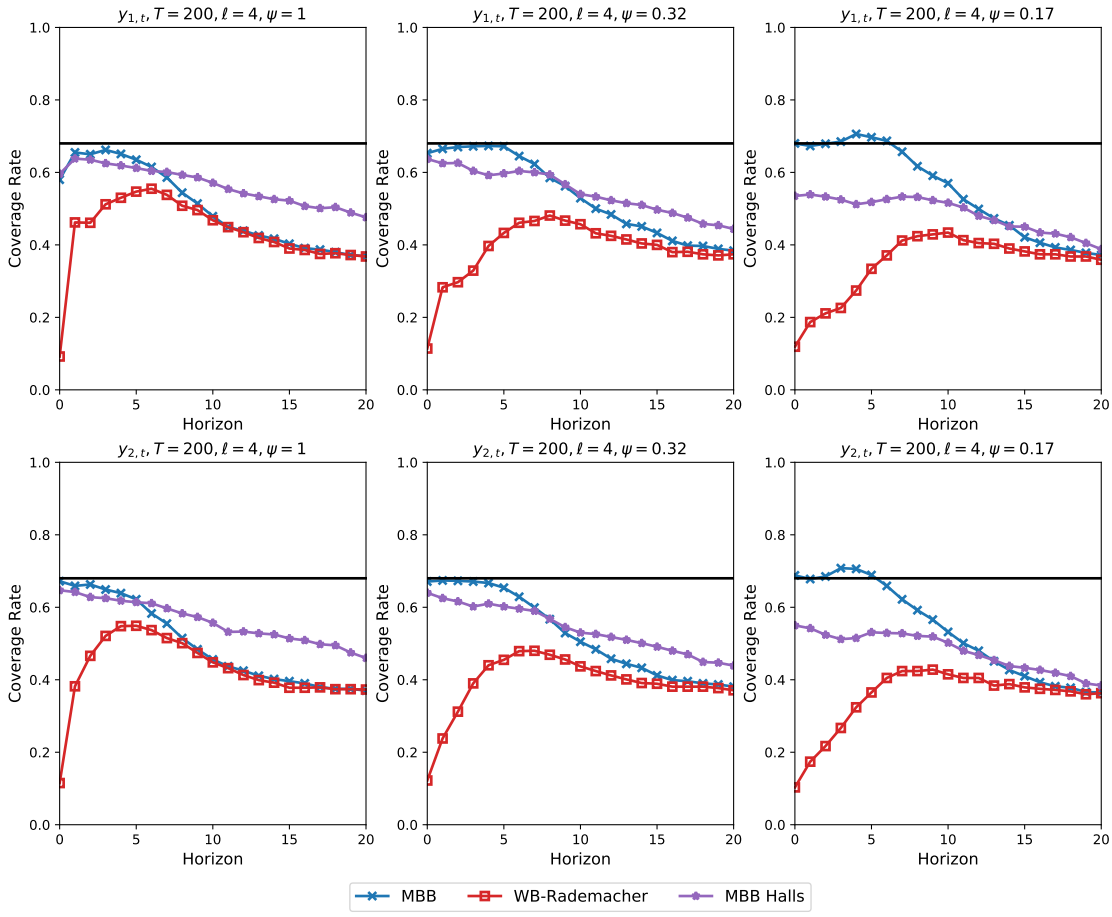

Figure G.5: Coverage rates of 68% confidence intervals for one standard deviation IRFs. The solid horizontal line shows the 0.68 target level.

As the proxy variable gets weaker, the coverage rates from Hall's percentile intervals become lower at shorter horizons. For $\psi = 0.17$, Hall's percentile intervals has coverage rates that are generally lower than those from the standard percentile intervals.

Overall, Figure G.5 shows us that which intervals, standard percentile or Hall's percentile, have coverage rates that are closer to target can vary with impulse horizon and proxy strength. Taken together with the coverage rates in Figures G.2 and G.4, our general finding is that standard percentile intervals do not dominate Hall's percentile intervals in terms of coverage rates nor vice versa. Rather, which interval have coverage rates that are closer to target can depend on the statistic that is begin computed, the variable in the VAR, the horizon, and the proxy strength.

G.3 95% Confidence Intervals

In this section, we show the results from the paper but with 95% confidence intervals instead of 68% confidence intervals. The overall pattern of results does not change much when switching from 68% confidence intervals to 95% confidence intervals. Because of this, we simply provide descriptions of this figures without further discussion. Figure G.6 shows the coverage rates of 95% confidence intervals for normalized IRFs. Figure G.7 shows the average length of 95% confidence intervals for normalized IRFs. Figure G.8 shows the coverage rates of 95% confidence intervals for FEVDs. Figure G.9 shows the coverage rates of 95% confidence intervals for one standard deviation IRFs.

G.4 Results with $T = 2000$

In this section, we show results based on simulations with an effective sample size of $T = 2000$ in order to highlight certain asymptotic results. Because we increase the sample size, we increase the block length to $\ell = 8$[2]. As in the paper, we only show results for 68% confidence intervals and sets in this section. As in the paper, we show results for MBB percentile intervals, the grid MBB AR confidence sets, and MSW analytic confidence sets. Further, following the previous sections of this appendix, we show results for Rademacher wild bootstraps, the MBB with Hall's percentile intervals, and grid simulation AR confidence sets.

Because we increase the sample size, we also adjust the values of ψ used in the DGP. We continue to use $\psi = 1$ for a strong proxy setting. Pre-tests reject the presence of a weak proxy in every simulation with $\psi = 1$. For the other DGPs, we use the weak proxy equation $\psi_T = C/\sqrt{T}$. In the paper with $T = 200$, we use $\psi = 0.32$ and $\psi = 0.17$. Now with a sample size of $T = 2000$, we use $\psi = 0.32 \times \sqrt{200}/\sqrt{2000} \approx 0.101$ as our middle proxy setting and $\psi = 0.17 \times \sqrt{200}/\sqrt{2000} \approx 0.054$ as our weak proxy setting. With $\psi = 0.101$ and $T = 2000$, pre-tests reject the presence of a weak proxy in roughly half of the simulations, similar to the rejection rates with $\psi = 0.32$ and $T = 200$ in the paper. With $\psi = 0.054$ and $T = 2000$, pre-tests reject the presence of a weak proxy in less than 10% of the simulations, similar to the rejection rates with $\psi = 0.17$ and $T = 200$ in the paper.

Coverage Rates for normalized IRFs: Figure G.10 shows the coverage rates for normalized IRFs. For $\psi = 1$, the coverage rates of all of the confidence intervals and sets are all close to the target level and virtually indistinguishable at shorter horizons. The one

[2]We find that larger block sizes, such as $\ell = 12$ or $\ell = 16$, has only small effects on the results.

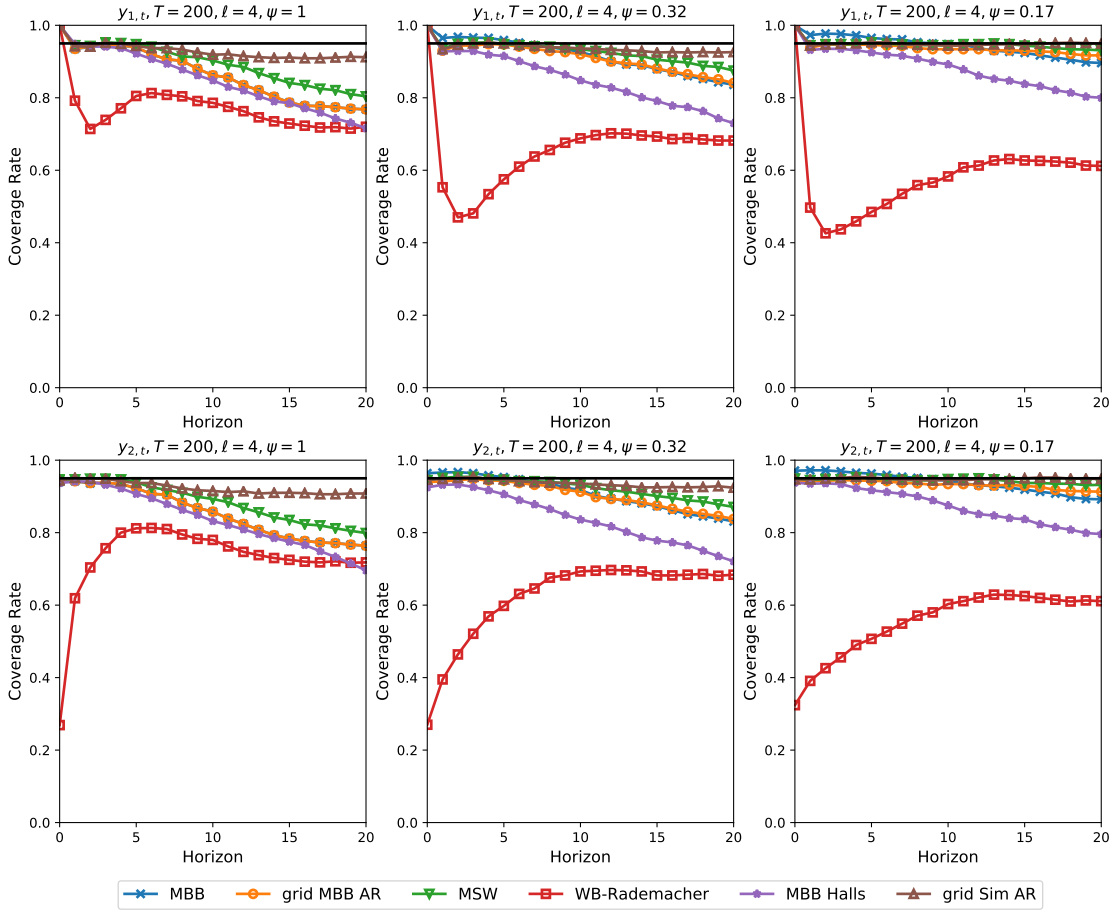

Figure G.6: Coverage rates of 95% confidence intervals for normalized IRFs. The solid horizontal line shows the 0.95 target level.

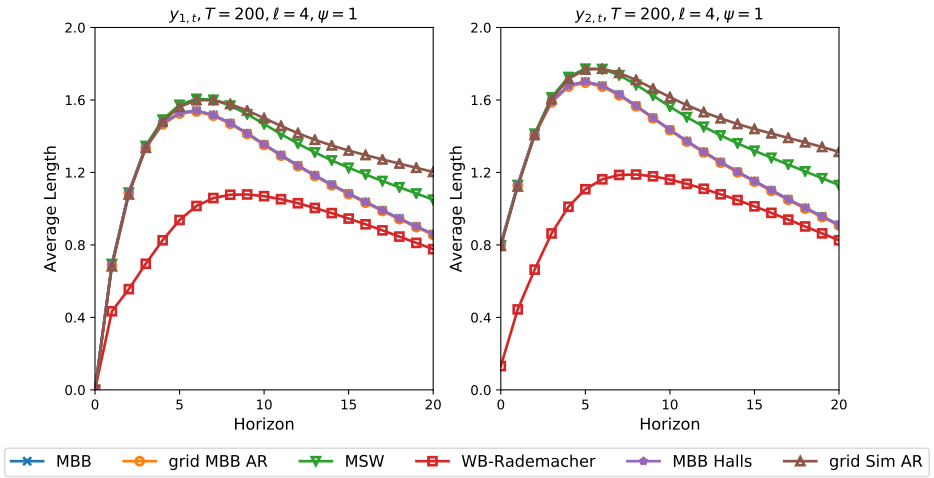

Figure G.7: Average lengths of 95% confidence intervals for normalized IRFs.

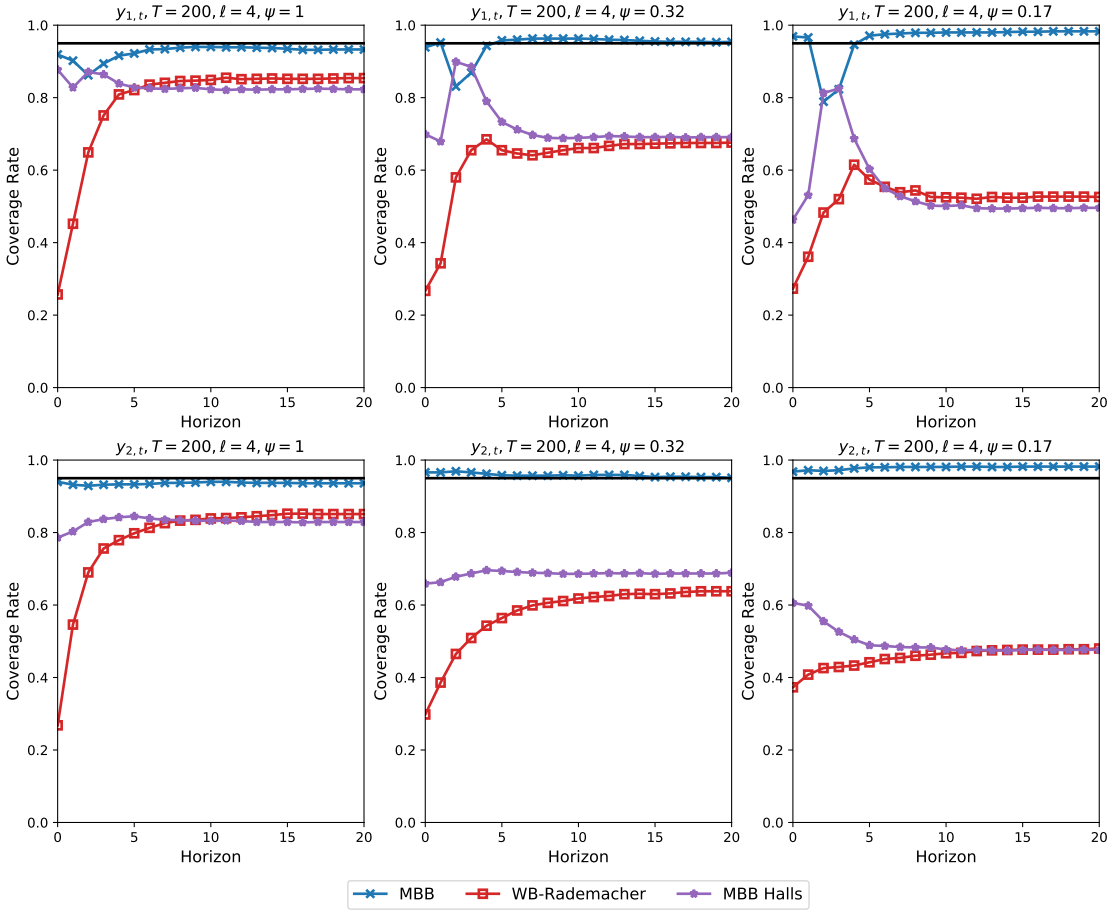

Figure G.8: Coverage rates of 95% confidence intervals for FEVDs. The solid horizontal line shows the 0.95 target level.

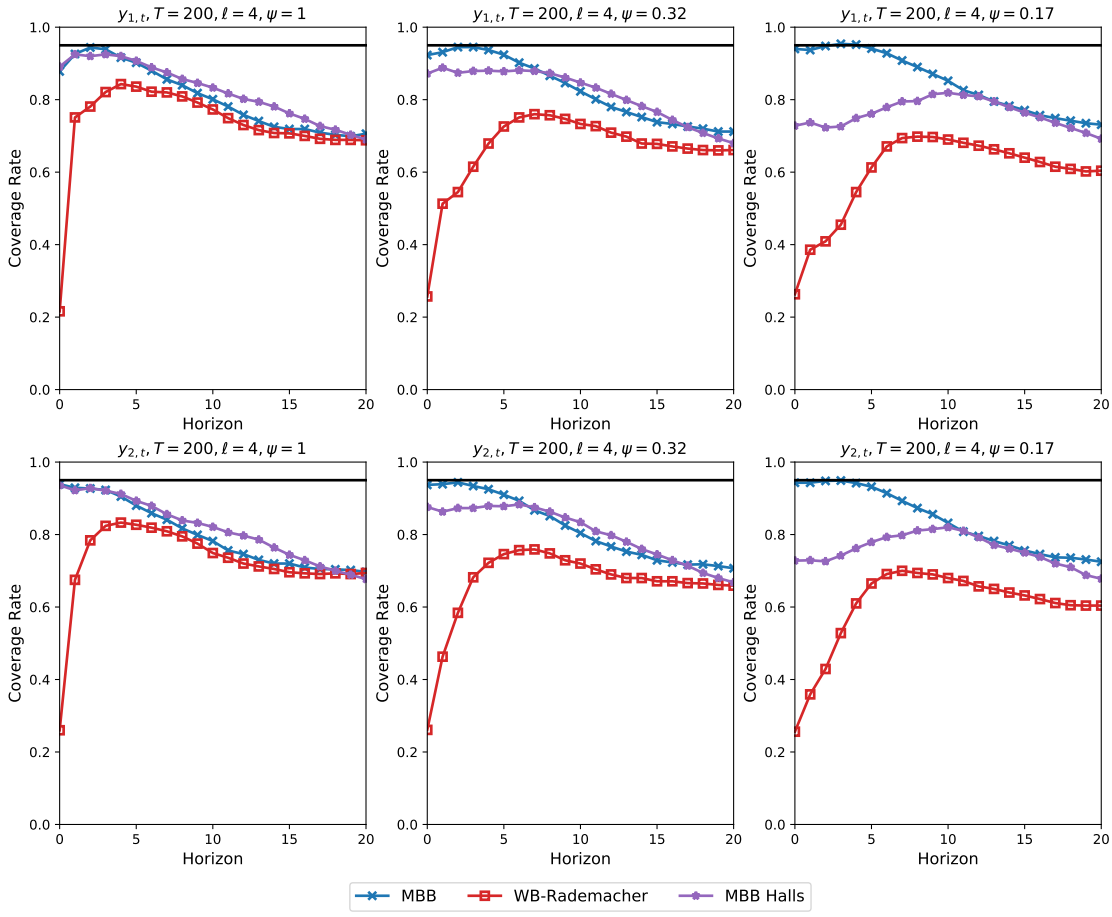

Figure G.9: Coverage rates of 95% confidence intervals for one standard deviation IRFs. The solid horizontal line shows the 0.95 target level.

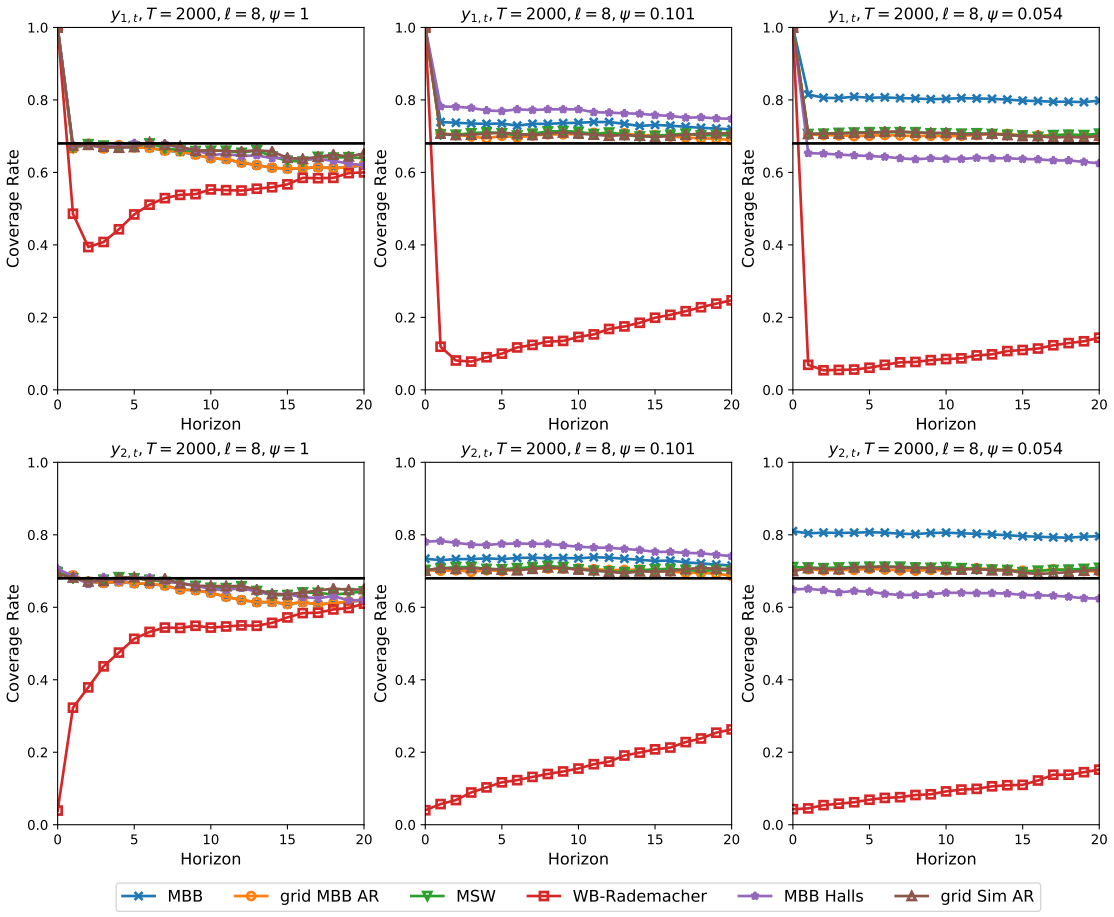

Figure G.10: Coverage rates of 68% confidence intervals for normalized IRFs. The solid horizontal line shows the 0.68 target level.

exception is the Rademacher wild bootstrap, which continues to have coverage rates that are below target. This highlights that the invalidity of the wild bootstrap is an asymptotic result and not just a feature of small sample sizes. With $\psi = 1$, the coverage rates of MBB percentile intervals and grid MBB AR confidence sets fall below the target level. However, these coverage rates are much closer to the target level than with $T = 200$ in the paper, and they are similar to the coverage rates from the other confidence intervals and sets.

As the proxy variable gets weaker, the coverage rates of MBB percentile intervals and Hall's percentile intervals get farther away from the target level. This highlights that both standard percentile and Hall's percentile intervals can have coverage rates that are away from the target with a weak proxy variable in with very large sample sizes. In contrast, coverage rates of both types of grid AR confidence sets and coverage rates of the MSW analytic confidence sets are near the target level with both $\psi = 0.101$ and $\psi = 0.054$.

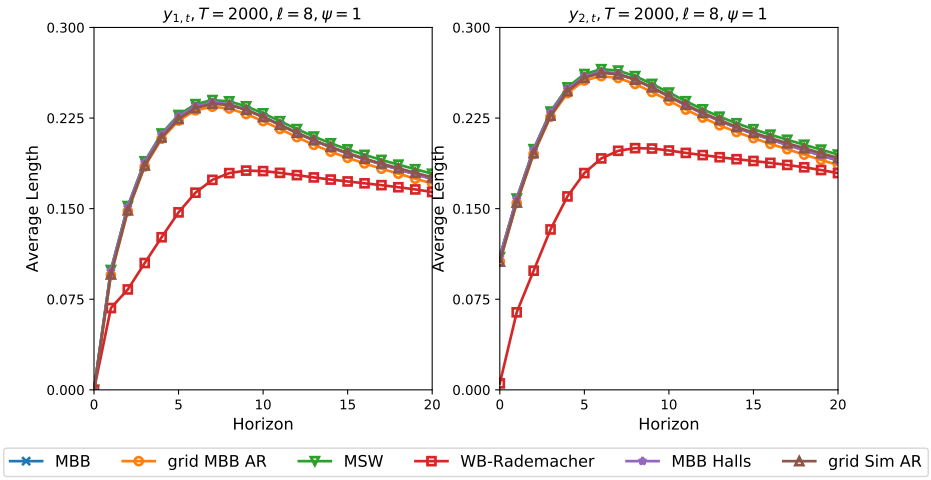

Figure G.11: Average length of 68% confidence intervals for normalized IRFs.

Finally, as the proxy variable gets weaker, the Rademacher wild bootstrap produces increasingly small coverage rates. With $\psi = 0.054$, the Rademacher wild bootstrap's coverage rate never exceeds 0.2.

Average lengths for normalized IRFs: Figure G.11 shows the average length of the confidence intervals and sets for normalized IRFs with $\psi = 1$. Except for the Rademacher wild bootstrap, which always has the shortest average length, the average length of all confidence intervals and sets are essentially the same. This is consistent with the very similar coverage rates shown in the left panels in Figure G.10.

Coverage rates for FEVDs: Figure G.12 shows the coverage rates for FEVDs. For $\psi = 1$, the coverage rates of the standard percentile intervals and Hall's percentile intervals from the MBB are close to the target level and essentially the same. For $\psi = 0.101$, the coverage rates of the standard percentile intervals and Hall's percentile intervals from the MBB move away from the target level. In general, the coverage rates of the standard percentile interval become too big and the coverage rates of Hall's percentile interval become too small. The exception is for $y_{1,t}$ and horizons 2 and 3. Again, we note that $y_{1,t}$ has a clear trough in Figure G.1 at horizons 2 and 3. For $\psi = 0.054$, the coverage rates of the standard percentile interval continue to become too big and the coverage rates of Hall's percentile interval continue to become too small. Again, the exception for $y_{1,t}$ and horizons 2 and 3. Overall, the results in Figure G.12 are similar to the patterns in Figure G.4 for $T = 200$.

Finally, as in Figure G.10, coverage rates from the Rademacher wild bootstrap are too low and become increasingly small as the proxy variable gets weaker.

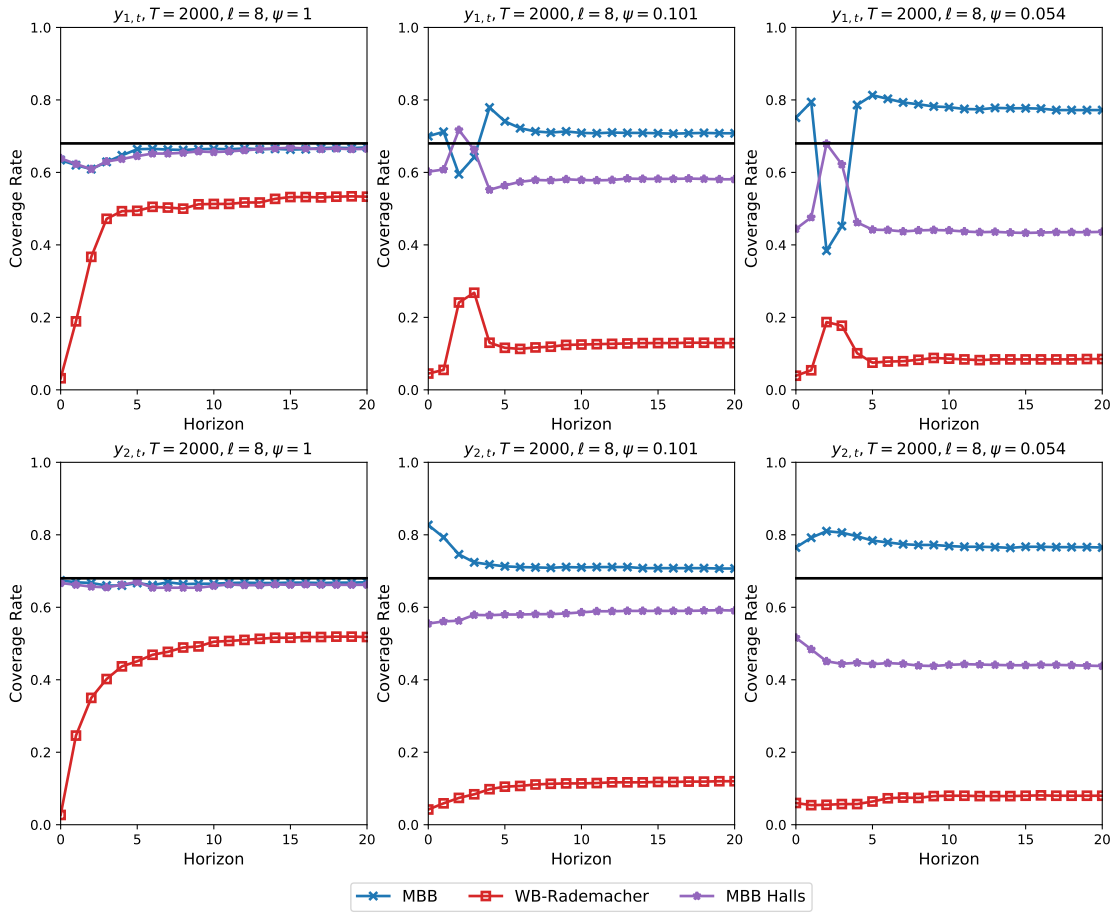

Figure G.12: Coverage rates of 68% confidence intervals for FEVDs. The solid horizontal line shows the 0.68 target level.

G.5 Less Persistent VAR Data

We write the companion matrix for the VAR slope coefficients as

$$\begin{bmatrix} A_1 & A_2 \\ I_2 & 0 \end{bmatrix}.$$

For the DGP used in the paper, the eignevalues of this companion matrix are 0.930, 0.634, $0.098 + 0.134i$, and $0.098 - 0.134i$, with $i = \sqrt{-1}$. The first eigenvalue, 0.930, is close to zero, yielding a high degree of persistence that is common in VARs with macroeconomic data. However, having an eigenvalue this close to 1 may cause small-sample biases in the estimates of A_1 and A_2, causing the coverage rates of confidence intervals to be away from their target levels. To study this further, we consider an alternative DGP with different values for A_1

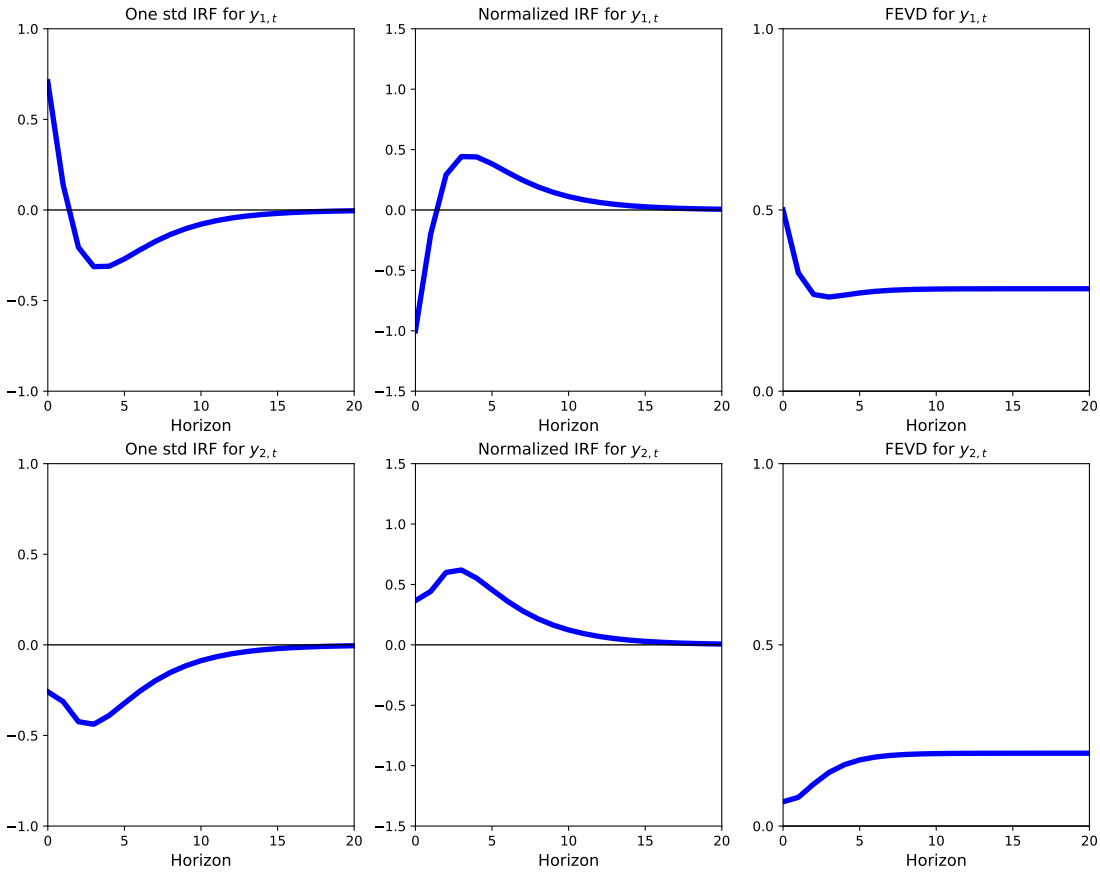

Figure G.13: The one standard deviation IRF, normalized IRF and FEVD based on the VAR slope coefficients in (G.1) and used only in Subsection G.5.

and A_2 than what is used in the body of the paper. We use

$$A_1 = \begin{bmatrix} 0.382 & 0.504 \\ -0.053 & 1.065 \end{bmatrix} \quad \text{and} \quad A_2 = \begin{bmatrix} -0.144 & 0 \\ -0.144 & -0.072 \end{bmatrix}. \tag{G.1}$$

The eigenvalues of the corresponding companion matrix are 0.746, 0.505, $0.098 + 0.134i$, and $0.098 - 0.134i$. The first two eigenvalue are about 0.8 times the first two eigenvalues for the DGP used in the paper. The complex eigenvalues are unchanged.

Figure G.13 shows the IRFs and FEVDs implied by this new DGP. The IRFs and FEVDs continue to be hump-shaped as in Figure G.1. However, the IRFs in Figure G.13, revert to zero more quickly than in Figure G.1.

Coverage rates for normalized IRFs: Figure G.14 shows the coverage rates of 68% confidence intervals and sets for the normalized IRFs. This figure parallels Figure G.2, but

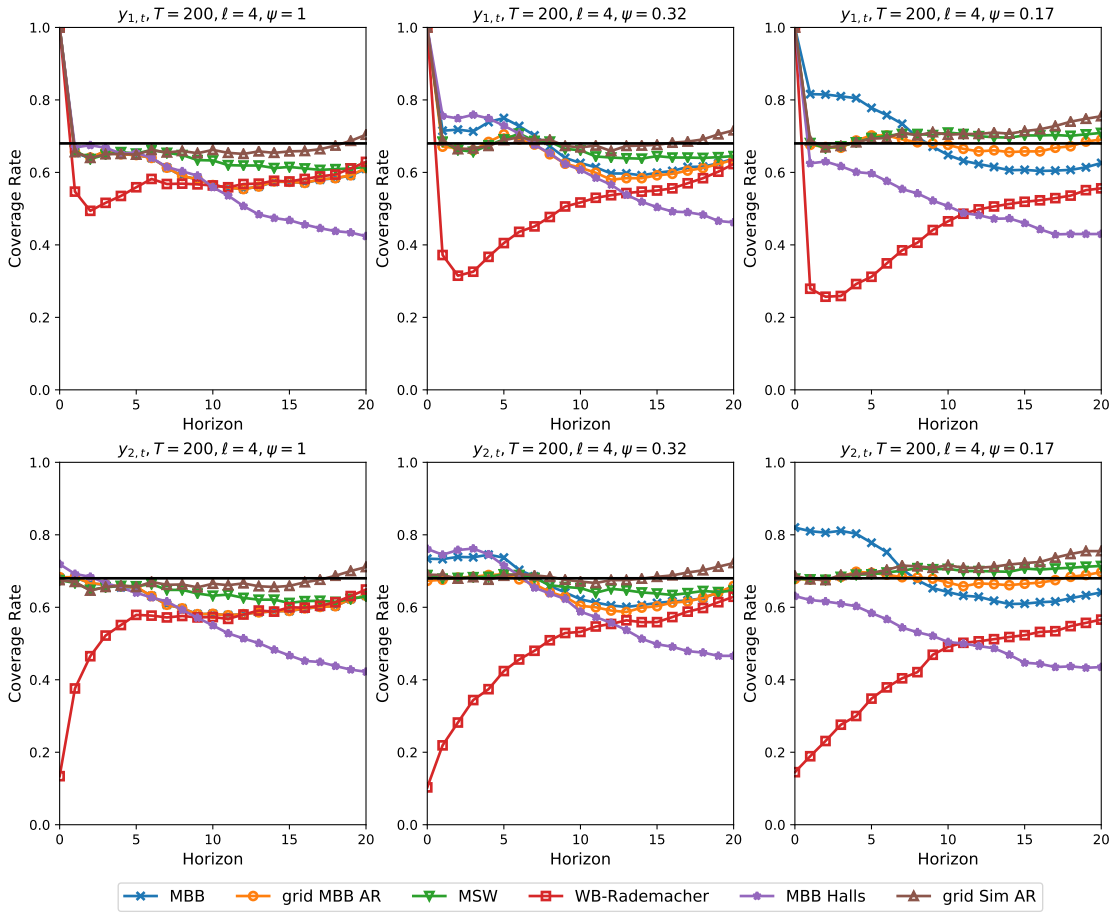

Figure G.14: Coverage rates of 68% confidence intervals and sets for normalized IRFs. The solid horizontal line shows the 0.68 target level.

with the less persistent DGP in (G.1). The top and bottom panels show the coverage rates for $y_{1,t}$ and $y_{2,t}$, respectively. The columns show coverage rates for different values of ψ. We normalize $y_{1,t}$ to fall by 1 on impact and do this within every bootstrap loop. Hence, the coverage rates for $y_{1,t}$ are always 1 at horizon 0.

For $\psi = 1$, the coverage rates of the MBB percentile intervals, the grid MBB AR confidence sets, the MSW confidence sets, and the grid simulation AR confidence sets are slightly below target at short horizons. At longer horizons, these confidence intervals and sets remain slightly below the target level. However, compared to the coverage rates from the more persistent DGP in Figure G.2, the coverage rates of the MBB percentile intervals, the grid MBB AR confidence sets, the MSW confidence sets, and the grid simulation AR confidence sets in Figure G.14 are much closer to target at longer horizons. Further, the coverage rates of the MBB percentile intervals, the grid MBB AR confidence sets and the MSW confidence sets

are essentially the same at horizons 19 and 20. These results suggest that the low coverage rates at long horizons in the body of the paper and in G.2 are due to small-sample biases caused by a persistent DGP. See Kilian (1998) for further discussion.

Another results worth noting for $\psi = 1$ is that the coverage rates of Hall's percentile intervals are below those of standard percentile intervals at longer horizons in Figure G.14. This is the opposite of what we found for the more persistent DGP in Figure G.2. Again, this highlights that neither Hall's nor standard percetile intervals appear to dominate one another in terms of coverage rates. Rather, which interval is better appears to vary.

For weaker proxy variables, MBB percentile intervals become oversized at short horizons as in the body of the paper and in G.2. In contrast, the coverage rates of the grid MBB AR confidence sets, the MSW confidence sets and the grid simulation AR confidence sets are close to target at short horizons. At longer horizons, the coverage rates of the grid MBB AR confidence sets, the MSW confidence set and the grid simulation AR confidence sets stay closer to target than is the case in the body of the paper and in Figure G.2. With $\psi = 0.17$, the MSW confidence sets and the grid simulation AR confidence sets can be oversized at long horizons while the grid MBB AR confidence sets stay close to the target size.

The Rademacher wild bootstrap has coverage rates that are too low for both variables, all proxy strengths, and every horizon. However, with this less persistent DGP, the coverage rates of the Rademacher wild bootstrap are closer to target at longer horizons compared to Figure G.2.

Average lengths for normalized IRFs: Figure G.15 shows the average lengths of 68% confidence intervals and sets for the normalized IRFs. This figure parallels Figure G.3, but with the less persistent DGP in (G.1). The left and right panels show average lengths for $y_{1,t}$ and $y_{2,t}$, respectively. Because we normalize $y_{1,t}$ to fall by 1 on impact, the average length for $y_{1,t}$ is always 0 at horizon 0.

Overall, the average lengths of all of the confidence intervals and sets are smaller in Figure G.15 with the less persistent DGP compared to Figure G.3. It is also the case that the average lengths of the MBB percentile intervals and the grid MBB AR confidence sets stay closer to the average lengths of the MSW confidence sets and the grid simulation AR confidence sets at longer horizons in Figure G.15 compared to Figure G.3. Finally, the Rademacher wild bootstrap's average lengths are generally the smallest but become similar to the average lengths of the other confidence intervals and sets at horizons 10 and longer in Figure G.15.

Coverage rates for FEVDs: Figure G.16 shows coverage rates of 68% confidence intervals for the FEVDs. This figure parallels Figure G.4, but with the less persistent DGP

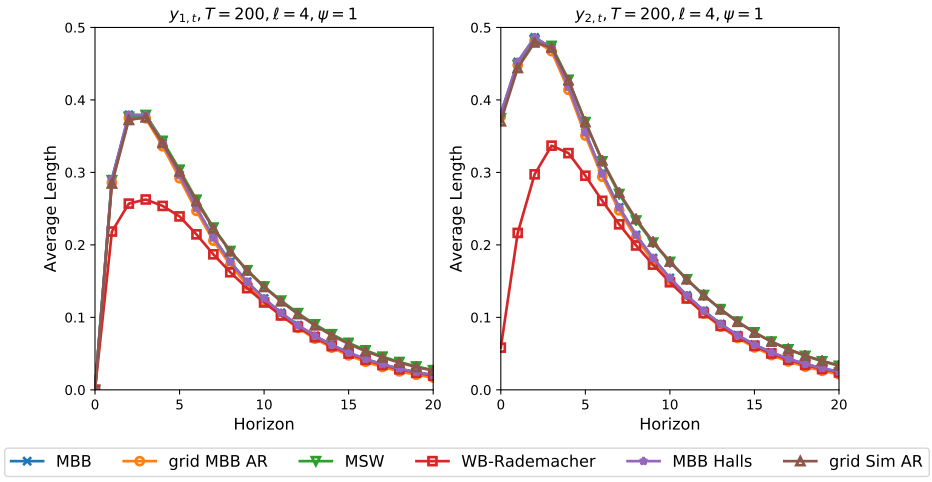

Figure G.15: Average lengths of 68% confidence intervals for normalized IRFs.

in (G.1). The top and bottom panels show the coverage rates for $y_{1,t}$ and $y_{2,t}$, respectively. The columns show coverage rates for different values of ψ.

For $y_{1,t}$, the MBB percentile interval's coverage rates are generally lower in Figure G.16 compared to Figure G.4. For $\psi = 0.17$, the MBB percentile interval's coverage rates are much lower for $y_{1,t}$ in Figure G.16 than in Figure G.4. In contrast, the general pattern of coverage rates for $y_{2,t}$ in Figure G.16 is similar to the pattern in Figure G.4. These results show that the interaction of less VAR persistence and a weak proxy variable can materially change, but will not necessarily change, the coverage rates for FEVDs.

G.6 More Persistent Stochastic Volatility

We now return to using the VAR slope coefficients from the body of the paper, but change the stochastic volatility parameters of the structural shocks. In the body of the paper, we model the structural shocks with $\epsilon_{i,t} = \sigma_{i,t} w_{i,t}$ for $i = 1, 2$ with $w_{i,t} \overset{iid}{\sim} \mathcal{N}(0,1)$, $\ln \sigma_{i,t} = (1-\rho) \ln \bar{\sigma} + \rho \ln \sigma_{i,t-1} + \sigma_e e_{i,t}$, and $e_{i,t} \overset{iid}{\sim} \mathcal{N}(0,1)$. The equation $\ln \sigma_{i,t} = (1-\rho) \ln \bar{\sigma} + \rho \ln \sigma_{i,t-1} + \sigma_e e_{i,t}$ then governs the persistence and volatility of our stochastic volatility. In the paper, we use $\rho = 0.85$ and $\sigma_e = 0.15$ and impose $\ln \bar{\sigma} = -\sigma_e^2/(1-\rho^2)$. With these settings, the unconditional distribution of $\ln \sigma_{i,t}$ is $\mathcal{N}(-0.081, 0.081)$.

In this section, we make the stochastic volatility more persistent, choosing $\rho = 0.95$. This matches the stochastic volatility persistence of the shock with the most persistent stochastic volatility in Fernández-Villaverde, Guerrón-Quintana, and Rubio-Ramírez (2010). To isolate the effect of changing the degree of persistence, we then choose $\sigma_e = 0.09$. This

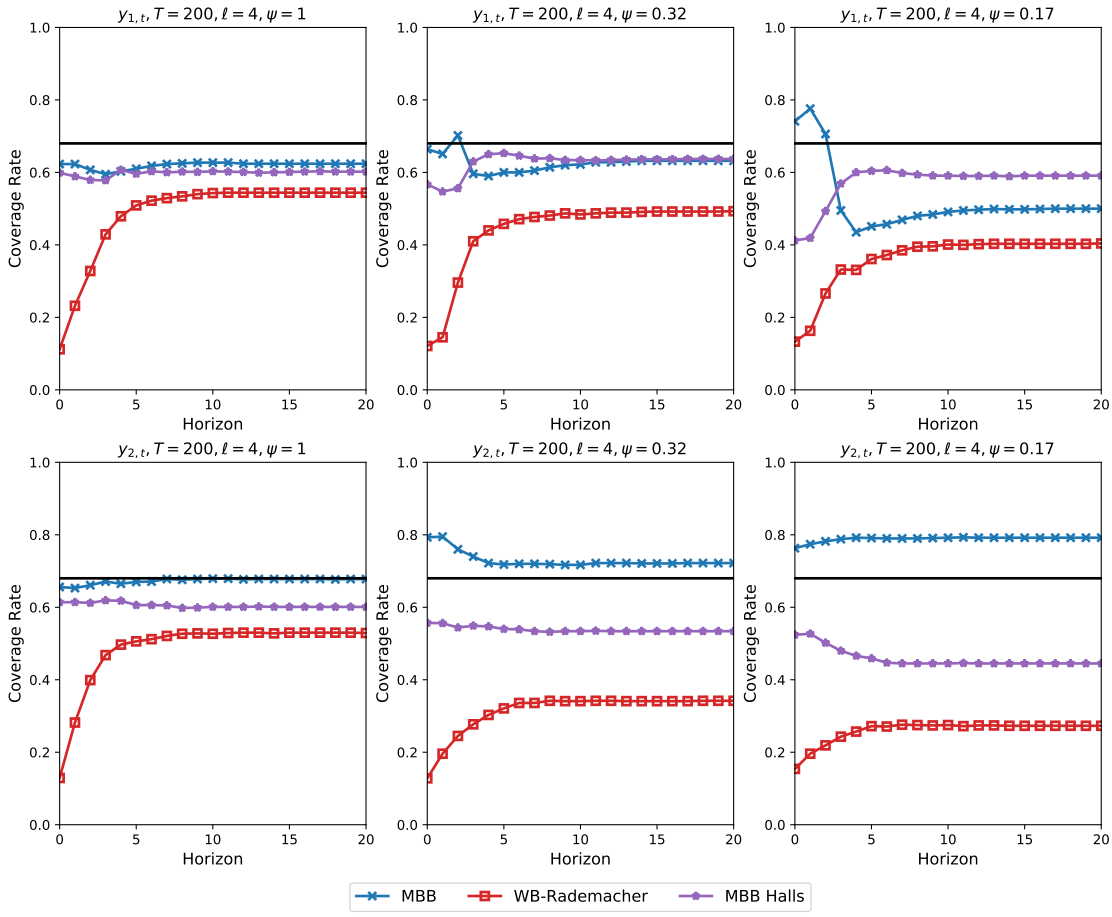

Figure G.16: Coverage rates of 68% confidence intervals for FEVDs. The solid horizontal line shows the 0.68 target level.

implies that the unconditional distribution of $\ln \sigma_{i,t}$ is $\mathcal{N}(-0.083, 0.083)$, closely matching the unconditional distribution from the paper.

We show the coverage rates of 68% confidence intervals and sets for normalized IRFs in Figure G.17, the average lengths of 68% confidence intervals and sets for the normalized IRFs in Figure G.18, and the coverage rates of 68% confidence intervals for FEVDs in Figure G.19. Overall, we find that the results for coverage rates and average lengths of confidence intervals and sets for normalized IRFs are little changed with the increased persistence in stochastic volatility. That is, the results in Figures G.17 and G.18 are very similar to the results in Figures G.2 and G.3. Turning to coverage rates for confidence intervals for FEVDs, we find that the coverage rates of the MBB percentile intervals for $\psi = 1$ are slightly lower in Figure G.19 with the increased stochastic volatility persistence than in Figure G.4. Otherwise, the results in the two figures are generally similar.

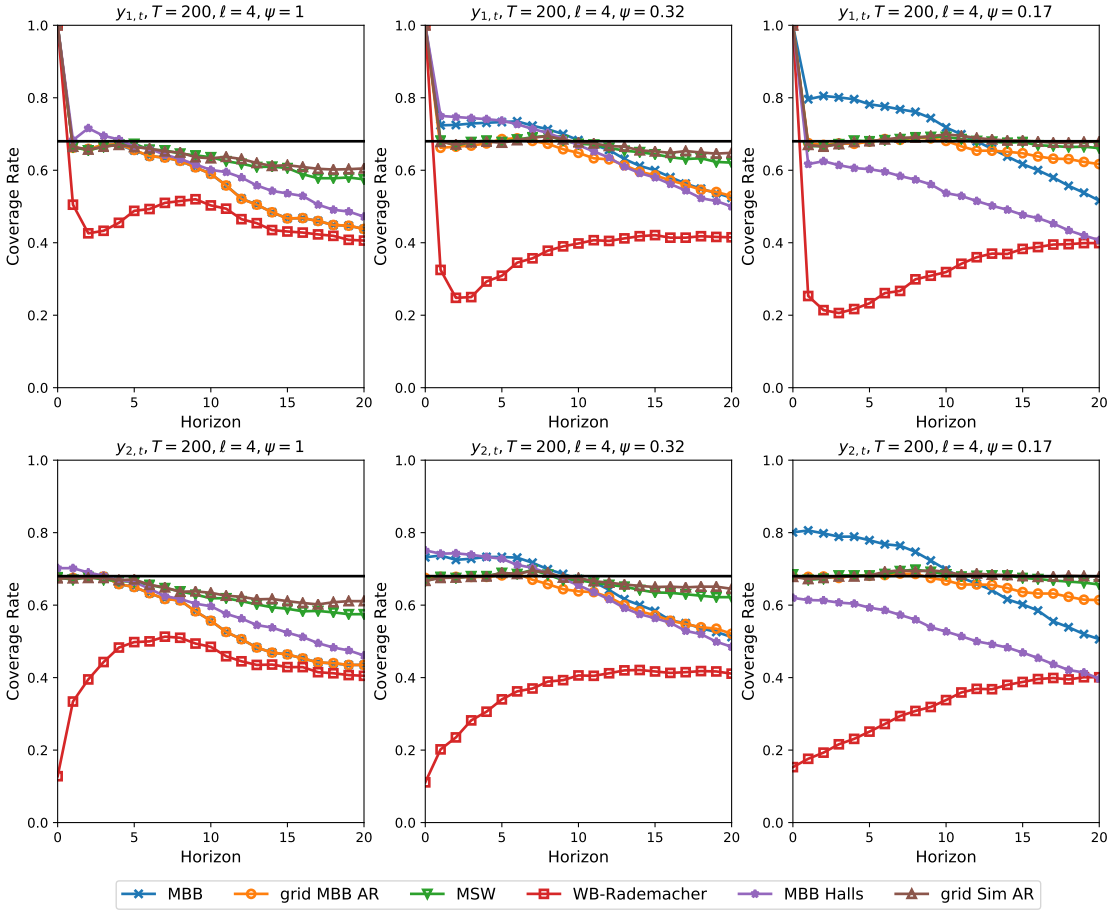

Figure G.17: Coverage rates of 68% confidence intervals and sets for normalized IRFs. The solid horizontal line shows the 0.68 target level.

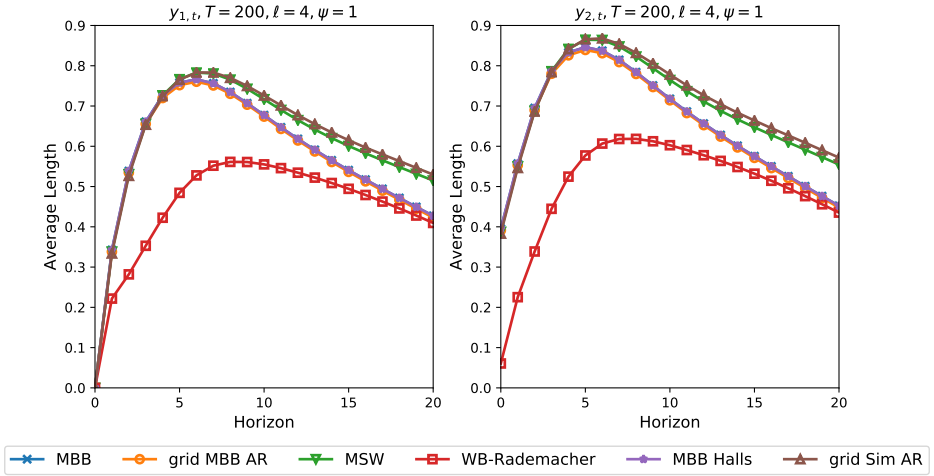

Figure G.18: Average lengths of 68% confidence intervals for normalized IRFs.

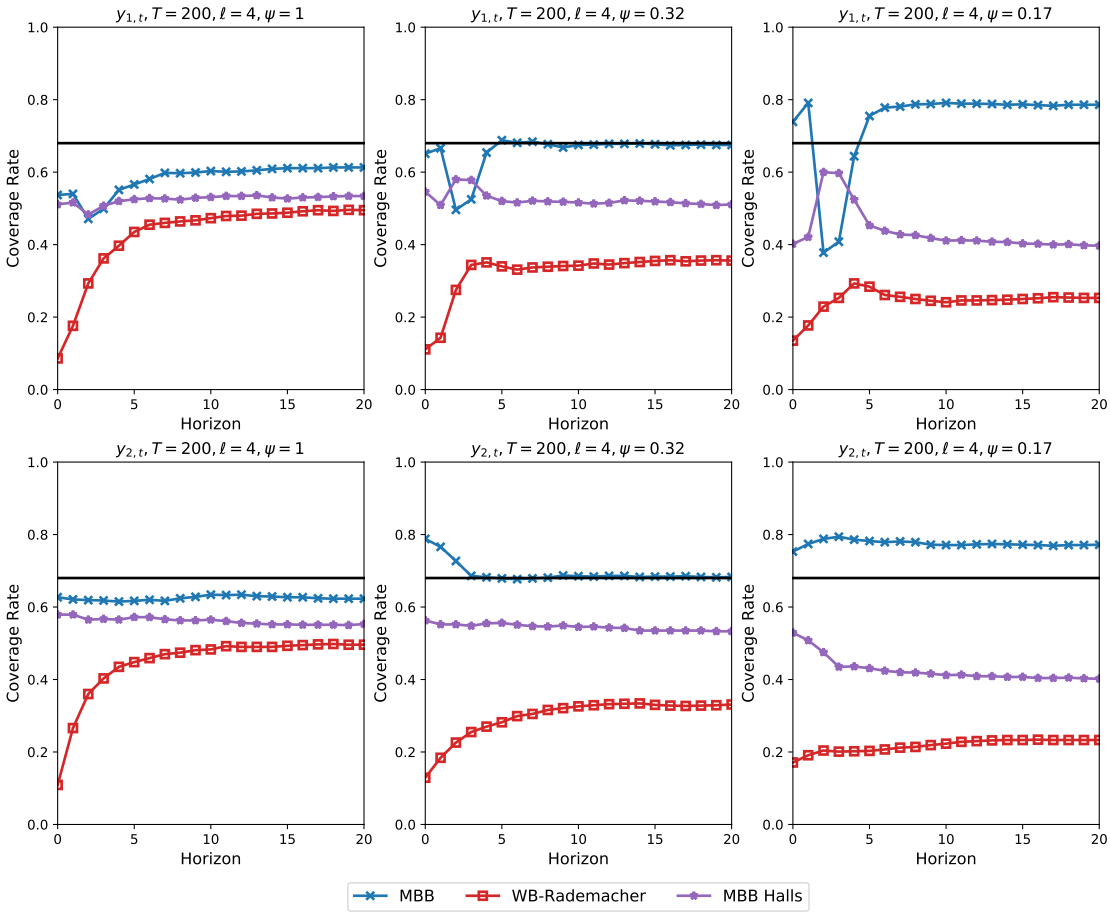

Figure G.19: Coverage rates of 68% confidence intervals for FEVDs. The solid horizontal line shows the 0.68 target level.

G.7 A DGP with Censoring

In the body of the paper, we model the proxy variable with $m_t = \psi\epsilon_{1,t} + v_t$ with $v_t \overset{iid}{\sim} \mathcal{N}(0,1)$. In this section, we instead use $m_t = d_t(\psi\epsilon_{1,t} + v_t)$ in which $d_t \in \{0,1\}$ is a random scalar process that either censors m_t to be zero or allows m_t to take the value $\psi\epsilon_{1,t} + v_t$. In this section, we use $\Pr(d_t = 0) = 0.4$. This value falls in between Gertler and Karadi (2015), in which $\Pr(d_t = 0)$ is roughly 0.2, and Mertens and Ravn (2013), in which $\Pr(d_t = 0)$ is greater than 0.9.

In order to isolate the effects of censoring without weakening the proxy variable, we change the values ψ in each DGP. We use $\psi = 2.2$ as our strong proxy DGP, $\psi = 0.42$ as our middle proxy DGP, and $\psi = 0.22$ as our weak proxy DGP. We choose these values to roughly match the distribution of F statistics that serve as pre-tests of proxy strength from the body of the paper. With $\psi = 2.2$, the pre-tests reject the presence of a weak proxy in every simulation. With $\psi = 0.43$, the pre-tests reject the presence of a weak proxy in roughly half of the simulations. With $\psi = 0.22$, the pre-tests reject the presence of a weak proxy in less than 10% of the simulations.

We show the coverage rates of 68% confidence intervals and sets for normalized IRFs in Figure G.20, the average lengths of 68% confidence intervals and sets for the normalized IRFs in Figure G.21, and the coverage rates of 68% confidence intervals for FEVDs in Figure G.22. Overall, we find that censoring has little effect on the results. The coverage rates of confidence intervals and sets for normalized IRFs are slightly lower at short horizons in Figure G.20 compared to Figure G.2. Otherwise, the general pattern of coverage rates is quite similar. We also find that the average lengths of confidence intervals and sets for the normalized IRFs in Figure G.21 are similar to those in Figure G.3. Finally, there is very little difference in the coverage rates of confidence intervals for FEVDs in Figures G.22 and G.4.

G.8 Results with Lag Augmentation

In this section, we follow the spirit of Inoue and Kilian (2020) and use lag augmentation to produce IRFs and FEVDs with the MBB. We make two notes before proceeding. First, Inoue and Kilian (2020) study autoregressions – not vector autoregressions. Second, they assume that the autoregression innovations are iid – not α-mixing. Because of this, we do not have theory proving that lag augmentation is valid for proxy SVARs under α-mixing. Rather, this section simply provides some simulation evidence that lag augmentation in the spirit of Inoue and Kilian (2020) may potentially be extended to proxy SVARs.

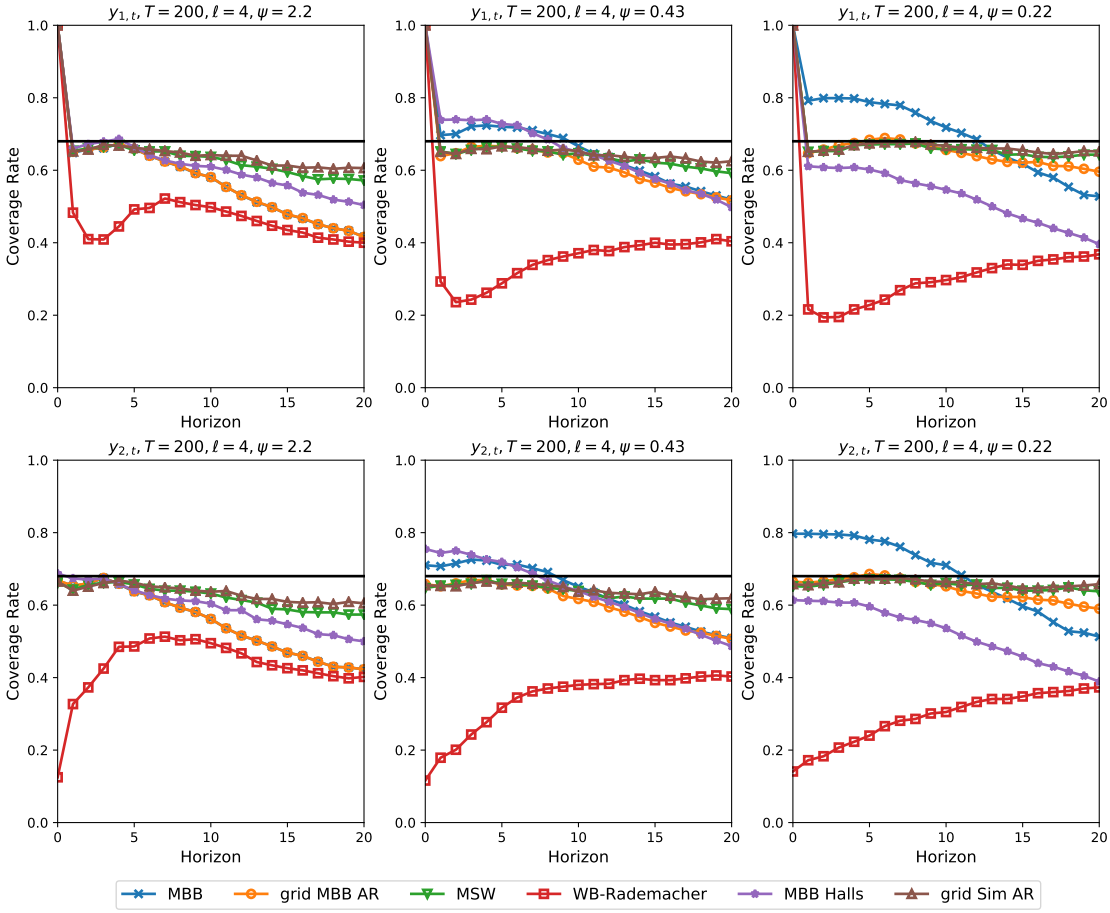

Figure G.20: Coverage rates of 68% confidence intervals and sets for normalized IRFs. The solid horizontal line shows the 0.68 target level.

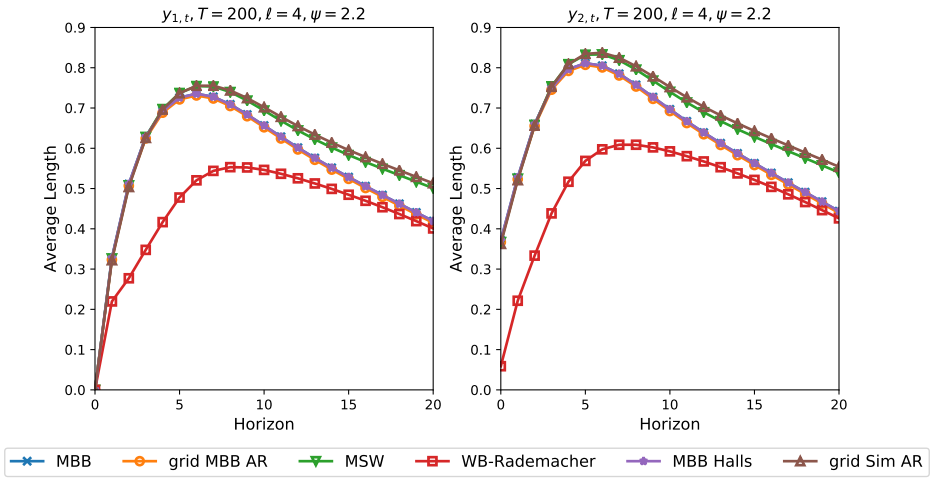

Figure G.21: Average lengths of 68% confidence intervals for normalized IRFs.

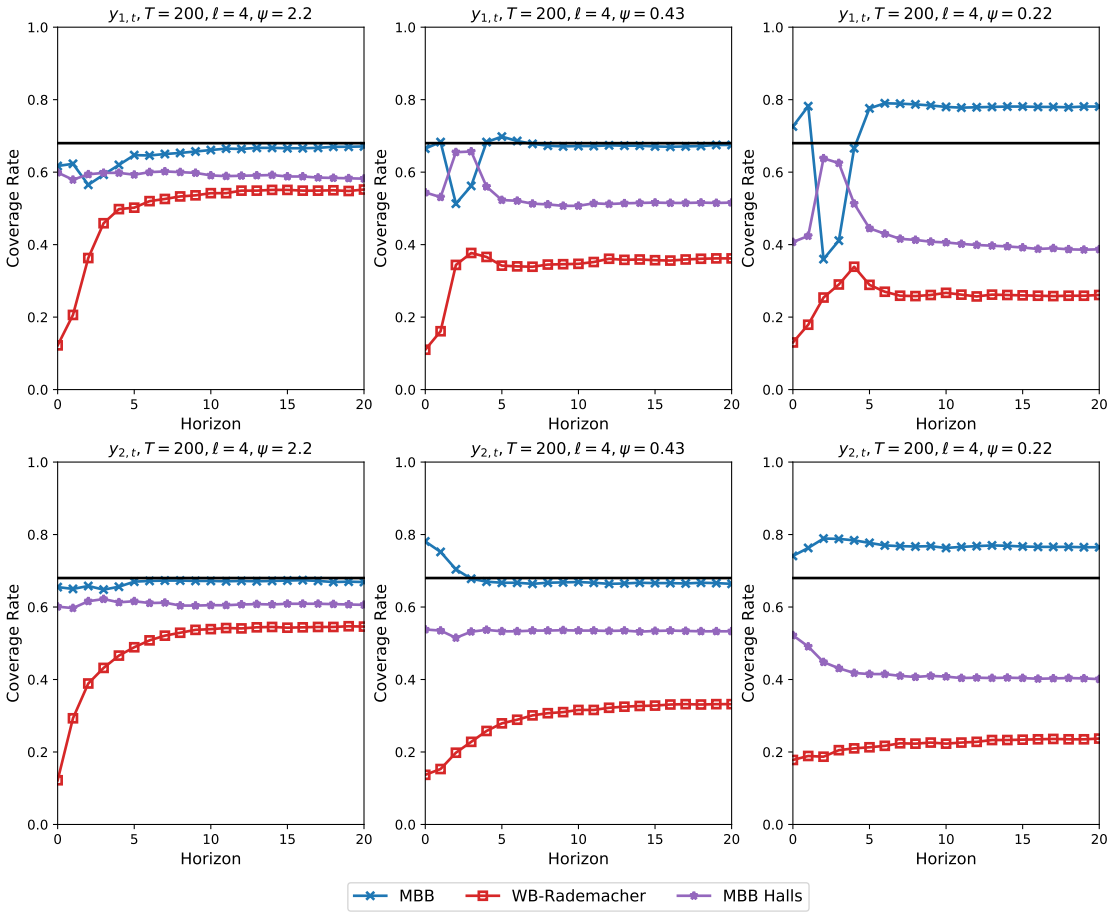

Figure G.22: Coverage rates of 68% confidence intervals for FEVDs. The solid horizontal line shows the 0.68 target level.

We begin by describing the MBB with lag augmentation. To initialize the algorithm, we estimate the VAR intercepts and slope coefficients in

$$y_t = \nu + A_1 y_{t-1} + \cdots A_p y_{t-p} + A_{p+1} y_{t-p-1} + u_t$$

with least squares. We denote the estimates with $\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_{p+1}$. We note that this VAR includes $p+1$ lags and not p lags. This additional lag is the lag augmentation from Inoue and Kilian (2020). Because the assumption in the paper is that the data generating process is a VAR(p), the true value of A_{p+1} is $O_{K\times K}$ and the asymptotic value of $\widehat{A}_{p+1}$ is also $O_{K\times K}$.

Next, as part of the initialization, we compute the VAR residuals using the estimated slope coefficients for the $p+1$ lags: $\widehat{u}_t = y_t - \widehat{\nu} + \widehat{A}_1 y_{t-1} + \cdots + \widehat{A}_{p+1} y_{t-p-1}$. Under the maintained assumption that we observe the data sample $(y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T)$, we then estimate $T-1$ observations of u_t from $t=2$ to $t=T$.

Finally, as part of the initialization, choose a block length ℓ and compute $N = [(T-1)/\ell]$, where $[\cdot]$ rounds up to the nearest integer so that $N\ell \geq T-1$. Next, collect the $(K\times\ell)$ blocks $\widehat{\mathcal{U}}_i = (\widehat{u}_i, \ldots, \widehat{u}_{i+\ell-1})$ for $i = 2, \ldots, T-\ell+1$ and the $(r\times\ell)$ blocks $\mathcal{M}_i = (m_i, \ldots, m_{i+\ell-1})$ for $i = 2, \ldots, T-\ell+1$.

With this initialization, the lag-augmented MBB is as follows

1. Independently draw N integers with replacement from the set $\{2, \ldots, T-\ell-1\}$, putting equal probability on each element of the set. Denote these integers as $i_1, \ldots, i_N$.

2. Collect the blocks $(\widehat{\mathcal{U}}_{i_1}, \ldots, \widehat{\mathcal{U}}_{i_N})$ and $(\mathcal{M}_{i_1}, \ldots, \mathcal{M}_{i_N})$ and drop the last $N\ell - (T-1)$ elements to produce $(\widetilde{u}_2^*, \ldots, \widetilde{u}_T^*)$ and $(m_2^*, \ldots, m_T^*)$. As with the MBB in the paper, we also use " * " to denote bootstrap quantities from this lag-augmented MBB.

3. Center $(\widetilde{u}_2^*, \ldots, \widetilde{u}_T^*)$ according to $u_{j\ell+s}^* = \widetilde{u}_{j\ell+s}^* - \frac{1}{T-\ell}\sum_{r=0}^{T-1-\ell} \widehat{u}_{s+r}$ for $s = 2, \ldots, \ell+1$ and $j = 0, 1, \ldots, N-1$ in order to produce $(u_2^*, \ldots, u_T^*)$ to assure that u_t^*'s are centered conditionally on the data.

4. Set the initial condition $(y_{-p+1}^*, \ldots, y_0^*, y_1^*) = (y_{-p+1}, \ldots, y_0, y_1)$. Use the initial condition along with $\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_{p+1}$ and u_t^* to recursively produce $(y_2^*, \ldots, y_T^*)$ with $y_t^* = \widehat{\nu} + \widehat{A}_1 y_{t-1}^* + \cdots + \widehat{A}_p y_{t-p}^* + \widehat{A}_{p+1} y_{t-p-1}^* + u_t^*$.

5. Estimate $\widehat{\nu}^*, \widehat{A}_1^*, \ldots, \widehat{A}_{p+1}^*$ by least squares from the bootstrap sample $(y_{-p+1}^*, \ldots, y_T^*)$ and set $\widehat{u}_t^* = y_t^* - \widehat{\nu}^* + \widehat{A}_1^* y_{t-1}^* + \cdots + \widehat{A}_{p+1}^* y_{t-p-1}^*$.

6. Use $\widehat{u}_t^*$ and m_t^* for $t = 2, \ldots, T$ to estimate $\widehat{\Sigma}_u^* = \frac{1}{T-1}\sum_{t=2}^{T} \widehat{u}_t^* \widehat{u}_t^{*\prime}$ and $\widehat{H^{(1)}\Psi'}^* = \frac{1}{T-1}\sum_{t=2}^{T} \widehat{u}_t^* m_t^{*\prime}$.

7. Use $\widehat{A}_1^*, \ldots, \widehat{A}_p^*$, *but not* $\widehat{A}_{p+1}^*$, to produce the bootstrapped VMA coefficients, $\widehat{\Phi}_i^*$ for $i = 0, 1, \ldots$. Then, use the VMA coefficients, $\widehat{\Sigma}_u^*$ and $\widehat{H^{(1)}\Psi'}^*$ to produce the bootstrapped IRFs and FEVDs. In this step, use the same identification scheme as when computing the point estimates of the IRFs and FEVDs. This includes the same sign and scale normalizations.

We emphasize here that $\widehat{A}_{p+1}^*$ is effectively treated as $\mathbf{0}_{K\times K}$ in step 7 and not used when computing VMA coefficients, IRFs, or FEVDs. We repeat the algorithm a large number of times and save the bootstrapped IRFs and FEVDs. We produce our confidence intervals with a standard percentile interval by sorting the bootstrapped IRFs and FEVDs and keeping the $\alpha/2$ and $1-\alpha/2$ percentiles as the confidence interval.

For the grid MBB AR confidence sets, we use the computational approach in Section F. The lag-augmented algorithm is

1. Construct a grid $\boldsymbol{\xi} = \{\xi_1, \ldots \xi_G\}$, with ξ_g denoting one grid point, of null hypotheses $\Xi_{j1,i}(s; m, 1) = \xi_g$.

2. Follow steps 1 through 7 of the lag-augmented residual-based MBB algorithm to compute a large number of the bootstrapped estimates $\widehat{A}_1^*, \ldots, \widehat{A}_p^*, \widehat{A}_{p+1}^*$ and $\widehat{\boldsymbol{\varphi}}^* = \widehat{H^{(1)}\psi}^*$. Use $\widehat{A}_1^*, \ldots, \widehat{A}_p^*$, *but not* $\widehat{A}_{p+1}^*$, to compute $\widehat{\Phi}_i^*$ for $i = 0, 1, \ldots$.

3. For each grid point, use the bootstrapped estimates to compute $\sqrt{T}(se_j'\widehat{\Phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$ for each bootstrap loop. Compute the fraction of the bootstrapped values of $\sqrt{T}(se_j'\widehat{\phi}_i^* - \xi_g e_m')\widehat{\boldsymbol{\varphi}}^*$ that are less than or equal to zero.

4. Construct the confidence interval for $\Xi_{j1,i}(s; m, 1)$ by including any grid point, $\xi_g \in \boldsymbol{\xi}$, with the property that the fraction in the previous step is between $\alpha/2$ and $1-\alpha/2$.

We now show the simulation results for the lag-augmented MBB compared to the MBB provided in the paper. We how four sets of results: coverage rates for normalized IRFs, average lengths of confidence intervals for normalized IRFs, coverage rates for FEVDs, and average lengths of confidence intervals for FEVDs. Overall, we find that the lag-augmented MBB produces coverage rates that are closer to the target level, especially at longer horizons. We also find that the average lengths of the confidence intervals from the lag-augmented MBB

can be much bigger at longer horizons. We only show results for the strong proxy variable with $\psi = 1$.

Coverage rates for normalized IRFs: Figure G.23 shows the coverage rates of 68% MBB percentile intervals and grid MBB AR confidence sets with and without lag augmentation for normalized IRFs. The coverage rates without lag augmentation are as in the paper: near the target level at short horizons but too low at long horizons. With lag augmentation, the coverage rates at short horizons are essentially unchanged. However, at horizons longer than 10, the coverage rates with lag augmentation are generally closer to the target level and are very close to the target level at horizons 17 through 20.

Average lengths for normalized IRFs: Figure G.24 shows the average lengths of 68% MBB percentile intervals and grid MBB AR confidence sets with and without lag augmentation for normalized IRFs. The average lengths without lag augmentation are as in the paper. At short horizons, the average lengths with lag augmentation are quite similar to those without lag augmentation. However, the average lengths with lag augmentation are much bigger at longer horizons – roughly 10 times bigger.

Coverage rates for FEVDs: Figure G.25 shows the coverage rates of 68% MBB percentile intervals with and without lag augmentation for FEVDs. The coverage rates without lag augmentation are as in the paper. With lag augmentation, the coverage rates of $y_{1,t}$ are slightly closer to the target level. For $y_{2,t}$, the coverage rates with lag augmentation are essentially the same as without lag augmentation.

Average lengths for FEVDs: Figure G.26 shows the average lengths of 68% MBB percentile intervals with and without lag augmentation for FEVDs. The average lengths with lag augmentation are generally equal to or larger the average lengths without lag augmentation; however, the differences are small. Overall, in contrast to the normalized IRFs, we find that lag augmentation only has small effects on coverage rates and average lengths of percentile intervals for FEVDs.

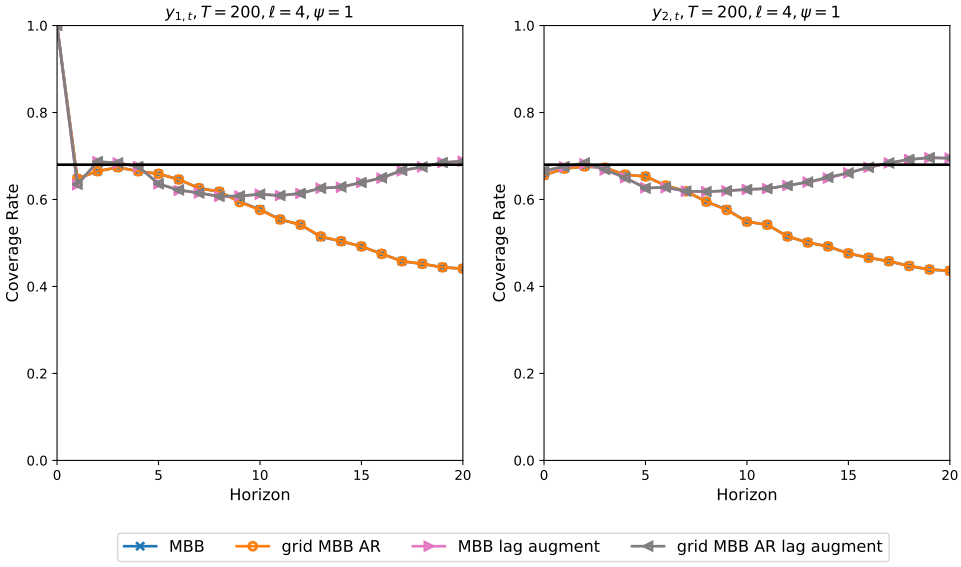

Figure G.23: Coverage rates of 68% confidence intervals for normalized IRFs. The solid horizontal line shows the 0.68 target level.

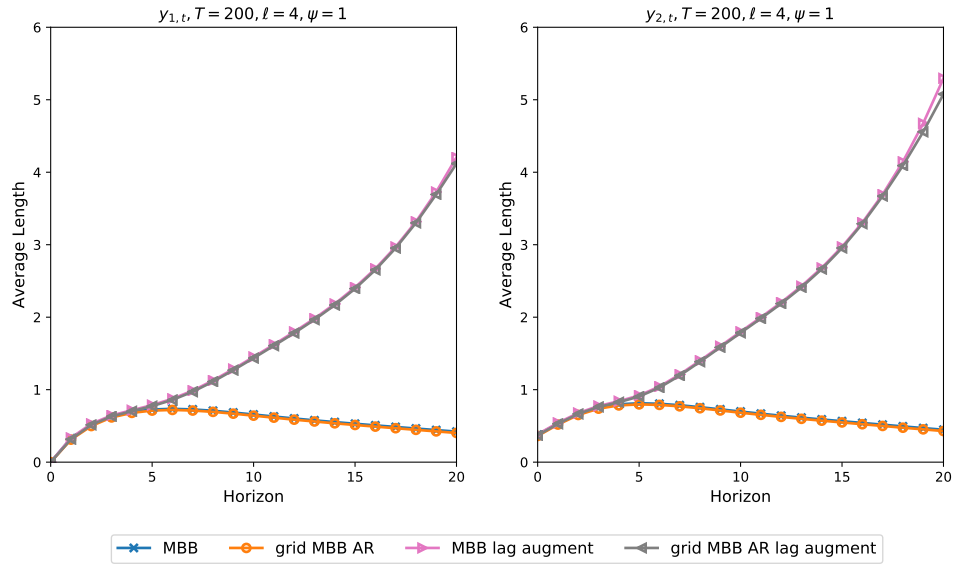

Figure G.24: Average lengths of 68% confidence intervals for normalized IRFs.

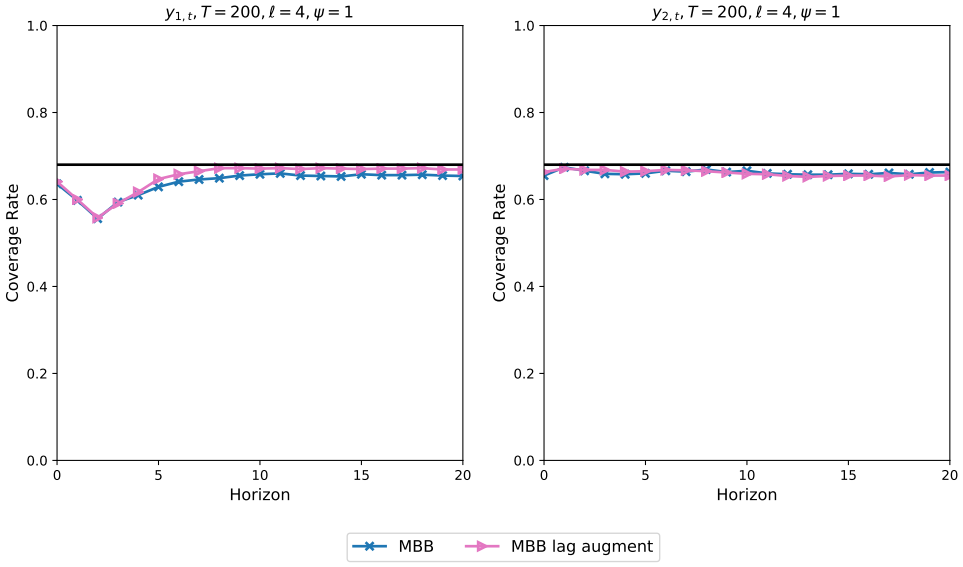

Figure G.25: Coverage rates of 68% confidence intervals for FEVDs. The solid horizontal line shows the 0.68 target level.

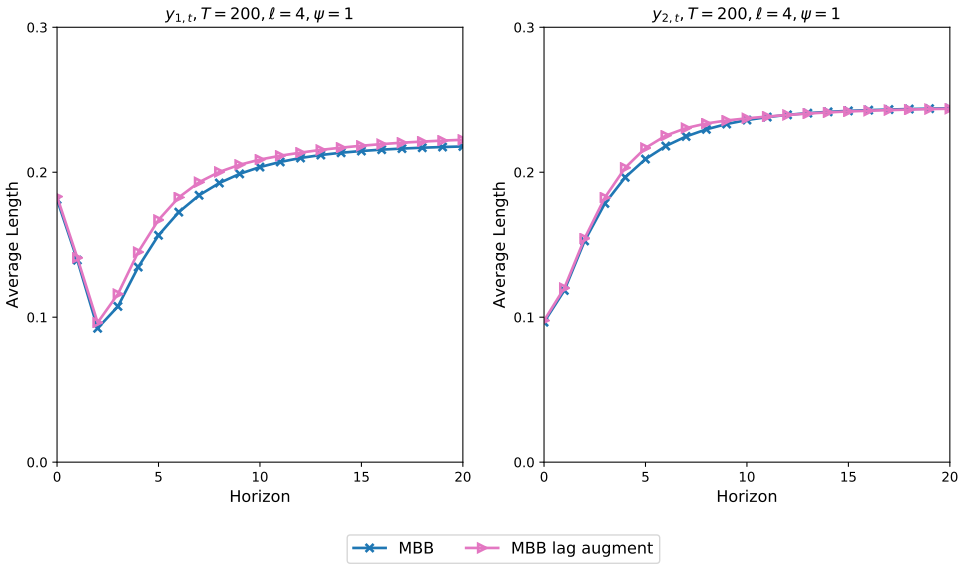

Figure G.26: Average lengths of 68% confidence intervals for FEVDs.

References

Anderson, T. W. and Herman Rubin. 1949. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations." *Annals of Mathematical Statistics* 20 (1):46–63.

Brockwell, Peter J. and Richard A. Davis. 1991. *Time Series: Theory and Methods, Second Edition*. Springer New York.

Brüggemann, Ralf, Carsten Jentsch, and Carsten Trenkler. 2014. "Inference in VARs with Conditional Heteroskedasticity of Unknown Form." University of Konstanz working paper 2014-13.

———. 2016. "Inference in VARs with Conditional Heteroskedasticity of Unknown Form." *Journal of Econometrics* 191 (1):69–85.

Fernández-Villaverde, Jesús, Pablo Guerrón-Quintana, and Juan F. Rubio-Ramírez. 2010. "Reading the Recent Monetary History of the United States, 1959-2007." Federal Reserve Bank of St. Louis *Review*, 92(4): 311-338.

Gertler, Mark and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1):44–76.

Gonçalves, Sílvia and Lutz Kilian. 2004. "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form." *Journal of Econometrics* 123 (1):89–120.

Hall, Peter. 1992. *The Bootstrap and Edgeworth Expansion*. Springer, New York.

Inoue, Atsushi and Lutz Kilian. 2020. "The Uniform Validity of Impulse Response Inference in Autoregressions." *Journal of Econometrics* 215 (2):450–472.

Jentsch, Carsten and Kurt G. Lunsford. 2016. "Proxy SVARs: Asymptotic Theory, Bootstrap Inference, and the Effects of Income Tax Changes in the United States." Federal Reserve Bank of Cleveland, Working Paper no. 16-19.

———. 2019a. "Asymptotically Valid Bootstrap Inference for Proxy SVARs." Federal Reserve Bank of Cleveland, Working Paper no. 19-08.

———. 2019b. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment." *American Economic Review* 109 (7):2655–2678.

Kilian, Lutz. 1998. "Small-Sample Confidence Intervals for Impulse Response Functions." *Review of Economics and Statistics* 80 (2):218–230.

Lunsford, Kurt G. 2015. "Identifying Structural VARs with a Proxy Variable and a Test for a Weak Proxy." Federal Reserve Bank of Cleveland Working Paper no. 15-28.

Lütkepohl, Helmut. 2005. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag.

Mertens, Karel and Morten O. Ravn. 2013. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States." *American Economic Review* 103 (4):1212–1247.

Montiel Olea, José Luis, James H. Stock, and Mark W. Watson. forthcoming. "Inference in Structural Vector Autoregressions Identified with an External Instrument." In Press for *Journal of Econometrics*, DOI:https://doi.org/10.1016/j.jeconom.2020.05.014.

Stock, James H. and Mark W. Watson. 2012. "Disentangling the Channels of the 2007-09 Recession." *Brookings Papers on Economic Activity* Spring:81–135.