working paper

19 08

Asymptotically Valid Bootstrap Inference for Proxy SVARs

Carsten Jentsch and Kurt G. Lunsford

FEDERAL RESERVE BANK OF CLEVELAND

ISSN: 2573-7953

Working Paper 19-08 May 2019

Asymptotically Valid Bootstrap Inference for Proxy SVARs

Carsten Jentsch and Kurt G. Lunsford

Proxy structural vector autoregressions identify structural shocks in vector autoregressions with external variables that are correlated with the structural shocks of interest but uncorrelated with all other structural shocks. We provide asymptotic theory for this identification approach under mild α-mixing conditions that cover a large class of uncorrelated, but possibly dependent innovation processes, including conditional heteroskedasticity. We prove consistency of a residual-based moving block bootstrap for inference on statistics such as impulse response functions and forecast error variance decompositions. Wild bootstraps are proven to be generally invalid for these statistics and their coverage rates can be badly and persistently mis-sized.

JEL classification: C30, C32.

Keywords: External Instruments, Mixing, Proxy Variables, Residual-Based Moving Block Bootstrap, Structural Vector Autoregression, Wild Bootstrap.

Suggested citation: Jentsch, Carsten, and Kurt G. Lunsford. 2019. "Asymptotically Valid Bootstrap Inference for Proxy SVARs." Federal Reserve Bank of Cleveland, Working Paper no. 19-08. https://doi.org/10.26509/frbc-wp-201908.

Carsten Jentsch is at TU Dortmund University (jentsch@statistik.tu-dortmund.de). Kurt G. Lunsford is at the Federal Reserve Bank of Cleveland (kurt.lunsford@clev.frb.org). Some results in this paper originally circulated in a working paper titled, "Proxy SVARs: Asymptotic Theory, Bootstrap Inference, and the Effects of Income Tax Changes in the United States." The authors are grateful to Todd E. Clark, Lutz Kilian, Helmut Lütkepohl, Johannes Pfeifer, Carsten Trenkler, and participants at the IAAE 2016 Annual Conference and the CFE 2018 conference for helpful comments and conversations. Large parts of this research were conducted while Carsten Jentsch held a position at the University of Mannheim, where he was financially supported by the German Research Foundation DFG via the Collaborative Research Center SFB 884 "Political Economy of Reforms" (Project B6) and the Baden-Württemberg-Stiftung via the Eliteprogram for Postdocs. Further support by the Collaborative Research Center SFB 823 "Modeling of Nonlinear Dynamic Processes" (Project A1) is acknowledged.

1 Introduction

Stock and Watson (2008, 2012), Montiel Olea, Stock, and Watson (2012), and Mertens and Ravn (2013) have developed a method for identification in structural vector autoregressions (SVARs) that uses external instruments or proxy variables.[1] These variables are external to the reduced-form vector autoregression (VAR). They are also assumed to be correlated with the structural shocks of interest but uncorrelated with the other structural shocks. This identification approach has proven to be very useful and it has been discussed in the recent handbook chapters by Ramey (2016) and Stock and Watson (2016). Further discussion and examples of applications are also provided in the citations in the next paragraph.

In this paper, we derive the joint limiting distribution of parameter estimators in proxy SVARs and provide econometric theory for residual-based bootstrap algorithms that provide inference on statistics, such as impulse response functions (IRFs) and forecast error variance decompositions (FEVDs), estimated from proxy SVARs.[2] Our focus on residual-based bootstraps is due to their popularity in time series econometrics to do statistical inference. In the proxy SVAR literature, residual-based wild bootstraps are very popular and have been used to produce confidence intervals for IRFs in, for example, Mertens and Ravn (2013, 2014), Carriero et al. (2015), Gertler and Karadi (2015), Passari and Rey (2015), Miranda-Agrippino (2016), Rey (2016), Cesa-Bianchi and Sokol (2017), Hachula and Nautz (2018), Piffer and Podstawski (2018), and Kerssenfischer (2019).

Despite the general popularity of residual-based bootstraps for SVARs, to the best of our knowledge, no econometric theory exists to guide researchers as to which residual-based bootstraps are asymptotically valid for proxy SVARs and, hence, can be recommended in practice. To fill this gap in the literature, this paper makes three contributions. First, we provide a joint central limit theorem (CLT) for estimators of the VAR coefficients, the (unconditional) covariance matrix of the VAR innovations, and the (unconditional) covariance matrix of the VAR innovations with the proxy variables under mild α-mixing conditions that cover a large class of uncorrelated, but possibly dependent innovation processes, including conditional heteroskedasticity. This result allows us to also derive the limiting distributions of IRFs and FEVDs. Such results extend Theorem 3.1 and Corollary 5.1 achieved under

[1] "Proxy variable" and "external instrument" are different names for the same variable in this literature. SVARs identified with these variables have been called "proxy SVARs" or "external instrument SVARs." To simplify communication, we use "proxy variable" and "proxy SVAR" going forward. However, it should be understood that we could equivalently have used "external instrument" or "external instrument SVARs."

[2] Throughout the paper, "IRF" refers to structural impulse response functions from orthogonalized economic shocks. It does not refer to forecast error impulse response functions as in Lütkepohl (2005, p. 52).

α-mixing conditions for Cholesky-based identification in SVARs in Brüggemann, Jentsch, and Trenkler (2016) to the proxy SVAR setup. In particular, conditional heteroskedasticity has been observed in many economic time series data and results achieved under more general conditions imposed on the error processes, such as α-mixing, are required. Conditional heteroskedasticity has been noted previously in the proxy SVAR literature (Mertens and Ravn, 2013; Gertler and Karadi, 2015) and our theory allows for statistical inference that is robust against conditional heteroskedasticity.

Second, we prove that the popular residual-based wild bootstrap is not asymptotically valid for inference on IRFs and FEVDs. While wild bootstraps can be valid for inference on estimates of the VAR slope coefficients under conditional heteroskedasticity,[3] they are not valid for inference on estimates of the covariance matrix of the VAR innovations, and the covariance of the VAR innovations with the proxy variables. This is because wild bootstraps do not replicate the relevant fourth-order dependence structure of the VAR innovations. Wild bootstraps even fail when these innovations are independent and identically distributed (iid). This causes wild bootstraps to be generally invalid also for smooth functions of the covariance matrices, such as IRFs and FEVDs.

As a corollary to this result, we show that the Rademacher distribution, which takes the values 1 or -1 with equal probability of one half and is used in many of the papers cited above, is particularly problematic for producing wild bootstrap multipliers.[4] Specifically, it gives an asymptotic variance of zero for the estimates of the covariance matrix of the VAR innovations and the covariance between the VAR residuals and the proxy variables. This is because the bootstrap multipliers effectively drop out of the bootstrap algorithm when computing the covariance matrices.

In practice, confidence intervals for IRFs and FEVDs that are produced from Rademacher wild bootstraps can be badly undersized. In Monte Carlo simulations, we show that the Rademacher wild bootstrap's 95 percent confidence intervals only include the true initial impulse response in 20 percent or less of the simulations with a sample size of 400. This is not a finite sample effect. Instead, these coverage rates shrink as the sample size increases. Further, our simulations show that the Rademacher wild bootstrap's low coverage rates can

[3]This result holds under a martingale difference condition imposed on the error process. See Gonçalves and Kilian (2004) for univariate AR models and Brüggemann, Jentsch, and Trenkler (2014, 2016) for the multivariate case.

[4]We describe the residual-based wild bootstrap algorithm in detail in Section 3.1.1. For exposition, we note that this bootstrap resamples VAR residuals by multiplying these residuals with a random variable. The Rademacher distribution is one distribution from which these random multipliers can be drawn.

persist even as the IRF horizon increases. This appears particularly true when the proxy variable's covariance with the structural shock of interest is low. In our simulations, we also find that these low coverage rates are more persistent for normalized IRFs than for one standard deviation IRFs. Finally, the Rademacher wild bootstrap's confidence intervals appear persistently undersized for FEVDs at long forecast horizons. The asymptotic theory established in this paper allows us to explain this observed pattern of persistently undersized confidence intervals for normalized IRFs and FEVDs in contrast to one standard deviation IRFs.

To replace the residual-based wild bootstrap, our third contribution is to prove that a modified version of the residual-based moving block bootstrap (MBB) introduced and studied by Brüggemann, Jentsch, and Trenkler (2016) is asymptotically valid for inference on statistics, such as IRFs and FEVDs, that are smooth functions of the VAR coefficients, the covariance matrix of the VAR innovations, and the covariance of the VAR innovations with the proxy variables. The modification to Brüggemann, Jentsch, and Trenkler's (2016) MBB incorporates the proxy variables into the block resampling. Importantly, the MBB is capable of mimicking the joint fourth-order dependence structure of the VAR innovations and the proxy variables.

In our Monte Carlo simulations, the MBB generally performs well. However, in small samples, the MBB's confidence intervals can become undersized for IRFs at longer horizons. Also, when the proxy variable's covariance with the structural shocks of interest is low, the MBB's confidence intervals may become mis-sized around humps in FEVDs. Despite this, the MBB often performs much better and never performs worse than the Rademacher wild bootstrap in terms of statistical size.

In the proxy SVAR literature, it is common for the proxy variables to have observations that get censored to zero. In Mertens and Ravn (2013), this occurs in periods where no tax legislation is passed. In Gertler and Karadi (2015), this occurs in periods with no Federal Open Market Committee meetings. We note that our CLT can accommodate data generating processes where the proxy variables have observations that are randomly censored to zero. The MBB will generally remain asymptotically valid under the mild condition that the censoring mechanism preserves stationarity. In particular, this is true for the model that Mertens and Ravn (2013) propose to include censoring of the proxy variable.[5]

In a related paper (Jentsch and Lunsford, 2019), we demonstrate the practical significance of using the asymptotically valid MBB instead of the invalid wild bootstrap. We apply the

[5]See Equation (8) in Mertens and Ravn (2013) and our Equation (21).

MBB to Mertens and Ravn (2013), who study the effects of tax changes on macroeconomic aggregates in the US. We use the MBB to produce confidence intervals for their IRFs, which were originally produced with a Rademacher wild bootstrap. With the MBB, all of the confidence intervals become much larger, and several of the results thought to be statistically significant at the 95 percent level were no longer significant even at the 68 percent level. Hence, this application and our Monte Carlo results provide evidence that the Rademacher wild bootstrap dramatically underestimates the uncertainty surrounding estimates of IRFs. For further discussion of this application, see Mertens and Ravn (2019).

While this paper studies the residual-based bootstrap methods for inference that are currently popular in the proxy SVAR literature, other methods for inference exist. Stock and Watson (2018) use a bootstrap algorithm, but their bootstrap is not residual-based like the MBB or the wild bootstrap. Rather, it treats the innovations to the VAR and proxy variables as iid normal and simulates these innovations using the corresponding estimates of the covariance matrices. These iid simulations imply that their bootstrap is not robust against conditional heteroskedasticity or, more generally, against the α-mixing conditions that we study. In an important contribution, Montiel Olea, Stock, and Watson (2018) provide confidence intervals for IRFs that are robust when the correlation between the proxy variable and the structural shock of interest is near zero, which is similar to the problem of weak instruments in linear IV regressions (Staiger and Stock, 1997). However, their confidence intervals only apply to the case where one proxy variable identifies one structural shock and cannot be used when multiple proxy variables identify multiple structural shocks as in Mertens and Ravn (2013) and Jentsch and Lunsford (2019). Further, their confidence intervals are constructed specifically for normalized IRFs. Unlike the MBB, they do not provide confidence intervals for one standard deviation IRFs or FEVDs. Finally, Caldara and Herbst (2019), Jarociński and Karadi (2018) and Arias, Rubio-Ramírez, and Waggoner (2018) provide Bayesian approaches to estimation and inference.

This paper proceeds as follows. Section 2 describes the proxy SVAR methodology and provides joint asymptotic theory in the form of CLTs. Section 3 describes the bootstrap algorithms and gives the theorems for the asymptotic invalidity of wild bootstraps and the asymptotic validity of the MBB. Section 4 studies the bootstrap coverage rates with Monte Carlo simulations, and Section 5 concludes. Appendix A provides additional details about proxy SVAR identification, and Appendix B contains the proofs. An online supplemental appendix has additional Monte Carlo results.

2 Proxy Structural Vector Autoregressions

2.1 An Overview of Proxy Structural Vector Autoregressions

The proxy SVAR methodology begins with a standard SVAR setup. We observe a data sample $(y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T)$ of sample size T with p pre-sample values from the following data generating process (DGP) for the K-dimensional time series $y_t = (y_{1,t}, \ldots, y_{K,t})'$,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad t \in \mathbb{Z}, \tag{1}$$

where $(u_t, t \in \mathbb{Z})$ is a K-dimensional white noise sequence. A compact representation is given by $A(L)y_t = u_t$, where $A(L) = I_K - A_1 L - \cdots - A_p L^p$, $A_p \neq 0$, I_K is the K-dimensional identity matrix, and L is the lag operator such that $Ly_t = y_{t-1}$. We assume that the lag order p is known and that $\det(A(z))$ has all roots outside the unit circle so that the DGP is a stable (invertible and causal) VAR model of order p. In addition, there is a K-dimensional sequence of structural shocks $(\epsilon_t, t \in \mathbb{Z})$ such that $\mathbb{E}(\epsilon_t \epsilon_t') = I_K$, which are related to the VAR innovations, u_t, according to

$$u_t = H\epsilon_t, \tag{2}$$

where H is an invertible $K \times K$ matrix. Hence, we have that

$$\mathbb{E}(u_t u_t') = HH' = \Sigma_u \tag{3}$$

is positive definite.

Our goal is to identify the effects of r of the structural shocks where $r < K$. To be precise, we partition the structural shocks into $\epsilon_t = (\epsilon_t^{(1)\prime}, \epsilon_t^{(2)\prime})'$, where $\epsilon_t^{(1)}$ is the r-dimensional vector that contains the structural shocks of interest and $\epsilon_t^{(2)}$ is the $(K-r)$-dimensional vector of other structural shocks. We then partition H into $H = [H^{(1)}, H^{(2)}]$, where $H^{(1)}$ is the $K \times r$ matrix of coefficients that correspond to the structural shocks of interest and $H^{(2)}$ is the $K \times (K-r)$ matrix of coefficients that correspond to the other shocks. Then, our goal is to identify $H^{(1)}$ and functions of $H^{(1)}$, such as IRFs and FEVDs.

The difficulty in identifying $H^{(1)}$ is that $\epsilon_t^{(1)}$ is unobserved and Equation (3) only provides $(K+1)K/2$ moment restrictions for the K^2 elements of H. To provide additional moment restrictions, Stock and Watson (2008, 2012), Montiel Olea, Stock, and Watson (2012), and Mertens and Ravn (2013) use the proxy variable approach. They assume that there exists a sequence of r-dimensional vectors of proxy variables, denoted by $(m_t, t \in \mathbb{Z})$, taken from outside of the VAR. Without loss of generality, these proxy variables are mean zero, $\mathbb{E}(m_t) = 0$. In addition, they are relevant for identifying the structural shocks of interest. That is,

$$\mathbb{E}(m_t \epsilon_t^{(1)\prime}) = \Psi, \tag{4}$$

where Ψ is $r \times r$ and invertible. They are also exogenous to the other structural shocks:

$$\mathbb{E}(m_t \epsilon_t^{(2)\prime}) = 0. \tag{5}$$

The relevance and exogeneity assumptions along with the partition of H imply

$$\mathbb{E}(m_t u_t') = \Psi H^{(1)\prime}, \tag{6}$$

which provides additional moment restrictions used for identification. Finally, we assume that the proxy variables and lags of y_t are uncorrelated. That is, $\mathbb{E}(m_t y_{t-j}') = 0$ for all $j = 1, \ldots, p$.[6]

When $r = 1$, Lunsford (2015) shows that $H^{(1)}$ is identified up to a sign restriction with

$$H^{(1)} = \pm \mathbb{E}(u_t m_t') \{\mathbb{E}(m_t u_t') [\mathbb{E}(u_t u_t')]^{-1} \mathbb{E}(u_t m_t')\}^{-1/2}. \tag{7}$$

Here, the sign restriction can be determined by whether the proxy variable is intended to be positively or negatively correlated with the structural shock of interest. When $r > 1$, additional restrictions are needed. We give two examples of zero restrictions in Appendix A. The first example generalizes (7), and the second example follows Mertens and Ravn (2013).

With $H^{(1)}$ identified, IRFs and FEVDs can be constructed. Since the process y_t is stable, it has a vector moving average (VMA) representation

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j}, \quad t \in \mathbb{Z}, \tag{8}$$

where Φ_j, $j \in \mathbb{N}_0$, is a sequence of (exponentially fast decaying) $K \times K$ coefficient matrices with $\Phi_0 = I_K$ and $\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j$, $i = 1, 2, \ldots$. Further, we define the $(Kp \times K)$ matrices $C_j = (\Phi_{j-1}', \ldots, \Phi_{j-p}')'$ and the $(Kp \times Kp)$ matrix $\Gamma = \sum_{j=1}^{\infty} C_j \Sigma_u C_j'$. Then, the impulse

[6] As discussed in Mertens and Ravn (2013), this is not a restrictive assumption. To ensure it holds, one can always regress the proxies on the lags of y_t and keep the residuals as the new proxies.

response of $y_{j,t+i}$ to a one standard deviation shock to $\epsilon_{k,t}$ is

$$\Theta_{jk,i} = e_j' \Phi_i H e_k, \tag{9}$$

where e_j and e_k are the jth and kth columns of I_K. We limit $k \leq r$ so that He_k is one of the identified columns of $H^{(1)}$.

In addition to this one standard deviation IRF, researchers also construct impulse responses to $y_{j,t+i}$ to a shock to $\epsilon_{k,t}$ that is normalized to give a certain response on impact. For example, it is common in studies of monetary policy shocks to normalize the response of the federal funds rate to a monetary policy shock to be 0.25 percent or 1.00 percent on impact. This normalized IRF is given by

$$\Xi_{jk,i}(s; m, n) = s\Theta_{jk,i}/(e_m' H e_n). \tag{10}$$

We limit $n \leq r$ so that He_n is one of the identified columns of $H^{(1)}$, and we assume that $e_m' H e_n \neq 0$ if estimating normalized IRFs. Note that $\Xi_{mn,0}(s; m, n) = s\Theta_{mn,0}/(e_m' H e_n) = s$. In words, this normalizes the response of variable m to structural shock n to be s on impact.

Finally, researchers may produce FEVDs, which measure how much the shocks of interest explain the variation in unexpected changes in $y_{j,t}$ over an h period horizon. FEVDs have long been used in SVAR analysis to assess which shocks are important for understanding macroeconomic fluctuations. See Sims (1980) for an early example and the last section of Ramey (2016) for a recent discussion. The total forecast error of $y_{j,t+h}$ is given by $y_{j,t+h} - \mathbb{E}_t(y_{j,t+h}) = e_j' \sum_{i=0}^{h-1} \Phi_i u_{t+h-i}$, and the forecast error of $y_{j,t+h}$ attributable to structural shock k is given by $e_j' \sum_{i=0}^{h-1} \Phi_i H e_k \epsilon_{k,t+h-i}$. Then, the fraction of the forecast error variance of $y_{j,t+h}$ attributable to structural shock k is given by

$$\omega_{jk,h} = \frac{\sum_{i=0}^{h-1} \Theta_{jk,i}^2}{MSE_j(h)} \tag{11}$$

where $MSE_j(h) = \sum_{i=0}^{h-1} e_j' \Phi_i \Sigma_u \Phi_i' e_j$ is the mean squared error of the forecast of $y_{j,t+h}$.

2.2 Estimation

To estimate the proxy SVAR and the corresponding IRFs and FEVDs, we focus on estimators for the VAR coefficients $A_1, \ldots, A_p$, the innovation covariance matrix Σ_u, and the $r \times K$ matrix $\Psi H^{(1)\prime}$. We introduce the following notation, where the dimensions of the defined quantities are given in parentheses:

$$\begin{aligned} \mathbf{y} &= \text{vec}(y_1, \dots, y_T)\ (KT \times 1), \quad & Z_t &= \text{vec}(y_t, \dots, y_{t-p+1})\ (Kp \times 1) \\ Z &= (Z_0, \dots, Z_{T-1})\ (Kp \times T), \quad & \boldsymbol{\beta} &= \text{vec}(A_1, \dots, A_p)\ (K^2 p \times 1) \\ \mathbf{u} &= \text{vec}(u_1, \dots, u_T)\ (KT \times 1), \quad & \boldsymbol{\varphi} &= \text{vec}(\Psi H^{(1)\prime})\ (Kr \times 1), \end{aligned} \tag{12}$$

where 'vec' denotes the column stacking operator.

The parameter $\boldsymbol{\beta}$ is estimated by $\widehat{\boldsymbol{\beta}} = \text{vec}(\widehat{A}_1, \dots, \widehat{A}_p)$ via multivariate least squares so that $\widehat{\boldsymbol{\beta}} = (ZZ')^{-1}(Z \otimes I_K)\mathbf{y}$ (Lütkepohl, 2005, p. 71). Here, $A \otimes B = (a_{ij}B)_{ij}$ denotes the Kronecker product of matrices $A = (a_{ij})$ and $B = (b_{ij})$. The standard estimator of Σ_u is

$$\widehat{\Sigma}_u = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_t \widehat{u}_t', \tag{13}$$

where $\widehat{u}_t = y_t - \widehat{A}_1 y_{t-1} - \dots - \widehat{A}_p y_{t-p}$ are the residuals from the estimated VAR(p) model. We set $\boldsymbol{\sigma} = \text{vech}(\Sigma_u)$ and $\widehat{\boldsymbol{\sigma}} = \text{vech}(\widehat{\Sigma}_u)$. The 'vech' operator stacks the elements on and below the main diagonal of a square matrix columnwise. Further, let $\widehat{\boldsymbol{\varphi}} = \text{vec}(\widehat{\Psi H^{(1)\prime}})$, where

$$\widehat{\Psi H^{(1)\prime}} = \frac{1}{T} \sum_{t=1}^{T} m_t \widehat{u}_t'. \tag{14}$$

Next, we estimate $H^{(1)}$. For the general case of $r \geq 1$, identification may require additional restrictions to the proxy SVAR model as noted above. The estimator for $H^{(1)}$ will also depend on these additional restrictions. For the identification methods in Appendix A, we estimate $H^{(1)}$ by plugging $\widehat{\Sigma}_u$ in for $\mathbb{E}(u_t u_t') = \Sigma_u$ and $\widehat{\Psi H^{(1)\prime}}$ in for $\mathbb{E}(m_t u_t') = \Psi H^{(1)\prime}$ in the identification equations. For any other identification schemes where $H^{(1)}$ is a smooth function of $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, we assume that similar plug-in estimators can be used.

Following this approach for $r = 1$, we have that $\Psi H^{(1)\prime}$ is $(1 \times K)$ and the notation in (12) implies that $\boldsymbol{\varphi} = \text{vec}(\Psi H^{(1)\prime}) = (\Psi H^{(1)\prime})' = H^{(1)}\Psi'$ is $(K \times 1)$. Hence, Equation (7) reads

$$H^{(1)} = \pm \mathbb{E}(u_t m_t') \{\mathbb{E}(m_t u_t')[\mathbb{E}(u_t u_t')]^{-1}\mathbb{E}(u_t m_t')\}^{-1/2} = \pm \boldsymbol{\varphi}(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-1/2}. \tag{15}$$

which implies the plug-in estimator

$$\widehat{H}^{(1)} = \pm \widehat{H^{(1)}\Psi'}(\widehat{\Psi H^{(1)\prime}}\widehat{\Sigma}_u^{-1}\widehat{H^{(1)}\Psi'})^{-1/2} = \pm \widehat{\boldsymbol{\varphi}}(\widehat{\boldsymbol{\varphi}}'\widehat{\Sigma}_u^{-1}\widehat{\boldsymbol{\varphi}})^{-1/2} \tag{16}$$

and that a sign restriction must be chosen.

Finally, we estimate IRFs and FEVDs with plug-in estimators. We start with $\widehat{\Phi}_0 = I_K$ and $\widehat{\Phi}_i = \sum_{j=1}^{i} \widehat{\Phi}_{i-j}\widehat{A}_j$, $i = 1, 2, \ldots$. Next, for the general case of $r \geq 1$, we use the notation $\widehat{H}e_k$ and $\widehat{H}e_n$ with $k, n \leq r$ to denote the relevant column of $\widehat{H}^{(1)}$. Then, we have $\widehat{\Theta}_{jk,i} = e_j'\widehat{\Phi}_i\widehat{H}e_k$, $\widehat{\Xi}_{jk,i}(s; m, n) = s\widehat{\Theta}_{jk,i}/(e_m'\widehat{H}e_n)$, such that $\widehat{\Xi}_{mn,0}(s; m, n) = s\widehat{\Theta}_{mn,0}/(e_m'\widehat{H}e_n) = s$, and $\widehat{\omega}_{jk,h} = \left(\sum_{i=0}^{h-1} \widehat{\Theta}_{jk,i}^2\right)/\widehat{MSE}_j(h)$ with $\widehat{MSE}_j(h) = \sum_{i=0}^{h-1} e_j'\widehat{\Phi}_i\widehat{\Sigma}_u\widehat{\Phi}_i'e_j$. When $r = 1$, we can simplify notation to $\widehat{\Theta}_{j1,i} = e_j'\widehat{\Phi}_i\widehat{H}^{(1)}$ and $\widehat{\Xi}_{j1,i}(s; m, 1) = s\widehat{\Theta}_{j1,i}/(e_m'\widehat{H}^{(1)})$.

2.3 Assumptions and Asymptotic Inference

In addition to the setup described in Section 2.1, we make use of the following assumptions:

Assumption 2.1 (Mixing Conditions)

(i) *Let $x_t = (u_t', m_t')'$ and assume that the $(K+r)$-dimensional process $(x_t, t \in \mathbb{Z})$ is strictly stationary.*

(ii) *Let $\alpha(n) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_n^\infty} |P(A \cap B) - P(A)P(B)|$, $n = 1, 2, \ldots$, denote the α-mixing coefficients of the process $(x_t, t \in \mathbb{Z})$, where $\mathcal{F}_{-\infty}^0 = \sigma(\ldots, x_{-2}, x_{-1}, x_0)$, $\mathcal{F}_n^\infty = \sigma(x_n, x_{n+1}, \ldots)$. For some $\delta > 0$, we have*

$$\sum_{n=1}^{\infty} (\alpha(n))^{\delta/(2+\delta)} < \infty \tag{17}$$

and that $\mathbb{E}|x_t|_{4+2\delta}^{4+2\delta}$ is bounded, where $|A|_p = (\sum_{i,j} |a_{ij}|^p)^{1/p}$ for some matrix $A = (a_{ij})$.

(iv) *For $a, b, c \in \mathbb{Z}$ define $(K^2 \times K^2)$ matrices*

$$\begin{aligned} \tau_{a,b,c} &= \mathbb{E}\left(vec(u_t u_{t-a}')vec(u_{t-b}u_{t-c}')'\right), & (18)\\ \nu_{a,b,c} &= \mathbb{E}\left(vec(m_t u_{t-a}')vec(u_{t-b}u_{t-c}')'\right), & (19)\\ \zeta_{a,b,c} &= \mathbb{E}\left(vec(m_t u_{t-a}')vec(m_{t-b}u_{t-c}')'\right), & (20) \end{aligned}$$

use $\widetilde{K} = K(K+1)/2$ and assume that the $(K^2m + \widetilde{K} + Kr \times K^2m + \widetilde{K} + Kr)$ matrix Ω_m defined in Equation (B.4) exists and is eventually positive definite for sufficiently large $m \in \mathbb{N}$.

Instead of the common iid assumption for the white noise process $(u_t, t \in \mathbb{Z})$, the less restrictive mixing condition in Assumption 2.1 covers a large class of uncorrelated, but possibly dependent stationary innovation processes, allowing, among other forms of weak white noise, for example, for conditional heteroskedasticity. In addition, the proxy variables $(m_t, t \in \mathbb{Z})$ may show rather general serial dependence that can go far beyond iid-ness or conditional heteroskedasticity.

We state the following central limit theorem (CLT) for our estimators. It is an extension of Theorem 3.1 in Brüggemann, Jentsch, and Trenkler (2016) to the proxy SVAR setup.

Theorem 2.1 (Joint CLT for $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$) *Under Assumption 2.1, we have*

$$\sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} \\ \widehat{\boldsymbol{\sigma}}-\boldsymbol{\sigma} \\ \widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi} \end{pmatrix} \overset{\mathcal{D}}{\to} \mathcal{N}(0,V), \quad V = \begin{pmatrix} V^{(1,1)} & V^{(2,1)\prime} & V^{(3,1)\prime} \\ V^{(2,1)} & V^{(2,2)} & V^{(3,2)\prime} \\ V^{(3,1)} & V^{(3,2)} & V^{(3,3)} \end{pmatrix}$$

where '$\overset{\mathcal{D}}{\to}$' denotes convergence in distribution, with

$$V^{(1,1)} = (\Gamma^{-1}\otimes I_K)\left(\sum_{i,j=1}^{\infty}(C_i\otimes I_K)\sum_{h=-\infty}^{\infty}\tau_{i,h,h+j}(C_j\otimes I_K)'\right)(\Gamma^{-1}\otimes I_K)',$$

$$V^{(2,1)} = L_K\left(\sum_{j=1}^{\infty}\sum_{h=-\infty}^{\infty}\tau_{0,h,h+j}(C_j\otimes I_K)'\right)(\Gamma^{-1}\otimes I_K)',$$

$$V^{(2,2)} = L_K\left(\sum_{h=-\infty}^{\infty}\left\{\tau_{0,h,h} - D_K\boldsymbol{\sigma}\boldsymbol{\sigma}'D_K'\right\}\right)L_K',$$

$$V^{(3,1)} = \left(\sum_{j=1}^{\infty}\sum_{h=-\infty}^{\infty}\nu_{0,h,h+j}(C_j\otimes I_K)'\right)(\Gamma^{-1}\otimes I_K)',$$

$$V^{(3,2)} = \left(\sum_{h=-\infty}^{\infty}\left\{\nu_{0,h,h} - \boldsymbol{\varphi}\boldsymbol{\sigma}'D_K'\right\}\right)L_K',$$

$$V^{(3,3)} = \sum_{h=-\infty}^{\infty}\left\{\zeta_{0,h,h} - \boldsymbol{\varphi}\boldsymbol{\varphi}'\right\}.$$

and L_K is the $(K(K+1)/2 \times K^2)$ elimination matrix such that $vech(A) = L_K vec(A)$ for any $(K \times K)$ matrix A and D_K is the $(K^2 \times K(K+1)/2)$ duplication matrix such that $vec(A) = D_K vech(A)$ for any symmetric $(K \times K)$ matrix A.

Some of the sub-matrices of V simplify if we impose additional structure on the joint process of innovations and proxy variables $x_t = (u_t', m_t')'$. The following corollary summarizes the results of imposing either a martingale difference sequence (mds), which covers, for example, the popular class of GARCH processes to model conditional heteroskedasticity, or an iid structure. We denote the sub-matrices $V^{(i,j)}$ of V under an mds and an iid structure by $V_{mds}^{(i,j)}$ and $V_{iid}^{(i,j)}$, respectively.

Corollary 2.1 (V under Additional Structure)

(i) If in addition to Assumption 2.1, $x_t = (u_t', m_t')'$ is an mds with $\mathbb{E}(x_t|\mathcal{F}_{t-1}) = 0$ a.s., where $\mathcal{F}_{t-1} = \sigma(x_{t-1}, x_{t-2}, \ldots)$, we have $V_{mds}^{(i,j)} = V^{(i,j)}$, $i, j = 2, 3$ and

$$
\begin{aligned}
V_{mds}^{(1,1)} &= (\Gamma^{-1} \otimes I_K) \left(\sum_{i,j=1}^{\infty} (C_i \otimes I_K) \tau_{i,0,j} (C_j \otimes I_K)' \right) (\Gamma^{-1} \otimes I_K)', \\
V_{mds}^{(2,1)} &= L_K \left(\sum_{j=1}^{\infty} \sum_{h=0}^{\infty} \tau_{0,h,h+j} (C_j \otimes I_K)' \right) (\Gamma^{-1} \otimes I_K)', \\
V_{mds}^{(3,1)} &= \left(\sum_{j=1}^{\infty} \sum_{h=0}^{\infty} \nu_{0,h,h+j} (C_j \otimes I_K)' \right) (\Gamma^{-1} \otimes I_K)'.
\end{aligned}
$$

(ii) If in addition to Assumption 2.1, $x_t = (u_t', m_t')'$ are iid, we have $V_{iid}^{(2,1)} = 0$, $V_{iid}^{(3,1)} = 0$ and

$$
\begin{aligned}
V_{iid}^{(1,1)} &= (\Gamma^{-1} \otimes I_K) \left(\sum_{i=1}^{\infty} (C_i \otimes I_K) \tau_{i,0,i} (C_i \otimes I_K)' \right) (\Gamma^{-1} \otimes I_K)' = (\Gamma^{-1} \otimes \Sigma_u), \\
V_{iid}^{(2,2)} &= L_K \left(\tau_{0,0,0} - D_K \boldsymbol{\sigma} \boldsymbol{\sigma}' D_K' \right) L_K', \\
V_{iid}^{(3,2)} &= \left(\nu_{0,0,0} - \boldsymbol{\varphi} \boldsymbol{\sigma}' D_K' \right) L_K', \\
V_{iid}^{(3,3)} &= \zeta_{0,0,0} - \boldsymbol{\varphi} \boldsymbol{\varphi}'.
\end{aligned}
$$

We note here that in the cases of α-mixing, mds, and iid, the proxy variables depend via $u_t = H\epsilon_t$ on the process (ϵ_t). This includes DGPs where the proxy variables are censored to zero, including the DGP proposed by Mertens and Ravn (2013):

$$
m_t = d_t(\Pi \epsilon_t^{(1)} + v_t). \tag{21}
$$

Here, $(d_t, t \in \mathbb{Z})$ is a sequence of scalar dummy variables taking values in $\{0, 1\}$, $(v_t, t \in \mathbb{Z})$ is an r-dimensional white noise process, and Π is an $(r \times r)$ matrix. Further, Mertens and

Ravn (2013) assume $\mathbb{E}(v_t \epsilon_t^{(1)\prime}) = 0$ and $\mathbb{E}(d_t v_t \epsilon_t^{(1)\prime}) = 0$. In particular, this DGP in (21) is covered by Assumption 2.1 if the joint process $((\epsilon_t', v_t', d_t')', t \in \mathbb{Z})$ is strictly stationary and fulfills mixing and moment conditions corresponding to Assumption 2.1(ii). If the process $((\epsilon_t', v_t', d_t')', t \in \mathbb{Z})$ is iid, the sequences $(m_t, t \in \mathbb{Z})$ and $(x_t, t \in \mathbb{Z})$ will also be iid and part (ii) of Corollary 2.1 applies.

The limiting variances in Corollary 2.1 and particularly, without any simplifying assumptions, those in Theorem 2.1 can be very complicated. Hence, bootstrap methods for inference are generally desired and very popular in practice. Before we discuss suitable residual-based bootstrap methods for inference in the next section, we provide explicit limiting results for IRFs and FEVDs in proxy SVARs for $r = 1$ based on the Delta method.

Corollary 2.2 (CLTs for IRFs and FEVDs) *Let $r = 1$. Under Assumption 2.1, for any $s \in \mathbb{R}$, $j, m \in \{1, \dots, K\}$ and $i, h \in \{0, 1, 2, \dots\}$, we have*

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

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

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

where the exact formulas for the limiting variances can be found in Equations (B.12), (B.13), and (B.14), respectively.

Remark 2.1 *Equations (B.12) and (B.13) 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 $(s/(e_m' H^{(1)}))^2$. This is a natural scaling as $s/(e_m' H^{(1)})$ 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 $(I_K - H^{(1)} e_m' / (e_m' H^{(1)}))$ 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 (9), (10), and (15) 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 $\boldsymbol{\sigma}$.*

3 Bootstrap Inference for Proxy SVARs

In this section, we study bootstrap algorithms for inference for proxy SVARs. We focus on inference for statistics that are functions of estimators of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$. These statistics include the estimate $H^{(1)}$ as well as estimates of IRFs and FEVDs.

This section proceeds as follows. In Section 3.1 we describe two residual-based bootstrap algorithms that have been proposed for inference in the SVAR literature. The first is the residual-based recursive-design wild bootstrap. The second is the residual-based MBB proposed by Brüggemann, Jentsch, and Trenkler (2016), which we modify to include moving blocks of the proxy variables. In Section 3.2, we prove the invalidity of the wild bootstrap and establish the asymptotic validity of the MBB.

3.1 Bootstrap Algorithms

3.1.1 Residual-based Wild Bootstrap

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 $\widehat{\Psi H^{(1)\prime}}^+ = T^{-1} \sum_{t=1}^T m_t^+ \widehat{u}_t^{+\prime}$.

6. Use $\widehat{A}_1^+, \ldots, \widehat{A}_p^+$, $\widehat{\Sigma}_u^+$ and $\widehat{\Psi H^{(1)\prime}}^+$ to produce the bootstrapped statistics of interest.

Repeat the algorithm a large number of times and collect the bootstrapped statistics of interest. We produce our confidence intervals with a standard percentile interval by sorting the bootstrapped statistics of interest and keeping the $\alpha/2$- and $1-\alpha/2$-percentiles as the confidence interval, where α is the level of significance. We use these intervals because they are the most popular in the proxy SVAR literature. Using Hall's percentile intervals (Hall (1992) and Lütkepohl (2005, Appendix D)) does not change the asymptotic results below.[7]

As discussed above, the Rademacher distribution where $\eta_t = 1$ with probability 0.5 and $\eta_t = -1$ with probability 0.5 is most common in the proxy SVAR literature. Because of this popularity, we focus on the Rademacher distribution in this paper. Another option is to draw $(\eta_t, t \in \mathbb{Z})$ from a standard normal distribution. We provide Monte Carlo results with the normal distribution in the appendix.

3.1.2 Residual-based Moving Block Bootstrap

The algorithm for the residual-based MBB is as follows. First, to initialize the algorithm, we choose a block length ℓ and compute $N = [T/\ell]$, where $[\cdot]$ rounds up to the nearest integer so that $N\ell \geq T$. Next, collect the $K \times \ell$ blocks $\mathcal{U}_i = (\widehat{u}_i, \dots, \widehat{u}_{i+\ell-1})$ for $i = 1, \dots T-\ell+1$ and the $r \times \ell$ blocks $\mathcal{M}_i = (m_i, \dots, m_{i+\ell-1})$ for $i = 1, \dots T-\ell+1$. Then,

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

2. Collect the blocks $(\mathcal{U}_{i_1}, \dots, \mathcal{U}_{i_N})$ and $(\mathcal{M}_{i_1}, \dots, \mathcal{M}_{i_N})$ and drop the last $N\ell - T$ elements to produce $(\widetilde{u}_1^*, \dots, \widetilde{u}_T^*)$ and $(\widetilde{m}_1^*, \dots, \widetilde{m}_T^*)$. Here, we use " * " to denote bootstrap quantities obtained from the MBB.

3. Center $(\widetilde{u}_1^*, \dots, \widetilde{u}_T^*)$ according to

$$u_{j\ell+s}^* = \widetilde{u}_{j\ell+s}^* - \frac{1}{T-\ell+1} \sum_{r=1}^{T-\ell} \widehat{u}_{s+r-1} \tag{22}$$

   for $s = 1, \dots, \ell$ and $j = 0, 1, \dots, N-1$ in order to produce $(u_1^*, \dots, u_T^*)$ to assure that all u_t^*'s are centered conditionally on the data.

[7] We provide Monte Carlo results for Hall's percentile intervals in the online supplemental appendix. As in Kilian (1999), we find that Hall's percentile intervals are not systematically better than standard percentile intervals for inference on IRFs. Rather, the coverage rates from the two intervals are usually similar for IRFs. For FEVDs, coverage rates from Hall's intervals are often worse than those from standard percentile intervals.

4. Center $(\widetilde{m}_1^*, \ldots, \widetilde{m}_T^*)$ similarly to the VAR errors in (22) to produce $(m_1^*, \ldots, m_T^*)$.[8]

5. 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

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

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

7. 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 $\widehat{\Psi H^{(1)\prime}}^* = T^{-1} \sum_{t=1}^T m_t^* \widehat{u}_t^{*\prime}$.

8. Use $\widehat{A}_1^*, \ldots, \widehat{A}_p^*$, $\widehat{\Sigma}_u^*$ and $\widehat{\Psi H^{(1)\prime}}^*$ to produce the bootstrapped statistics of interest.

As with the wild bootstrap, repeat the algorithm a large number of times, collect the bootstrap statistics, and produce confidence intervals with a standard percentile interval.

This algorithm is similar to the residual-based MBB studied in Brüggemann, Jentsch, and Trenkler (2016). In order to apply it to the proxy SVAR method, we added the re-sampling and centering of the proxy variables along with the computing of $\widehat{\Psi H^{(1)\prime}}^*$.

We will establish the asymptotic validity of this MBB in the next subsection. However, there is one potential issue with the MBB in small samples. If a large number of the observations of m_t are censored to zero, as in Mertens and Ravn (2013), for example, then $(m_1^*, \ldots, m_t^*)$ might contain only zeros. In contrast, it will never be the case that $(m_1^*, \ldots, m_t^*)$ contains only zeros with the wild bootstrap method. However, as discussed in Jentsch and Lunsford (2019), this is not a relevant issue in practice.

3.2 Asymptotic Bootstrap Theory for Proxy SVARs

We next study the asymptotic properties of the bootstrap algorithms described in the previous subsection. To derive the theory, we make the following additional assumption.

Assumption 3.1 (cumulants) *The $(K+r)$-dimensional process $(x_t, t \in \mathbb{Z})$ (as defined in Assumption 2.1) has absolutely summable cumulants up to order eight. More precisely, we*

[8]When the proxy variables have some censoring, we only apply the centering to the non-censored observations and leave the censored proxy variables with a value of zero.

have for all $j = 2, \ldots, 8$ *and* $a_1, \ldots, a_j \in \{1, \ldots, K\}$, $\mathbf{a} = (a_1, \ldots, a_j)$ *that*

$$\sum_{h_2,\ldots,h_j=-\infty}^{\infty} |\mathrm{cum}_{\mathbf{a}}(0, h_2, \ldots, h_j)| < \infty \tag{23}$$

holds, where $\mathrm{cum}_{\mathbf{a}}(0, h_2, \ldots, h_j)$ *denotes the* j*th joint cumulant of* $x_{0,a_1}, x_{h_2,a_2}, \ldots, x_{h_j,a_j}$*; see, for example, Brillinger (1981). This condition includes the existence of eight moments of* $(x_t, t \in \mathbb{Z})$*.*

Such a condition has been imposed in Gonçalves and Kilian (2007) to prove the consistency of wild and pairwise bootstrap methods for univariate AR(∞) processes and in Brüggemann, Jentsch, and Trenkler (2016) to prove the consistency of a residual-based moving block bootstrap for VAR(p) models. In terms of α-mixing conditions, Assumption 3.1 is implied by

$$\sum_{n=1}^{\infty} n^{m-2}(\alpha_x(n))^{\delta/(2m-2+\delta)} < \infty \tag{24}$$

for $m = 8$ if all moments up to order eight of $(x_t, t \in \mathbb{Z})$ exist; see Künsch (1989). For example, the popular class of generalized autoregressive conditional heteroskedastic (GARCH) processes is geometrically strong mixing under mild assumptions on the conditional distribution such that the summability condition in Equation (24) always holds.

3.2.1 Inconsistency of the Wild Bootstrap

In this section, we show that the wild bootstrap is generally not consistent for inference on $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$ and, consequently, also for statistics that are functions of these estimators. Define $\widehat{\boldsymbol{\beta}}^+ = \mathrm{vec}(\widehat{A}_1^+, \ldots, \widehat{A}_p^+)$, $\widehat{\boldsymbol{\sigma}}^+ = \mathrm{vech}(\widehat{\Sigma}_u^+)$, and $\widehat{\boldsymbol{\varphi}}^+ = \mathrm{vec}(\widehat{\Psi H^{(1)\prime}}^{+})$ to be the estimators from the wild bootstrap that correspond to $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, respectively. We derive the joint limiting variance of $\sqrt{T}((\widehat{\boldsymbol{\beta}}^+ - \widehat{\boldsymbol{\beta}})', (\widehat{\boldsymbol{\sigma}}^+ - \widehat{\boldsymbol{\sigma}})', (\widehat{\boldsymbol{\varphi}}^+ - \widehat{\boldsymbol{\varphi}})')'$ in the following theorem.

Theorem 3.1 (Residual-based Wild Bootstrap Limiting Variance) *Suppose Assumptions 2.1 and 3.1 hold and the residual-based wild bootstrap from Section 3.1.1 is used to compute bootstrap statistics* $\widehat{\boldsymbol{\beta}}^+, \widehat{\boldsymbol{\sigma}}^+$ *and* $\widehat{\boldsymbol{\varphi}}^+$*. Then, we have*

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

As $V_{WB} \neq V$ for V as defined in Theorem 2.1, a consequence of Theorem 3.1 is that the residual-based wild bootstrap is generally inconsistent for statistics that are functions of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$. However, as $V_{WB}^{(1,1)} = V_{mds}^{(1,1)}$ holds, the only exclusion is the case where the statistic of interest is a (smooth) function of $\widehat{\boldsymbol{\beta}}$ only under an additional mds assumption; compare also Corollary 2.1.[9] The latter framework 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 $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\sigma}}$ 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 2.1(ii).

If the 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 3.1 (Residual-based Rademacher Wild Bootstrap Limiting Variance) *Under the assumptions of Theorem 3.1 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*

$$V_{WB} = \begin{pmatrix} V_{mds}^{(1,1)} & O_{K^2p \times \widetilde{K}+Kr} \\ O_{\widetilde{K}+Kr \times K^2p} & O_{\widetilde{K}+Kr \times \widetilde{K}+Kr} \end{pmatrix}. \tag{25}$$

A comparison of V_{WB} in Equation (25) with V from Theorem 2.1 leads to the conclusion that a considerable amount of estimation uncertainty caused by estimating Σ_u and $\Psi H^{(1)\prime}$ 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 4, 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 $\Psi H^{(1)\prime}$, 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 $\Psi H^{(1)\prime}$ from Equation

[9] 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 $i, j = 1, 2$.

(14) is given by

$$\widehat{\Psi H^{(1)\prime}}^{+} = T^{-1}\sum_{t=1}^{T} m_t^{+} u_t^{+\prime}.$$

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

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

That is, when u_t is directly observable, the Rademacher wild bootstrap estimator is simply the non-bootstrapped sample estimate and $\widehat{\Psi H^{(1)\prime}}^{+} = \widehat{\Psi H^{(1)\prime}}$ holds for every bootstrap replication. This implies that the uncertainty in the estimation of the covariance $\Psi H^{(1)\prime}$ 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 $\widehat{H}^{(1)+} = \widehat{H}^{(1)}$ holds exactly, but $\widehat{H}^{(1)+} = \widehat{H}^{(1)} + o_{P^+}(1)$ as $T \to \infty$, where P^+ denotes the probability measure induced by the Rademacher wild bootstrap. Hence, although the bootstrapped variance of $H^{(1)}$ will generally not be zero in finite samples with the Rademacher wild bootstrap, it will converge to zero as the sample size increases.

3.2.2 Consistency of the Moving Block Bootstrap

Next, we show that the MBB can approximate the limiting distribution of $\sqrt{T}((\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})', (\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})', (\widehat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})')'$ derived in Theorem 2.1. We define $\widehat{\boldsymbol{\beta}}^* = \text{vec}(\widehat{A}_1^*, \ldots, \widehat{A}_p^*)$, $\widehat{\boldsymbol{\sigma}}^* = \text{vech}(\widehat{\Sigma}_u^*)$, and $\widehat{\boldsymbol{\varphi}}^* = \text{vec}(\widehat{\Psi H^{(1)\prime}}^{*})$ to be the estimators from the MBB that correspond to $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, respectively. We get the following theorem.

Theorem 3.2 (Residual-based MBB Consistency) *Suppose Assumptions 2.1 and 3.1 hold and the residual-based MBB from Section 3.1.2 is used to compute bootstrap statistics $\widehat{\boldsymbol{\beta}}^*, \widehat{\boldsymbol{\sigma}}^*$ and $\widehat{\boldsymbol{\varphi}}^*$. If $\ell \to \infty$ such that $\ell^3/T \to 0$ as $T \to \infty$, we have*

$$\sup_{x\in\mathbb{R}^{\bar{K}}}\left|P^*\left(\sqrt{T}\left((\widehat{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}})', (\widehat{\boldsymbol{\sigma}}^*-\widehat{\boldsymbol{\sigma}})', (\widehat{\boldsymbol{\varphi}}^*-\widehat{\boldsymbol{\varphi}})'\right)' \le x\right) - P\left(\sqrt{T}\left((\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})', (\widehat{\boldsymbol{\sigma}}-\boldsymbol{\sigma})', (\widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi})'\right)' \le x\right)\right| \to 0$$

in probability, where P^* denotes the probability measure induced by the residual-based MBB and $\bar{K} = K^2p + (K^2+K)/2 + Kr$. The shorthand $x \le y$ for some $x, y \in \mathbb{R}^d$ is used to *denote $x_i \leq y_i$ for all $i = 1, \ldots, d$.*

As defined below Equation (8), $\Theta_{jk,i}$ denotes the one standard deviation IRF of the the j-th variable to the k-th structural shock of interest that occurred i periods ago for $j = 1, \ldots, K$ and $k = 1, \ldots, r$. $s\Theta_{jk,i}/(e_l' H e_n)$ is the corresponding normalized IRF, and $\omega_{jk,i}$ is the fraction of the forecast error variance of $y_{j,t+i}$ attributable to structural shock k. To simplify notation we suppress the subscripts in the following and simply use Θ and $\widehat{\Theta}$ to represent a specific (one standard deviation or normalized) impulse response coefficient and its estimator, respectively. We also use ω and $\widehat{\omega}$ to denote a specific forecast error variance decomposition and its estimator, respectively. One standard deviation IRFs, normalized IRFs, and FEVDs are continuously differentiable functions of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$. Hence, the asymptotic validity of the residual-based MBB scheme to construct confidence intervals for the IRFs and FEVDs in the proxy SVAR framework is implied by Theorem 3.1 and the Delta method to get the following corollary.

Corollary 3.2 (Asymptotic Validity of Bootstrap IRFs and FEVDs in proxy SVARs) *Under Assumptions 2.1 and 3.1 and if $\ell \to \infty$ such that $\ell^3/T \to 0$ as $T \to \infty$, we have*

$$(i) \quad \sup_{x \in \mathbb{R}} \left| P^* \left(\sqrt{T} \left(\widehat{\Theta}^* - \widehat{\Theta} \right)' \leq x \right) - P \left(\sqrt{T} \left(\widehat{\Theta} - \Theta \right)' \leq x \right) \right| \to 0,$$

$$(ii) \quad \sup_{x \in \mathbb{R}} \left| P^* \left(\sqrt{T} \left(\widehat{\omega}^* - \widehat{\omega} \right)' \leq x \right) - P \left(\sqrt{T} \left(\widehat{\omega} - \omega \right)' \leq x \right) \right| \to 0$$

in probability, respectively.

4 Monte Carlo Simulations

To study the wild bootstrap and the MBB, we use Monte Carlo simulations with three different DGPs: two with iid innovations and one with innovations that follow GARCH(1,1) processes. In all three DGPs, we simulate y_t with the bivariate VAR(2) process

$$y_t = \begin{bmatrix} 0.44 & 0.66 \\ -0.11 & 1.32 \end{bmatrix} y_{t-1} + \begin{bmatrix} -0.18 & 0 \\ -0.18 & -0.09 \end{bmatrix} y_{t-2} + u_t, \quad \mathbb{E}(u_t u_t') = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}. \tag{26}$$

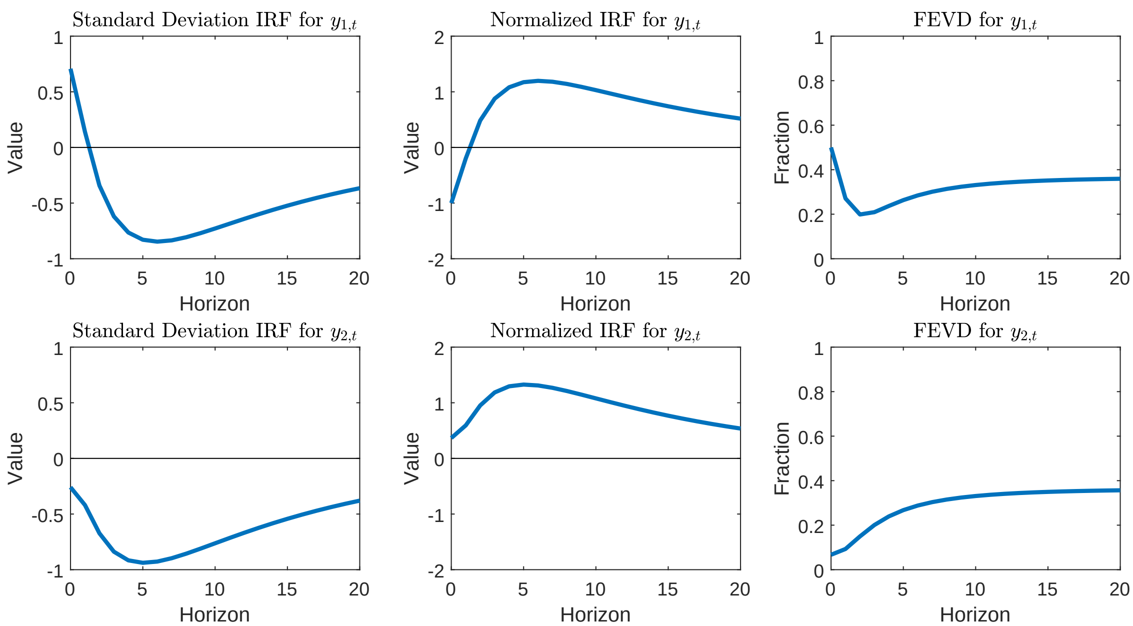

Figure 1: The top panels show the IRFs and FEVDs of $y_{1,t}$ generated from Equations (26) and (27). For normalized IRFs, $y_{1,t}$ is normalized to fall by 1 on impact. The bottom panels show the corresponding IRFs and FEVDs of $y_{2,t}$.

To simulate the VAR innovations, u_t follows Equation (2) with

$$H = \begin{bmatrix} 0.707 & 0.707 \\ -0.259 & 0.966 \end{bmatrix}. \tag{27}$$

In the simulations, $\epsilon_t^{(1)}$ is the structural shock of interest so that the structural IRFs are produced from the first column of H. This implies that $r = 1$ so that the proxy variable is a scalar. To simulate the proxy variable, we use $m_t = \Psi\epsilon_t^{(1)} + v_t$ with $v_t \sim \mathcal{N}(0, 1)$. This corresponds to model (21) without censoring.[10]

Equation (26) is similar to the DGP in Brüggemann, Jentsch, and Trenkler (2016), and a shock to $\epsilon_t^{(1)}$ produces persistent hump-shaped IRFs that are common in macroeconomic studies. Figure 1 shows the one standard deviation IRFs for each element of y_t, denoted by $y_{1,t}$ and $y_{2,t}$, out to a horizon of 20. It also shows IRFs normalized so that $y_{1,t}$ falls by 1

[10]We provide Monte Carlo simulations with censored proxy variables in the online supplemental appendix. In general, we find that the MBB's coverage rates are very similar with and without censoring. This is consistent with Jentsch and Lunsford (2019), who have Monte Carlo simulations with censoring for a different DGP. We note that for smaller values of Ψ, censoring may cause the MBB's coverage rates to fall slightly.

on impact and the FEVDs for each element of y_t. In the simulations below, we will study coverage rates of confidence intervals for each of these six objects. An important difference from Brüggemann, Jentsch, and Trenkler (2016) is that our impact matrix in Equation (27) is not triangular and cannot be estimated with a Cholesky decomposition. Rather, we show below that the proxy SVAR methodology with the MBB produces good coverage rates, especially at horizon 0, when the IRFs and FEVDs are only functions of H.

In the iid simulations, each element of ϵ_t is an independent standard normal random variable. Our two DGPs with iid innovations have two different covariances between m_t and $\epsilon_t^{(1)}$. In DGP1, we set $\Psi = 0.5$. In DGP2, we set $\Psi = 0.2$.

DGP3 is a GARCH(1,1) simulation. The elements $\epsilon_t^{(1)}$ and $\epsilon_t^{(2)}$ of ϵ_t are independent and follow $\epsilon_t^{(i)} = g_t^{(i)} w_t^{(i)}$ and $(g_t^{(i)})^2 = \gamma_0 + \gamma_1(\epsilon_{t-1}^{(i)})^2 + \gamma_2(g_{t-1}^{(i)})^2$, for $i = 1, 2$. Here, $w_t^{(i)}$ for $i = 1, 2$ are independent standard normal random variables, $\gamma_1 = 0.05$, $\gamma_2 = 0.93$, and $\gamma_0 = 1 - \gamma_1 - \gamma_2$. This specification lies between cases G1 and G2 in Brüggemann, Jentsch, and Trenkler (2016). We also set $\Psi = 0.5$.

For each DGP, we produce confidence intervals for IRFs with the residual-based MBB and the residual-based Rademacher wild bootstrap. We run the Monte Carlo simulations with effective sample sizes of 400 and 2000. For the residual-based MBB, we use block lengths of 22 and 34 for the sample sizes 400 and 2000, respectively.[11] Note that fixing a block length of $\ell = 1$ gives an iid bootstrap design that extends Runkle (1987) to the proxy SVAR case. This is sufficient for inference for DGP1 and DGP2 as both rely on iid VAR innovations and proxy variables. Further, this iid bootstrap gains on the MBB in terms of efficiency for DGP1 and DGP2, leading to slightly better results compared to what is shown below for these DGPs. In contrast, $\ell > 1$ is necessary to capture the nonlinear dependence caused by the GARCH processes in DGP3, and an iid bootstrap would be asymptotically invalid. We use the MBB for all three DGPs to highlight that it generally performs well and is never worse than the Rademacher wild bootstrap for both iid and GARCH innovations. Thus, the MBB can be viewed as robust for inference when researchers are unsure if their VAR innovations are truly iid or may have some nonlinear dependence.

For each simulation, we draw ϵ_t and compute u_t of sample size $T + 1000$, where T is the relevant effective sample size. In DGP1 and DGP2, we use $y_0 = y_{-1} = 0$ and u_t to recursively generate y_t for $t = 1, \ldots, T + 1000$. In DGP3, we use $y_0 = y_{-1} = 0$, $(g_0^{(i)})^2 = 1$ for $i = 1, 2$, and $(\epsilon_0^{(i)})^2 = 1$ for $i = 1, 2$ to recursively generate ϵ_t and y_t for $t = 1, \ldots, T + 1000$. We then

[11] From Theorem 3.2, we need $\ell \to \infty$ and $\ell^3/T \to 0$ as $T \to \infty$. Following Jentsch and Lunsford (2016, 2019), we use the rule $\ell = \kappa T^{1/4}$ with $\kappa = 5.03$, which normalizes $\ell = 20$ when $T = 250$.

drop the first 998 observations of y_t to get a sample of length T plus two pre-sample values. For each DGP, we produce 1000 simulations and use 2000 bootstrap replications. Then, we compute the coverage rate of a confidence interval to be the fraction of simulations where the true IRF or FEVD lies within the confidence interval. We show coverage rates for 95 percent confidence intervals. Coverage rates for 68 percent confidence intervals are in the online supplemental appendix.

4.1 Results for DGP1

Figure 2 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for the one standard deviation IRFs under DGP1. All panels of Figure 2 show that the coverage rates of the MBB are close to the intended levels on impact, displayed as horizon 0. In contrast, the coverage rates for the wild bootstrap are much too small on impact. They are 0.2 or less for $T = 400$. For $T = 2000$, these rates shrink to less than 0.08. Because the IRF at horizon 0 is just $H^{(1)}$, which is a function of $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$ but not $\boldsymbol{\beta}$, this shrinking of the coverage rates is consistent with Corollary 3.1.

For IRF horizons 1 through 20, Figure 2 shows that the MBB's confidence intervals gradually become undersized. However, this is much less of a problem with large sample sizes. Conversely, the coverage rates of the wild bootstrap's confidence intervals gradually rise and become similar to those of the MBB. Under each of our DGPs, the wild bootstrap is asymptotically valid for inference on the VAR slope coefficients, $\boldsymbol{\beta}$. Hence, Figure 2 shows that the uncertainty about the covariances, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, dies out as the IRF horizon increases under DGP1 and the uncertainty about the slope coefficients becomes more relevant. Jentsch and Lunsford (2019) also discuss this. Kilian and Lütkepohl (2017, p. 341) raise a similar point for IRFs that are functions of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$, such as in Cholesky-identified SVARs.

While some papers in the proxy SVAR literature use one standard deviation IRFs (Gertler and Karadi, 2015), others use normalized IRFs (Mertens and Ravn, 2013). Figure 3 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for normalized IRFs under DGP1. We reiterate that $y_{1,t}$ is normalized to fall by 1 on impact. We make this normalization within every bootstrap loop, causing the coverage rates for $y_{1,t}$ to be 1 at horizon 0 for both the MBB and the wild bootstrap. For $y_{2,t}$, the MBB is appropriately sized on impact. However, the wild bootstrap is very undersized on impact for $y_{2,t}$, and the coverage rates of its confidence intervals shrink when the sample size increases.

For horizons 1 through 20 with $T = 400$, the coverage rates for the MBB fall. However,

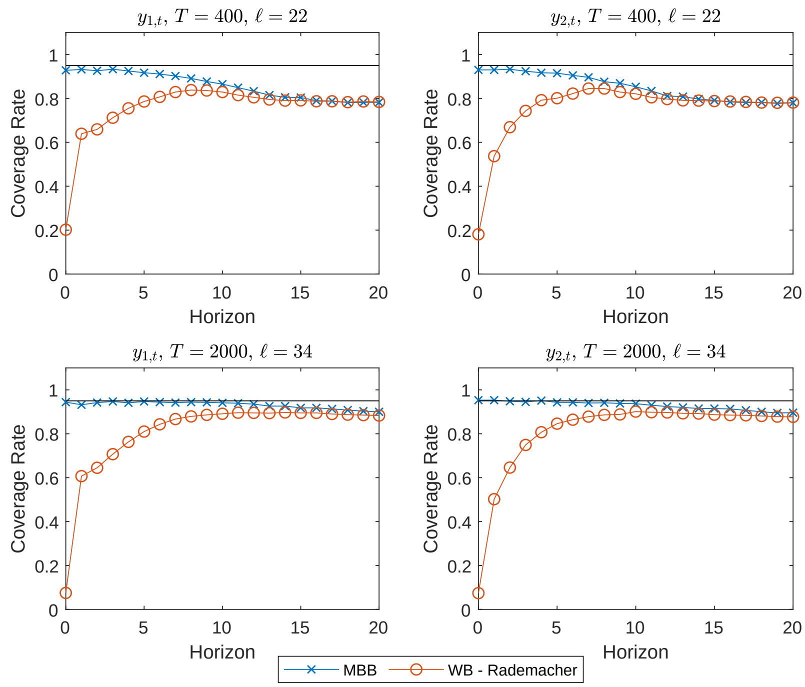

Figure 2: Coverage rates of 95 percent confidence intervals for one standard deviation IRFs under DGP1. The solid horizontal line shows the 0.95 target level.

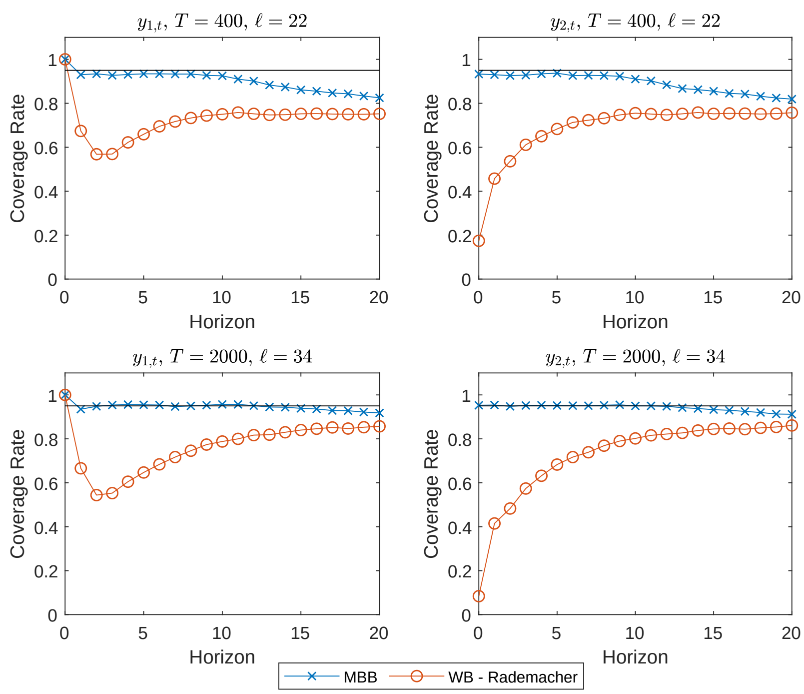

Figure 3: Coverage rates of 95 percent confidence intervals for normalized IRFs under DGP1. The solid horizontal line shows the 0.95 target level.

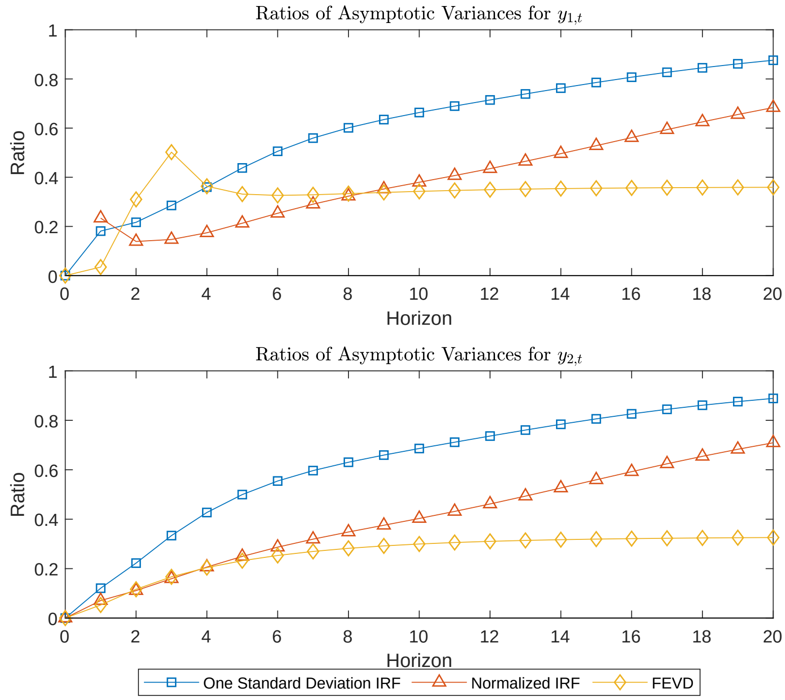

Figure 4: Ratios of the asymptotic variances of bootstrapped one standard deviation IRFs, normalized IRFs, and FEVDs using the Rademacher wild bootstrap to the true asymptotic variances of one standard deviation IRFs, normalized IRFs, and FEVDs under DGP1.

this fall is less noticeable at earlier horizons than for one standard deviation IRFs. For $T = 2000$, the MBB performs well. In contrast, the wild bootstrap's confidence intervals are undersized at both sample sizes. Further, unlike with the one standard deviation IRFs, the wild bootstrap's coverage rates do not converge with those of the MBB by horizon 20.

To better understand the wild bootstrap's differing coverage rates for one standard deviation IRFs and normalized IRFs, we use the results in Corollaries 2.2 and 3.1 to compute the asymptotic variances of bootstrapped one standard deviation IRFs and normalized IRFs that are produced by the Rademacher wild bootstrap, denoted by $\Sigma_{\widehat{\Theta}^{+}_{\bullet 1,i}}$ and $\Sigma_{\widehat{\Xi}^{+}_{\bullet 1,i}(s;m,1)}$, respectively, under DGP1. The top panel of Figure 4 shows the ratios of the (1,1) elements of $\Sigma_{\widehat{\Theta}^{+}_{\bullet 1,i}}$ and $\Sigma_{\widehat{\Xi}^{+}_{\bullet 1,i}(s;m,1)}$ to the (1,1) elements of the true asymptotic variances, $\Sigma_{\widehat{\Theta}_{\bullet 1,i}}$ and $\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$, at each horizon. The bottom panel displays the ratios of the (2,2) elements of these matrices.[12] Figure 4 shows that the Rademacher wild bootstrap implies zero asymptotic variance for one standard deviation IRFs and for the normalized IRF of $y_{2,t}$ at horizon

[12]Corollary 3.2 implies $\Sigma_{\widehat{\Theta}^{*}_{\bullet 1,i}} = \Sigma_{\widehat{\Theta}_{\bullet 1,i}}$ and $\Sigma_{\widehat{\Xi}^{*}_{\bullet 1,i}(s;m,1)} = \Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$ for the MBB. Hence, these same ratios for the MBB would all be 1, indicating its bootstrap consistency, and we do not display them.

zero.[13] This is consistent with the Rademacher wild bootstrap's very small coverage rates at horizon zero in Figures 2 and 3. As the horizon increases, the ratios in Figure 4 typically increase. This result is consistent with the uncertainty about the covariances dying out as the IRF horizon increases, discussed by Jentsch and Lunsford (2019) and Kilian and Lütkepohl (2017, p. 341). However, the ratios in Figure 4 are typically smaller for the normalized IRF than for the one standard deviation IRF. Hence, the reason that the Rademacher wild bootstrap's coverage rates do not converge with the MBB's for the normalized IRFs, while they do for the one standard deviation IRFs, is that the asymptotic variance implied by the Rademacher wild bootstrap is persistently smaller for the normalized IRF as a fraction of the true asymptotic variance. As discussed in Remark 2.1, switching from a one standard deviation IRF to a normalized IRF changes how the VMA coefficients interact with the variances and covariances of $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$. For DGP1, this implies that uncertainty about these covariances dies out more slowly for normalized IRFs than for one standard deviation IRFs.

Figure 4 also shows the ratios of the asymptotic variances of the FEVDs implied by the Rademacher wild bootstrap to the true asymptotic variances, with the top panel showing $\Sigma_{\widehat{\omega}^+_{11,h}}/\Sigma_{\widehat{\omega}_{11,h}}$ and the bottom panel showing $\Sigma_{\widehat{\omega}^+_{21,h}}/\Sigma_{\widehat{\omega}_{21,h}}$. Consistent with the one standard deviation IRFs, the Rademacher wild bootstrap implies zero asymptotic variance for the FEVDs at horizon 0. Further, the ratios for the FEVDs are typically lower than for the one standard deviation IRFs and become lower than the ratios for the normalized IRFs at long horizons. Because the FEVDs are functions of cumulative sums, uncertainty about the covariances, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, remains even as the horizon increases and causes the asymtptotic variances implied by the wild bootstrap to be persistently too small. Figure 4 shows an exception to these general patterns for $y_{1,t}$ for horizons 2 through 4. Comparing this exception to Figure 1, we see that the FEVD has a hump-shaped trough for $y_{1,t}$ at these horizons, causing the wild bootstrap's underestimation of the asymptotic variance to be less severe.

Figure 5 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for the FEVDs under DGP1. All panels of Figure 5 show that the coverage rates of the MBB are generally close to the intended levels, but may be slightly too small with smaller sample sizes. In contrast, all panels of Figure 5 show that the coverage rates of the wild bootstrap are much too small. Similar to the IRFs, the coverage rates for the wild bootstrap are smallest at horizon 0 when the FEVDs are only functions of $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$ and

[13] The normalized IRF for $y_{1,t}$ at $i = 0$ has an asymptotic variance of 0 because it is always normalized equal to -1. This also gives it an asymptotic variance of 0 under the wild bootstrap. Hence, the ratio here is 0/0, which we do not display in Figure 4.

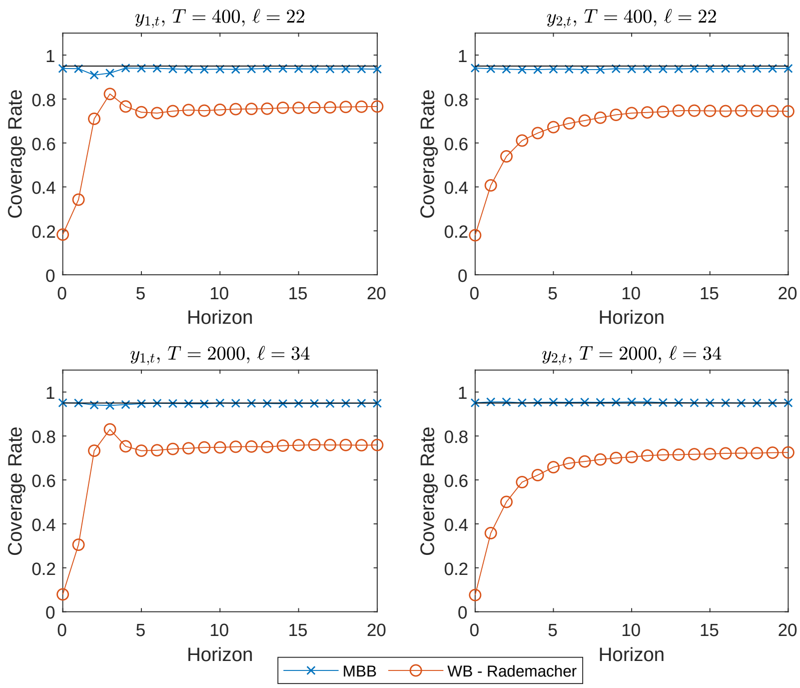

Figure 5: Coverage rates of 95 percent confidence intervals for FEVDs under DGP1. The solid horizontal line shows the 0.95 target level.

shrink when the sample size increases. Consistent with the ratios of the asymptotic variances in Figure 4, the coverage rates of the wild bootstrap generally rise with the horizon, but do not converge to the MBB at a horizon of 20.

4.2 Results for DGP2

In proxy SVARs, IRFs and FEVDs are functions of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$. Hence, coverage rates of confidence intervals from bootstrap algorithms can depend on all three parameters. This is different than in other bootstrap analysis, such as Brüggemann, Jentsch, and Trenkler (2016) and Kilian and Lütkepohl (2017, ch. 12), where IRFs and FEVDs are only functions of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$. To highlight the importance of changes to $\boldsymbol{\varphi}$, we study DGP2. It is the same as DGP1 in all respects, except that the covariance between m_t and $\epsilon_t^{(1)}$, Ψ, is reduced from 0.5 to 0.2, also reducing $\boldsymbol{\varphi}$.

Figure 6 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for the one standard deviation IRFs under DGP2. Relative to DGP1, the proxy variable provides less of a signal for the structural shock of interest. Despite this weaker signal, the MBB provides very similar coverage rates when compared to DGP1 in

Figure 2. On impact, the coverage rates are close to the intended levels. With $T = 400$, the coverage rates gradually decline to about 0.78 at a horizon of 20 under DGP1 and DGP2. Under both DGP1 and DGP2, the MBB stays close to the intended levels with $T = 2000$.

In contrast to DGP1, the wild bootstrap coverage rates do not fully converge with the MBB converge rates under DGP2. With $T = 400$, the coverage rates only rise to about 0.70 at a horizon of 20 under DGP2, while they rise to about 0.78 under DGP1. While the coverage rates for $T = 2000$ rise to slightly higher levels, they remain 0.07 to 0.09 lower than the MBB. This shows that although the uncertainty about the covariances, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, dies out as the IRF horizon increases, it can be quite persistent when the covariance of m_t and $\epsilon_t^{(1)}$ is relatively low. In this case, the estimate of $H^{(1)}$ will be more uncertain because the proxy variable is giving less information about the structural shock of interest. However, the Rademacher wild bootstrap asymptotically ignores the uncertainty around the estimate of $H^{(1)}$, causing its coverage rates to become even more undersized.

Figure 7 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for normalized IRFs under DGP2. As with the normalized IRFs under DGP1, the coverage rates for $y_{1,t}$ are 1 at horizon 0 for both the MBB and the wild bootstrap. Aside from this, the MBB's coverage rates are very close to the target level at low horizons. For $T = 400$, the MBB's coverage rates fall slightly at longer horizons.

In Figure 7, the wild bootstrap's coverage rates stay persistently undersized out to horizon 20, and these rates do not rise above 0.68 when $T = 400$. This suggests that for conventional sample sizes, it is possible that 95 percent confidence intervals from a wild bootstrap will not even give 68 percent coverage for normalized IRFs. This is consistent with the findings of Jentsch and Lunsford (2019). In that paper, we study the effects of tax changes that were originally studied by Mertens and Ravn (2013). We show that many of the results thought to be statistically significant at the 95 percent level were no longer significant at the 68 percent level when using the MBB instead of the wild bootstrap.

To better understand the results in Figures 6 and 7, the top panel of Figure 8 shows the ratios of the (1,1) elements of $\Sigma_{\widehat{\Theta}^{+}_{\bullet 1,i}}$ and $\Sigma_{\widehat{\Xi}^{+}_{\bullet 1,i}(s;m,1)}$ to the (1,1) elements of $\Sigma_{\widehat{\Theta}_{\bullet 1,i}}$ and $\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)}$ at each horizon under DGP2. The bottom panel displays the ratios of the (2,2) elements of these matrices. Compared to Figure 4, Figure 8 shows that the Rademacher wild bootstrap captures even less of the true asymptotic variance under DGP2 than under DGP1. In particular, the ratios for the normalized IRFs in Figure 8 do not rise to half of the levels in Figure 4. Hence, the wild bootstrap's persistently low coverage rates in Figure 7 are not due to small sample sizes. Rather, under DGP2 when the correlation between the

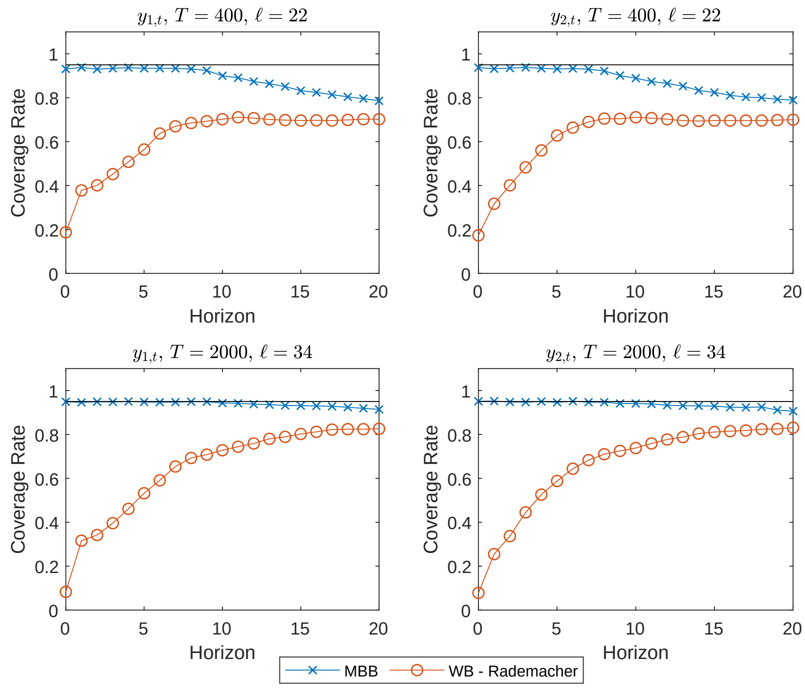

Figure 6: Coverage rates of 95 percent confidence intervals for one standard deviation IRFs under DGP2. The solid horizontal line shows the 0.95 target level.

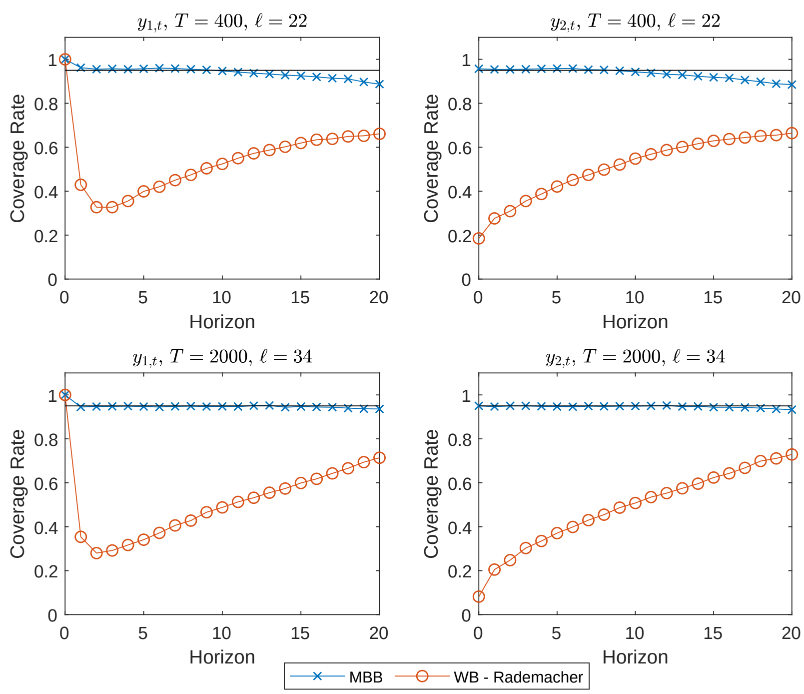

Figure 7: Coverage rates of 95 percent confidence intervals for normalized IRFs under DGP2. The solid horizontal line shows the 0.95 target level.

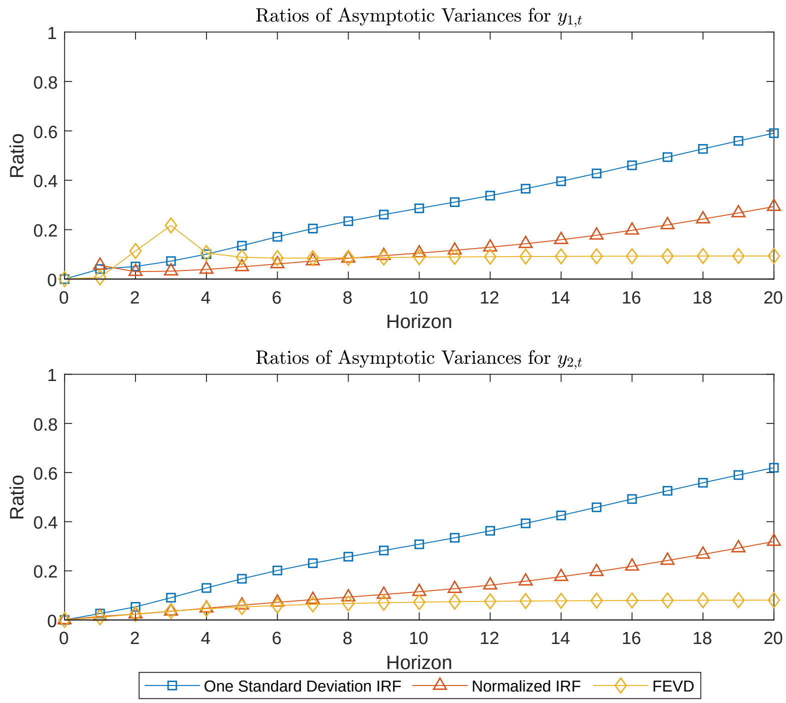

Figure 8: Ratios of the asymptotic variances of bootstrapped one standard deviation IRFs, normalized IRFs, and FEVDs using the Rademacher wild bootstrap to the true asymptotic variances of one standard deviation IRFs, normalized IRFs, and FEVDs under DGP2.

proxy variable and the structural shock of interest is low, the Rademacher wild bootstrap's implied asymptotic variance for normalized IRFs is much too low, even at long horizons.

Figure 9 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for the FEVDs under DGP2. While the MBB's coverage rates for IRFs were quite similar for DGP1 and DGP2, we see that the coverage rates for FEVDs deteriorate slightly under DGP2 relative to DGP1 when $T = 400$. This is particularly true for $y_{1,t}$ at horizons 2 through 4. Comparing this to Figure 1, we see that the FEVD is hump-shaped for $y_{1,t}$ at these horizons, and it appears that the confidence intervals for the MBB struggle to correctly capture the uncertainty around this hump. However, this is just a small sample result and the MBB does much better when $T = 2000$.

Figure 9 shows that the wild bootstrap generally performs worse under DGP2 than under DGP1 in Figure 5. This is consistent with the wild bootstrap's worse performance for the IRFs under DGP2 relative to DGP1. In particular, the coverage rates for $y_{2,t}$ never rise to be even half of the intended levels in Figure 9. This is consistent with the ratios of the asymptotic variances $\Sigma_{\widehat{\omega}^+_{11,h}}/\Sigma_{\widehat{\omega}_{11,h}}$ and $\Sigma_{\widehat{\omega}^+_{21,h}}/\Sigma_{\widehat{\omega}_{21,h}}$ in the top and bottom panels of Figure 8,

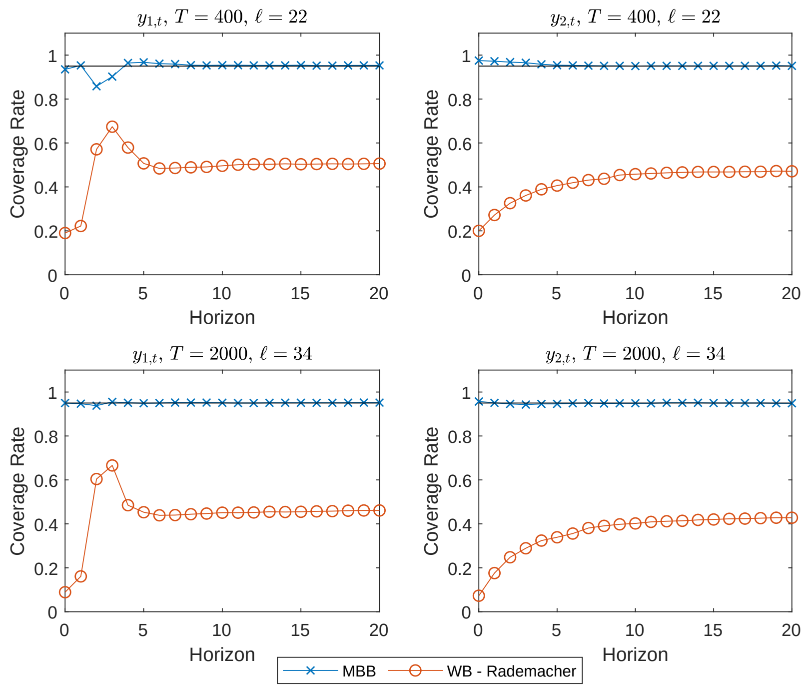

Figure 9: Coverage rates of 95 percent confidence intervals for FEVDs under DGP2. The solid horizontal line shows the 0.95 target level.

respectively. These ratios show that the asymptotic variances implied by the wild bootstrap for FEVDs do not capture even 10 percent of the true asymptotic variances of the FEVDs, except for $y_{1,t}$ at horizons 2 through 4.

Overall, we find that changing Ψ, and hence $\boldsymbol{\varphi}$, causes material changes to the performance of the bootstrap algorithms. For lower covariances between the proxy variables and the structural shocks of interest, the MBB's coverage rates for IRFs do not appear to change much. However, its coverage rates around humps in FEVDs may deteriorate. In contrast, the wild bootstrap's performance can noticeably worsen. The asymptotic variances implied by the wild bootstrap can shrink relative to the true asymptotic variances, causing its confidence intervals to become more persistently undersized.

4.3 Results for DGP3

We conclude our Monte Carlo analysis by studying DGP3, which has conditional heteroskedasticity in both the VAR innovations and the proxy variables. Robustness against conditional heteroskedasticity has been a motivating factor for using wild bootstraps in the proxy SVAR literature (Mertens and Ravn, 2013; Gertler and Karadi, 2015). However, as we proved, this is only true for the VAR slope coefficients under an mds assumption, and not for the covariances of the VAR innovations and the proxy variables. Rather, the MBB is asymptotically valid for all three sets of parameters under conditional heteroskedasticity.

Figure 10 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for the one standard deviation IRFs under DGP3. The coverage rates for the MBB are generally undersized. In contrast to DGP1 and DGP2, the coverage rates may even be undersized on impact. This is consistent with the findings of Brüggemann, Jentsch, and Trenkler (2016), who show that the MBB's coverage can be undersized in small samples for Cholesky-identified SVARs. However, Figure 10 shows a noticeable improvement in coverage rates from $T = 400$ to $T = 2000$. Further, despite its low coverage, the MBB usually performs much better and never worse than the wild bootstrap in terms of statistical size under DGP3. As with the previous DGPs, the wild bootstrap's coverage rates can be much too small, especially at low horizons.

Figure 11 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for normalized IRFs under DGP3. The MBB's coverage rates are good at low horizons, in contrast to its coverage rates for one standard deviation IRFs. While the MBB does become undersized at long horizons for $T = 400$, it performs well for $T = 2000$. In contrast, the wild bootstrap is very undersized at low horizons and persistently undersized at long horizons, consistent with its coverage of the normalized IRFs under DGP1 and DGP2.

Figure 12 shows the confidence interval coverage rates from the MBB and the Rademacher wild bootstrap for the FEVDs under DGP3. The MBB's coverage rates are consistently too low at every horizon with rates as low as 0.79. These rates change little from $T = 400$ to $T = 2000$. Despite these low coverage rates, the MBB consistently dominates the wild bootstrap in terms of statistical size under DGP3. As with the previous DGPs, the wild bootstrap is persistently very undersized.

5 Conclusions

Recent research has shown how to use external instruments or proxy variables to identify the effects of structural shocks in SVARs. Residual-based bootstrap methods, especially the residual-based wild bootstrap, have been popular for providing inference for IRFs from these proxy SVARs. In this paper, we make three contributions to improve the understanding of these residual-based bootstrap methods. First, we provide a joint CLT for the VAR coefficients, the covariance matrix of the VAR innovations, and the covariance matrix of the

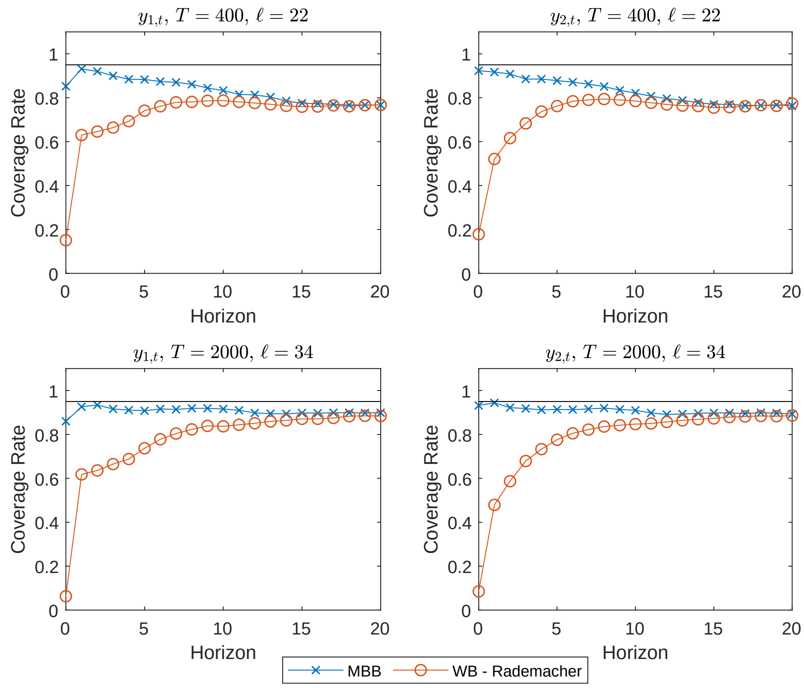

Figure 10: Coverage rates of 95 percent confidence intervals for one standard deviation IRFs under DGP3. The solid horizontal line shows the 0.95 target level.

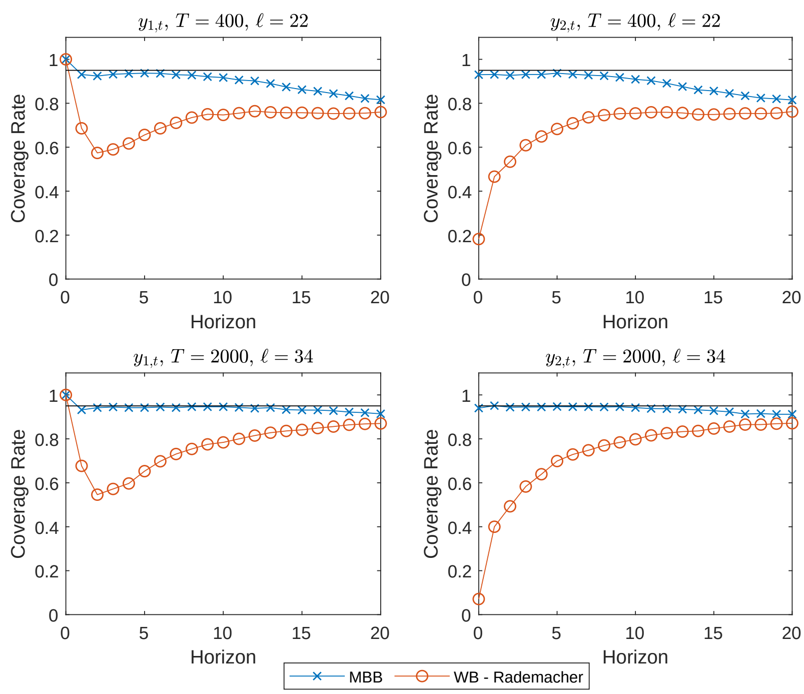

Figure 11: Coverage rates of 95 percent confidence intervals for normalized IRFs under DGP3. The solid horizontal line shows the 0.95 target level.

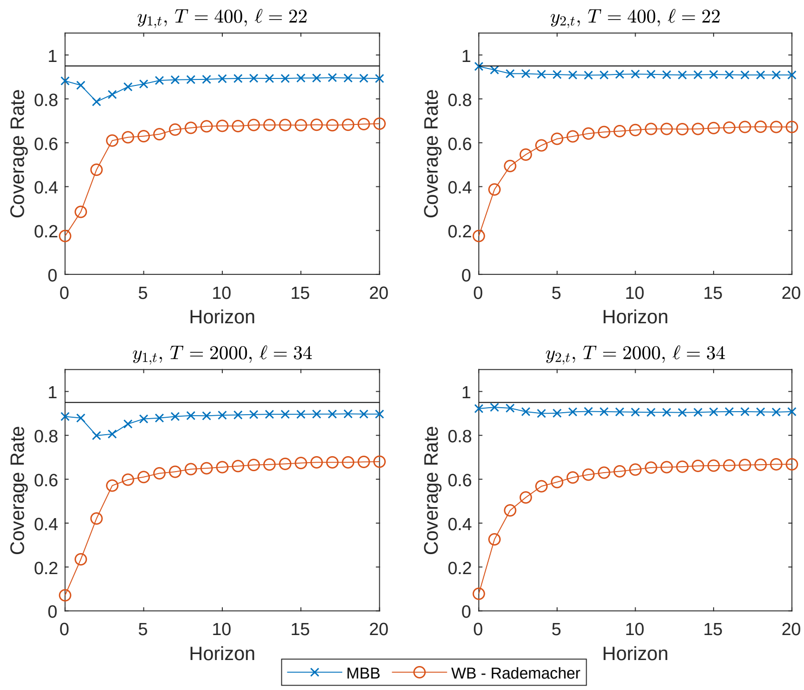

Figure 12: Coverage rates of 95 percent confidence intervals for FEVDs under DGP3. The solid horizontal line shows the 0.95 target level.

VAR innovations with the proxy variables under mild α-mixing conditions that cover a large class of uncorrelated, but possibly dependent innovation processes, including conditional heteroskedasticity. We extend this result to also derive the limiting distributions of IRFs and FEVDs. Second, we prove that the residual-based wild bootstrap is not asymptotically valid for inference on IRFs and FEVDs. This is true even when the VAR innovations and proxy variables are iid. As a corollary to this result, we show that the commonly used Rademacher distribution is particularly problematic because its asymptotics imply that there is no uncertainty surrounding the estimates of the VAR innovation covariance matrix and the covariance matrix of the VAR innovations with the proxy variables. This will cause the Rademacher wild bootstrap to produce confidence intervals that are very undersized. Our Monte Carlo simulations show that these confidence intervals can remain undersized at long forecast horizons, which is consistent with our limiting distributions for IRFs and FEVDs. To replace the residual-based wild bootstrap, our third contribution is to prove that a residual-based MBB is asymptotically valid for inference on IRFs and FEVDs from proxy SVARs. Our Monte Carlo simulations show that the MBB generally produces confidence intervals that are appropriately sized.

A Identification When $r > 1$

We show two methods to identify $H^{(1)}$ when $r > 1$. These methods use zero restrictions in the form of assuming that certain square matrices are lower triangular. The first method is a generalization of Equation (7). First, note that Equations (3) and (6) imply

$$\mathbb{E}(m_t u_t')[\mathbb{E}(u_t u_t')]^{-1}\mathbb{E}(u_t m_t') = \Psi\Psi'. \tag{A.1}$$

We assume that Ψ is lower triangular so that a Cholsky decomposition can identify Ψ. Similar to Equation (7), sign restrictions are also needed in the diagonal elements of the Ψ. Then, Equations (6) and (A.1) and the lower triangular assumption on Ψ imply

$$H^{(1)} = \mathbb{E}(u_t m_t') chol\{\mathbb{E}(m_t u_t')[\mathbb{E}(u_t u_t')]^{-1}\mathbb{E}(u_t m_t')\}^{-1}, \tag{A.2}$$

where *chol* is the upper triangular Cholesky function so that $chol(\Psi\Psi') = \Psi'$.

The second method to identify $H^{(1)}$ when $r > 1$ comes from Mertens and Ravn (2013). First, partition u_t and further partition H such that Equation (2) can be rewritten as

$$\begin{bmatrix} u_t^{(1)} \\ (r \times 1) \\ u_t^{(2)} \\ (K-r \times 1) \end{bmatrix} = \begin{bmatrix} H^{(1,1)} & H^{(1,2)} \\ (r \times r) & (r \times K-r) \\ H^{(2,1)} & H^{(2,2)} \\ (K-r \times r) & (K-r \times K-r) \end{bmatrix} \begin{bmatrix} \epsilon_t^{(1)} \\ (r \times 1) \\ \epsilon_t^{(2)} \\ (K-r \times 1) \end{bmatrix}, \tag{A.3}$$

where $H^{(1,1)}$ and $H^{(2,2)}$ are additionally assumed to be non-singular. Rearranging yields

$$u_t^{(1)} = H^{(1,2)} H^{(2,2)-1} u_t^{(2)} + S^{(1)} \epsilon_t^{(1)} \tag{A.4}$$

and

$$u_t^{(2)} = H^{(2,1)} H^{(1,1)-1} u_t^{(1)} + S^{(2)} \epsilon_t^{(2)}, \tag{A.5}$$

where $S^{(1)} = (I_r - H^{(1,2)} H^{(2,2)-1} H^{(2,1)} H^{(1,1)-1}) H^{(1,1)}$ and $S^{(2)} = (I_{K-r} - H^{(2,1)} H^{(1,1)-1} H^{(1,2)} H^{(2,2)-1}) H^{(2,2)}$. Then, Equations (4), (5), and (A.3) imply $\mathbb{E}(m_t u_t^{(1)\prime}) = \Psi H^{(1,1)\prime}$ and $\mathbb{E}(m_t u_t^{(2)\prime}) = \Psi H^{(2,1)\prime}$, which jointly yield

$$H^{(2,1)} H^{(1,1)-1} = \left([\mathbb{E}(m_t u_t^{(1)\prime})]^{-1} \mathbb{E}(m_t u_t^{(2)\prime}) \right)'. \tag{A.6}$$

From Equation (A.5), this is IV identification for the regression of $u_t^{(2)}$ on $u_t^{(1)}$, using m_t as instruments. Identification of $H^{(1)}$ then proceeds as follows. The definition of $S^{(1)}$ implies

$$\begin{aligned} &S^{(1)}S^{(1)\prime} \\ &= (I_r - H^{(1,2)}H^{(2,2)-1}H^{(2,1)}H^{(1,1)-1})H^{(1,1)}H^{(1,1)\prime}(I_r - H^{(1,2)}H^{(2,2)-1}H^{(2,1)}H^{(1,1)-1})' \end{aligned} \tag{A.7}$$

and

$$H^{(1,1)} = (I_r - H^{(1,2)}H^{(2,2)-1}H^{(2,1)}H^{(1,1)-1})^{-1}S^{(1)}. \tag{A.8}$$

Next, impose that $S^{(1)}$ is lower triangular, again imposing sign restrictions on the diagonal elements. Then, use $S^{(1)}$ to identify $H^{(1,1)}$ from Equation (A.8), and then use $H^{(1,1)}$ to identify $H^{(2,1)}$ in Equation (A.6). Finally, $H^{(1)} = [H^{(1,1)\prime}, H^{(2,1)\prime}]'$. See Mertens and Ravn (2013) for a discussion on how to interpret the assumption that $S^{(1)}$ is lower triangular.

All that is left to implement this approach is to identify $H^{(1,1)}H^{(1,1)\prime}$ and $H^{(1,2)}H^{(2,2)-1}$, which can be done as follows. First, for the purposes of notation, define $\mathbb{E}(u_t^{(1)}u_t^{(1)\prime}) = \Sigma_u^{(1,1)}$, $\mathbb{E}(u_t^{(2)}u_t^{(1)\prime}) = \Sigma_u^{(2,1)}$, and $\mathbb{E}(u_t^{(2)}u_t^{(2)\prime}) = \Sigma_u^{(2,2)}$. Second, Equations (3) and (A.3) imply

$$\Sigma_u^{(1,1)} = H^{(1,1)}H^{(1,1)\prime} + H^{(1,2)}H^{(1,2)\prime}, \tag{A.9}$$

$$\Sigma_u^{(2,1)} = H^{(2,1)}H^{(1,1)\prime} + H^{(2,2)}H^{(1,2)\prime}, \tag{A.10}$$

$$\Sigma_u^{(2,2)} = H^{(2,1)}H^{(2,1)\prime} + H^{(2,2)}H^{(2,2)\prime}. \tag{A.11}$$

Using Equations (A.9) and (A.10), it is the case that

$$\Sigma_u^{(2,1)} - H^{(2,1)}H^{(1,1)-1}\Sigma_u^{(1,1)} = (H^{(2,2)} - H^{(2,1)}H^{(1,1)-1}H^{(1,2)})H^{(1,2)\prime}. \tag{A.12}$$

Third, define

$$Z = (H^{(2,2)} - H^{(2,1)}H^{(1,1)-1}H^{(1,2)})(H^{(2,2)} - H^{(2,1)}H^{(1,1)-1}H^{(1,2)})'. \tag{A.13}$$

In the appendix of Jentsch and Lunsford (2019), we show that Z is also given by

$$\begin{aligned} Z = \Sigma_u^{(2,2)} &- H^{(2,1)}H^{(1,1)-1}\Sigma_u^{(2,1)\prime} - \Sigma_u^{(2,1)}(H^{(2,1)}H^{(1,1)-1})' \\ &+ H^{(2,1)}H^{(1,1)-1}\Sigma_u^{(1,1)}(H^{(2,1)}H^{(1,1)-1})'. \end{aligned} \tag{A.14}$$

Hence, Z is identified. We also show that Equations (A.12) and (A.13) imply

$$H^{(1,2)}H^{(1,2)\prime} = (\Sigma_u^{(2,1)} - H^{(2,1)}H^{(1,1)-1}\Sigma_u^{(1,1)})'Z^{-1}(\Sigma_u^{(2,1)} - H^{(2,1)}H^{(1,1)-1}\Sigma_u^{(1,1)}) \tag{A.15}$$

so that $H^{(1,2)}H^{(1,2)\prime}$ is identified. Fifth, Equation (A.9) yields

$$H^{(1,1)}H^{(1,1)\prime} = \Sigma_u^{(1,1)} - H^{(1,2)}H^{(1,2)\prime},$$

which can then produce

$$H^{(2,2)}H^{(2,2)\prime} = \Sigma_u^{(2,2)} - H^{(2,1)}H^{(1,1)-1}H^{(1,1)}H^{(1,1)\prime}(H^{(2,1)}H^{(1,1)-1})'$$

from Equation (A.11). Then Equation (A.10) implies

$$H^{(1,2)}H^{(2,2)-1} = [\Sigma_u^{(2,1)\prime} - H^{(1,1)}H^{(1,1)\prime}(H^{(2,1)}H^{(1,1)-1})'](H^{(2,2)}H^{(2,2)\prime})^{-1}$$

To summarize, the IV identification in Equation (A.6) can be used to identify Z, $H^{(1,2)}H^{(1,2)\prime}$, $H^{(1,1)}H^{(1,1)\prime}$, $H^{(2,2)}H^{(2,2)\prime}$, and $H^{(1,2)}H^{(2,2)-1}$. Given $H^{(1,1)}H^{(1,1)\prime}$ and $H^{(1,2)}H^{(2,2)-1}$, we can identify $S^{(1)}S^{(1)\prime}$, impose the lower triangularity on $S^{(1)}$, and identify $H^{(1,1)}$ and $H^{(2,1)}$.

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 $\widetilde{\boldsymbol{\varphi}} = \text{vec}(\widetilde{\Psi H^{(1)\prime}})$, where $\widetilde{\Psi H^{(1)\prime}} = \frac{1}{T}\sum_{t=1}^{T} m_t u_t'$. Due to $\sqrt{T}(\widehat{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}) = o_P(1)$ and $\sqrt{T}(\widehat{\boldsymbol{\varphi}} - \widetilde{\boldsymbol{\varphi}}) = o_P(1)$ by standard arguments using ergodicity and $\mathbb{E}(m_t y_{t-j}') = 0$, $j = 1, \ldots, p$, we can replace $\widehat{\boldsymbol{\sigma}}$ by $\widetilde{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$ by $\widetilde{\boldsymbol{\varphi}}$ in the following calculations. Furthermore, by using

$$Z_{t-1} = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix} = \sum_{j=0}^{\infty} \begin{pmatrix} \Phi_j u_{t-1-j} \\ \vdots \\ \Phi_j u_{t-p-j} \end{pmatrix} = \sum_{j=1}^{\infty} \begin{pmatrix} \Phi_{j-1} u_{t-j} \\ \vdots \\ \Phi_{j-p} u_{t-j} \end{pmatrix} = \sum_{j=1}^{\infty} C_j u_{t-j}, \tag{B.1}$$

it can be shown that

$$\sqrt{T}\begin{pmatrix}\widehat{\boldsymbol\beta}-\boldsymbol\beta\\ \widetilde{\boldsymbol\sigma}-\boldsymbol\sigma\\ \widetilde{\boldsymbol\varphi}-\boldsymbol\varphi\end{pmatrix} = \begin{pmatrix}\left\{(\frac{1}{T}ZZ')^{-1}\otimes I_K\right\}\sum_{j=1}^{\infty}(C_j\otimes I_K)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\mathrm{vec}(u_tu_{t-j}')\right\}\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}L_K\left\{\mathrm{vec}(u_tu_t')-\mathrm{vec}(\Sigma_u)\right\}\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\mathrm{vec}(m_tu_t')-\mathrm{vec}(\Psi H^{(1)\prime})\right\}\end{pmatrix} \tag{B.2}$$
$$= A_m+(A-A_m),$$

where A denotes the right-hand side of Equation (B.2) and A_m is the same expression, but with $\sum_{j=1}^{\infty}$ replaced by $\sum_{j=1}^{m}$ for some $m\in\mathbb{N}$. In the following, we make use of Proposition 6.3.9 of Brockwell and Davis (1991) and it suffices to show

$$(a)\quad A_m\xrightarrow{\mathcal{D}}\mathcal{N}(0,V_m)\text{ as }T\to\infty$$
$$(b)\quad V_m\to V\text{ as }m\to\infty$$
$$(c)\quad \forall\,\delta>0:\ \lim_{m\to\infty}\limsup_{T\to\infty}P(|A-A_m|_1>\delta)=0.$$

To prove (a), setting $\widetilde{K}=K(K+1)/2$, we can write

$$A_m = \begin{pmatrix}(\frac{1}{T}ZZ')^{-1}\otimes I_K & O_{K^2p\times\widetilde{K}} & O_{K^2p\times Kr}\\ O_{\widetilde{K}\times K^2p} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr}\\ O_{Kr\times K^2p} & O_{Kr\times\widetilde{K}} & I_{Kr}\end{pmatrix}\begin{pmatrix}C_1\otimes I_K & \cdots & C_m\otimes I_K & O_{K^2p\times\widetilde{K}} & O_{K^2p\times Kr}\\ O_{\widetilde{K}\times K^2} & \cdots & O_{\widetilde{K}\times K^2} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr}\\ O_{Kr\times K^2} & \cdots & O_{Kr\times K^2} & O_{Kr\times\widetilde{K}} & I_{Kr}\end{pmatrix}$$
$$\times\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\mathrm{vec}(u_tu_{t-1}')\\ \vdots\\ \mathrm{vec}(u_tu_{t-m}')\\ L_K\left\{\mathrm{vec}(u_tu_t')-\mathrm{vec}(\Sigma_u)\right\}\\ \mathrm{vec}(m_tu_t')-\mathrm{vec}(\Psi H^{(1)\prime})\end{pmatrix}$$
$$= \widehat{Q}_TR_m\frac{1}{\sqrt{T}}\sum_{t=1}^{T}W_{t,m}$$

with an obvious notation for the $(K^2p+\widetilde{K}+Kr\times K^2p+\widetilde{K}+Kr)$ matrix $\widehat{Q}_T$, the $(K^2p+\widetilde{K}+Kr\times K^2m+\widetilde{K}+Kr)$ matrix R_m, and the $K^2m+\widetilde{K}+Kr$-dimensional vector $W_{t,m}$. By Lemma A.2 in Brüggemann, Jentsch, and Trenkler (2016), we have that $\widehat{Q}_T\to Q$ in probability, where $Q=\mathrm{diag}(\Gamma^{-1}\otimes I_K, I_{\widetilde{K}}, I_{Kr})$. Now, the CLT required for part (a) follows from Lemma B.1 with

$$V_m = \begin{pmatrix} V_m^{(1,1)} & V_m^{(1,2)} & V_m^{(1,3)} \\ V_m^{(2,1)} & V^{(2,2)} & V^{(2,3)} \\ V_m^{(3,1)} & V^{(3,2)} & V^{(3,3)} \end{pmatrix} = QR_m\Omega_m R_m' Q', \tag{B.3}$$

which leads to $V^{(i,j)} = \Omega^{(i,j)}$, $i,j \in \{2,3\}$ as defined in Equations (B.5), (B.8) and (B.9), $V_m^{(i,1)} = V_m^{(1,i)\prime}$, $i \in \{2,3\}$ and

$$\begin{aligned}
V_m^{(1,1)} &= (\Gamma^{-1} \otimes I_K) \left(\sum_{i,j=1}^{m} (C_i \otimes I_K) \sum_{h=-\infty}^{\infty} \tau_{i,h,h+j} (C_j \otimes I_K)' \right) (\Gamma^{-1} \otimes I_K)', \\
V_m^{(2,1)} &= L_K \left(\sum_{j=1}^{m} \sum_{h=-\infty}^{\infty} \tau_{0,h,h+j} (C_j \otimes I_K)' \right) (\Gamma^{-1} \otimes I_K)', \\
V_m^{(3,1)} &= \left(\sum_{j=1}^{m} \sum_{h=-\infty}^{\infty} \nu_{0,h,h+j} (C_j \otimes I_K)' \right) (\Gamma^{-1} \otimes I_K)'.
\end{aligned}$$

Part (b) follows from Assumption 3.1 and due to $\sum_{i=1}^{\infty} \|C_i \otimes I_K\| < \infty$. The second and third parts of $A - A_m$ in Equation (B.2) are zero and it suffices to show (c) for the first part, ignoring the factor $\widehat{Q}_T$. Let $\lambda \in \mathbb{R}^{K^2 p}$ and $\delta > 0$, then (c) follows with Markov inequality and $\|V^{(1,1)}\| < \infty$ from

$$\begin{aligned}
& P\left(\left| \sum_{j=m+1}^{\infty} \lambda' (C_j \otimes I_K) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \operatorname{vec}(u_t u_{t-j}') \right| > \delta \right) \\
\leq\ & \frac{1}{\delta^2 T} \mathbb{E}\left(\left| \sum_{j=m+1}^{\infty} \lambda' (C_j \otimes I_K) \sum_{t=1}^{T} \operatorname{vec}(u_t u_{t-j}') \right|^2 \right) \\
=\ & \frac{1}{\delta^2} \sum_{i,j=m+1}^{\infty} \lambda' (C_i \otimes I_K) \left\{ \frac{1}{T} \sum_{t_1,t_2=1}^{T} \mathbb{E}\left(\operatorname{vec}(u_{t_1} u_{t_1-i}') \operatorname{vec}(u_{t_2} u_{t_2-j}')' \right) \right\} (C_j \otimes I_K)' \lambda \\
=\ & \frac{1}{\delta^2} \sum_{i,j=m+1}^{\infty} \lambda' (C_i \otimes I_K) \left(\sum_{h=-(T-1)}^{T-1} \left(1 - \frac{|h|}{T} \right) \tau_{i,h,h+j} \right) (C_j \otimes I_K)' \lambda \\
\underset{T\to\infty}{\to}\ & \frac{1}{\delta^2} \sum_{i,j=m+1}^{\infty} \lambda' (C_i \otimes I_K) \sum_{h=-\infty}^{\infty} \tau_{i,h,h+j} (C_j \otimes I_K)' \lambda \\
\underset{m\to\infty}{\to}\ & 0.
\end{aligned}$$

$\square$

Lemma B.1 (CLT for innovations) *Let* $W_{t,m} = (W_{t,m}^{(1)\prime}, W_{t,m}^{(2)\prime}, W_{t,m}^{(3)\prime})'$, *where*

$$\begin{aligned} W_{t,m}^{(1)} &= (\operatorname{vec}(u_t u_{t-1}')', \dots, \operatorname{vec}(u_t u_{t-m}')')' \\ W_{t,m}^{(2)} &= L_K\{\operatorname{vec}(u_t u_t') - \operatorname{vec}(\Sigma_u)\} = \operatorname{vech}(u_t u_t') - \operatorname{vech}(\Sigma_u) \\ W_{t,m}^{(3)} &= \operatorname{vec}(m_t u_t') - \operatorname{vec}(\Psi H^{(1)\prime}) \end{aligned}$$

Under Assumption 2.1, for sufficiently large m, *we have*

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

where Ω_m *is a* $(K^2 m + \widetilde{K} + Kr \times K^2 m + \widetilde{K} + Kr)$ *block matrix*

$$\Omega_m = \begin{pmatrix} \Omega_m^{(1,1)} & \Omega_m^{(1,2)} & \Omega_m^{(1,3)} \\ \Omega_m^{(2,1)} & \Omega^{(2,2)} & \Omega^{(2,3)} \\ \Omega_m^{(3,1)} & \Omega^{(3,2)} & \Omega^{(3,3)} \end{pmatrix}. \tag{B.4}$$

Here, $\Omega_m^{(1,1)} = (\sum_{h=-\infty}^{\infty} \tau_{i,h,h+j})_{i,j=1,\dots,m}$ *is a block matrix with* $\tau_{i,h,h+j}$ *defined in Equation (19) and the* $(\widetilde{K} \times \widetilde{K})$, $(\widetilde{K} \times K^2 m)$, $(Kr \times K^2 m)$, $(Kr \times \widetilde{K})$ *and* $(Kr \times Kr)$ *matrices*

$$\Omega^{(2,2)} = L_K \left(\sum_{h=-\infty}^{\infty} \{\tau_{0,h,h} - \operatorname{vec}(\Sigma_u) \operatorname{vec}(\Sigma_u)'\} \right) L_K', \tag{B.5}$$

$$\Omega_m^{(2,1)} = L_K \left(\sum_{h=-\infty}^{\infty} (\tau_{0,h,h+1}, \dots, \tau_{0,h,h+m}) \right), \tag{B.6}$$

$$\Omega_m^{(3,1)} = \sum_{h=-\infty}^{\infty} (\nu_{0,h,h+1}, \dots, \nu_{0,h,h+m}) \tag{B.7}$$

$$\Omega^{(3,2)} = \left(\sum_{h=-\infty}^{\infty} \left\{ \nu_{0,h,h} - \operatorname{vec}(\Psi H^{(1)\prime}) \operatorname{vec}(\Sigma_u)' \right\} \right) L_K' \tag{B.8}$$

$$\Omega^{(3,3)} = \sum_{h=-\infty}^{\infty} \left\{ \zeta_{0,h,h} - \operatorname{vec}(\Psi H^{(1)\prime}) \operatorname{vec}(\Psi H^{(1)\prime})' \right\}, \tag{B.9}$$

respectively.

Proof.
The result follows analogously to the proof of Lemma A.1 (ii) in Brüggemann, Jentsch, and Trenkler (2014) extended to the proxy SVAR setup.

B.2 Proof of Corollary 2.2

As $r = 1$, we can make use of the identification scheme (up to sign restriction) given in Equations (15) and (16). It becomes apparent that $H^{(1)}$ 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 Φ_i's are smooth functions of $A_1, \ldots, A_p$. Hence, using the Delta method similar to (and borrowing some of the notation from) Lütkepohl (2005, Proposition 3.6), we get for any $q \in \{0, 1, \ldots\}$ that

$$\sqrt{T}\begin{pmatrix} \mathrm{vec}(\widehat{\Phi}_0) - \mathrm{vec}(\Phi_0) \\ \mathrm{vec}(\widehat{\Phi}_1) - \mathrm{vec}(\Phi_1) \\ \vdots \\ \mathrm{vec}(\widehat{\Phi}_q) - \mathrm{vec}(\Phi_q) \\ \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \\ \widehat{H}^{(1)} - H^{(1)} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, W\right), \quad W = \begin{pmatrix} W^{(1,1)} & W^{(2,1)\prime} & W^{(3,1)\prime} \\ W^{(2,1)} & W^{(2,2)} & W^{(3,2)\prime} \\ W^{(3,1)} & W^{(3,2)} & W^{(3,3)} \end{pmatrix} \quad \text{(B.10)}$$

holds, where

$$\begin{aligned}
W^{(1,1)} &= G_{0,q} V^{(1,1)} G'_{0,q} \\
W^{(2,1)} &= V^{(2,1)} G'_{0,q} \\
W^{(2,2)} &= V^{(2,2)} \\
W^{(3,1)} &= M_{\boldsymbol{\sigma}} V^{(2,1)} G'_{0,q} + M_{\boldsymbol{\varphi}} V^{(3,1)} G'_{0,q} \\
W^{(3,2)} &= M_{\boldsymbol{\sigma}} V^{(2,2)} + M_{\boldsymbol{\varphi}} V^{(3,2)} \\
W^{(3,3)} &= M_{\boldsymbol{\sigma}} V^{(2,2)} M'_{\boldsymbol{\sigma}} + M_{\boldsymbol{\varphi}} V^{(3,2)} M'_{\boldsymbol{\sigma}} + M_{\boldsymbol{\sigma}} V^{(3,2)\prime} M'_{\boldsymbol{\varphi}} + M_{\boldsymbol{\varphi}} V^{(3,3)} M'_{\boldsymbol{\varphi}}
\end{aligned}$$

with $V^{(i,j)}$ defined in Theorem 2.1 and $G'_{0,q} = [G'_0 : G'_1 : \cdots : G'_q]$ is a $(K^2 p \times K^2(q+1))$ matrix with block entries[14]

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

[14]Note that $\boldsymbol{\beta}$ here corresponds to $\boldsymbol{\alpha}$ in Lütkepohl (2005, Proposition 3.6) as no intercept term is included.

where $J = [I_K : 0_K : \cdots : 0_K]$ 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, by using standard rules for matrix differentiation from, for example, Sections A.12 and A.13 in Lütkepohl (2005),

$$M_{\boldsymbol{\sigma}} = \partial \text{vec}(H^{(1)})/\partial \boldsymbol{\sigma}' = \boldsymbol{\varphi}\left((1/2)(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-3/2}\right)\left(\boldsymbol{\varphi}' \otimes \boldsymbol{\varphi}'\right)\left((\Sigma_u^{-1})' \otimes (\Sigma_u^{-1})\right) D_K$$
$$M_{\boldsymbol{\varphi}} = \partial \text{vec}(H^{(1)})/\partial \boldsymbol{\varphi}' = -\boldsymbol{\varphi}\left((1/2)(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-3/2}\right)\left(\boldsymbol{\varphi}'((\Sigma_u^{-1})' + \Sigma_u^{-1})\right) + \left(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi}\right)^{-1/2} I_K.$$

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 $\widehat{H}^{(1)}$.

(i) For one standard deviation IRFs $\widehat{\Theta}_{\bullet 1,i} = \widehat{\Phi}_i \widehat{H}^{(1)}$, we can make use of the relevant parts of (B.10) as the basis result for another application of the Delta method. That is, we shall use

$$\sqrt{T}\begin{pmatrix} \text{vec}(\widehat{\Phi}_i) - \text{vec}(\Phi_i) \\ \widehat{H}^{(1)} - H^{(1)} \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \begin{pmatrix} G_i V^{(1,1)} G_i' & \left(M_{\boldsymbol{\sigma}} V^{(2,1)} G_i' + M_{\boldsymbol{\varphi}} V^{(3,1)} G_i'\right)' \\ M_{\boldsymbol{\sigma}} V^{(2,1)} G_i' + M_{\boldsymbol{\varphi}} V^{(3,1)} G_i' & W^{(3,3)} \end{pmatrix}\right) \quad \text{(B.11)}$$

Together with

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

Equation (B.11) 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

$$\begin{aligned} \Sigma_{\widehat{\Theta}_{\bullet 1,i}} &= \left(H^{(1)\prime} \otimes I_K\right) G_i V^{(1,1)} G_i' \left(H^{(1)} \otimes I_K\right) + \Phi_i \left(M_{\boldsymbol{\sigma}} V^{(2,1)} G_i' + M_{\boldsymbol{\varphi}} V^{(3,1)} G_i'\right)\left(H^{(1)} \otimes I_K\right) \\ &\quad + \left(H^{(1)\prime} \otimes I_K\right)\left(M_{\boldsymbol{\sigma}} V^{(2,1)} G_i' + M_{\boldsymbol{\varphi}} V^{(3,1)} G_i'\right)' \Phi_i' + \Phi_i W^{(3,3)} \Phi_i'. \end{aligned} \quad \text{(B.12)}$$

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

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

Equation (B.11) 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

$$\begin{aligned}
\Sigma_{\widehat{\Xi}_{\bullet 1,i}(s;m,1)} &= \left(\frac{s}{e_m'H^{(1)}}\right)^2 \Bigg[\left(H^{(1)\prime} \otimes I_K\right) G_i V^{(1,1)} G_i' \left(H^{(1)} \otimes I_K\right) \\
&+ \left(\Phi_i\left(I_K - H^{(1)}e_m'\left(\frac{1}{e_m'H^{(1)}}\right)\right)\right)\left(M_\sigma V^{(2,1)}G_i' + M_\varphi V^{(3,1)}G_i'\right)\left(H^{(1)} \otimes I_K\right) \\
&+ \left(H^{(1)\prime} \otimes I_K\right)\left(M_\sigma V^{(2,1)}G_i' + M_\varphi V^{(3,1)}G_i'\right)'\left(\Phi_i\left(I_K - H^{(1)}e_m'\left(\frac{1}{e_m'H^{(1)}}\right)\right)\right)' \\
&+ \left(\Phi_i\left(I_K - H^{(1)}e_m'\left(\frac{1}{e_m'H^{(1)}}\right)\right)\right) W^{(3,3)} \left(\Phi_i\left(I_K - H^{(1)}e_m'\left(\frac{1}{e_m'H^{(1)}}\right)\right)\right)'\Bigg].
\end{aligned} \tag{B.13}$$

Further, it is the case that $\left(I_K - H^{(1)}e_m'/(e_m'H^{(1)})\right)H^{(1)} = 0$. Then, $H^{(1)} = \boldsymbol{\varphi}(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-1/2}$ from (15) and after imposing a sign restriction implies $\left(I_K - H^{(1)}e_m'/(e_m'H^{(1)})\right)M_\sigma = 0$ and $\left(I_K - H^{(1)}e_m'/(e_m'H^{(1)})\right)M_\varphi = \left(I_K - \boldsymbol{\varphi}e_m'/(e_m'\boldsymbol{\varphi})\right)(\boldsymbol{\varphi}'\Sigma_u^{-1}\boldsymbol{\varphi})^{-1/2}$. It follows that

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

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

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

together with

$$F_{j1,h}(r) \quad := \quad \frac{\partial \text{vec}(\omega_{j1,h})}{\partial \text{vec}(\Phi_r)'} = -2 \frac{\sum_{i=0}^{h-1} (e_j' \Phi_i H^{(1)})^2}{(MSE_j(h))^2} (e_j' \Phi_r \Sigma_u \otimes e_j') + 2 \frac{e_j' \Phi_r H^{(1)}}{MSE_j(h)} (H^{(1)\prime} \otimes e_j')$$

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

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

Equation (B.10) 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

$$\Sigma_{\widehat{\omega}_{j1,h}} \quad = \quad [\boldsymbol{F}_{j1,h} : Q_{j1,h} : N_{j1,h}] \begin{pmatrix} W^{(1,1)} & W^{(2,1)\prime} & W^{(3,1)\prime} \\ W^{(2,1)} & W^{(2,2)} & W^{(3,2)\prime} \\ W^{(3,1)} & W^{(3,2)} & W^{(3,3)} \end{pmatrix} \begin{bmatrix} \boldsymbol{F}_{j1,h}' \\ Q_{j1,h}' \\ N_{j1,h}' \end{bmatrix} \qquad \text{(B.14)}$$

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

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

$$\mathbb{E}^+ \left(\text{vec}(u_t^+ u_{t-a}^{+\prime}) \text{vec}(u_{t-b}^+ u_{t-c}^{+\prime})' \right) = \text{vec}(\widehat{u}_t \widehat{u}_{t-a}') \text{vec}(\widehat{u}_{t-b} \widehat{u}_{t-c}')' \mathbb{E}\left(\eta_t \eta_{t-a} \eta_{t-b} \eta_{t-c}\right), \qquad \text{(B.15)}$$

where

$$\mathbb{E}\left(\eta_t\eta_{t-a}\eta_{t-b}\eta_{t-c}\right) = \begin{cases} \mathbb{E}(\eta_t^4), & a=b=c=0 \\ 1, & a=0\neq b=c \text{ or } b=0\neq a=c \text{ or } c=0\neq a=b\,. \\ 0, & \text{otherwise} \end{cases} \quad \text{(B.16)}$$

Note that analogous representations also hold for $\mathbb{E}^+\left(\mathrm{vec}(m_t^+ u_{t-a}^{+\prime})\mathrm{vec}(u_{t-b}u_{t-c}^{+\prime})'\right)$ as well as $\mathbb{E}^+\left(\mathrm{vec}(m_t^+ u_{t-a}^{+\prime})\mathrm{vec}(m_{t-b}u_{t-c}^{+\prime})'\right)$. Now, by using arguments similar to those used in the proof of Theorem 3.2 below, we can show that the variance of $\sqrt{T}((\widehat{\boldsymbol{\beta}}^+-\widehat{\boldsymbol{\beta}})', (\widehat{\boldsymbol{\sigma}}^+-\widehat{\boldsymbol{\sigma}})', (\widehat{\boldsymbol{\varphi}}^+-\widehat{\boldsymbol{\varphi}})')'$ converges to a quantity corresponding to V as defined in Theorem 2.1, 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. □

B.4 Proof of Theorem 3.2

By Polya's Theorem and by Lemma A.1 in Brüggemann, Jentsch, and Trenkler (2016) similarly to the proof of Theorem 2.1, it suffices to show that $\sqrt{T}((\widetilde{\boldsymbol{\beta}}^*-\widetilde{\boldsymbol{\beta}})', (\widetilde{\boldsymbol{\sigma}}^*-\widehat{\boldsymbol{\sigma}})', (\widetilde{\boldsymbol{\varphi}}^*-\widetilde{\boldsymbol{\varphi}})')'$ converges in distribution w.r.t. measure P^* to $\mathcal{N}(0,V)$ as obtained in Theorem 2.1, where $\widetilde{\boldsymbol{\beta}}^*-\widetilde{\boldsymbol{\beta}} := ((\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\widetilde{Z}^*\otimes I_K)\widetilde{\mathbf{u}}^*$, $\widetilde{\boldsymbol{\sigma}}^* = \mathrm{vech}(\widetilde{\Sigma}_u^*)$ with $\widetilde{\Sigma}_u^* = \frac{1}{T}\sum_{t=1}^T \widetilde{u}_t^*\widetilde{u}_t^{*\prime}$, $\widetilde{\boldsymbol{\sigma}} = \mathrm{vech}(\widetilde{\Sigma}_u)$ with $\widetilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^T u_t u_t'$, $\widetilde{\boldsymbol{\varphi}}^* = \mathrm{vec}(\widetilde{\Psi H^{(1)*\prime}})$ with $\widetilde{\Psi H^{(1)*\prime}} = \frac{1}{T}\sum_{t=1}^T m_t^*\widetilde{u}_t^{*\prime}$ and $\widetilde{\boldsymbol{\varphi}} = \mathrm{vec}(\widetilde{\Psi H^{(1)\prime}})$ with $\widetilde{\Psi H^{(1)\prime}} = \frac{1}{T}\sum_{t=1}^T m_t u_t'$. 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 $\widetilde{y}_t^* = A_1\widetilde{y}_{t-1}^* + \cdots + A_p\widetilde{y}_{t-p}^* + \widetilde{u}_t^*$, 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 moving block bootstrap procedure in Section 3.1, but from $u_1,\ldots,u_T$ instead of $\widehat{u}_1,\ldots,\widehat{u}_T$. Further, we use the notation $\widetilde{Z}_t^* = \mathrm{vec}(\widetilde{y}_t^*,\ldots,\widetilde{y}_{t-p+1}^*)\quad(Kp\times 1)$, $\widetilde{Z}^* = (\widetilde{Z}_0^*,\ldots,\widetilde{Z}_{T-1}^*)\quad(Kp\times T)$, and $\widetilde{\mathbf{u}}^* = \mathrm{vec}(\widetilde{u}_1^*,\ldots,\widetilde{u}_T^*)\quad(KT\times 1)$. Similarly to Equation (B.2), we get the representation

$$\begin{aligned} \sqrt{T}\begin{pmatrix} \widetilde{\boldsymbol{\beta}}^*-\widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\sigma}}^*-\widetilde{\boldsymbol{\sigma}} \\ \widetilde{\boldsymbol{\varphi}}^*-\widetilde{\boldsymbol{\varphi}} \end{pmatrix} &= \begin{pmatrix} \left\{(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\otimes I_K\right\}\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}(C_j\otimes I_K)\sum_{t=j+1}^T\left\{\mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_{t-j}^{*\prime})\right\} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T L_K\left\{\mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime})-\mathrm{vec}(u_tu_t')\right\} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T\left\{\mathrm{vec}(m_t^*\widetilde{u}_t^{*\prime})-\mathrm{vec}(m_tu_t')\right\} \end{pmatrix} \quad \text{(B.17)} \\ &= A_m^* + (A^*-A_m^*), \end{aligned}$$

where A^* denotes the right-hand side of Equation (B.10) and A_m^* is the same expression, but with $\sum_{j=1}^{T-1}$ replaced by $\sum_{j=1}^m$ for some fixed $m\in\mathbb{N}$, $m<T$. In the following, we make use

of Proposition 6.3.9 of Brockwell and Davis (1991) and it suffices to show

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

To prove (a), setting $\widetilde{K} = K(K+1)/2$, we can write

$$
\begin{aligned}
A_m^* &= \begin{pmatrix} (\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1} \otimes I_K & O_{K^2p\times\widetilde{K}} & O_{K^2p\times Kr} \\ O_{\widetilde{K}\times K^2p} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr} \\ O_{Kr\times K^2p} & O_{Kr\times\widetilde{K}} & I_{Kr} \end{pmatrix} \begin{pmatrix} C_1 \otimes I_K & \cdots & C_m \otimes I_K & O_{K^2p\times\widetilde{K}} & O_{K^2p\times Kr} \\ O_{\widetilde{K}\times K^2} & \cdots & O_{\widetilde{K}\times K^2} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr} \\ O_{Kr\times K^2} & \cdots & O_{Kr\times K^2} & O_{Kr\times\widetilde{K}} & I_{Kr} \end{pmatrix} \\
&\quad \times \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \operatorname{vec}(\widetilde{u}_t^*\widetilde{u}_{t-1}^{*\prime}) \\ \vdots \\ \operatorname{vec}(\widetilde{u}_t^*\widetilde{u}_{t-m}^{*\prime}) \\ L_K\left\{\operatorname{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}) - \operatorname{vec}(u_t u_t')\right\} \\ \operatorname{vec}(m_t^*\widetilde{u}_t^{*\prime}) - \operatorname{vec}(m_t u_t') \end{pmatrix} \\
&= \widetilde{Q}_T^* R_m \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widetilde{W}_{t,m}^*
\end{aligned}
$$

as $\widetilde{u}_t^* := 0$ for $t < 0$ and with an obvious notation for the $(K^2p + \widetilde{K} \times K^2p + \widetilde{K})$ matrix $\widetilde{Q}_T^*$ and the $(K^2m + \widetilde{K} + Kr)$-dimensional vector $\widetilde{W}_{t,m}^*$. By Lemma A.2 in Brüggemann, Jentsch, and Trenkler (2016), we have that $\widetilde{Q}_T^* \to Q$ with respect to P^*. By using a straightforward extension of Lemma A.3 in Brüggemann, Jentsch, and Trenkler (2016), the CLT required for part (a) follows with V_m defined in Equation (B.3). Part (b) follows from the summability of C_j and uniform boundedness of $\sum_{h=-\infty}^{\infty} \tau_{i,h,h+j}$ for $i, j \in \mathbb{N}$, which is implied by the cumulant condition of Assumption 3.1. As the factor $\widetilde{Q}_T^*$ can be ignored and the second and third parts of $A^* - A_m^*$ are zero, part (c) follows as in Theorem 4.1 in Brüggemann, Jentsch, and Trenkler (2016), which concludes the proof. □

References

Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner. 2018. "Inference in Bayesian Proxy-SVARs." Federal Reserve Bank of Atlanta Working Paper 2018-16. URL https://doi.org/10.29338/wp2018-16.

Brillinger, David R. 1981. *Time Series: Data Analysis and Theory*. Holden-Day.

Brockwell, Peter J. and Richard A. Davis. 1991. *Time Series: Theory and Methods, Second Edition*. Springer New York. URL https://doi.org/10.1007/978-1-4419-0320-4.

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. URL https://doi.org/10.1016/j.jeconom.2015.10.004.

Caldara, Dario and Edward Herbst. 2019. "Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs." *American Economic Journal: Macroeconomics* 11 (1):157–192. URL https://doi.org/10.1257/mac.20170294.

Carriero, Andrea, Haroon Mumtaz, Konstantinos Theodoridis, and Angeliki Theophilopoulou. 2015. "The Impact of Uncertainty Shocks under Measurement Error: A Proxy SVAR Approach." *Journal of Money, Credit and Banking* 47 (6):1223–1238. URL https://doi.org/10.1111/jmcb.12243.

Cesa-Bianchi, Ambrogio and Andrej Sokol. 2017. "Financial Shocks, Credit Spreads and the International Credit Channel." Bank of England Staff Working Paper No. 693. URL https://doi.org/10.2139/ssrn.3071661.

Gertler, Mark and Peter Karadi. 2015. "Monetary Policy Surprises, Credit Costs, and Economic Activity." *American Economic Journal: Macroeconomics* 7 (1):44–76. URL https://doi.org/10.1257/mac.20130329.

Gonçalves, Sílvia and Lutz Kilian. 2004. "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form." *Journal of Econometrics* 123 (1):89–120. URL https://doi.org/10.1016/j.jeconom.2003.10.030.

———. 2007. "Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity." *Econometric Reviews* 26 (6):609–641. URL https://doi.org/10.1080/07474930701624462.

Hachula, Michael and Dieter Nautz. 2018. "The Dynamic Impact of Macroeconomic News on Long-Term Inflation Expectations." *Economics Letters* 165:39–43. URL https://doi.org/10.1016/j.econlet.2018.01.015.

Hall, Peter. 1992. *The Bootstrap and Edgeworth Expansion*. Springer, New York. URL https://doi.org/10.1007/978-1-4612-4384-7.

Jarociński, Marek and Peter Karadi. 2018. "Deconstructing Monetary Policy Surprises - The Role of Information Shocks." CEPR Discussion Paper no. 12765. URL https://ssrn.com/abstract=3134416.

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. URL https://doi.org/10.26509/frbc-wp-201619.

———. 2019. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment." *American Economic Review*. Forthcoming. URL https://www.aeaweb.org/articles?id=10.1257/aer.20162011.

Kerssenfischer, Mark. 2019. "The Puzzling Effects of Monetary Policy in VARs: Invalid Identification or Missing Information?" *Journal of Applied Econometrics* 34 (1):18–25. URL https://doi.org/10.1002/jae.2647.

Kilian, Lutz. 1999. "Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Responses." *Review of Economics and Statistics* 81 (4):652–660. URL https://doi.org/10.1162/003465399558517.

Kilian, Lutz and Helmut Lütkepohl. 2017. *Structural Vector Autoregressive Analysis*. Cambridge University Press. URL https://doi.org/10.1017/9781108164818.

Künsch, Hans R. 1989. "The Jackknife and the Bootstrap for General Stationary Observations." *Annals of Statistics* 17 (3):1217–1241. URL https://doi.org/10.1214/aos/1176347265.

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. URL https://doi.org/10.26509/frbc-wp-201528.

Lütkepohl, Helmut. 2005. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag. URL https://doi.org/10.1007/3-540-27752-8.

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. URL https://doi.org/10.1257/aer.103.4.1212.

———. 2014. "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers." *Journal of Monetary Economics* 68 (Supplement):S1–S19. URL https://doi.org/10.1016/j.jmoneco.2013.04.004.

———. 2019. "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Reply to Jentsch and Lunsford." *American Economic Review*. Forthcoming.

Miranda-Agrippino, Silvia. 2016. "Unsurprising Shocks: Information, Premia, and the Monetary Transmission." Bank of England Staff Working Paper No. 626. URL https://doi.org/10.2139/ssrn.2865585.

Montiel Olea, José Luis, James H. Stock, and Mark W. Watson. 2012. "Inference in Structural VARs with External Instruments." Presentation Slides, Harvard University.

———. 2018. "Inference in Structural Vector Autoregressions Identified with an External Instrument." Working Paper, Columbia University.

Passari, Evgenia and Hélène Rey. 2015. "Financial Flows and the International Monetary System." *Economic Journal* 125 (584):675–698. URL https://doi.org/10.1111/ecoj.12268.

Piffer, Michele and Maximilian Podstawski. 2018. "Identifying Uncertainty Shocks Using the Price of Gold." *Economic Journal* 128 (616):3266–3284. URL https://doi.org/10.1111/ecoj.12545.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics*, edited by John B. Taylor and Harald Uhlig, chap. 2. Elsevier B.V., 71–162. URL https://doi.org/10.1016/bs.hesmac.2016.03.003.

Rey, Hélène. 2016. "International Channels of Transmission of Monetary Policy and the Mundellian Trilemma." *IMF Economic Review* 64 (1):6–35. URL https://doi.org/10.1057/imfer.2016.4.

Runkle, David E. 1987. "Vector Autoregressions and Reality." *Journal of Business & Economic Statistics* 5 (4):437–442. URL https://doi.org/10.2307/1391992.

Sims, Christopher A. 1980. "Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered." *American Economic Review* 70 (2):250–257. URL https://www.jstor.org/stable/1815476.

Staiger, Douglas and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3):557–586. URL https://doi.org/10.2307/2171753.

Stock, James H. and Mark W. Watson. 2008. "Lecture 7 - Recent Developments in Structural VAR Modeling." Presented at the NBER Summer Institute Econometrics Lectures: What's New in Economics - Time Series, July 15.

———. 2012. "Disentangling the Channels of the 2007-09 Recession." *Brookings Papers on Economic Activity* Spring:81–135. URL https://doi.org/10.1353%2Feca.2012.0005.

———. 2016. "Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics." In *Handbook of Macroeconomics*, edited by John B. Taylor and Harald Uhlig, chap. 8. Elsevier B.V., 415–525. URL https://doi.org/10.1016/bs.hesmac.2016.04.002.

———. 2018. "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments." *Economic Journal* 128 (610):917–948. URL https://doi.org/10.1111/ecoj.12593.