working paper

16 19

Proxy SVARs: Asymptotic Theory, Bootstrap Inference, and the Effects of Income Tax Changes in the United States

Carsten Jentsch and Kurt G. Lunsford

FEDERAL RESERVE BANK OF CLEVELAND

Working Paper 16-19 July 2016

Proxy SVARs: Asymptotic Theory, Bootstrap Inference, and the Effects of Income Tax Changes in the United States

Carsten Jentsch and Kurt G. Lunsford

Proxy structural vector autoregressions (SVARs) identify structural shocks in vector autoregressions (VARs) with external proxy variables that are correlated with the structural shocks of interest but uncorrelated with other structural shocks. We provide asymptotic theory for proxy SVARs when the VAR innovations and proxy variables are jointly α-mixing. We also prove the asymptotic validity of a residual-based moving block bootstrap (MBB) for inference on statistics that depend jointly on estimators for the VAR coefficients and for covariances of the VAR innovations and proxy variables. These statistics include structural impulse response functions (IRFs). Conversely, wild bootstraps are invalid, even when innovations and proxy variables are either independent and identically distributed or martingale difference sequences, and simulations show that their coverage rates for IRFs can be badly mis-sized. Using the MBB to re-estimate confidence intervals for the IRFs in Mertens and Ravn (2013), we show that inferences cannot be made about the effects of tax changes on output, labor, or investment.

Keywords: fiscal policy, mixing, residual-based moving block bootstrap, structural vector autoregression, tax shocks, wild bootstrap.

JEL Codes: C15, C32, E62, H24, H25, H31, H32.

Suggested citation: 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.

Carsten Jentsch is at the University of Mannheim (cjentsch@mail.uni-mannheim.de). Kurt G. Lunsford is at the Federal Reserve Bank of Cleveland (kurt.lunsford@clev.frb.org).
The authors are grateful to Todd E. Clark, Lutz Kilian, Helmut Lütkepohl, Johannes Pfeifer, Carsten Trenkler, and participants at the IAAE 2016 Annual Conference for helpful comments and conversations. The research of Carsten Jentsch was supported by the Research Center (SFB) 884 "Political Economy of Reforms" (Project B6), funded by the German Research Foundation (DFG).

1 Introduction

Since Sims (1980), estimating the dynamic effects of structural shocks in vector autoregressions (VARs) has been important for research in macroeconomics. In recent contributions, Stock and Watson (2008, 2012), Montiel Olea, Stock, and Watson (2012), and Mertens and Ravn (2013) have developed a method for estimating structural vector autoregressions (SVARs) that uses proxy variables. These variables are external from the VAR, and they act as proxies for the structural shocks of interest with the assumptions that they are correlated with structural shocks of interest but uncorrelated with the other structural shocks.[1]

This proxy SVAR approach has proven to be very useful. Mertens and Ravn (2013) use it to merge the SVAR literature on tax shocks (Blanchard and Perotti, 2002; Mountford and Uhlig, 2009) with the narrative approach of Romer and Romer (2010), Gertler and Karadi (2015) and Lunsford (2016) use it to study the effects monetary policy shocks, Carriero et al. (2015) use it to study the effects of uncertainty shocks, and Stock and Watson (2012) use it to study the effects of a large number of economic shocks, including oil shocks, productivity shocks, uncertainty shocks, and financial shocks. In addition, Mumtaz, Pinter, and Theodoridis (2015) show that it matches the effects of credit supply shocks from a dynamic stochastic general equilibrium model better than a Cholesky decomposition, and Drautzburg (2015) uses it to estimate a Bayesian VAR and a dynamic stochastic general equilibrium model. Mertens and Ravn (2014) show that it can reconcile the differences between structural VAR and narrative estimates of tax multipliers; however, Kliem and Kriwoluzky (2013) argue that it is not able to reconcile structural VAR and narrative estimates of monetary policy shocks. Finally, this proxy SVAR approach has been included as a standard method for identifying macroeconomic shocks in a recent handbook chapter (Ramey, 2016).

Despite the above research, inference for proxy SVARs has received little attention. To fill this gap in the literature, we derive limiting results for estimation in proxy SVARs, study the applicability of residual-based wild and moving block bootstrap (MBB) algorithms, and provide asymptotic bootstrap theory.

To produce the confidence intervals for their structural impulse response functions (IRFs), Mertens and Ravn (2013) (MR throughout the paper) use a recursive-design wild bootstrap on the VAR residuals, and they advertise three appealing features of this method. First, Gonçalves and Kilian (2004) show that this bootstrap design is robust against conditional heteroskedasticity in the univariate case. Second, it accounts for uncertainty in estimating the effects of structural shocks with proxy variables. Third, because MR's narrative proxy variables have many observations that are censored to zero, a bootstrap based on independent and identically distributed (iid) re-sampling would have a positive probability of drawing all zeros for a series of bootstrapped proxy variables. However, the wild bootstrap does not allow this event. Because of these benefits, MR set the precedent for producing confidence

[1] We follow the terminology of Mertens and Ravn (2013, 2014) by referring to the external variables as proxy variables. Because the covariance assumptions on these variables parallels those from the instrumental variables literature, they may also be called "external instruments."

intervals, and their wild bootstrap method was also used by Mertens and Ravn (2014), Gertler and Karadi (2015), Lunsford (2015a) and Nakamura and Steinsson (2015).

In contrast to MR's claims, the first two features of the wild bootstrap are not true as advertised. With regard to conditional heteroskedasticity, Gonçalves and Kilian (2004) only applies to estimators of AR coefficients, but not to the covariance matrix of the VAR innovations. This is problematic because structural IRFs are functions of both the VAR coefficients and the covariance matrix of the VAR innovations. Brüggemann, Jentsch, and Trenkler (2016) show that wild bootstraps do not replicate the relevant fourth moments of the VAR innovations even when these innovations are iid. Hence, wild bootstraps cannot be used for inference on structural IRFs.

With regard to accounting for the uncertainty of estimating the effects of structural shocks with proxy variables, a similar problem occurs. The wild bootstrap does not correctly replicate the relevant moments between the proxy variables and the VAR residuals. This is true for any distribution used to produce the iid sequence of wild bootstrap multipliers, such as a standard normal distribution. But the Rademacher distribution, which takes the values 1 or -1 with equal probability of one half and has become standard in the proxy SVAR literature, is especially problematic. This is because the bootstrap multipliers effectively drop out of the bootstrap algorithm when computing the covariance between the VAR residuals and the proxy variables. This causes the wild bootstrap to underestimate the uncertainty of the estimated structural shock. In a Monte Carlo simulation with a sample size of 250, the wild bootstrap's 68% confidence interval includes the true initial impulse response in only 5% to 8% of the simulations, and the 95% confidence interval includes the true initial impulse response in only 16% to 18% of simulations. Further, these coverage rates shrink as the sample size increases, showing that the Rademacher wild bootstrap produces confidence intervals that are generally too small.

To replace the wild bootstrap method, we proceed in two steps. First, we provide a joint central limit theorem (CLT) for 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 and independent innovation processes. This result extends Theorem 3.1 in Brüggemann, Jentsch, and Trenkler (2016) to the proxy SVAR setup.

Second, we prove that a modified version of the residual-based MBB studied by Brüggemann, Jentsch, and Trenkler (2016) is asymptotically valid for inference on statistics, such as structural IRFs, 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 that we make to Brüggemann, Jentsch, and Trenkler's (2016) MBB incorporates the proxy variables into the block resampling. In contrast to the wild bootstrap, the MBB is capable of mimicking the joint fourth order dependence structure of the VAR innovations and the proxy variables. In the same Monte Carlo simulation as the wild bootstrap, the MBB's 68% confidence interval includes the true initial impulse response in 60% to 63% of the simulations, and the 95% confidence interval includes the true initial impulse response in 91% to 92% of simulations. Although these coverage rates are slightly low, they improve as the sample size increases and are better than the coverage rates from either a normal or a Rademacher wild bootstrap.

As an application, we recreate a number of MR's IRFs with confidence intervals that are produced with the MBB. The primary result is that the confidence intervals for all of MR's IRFs become much larger. With the wild bootstrap, MR find that average personal income tax rate (APITR) and average corporate income tax rate (ACITR) cuts have statistically significant impacts on many economic variables at 90% and 95% confidence levels. However, many of MR's results are no longer inferable with the MBB, even at 68% confidence levels. Most importantly, cuts to neither the APITR nor the ACITR have statistically significant effects on output. Consistent with this, there are also no statistically significant effects on labor market variables nor on investment. While these results suggest that confidence intervals from proxy SVARs have the potential to be very large, it is not always the case that proxy SVARs yield no inferable results. For example, Lunsford (2015b) finds statistically significant results at the 90% level with the MBB when using Fernald's (2014) utilization-adjusted total factor productivities as proxy variables. Hence, proxy SVARs are useful for inferring the dynamic effects of structural shocks with the MBB.

To our knowledge, the only other paper that addresses inference in proxy SVARs is Montiel Olea, Stock, and Watson (2016), and our paper is complementary to theirs for several reasons. First, their confidence intervals only apply to the case where one proxy variable is used to identify one structural shock. In contrast, the MBB can be used to produce confidence intervals when multiple proxy variables are used to identify multiple structural shocks, as in MR and Drautzburg (2015). Second, their confidence intervals are constructed specifically for structural IRFs that have been normalized so that one of the VAR variables changes by a specified amount.[2] However, the MBB can be used for inference on any statistic that is a smooth function of the VAR coefficients, the covariance matrix of the VAR innovations, and the covariance matrix of the VAR innovations with the proxy variables. For example, the MBB can be used for inference on forecast error variance decompositions and on IRFs from a one standard deviation shock. Finally, Montiel Olea, Stock, and Watson (2016) develop confidence intervals for structural IRFs that are robust when a proxy variable is weakly correlated with the structural shock of interest in large samples, similar to the problem of a weak instrumental variable (Staiger and Stock, 1997). While our paper only applies when the proxy is strong, our asymptotic theory and MBB provide a foundation for inference in proxy SVARs under mild assumptions on the innovation and proxy processes.

The rest of the paper proceeds as follows. Section 2 describes the proxy SVAR methodology and provides the joint CLT. Section 3 describes the bootstrap algorithms, proves the asymptotic validity of the MBB, and evaluates the bootstrap coverage rates with Monte Carlo simulations. Section 4 recreates MR's results using the MBB, and Section 5 concludes.

[2]For example, initial government spending increases by 1% or the initial federal funds rate falls by 0.25%.

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.

The objective here is to identify the effects of r of the structural shocks where $r < K$. To be precise, 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. Further, we 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, the objective here is to estimate $H^{(1)}$.

The difficulty in estimating $H^{(1)}$ 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 MR introduce 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 an invertible $r \times r$ matrix. They are also exogenous from the other structural shocks. That is,

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

Furthermore, we assume that the proxy variables m_t and lags of y_t are uncorrelated. That is, $\mathbb{E}(m_t y_{t-j}^{\prime}) = 0$ for all $j = 1, \ldots, p$.[3]

When applying the proxy SVAR approach, m_t can come from a wide variety of sources. For example, MR follow the narrative approach of Romer and Romer (2009) to construct proxy variables for tax shocks, Gertler and Karadi (2015) follow the high frequency approach of Gürkaynak, Sack, and Swanson (2005) to construct proxy variables for monetary policy shocks, and Carriero et al. (2015) use Bloom's (2009) measure of uncertainty as a proxy for uncertainty shocks. Thus, the assumptions in Equations (4) and (5) allow for the merging of many different methods of identifying economic shocks with SVARs.

To implement the proxy SVAR approach, we partition u_t and further partition H such that Equation (2) can be re-written as

$$\begin{bmatrix} u_t^{(1)} \\ (r \times 1) \\ u_t^{(2)} \\ (K-r \times 1) \end{bmatrix} = \begin{bmatrix} H^{(1,1)} & H^{(1,2)} \\ (r \times r) & (r \times K-r) \\ H^{(2,1)} & H^{(2,2)} \\ (K-r \times r) & (K-r \times K-r) \end{bmatrix} \begin{bmatrix} \epsilon_t^{(1)} \\ (r \times 1) \\ \epsilon_t^{(2)} \\ (K-r \times 1) \end{bmatrix}, \tag{6}$$

where $H^{(1,1)}$ and $H^{(2,2)}$ are assumed to be non-singular. Then, the objective is to estimate $H^{(1)} = [H^{(1,1)\prime}, H^{(2,1)\prime}]^{\prime}$. To do this, first note that Equations (4), (5) and (6) imply

$$\mathbb{E}(m_t u_t^{(1)\prime}) = \Psi H^{(1,1)\prime} \tag{7}$$

and

$$\mathbb{E}(m_t u_t^{(2)\prime}) = \Psi H^{(2,1)\prime}. \tag{8}$$

Jointly, Equations (7) and (8) yield

$$H^{(2,1)} H^{(1,1)-1} = \left([\mathbb{E}(m_t u_t^{(1)\prime})]^{-1} \mathbb{E}(m_t u_t^{(2)\prime}) \right)^{\prime}. \tag{9}$$

Because the right-hand side of Equation (9) can be estimated from the data, it provides restrictions on parameters of the model in addition to those in Equation (3) that help estimate $H^{(1,1)}$ and $H^{(2,1)}$. Specifically, given an estimate of $H^{(2,1)} H^{(1,1)-1}$ from Equation (9), Equation (3) and the partitions in Equation (6) are sufficient to estimate $H^{(1,2)} H^{(1,2)\prime}$, $H^{(1,1)} H^{(1,1)\prime}$, $H^{(2,2)} H^{(2,2)\prime}$ and $H^{(1,2)} H^{(2,2)-1}$. Appendix A provides details of these estimations.

When $r = 1$, Lunsford (2015b) shows that $H^{(1,1)}$ and $H^{(2,1)}$ can be estimated up to a sign convention. However, in the general case of $r > 1$, as is the case in MR, we need additional

[3] As discussed in MR, 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.

restrictions. To get these restrictions, MR re-write the system in Equation (6) as

$$u_t^{(1)} = Q^{(1)} u_t^{(2)} + S^{(1)} \epsilon_t^{(1)} \tag{10}$$

and

$$u_t^{(2)} = Q^{(2)} u_t^{(1)} + S^{(2)} \epsilon_t^{(2)}, \tag{11}$$

where $Q^{(1)} = H^{(1,2)} H^{(2,2)-1}$, $Q^{(2)} = H^{(2,1)} H^{(1,1)-1}$, $S^{(1)} = (I_r - H^{(1,2)} H^{(2,2)-1} H^{(2,1)} H^{(1,1)-1}) H^{(1,1)}$, and $S^{(2)} = (I_{K-r} - H^{(2,1)} H^{(1,1)-1} H^{(1,2)} H^{(2,2)-1}) H^{(2,2)}$. Note that non-singularity of H, $H^{(1,1)}$ and $H^{(2,2)}$ also guarantees the same for $S^{(1)}$, which crops-up in the inverse of a partitioned matrix; compare e.g. Section A.10 in Lütkepohl (2005). Inversely, it is the case that

$$\begin{bmatrix} H^{(1,1)} \\ H^{(2,1)} \end{bmatrix} = \begin{bmatrix} I_r + Q^{(1)} (I_{K-r} - Q^{(2)} Q^{(1)})^{-1} Q^{(2)} \\ (I_{K-r} - Q^{(2)} Q^{(1)})^{-1} Q^{(2)} \end{bmatrix} S^{(1)} \tag{12}$$

so that $H^{(1,1)}$ and $H^{(2,1)}$ are smooth functions of $Q^{(1)}$, $Q^{(2)}$ and $S^{(1)}$. Here $Q^{(2)}$ can be estimated from Equation (9) and $Q^{(1)}$ can be estimated from Equations (3), (6) and (9), but $S^{(1)}$ cannot be estimated without additional restrictions. To get these restrictions, note that

$$\begin{aligned} &S^{(1)} S^{(1)\prime} \\ &= (I_r - H^{(1,2)} H^{(2,2)-1} H^{(2,1)} H^{(1,1)-1}) H^{(1,1)} H^{(1,1)\prime} (I_r - H^{(1,2)} H^{(2,2)-1} H^{(2,1)} H^{(1,1)-1})' \end{aligned} \tag{13}$$

and

$$H^{(1,1)} = (I_r - H^{(1,2)} H^{(2,2)-1} H^{(2,1)} H^{(1,1)-1})^{-1} S^{(1)}. \tag{14}$$

Equation (13) provides an estimate of $S^{(1)} S^{(1)\prime}$. Given this, we follow MR and impose that $S^{(1)}$ is the lower triangular Cholesky decomposition of $S^{(1)} S^{(1)\prime}$ with the normalization that the diagonal elements of $S^{(1)}$ are positive. Then, $H^{(1,1)}$ and $H^{(2,1)}$ can be estimated from Equations (14) and (9). This Cholesky decomposition provides the additional restrictions needed to estimate the model by restricting how the shocks of interest can interact. For example, MR estimate the effects of changes to both average personal income tax rates (APITRs) and average corporate income tax rates (ACITRs), and the Cholesky decomposition restricts how an APITR shock can effect the ACITR and vice versa. If the APITR is ordered before the ACITR in y_t, then the Cholesky decomposition implies that an APITR shock impacts the ACITR directly through $\epsilon_t^{(1)}$ and indirectly through $u_t^{(2)}$. In contrast, an ACITR shock only impacts the APITR indirectly though $u_t^{(2)}$.

2.2 Estimation

To estimate the proxy SVAR, we focus on estimators for the VAR coefficients $A_1, \ldots, A_p$, the innovation covariance matrix Σ_u, and for the $r \times K$ matrix $\Psi H^{(1)\prime}$, where $\Psi H^{(1)\prime} = \Psi[H^{(1,1)\prime}, H^{(2,1)\prime}]$ is the combination of Equations (7) and (8). We introduce the following notation, where the dimensions of the defined quantities are given in parentheses:

$$\begin{aligned} \mathbf{y} &= \text{vec}(y_1, \dots, y_T) \ (KT \times 1), \quad Z_t = \text{vec}(y_t, \dots, y_{t-p+1}) \ (Kp \times 1) \\ Z &= (Z_0, \dots, Z_{T-1}) \ (Kp \times T), \quad \boldsymbol{\beta} = \text{vec}(A_1, \dots, A_p) \ (K^2 p \times 1) \\ \mathbf{u} &= \text{vec}(u_1, \dots, u_T) \ (KT \times 1), \quad \boldsymbol{\varphi} = \text{vec}(\Psi H^{(1)\prime}) \ (Kr \times 1), \end{aligned} \tag{15}$$

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 $B = (b_{ij})$. Since the process is stable, y_t has a vector moving average (VMA) representation

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j}, \quad t \in \mathbb{Z}, \tag{16}$$

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 $\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j$, $i = 1, 2, \dots$. Further, we define the $(Kp \times K)$ matrices $C_j = (\Phi'_{j-1}, \dots, \Phi'_{j-p})'$ and the $(Kp \times Kp)$ matrix $\Gamma = \sum_{j=1}^{\infty} C_j \Sigma_u C'_j$. The standard estimator of Σ_u is

$$\widehat{\Sigma}_u = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_t \widehat{u}'_t, \tag{17}$$

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

$$\widehat{\Psi H^{(1)\prime}} = \frac{1}{T} \sum_{t=1}^{T} m_t \widehat{u}'_t. \tag{18}$$

After estimating $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, we can estimate $H^{(1,2)}H^{(1,2)\prime}$, $H^{(1,1)}H^{(1,1)\prime}$, $H^{(2,2)}H^{(2,2)\prime}$ and $H^{(1,2)}H^{(2,2)-1}$ by following the steps in Appendix A. Then, the estimate of $S^{(1)}S^{(1)\prime}$ follows from Equation (13), and $\widehat{S}^{(1)}$ is the Cholesky decomposition of $\widehat{S^{(1)}S^{(1)\prime}}$. Finally, Equations (14) and (9) give us $\widehat{H}^{(1,1)}$ and $\widehat{H}^{(2,1)}$, completing the estimation of $\widehat{H}^{(1)}$. Note that because H, $H^{(1,1)}$ and $H^{(2,2)}$ are assumed to be non-singular, $\widehat{H}^{(1)}$ is obtained as a sufficiently smooth function of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$.

Once $H^{(1)}$ is estimated, we can compute a number of statistics of economic interest. In this paper, we focus on structural IRFs that give the dynamic response of y_t to $\epsilon_t^{(1)}$. These IRFs are given by $\Theta_i = \Phi_i H^{(1)} \tilde{\epsilon}_t^{(1)}$, where $\tilde{\epsilon}_t^{(1)}$ is an $r \times 1$ vector that gives the size of the shock. One common way of producing these IRFs is to set the jth element of $\tilde{\epsilon}_t^{(1)}$ to 1 and the other elements of $\tilde{\epsilon}_t^{(1)}$ to zero, yielding responses to a one standard deviation change in the jth structural shock of interest. A second common way of producing these IRFs is to normalize the size of the structural shock so that one of the variables in y_t responds by s on impact. That is, the jth element of $\tilde{\epsilon}_t^{(1)}$ is set to $s/(e_i' H e_j)$, where e_i and e_j are the ith and jth columns of I_K, and the other elements of $\tilde{\epsilon}_t^{(1)}$ are zero. For example, MR normalize their shocks so that the APITR or ACITR fall by -1 on impact. It is important to note that both the one standard deviation and the normalized IRFs are smooth functions of $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$.

Despite our focus on structural IRFs, there are number of other statistics that our theory can be applied to. For example, many papers in the SVAR literature study forecast error variance decompositions, and Mertens and Ravn (2014) study the elements of H itself.

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*

$$\sum_{n=1}^{\infty} (\alpha(n))^{\delta/(2+\delta)} < \infty \tag{19}$$

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*

$$\tau_{a,b,c} = \mathbb{E}\left(vec(u_t u_{t-a}') vec(u_{t-b} u_{t-c}')' \right), \tag{20}$$

$$\nu_{a,b,c} = \mathbb{E}\left(vec(m_t u_{t-a}') vec(u_{t-b} u_{t-c}')' \right), \tag{21}$$

$$\zeta_{a,b,c} = \mathbb{E}\left(vec(m_t u_{t-a}') vec(m_{t-b} u_{t-c}')' \right), \tag{22}$$

use $\widetilde{K} = K(K+1)/2$ and assume that the $(K^2 m + \widetilde{K} + Kr \times K^2 m + \widetilde{K} + Kr)$ 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 dependent, but uncorrelated stationary innovation processes, allowing for conditional heteroskedasticity. In addition, the proxy variables $(m_t, t \in \mathbb{Z})$ may show rather general serial dependence.

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*

$$\sqrt{T}\begin{pmatrix}\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\\ \widehat{\boldsymbol{\sigma}}-\boldsymbol{\sigma}\\ \widehat{\boldsymbol{\varphi}}-\boldsymbol{\varphi}\end{pmatrix}\overset{\mathcal{D}}{\to}\mathcal{N}(0,V),$$

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

$$V=\begin{pmatrix}V^{(1,1)} & V^{(2,1)\prime} & V^{(3,1)\prime}\\ V^{(2,1)} & V^{(2,2)} & V^{(3,2)\prime}\\ V^{(3,1)} & V^{(3,2)} & V^{(3,3)}\end{pmatrix}$$

with

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

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

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

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

$$V^{(3,2)}=\left(\sum_{h=-\infty}^{\infty}\left\{\nu_{0,h,h}-vec(\Psi H^{(1)\prime})vec(\Sigma_u)'\right\}\right)L_K'$$

$$V^{(3,3)}=\sum_{h=-\infty}^{\infty}\left\{\zeta_{0,h,h}-vec(\Psi H^{(1)\prime})vec(\Psi H^{(1)\prime})'\right\}$$

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.

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) or an iid structure.

Corollary 2.1

(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

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

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

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

We note here that in the three cases of α-mixing, mds and iid, the proxy variables are are allowed to be dependent on the process (u_t). In particular, the DGP proposed by MR for the proxy variables,

$$m_t = d_t(\Pi \epsilon_t^{(1)} + v_t), \tag{23}$$

is also 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). 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, MR assume $\mathbb{E}(v_t \epsilon_t^{(1)\prime}) = 0$ and $\mathbb{E}(d_t v_t \epsilon_t^{(1)\prime}) = 0$. If the process $((\epsilon_t', v_t', d_t')', t \in \mathbb{Z})$ is iid, the sequences $(m_t, t \in \mathbb{Z})$ and consequently $(x_t, t \in \mathbb{Z})$ will also be iid sequences such that part (ii) of Corollary 2.1 applies.

Because the limiting distributions in Corollary 2.1 and particularly in Theorem 2.1 can be very complicated, we discuss suitable bootstrap methods for inference in the next section.

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 the VAR coefficients, the covariance matrix of the VAR innovations, and the covariance matrix of the proxy variables with the VAR innovations, $\mathbb{E}(m_t u_t') = \Psi H^{(1)\prime}$. These statistics include $H^{(1)}$ itself as well as structural IRFs and forecast error variance decompositions.

The rest of this section proceeds as follows. In Section 3.1 we describe two residual-based bootstrap algorithms used for inference. The first bootstrap algorithm is a recursive-design wild bootstrap, which was used by MR and has become standard in the proxy SVAR literature. The second bootstrap 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 discuss inconsistency of the wild bootstrap and establish the asymptotic validity of the MBB. In Section 3.3, we use Monte Carlo simulations to compare the wild bootstrap and the MBB for inference on IRFs from structural SVARs.

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 $(m_1^*, \ldots, m_T^*)$.
3. Set the initial condition $(y_{-p+1}^*, \ldots, y_0^*) = (y_{-p+1}, \ldots, y_0)$.[4]
4. Use the initial condition from the previous step along with $\widehat{A}_1, \ldots, \widehat{A}_p$ and u_t^* to recursively produce $(y_1^*, \ldots, y_T^*)$ with
$$y_t^* = \widehat{A}_1 y_{t-1}^* + \cdots + \widehat{A}_p y_{t-p}^* + u_t^*.$$
5. Estimate $\widehat{A}_1^*, \ldots, \widehat{A}_p^*$ by least squares from the bootstrap sample $(y_{-p+1}^*, \ldots, y_T^*)$ and set $\widehat{u}_t^* = y_t^* - \widehat{A}_1^* y_{t-1} - \cdots - \widehat{A}_p^* y_{t-p}$.
6. 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 $\widehat{\Psi H^{(1)\prime}}^* = T^{-1} \sum_{t=1}^T m_t^* \widehat{u}_t^{*\prime}$.
7. Use $\widehat{A}_1^*, \ldots, \widehat{A}_p^*$, $\widehat{\Sigma}_u^*$ and $\widehat{\Psi H^{(1)\prime}}^*$ to produce the bootstrap statistics of interest.

Repeat the algorithm a large number of times and collect the bootstrap statistics of interest. To be comparable to MR, we produce our confidence intervals with a standard percentile

[4] This is the initial condition in the algorithm used by Mertens and Ravn (2013), and we use it in order to keep our results comparable.

interval. That is, we sort the bootstrapped statistics of interest and keep the $\alpha/2$- and $1-\alpha/2$-percentiles as the confidence interval, where α is the level of significance. Because the statistic of interest in this paper is the structural IRF, we compute $\widehat{\Theta}_i^*$, described in Section 2.2, in step 7 of the bootstrap algorithm.

MR draw the sequence $(\eta_t, t \in \mathbb{Z})$ from a Rademacher distribution where $\eta_t = 1$ with probability 0.5 and $\eta_t = -1$ with probability 0.5. This is the standard wild bootstrap algorithm in the proxy SVAR literature. However, another common option is to draw $(\eta_t, t \in \mathbb{Z})$ from a standard normal distribution. We will study both of these methods further below.

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 $\mathcal{U}_i = (\widehat{u}_i, \ldots, \widehat{u}_{i+\ell-1})$ for $i = 1, \ldots T-\ell+1$ and the $r \times \ell$ blocks $\mathcal{M}_i = (m_i, \ldots, m_{i+\ell-1})$ for $i = 1, \ldots T-\ell+1$. Then,

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

2. Collect the blocks $(\mathcal{U}_{i_1}, \ldots, \mathcal{U}_{i_N})$ and $(\mathcal{M}_{i_1}, \ldots, \mathcal{M}_{i_N})$ and drop the last $N\ell - T$ elements to produce $(\widetilde{u}_1^*, \ldots, \widetilde{u}_T^*)$ and $(\widetilde{m}_1^*, \ldots, \widetilde{m}_T^*)$.

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

$$u_{j\ell+s}^* = \widetilde{u}_{j\ell+s}^* - \frac{1}{T-\ell+1} \sum_{r=1}^{T-\ell} \widehat{u}_{s+r-1} \tag{24}$$

   for $s = 1, \ldots, \ell$ and $j = 0, 1, \ldots, N-1$ in order to produce $(u_1^*, \ldots, u_T^*)$.

4. Center $(\widetilde{m}_1^*, \ldots, \widetilde{m}_T^*)$ similarly to the VAR errors in Equation (24) in order to produce $(m_1^*, \ldots, m_T^*)$.[5]

5. Set the initial condition $(y_{-p+1}^*, \ldots, y_0^*) = (y_{-p+1}, \ldots, y_0)$.

6. Use the initial condition from the previous step along with $\widehat{A}_1, \ldots, \widehat{A}_p$ and u_t^* to recursively produce $(y_1^*, \ldots, y_T^*)$ with

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

7. 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]Because of the censoring in the MR proxy variables, we only apply the centering to the non-censored observations and leave the censored proxy variables with a value of zero.

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

9. Use $\widehat{A}_1^*, \ldots, \widehat{A}_p^*$, $\widehat{\Sigma}_u^*$ and $\widehat{\Psi H^{(1)\prime}}^*$ to produce the bootstrap 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 is the case in MR, 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 we will discuss in Section 4, this is not a relevant issue in practice.

3.2 Asymptotic Bootstrap Theory for Proxy SVARs

In this subsection, we study the asymptotic properties of the bootstrap algorithms described in the previous subsection. Hence, 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 bootstrap estimators that correspond to $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\varphi}$, respectively.

To derive 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 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*

$$\sum_{h_2,\ldots,h_j=-\infty}^{\infty} |\text{cum}_{\mathbf{a}}(0, h_2, \ldots, h_j)| < \infty \tag{25}$$

holds, where $\text{cum}_{\mathbf{a}}(0, h_2, \ldots, h_j)$ *denotes the jth joint cumulant of $x_{0,a_1}, x_{h_2,a_2}, \ldots, x_{h_j,a_j}$, see e.g. Brillinger (1981). In particular, 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 consistency of wild and pairwise bootstrap methods for univariate AR(∞) processes and in Brüggemann, Jentsch, and Trenkler (2016) to prove consistency of a residual-based block bootstrap for VAR(p) models. In terms of α-mixing conditions, Assumption 3.1 is implied by

$$\sum_{n=1}^{\infty} n^{m-2}(\alpha_x(n))^{\delta/(2m-2+\delta)} < \infty \tag{26}$$

for $m = 8$ if all moments up to order eight of $(x_t, t \in \mathbb{Z})$ exist, see Künsch (1989). For example, GARCH processes are geometrically strong mixing under mild assumptions on the conditional distribution such that the summability condition in Equation (26) 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 $\widehat{\boldsymbol{\varphi}}$ or statistics that are functions of these estimators. To do so, 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*

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

in probability, where $\widetilde{K} = K(K+1)/2$ and $O_{j\times k}$ denotes the $j \times k$ zero matrix.

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.[6] 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). General asymptotic inconsistency of the residual-based wild bootstrap for functions of $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\sigma}}$ and $\widehat{\boldsymbol{\varphi}}$ as e.g. 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, as has been proposed by MR in the proxy SVAR setup, we have $\mathbb{E}^*(\eta_t^4) = \mathbb{E}(\eta_t^4) = 1$ which immediately leads to the following corollary.

Corollary 3.1 *Under the assumptions of Theorem 3.1 and if the (iid) bootstrap multipliers*

[6] The wild bootstrap also remains valid under mds assumptions in a very special and unrealistic scenario where $V^{(2,1)}$ and $V^{(3,1)}$ vanish and $\mathbb{E}(\eta_t^4)$ is accidently such that $V_{WB}^{(i,j)} = V^{(i,j)}$ holds for $i, j = 1, 2$, leading eventually to $V_{WB} = V$.

$(\eta_t, t \in \mathbb{Z})$ *follow a Rademacher distribution, that is* $P(\eta_t = -1) = P(\eta_t = 1) = 0.5$*, we get*

$$V_{WB} = \begin{pmatrix} V_{WB}^{(1,1)} & O_{K^2p\times\widetilde{K}+Kr} \\ O_{\widetilde{K}+Kr\times K^2p} & O_{\widetilde{K}+Kr\times\widetilde{K}+Kr} \end{pmatrix}. \quad (27)$$

A comparison of V_{WB} in Equation (27) 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 be also seen in the Monte Carlo simulations conducted in Section 3.3, the wild bootstrap clearly leads to considerable undercoverage of corresponding bootstrap confidence intervals for structural IRFs.

To see why the wild bootstrap underestimates e.g. the variance of $\Psi H^{(1)\prime}$, we temporarily consider a simpler specification than the VAR and assume that u_t can be observed directly and does not need to be estimated from the VAR. Then, the wild bootstrap estimate of $\Psi H^{(1,1)\prime}$ from Equation (7) is given by

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

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

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

which is simply the non-bootstrapped sample estimate. That is, when u_t is directly observable, the wild bootstrap yields $\widehat{\Psi H^{(1,1)\prime}}^* = \widehat{\Psi H^{(1,1)\prime}}$ for every bootstrap replication and implies that there is no bootstrap uncertainty in the estimate of this covariance at all.

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 such that the bootstrapped variance of $H^{(1)}$ will generally not be zero in finite samples. However, the bootstrap variance of $\widehat{H}^{(1)}$ 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 get the following theorem.

Theorem 3.2 (Residual-based MBB Consistency) *Suppose Assumptions 2.1 and 3.1 hold and the residual-based MBB bootstrap 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*

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

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

As noted above, $\Theta_i = (\theta_{jk,i})$ are the $K \times r$ structural IRFs of interest. In the following, we refer to the parameters Θ_i simply as IRFs. Further, let $\theta_{jk,i}$ be the response of the j-th variable to the k-th structural shock of interest that occurred i periods ago for $j = 1, \dots, K$ and $k = 1, \dots, r$. To simplify notation we suppress the subscripts in the following and simply use θ and $\widehat{\theta}$ to represent a specific structural impulse response coefficient and its estimator, respectively. For both the one standard deviation and normalized IRFs, Θ_i, $i = 0, 1, 2, \dots$, 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 in the proxy SVAR framework is easily implied by Theorem 3.1 and by the Delta method to get the following corollary.

Corollary 3.2 (Asymptotic Validity of Bootstrap IRFs 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*

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

in probability.

3.3 Monte Carlo Simulations

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

$$y_t = \begin{bmatrix} 0.2 & 0 \\ 0.5 & 0.5 \end{bmatrix} y_{t-1} + u_t, \quad \mathbb{E}(u_t u_t') = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}, \tag{28}$$

which closely follows the DGP used in Kilian (1998). To simulate the VAR innovations, u_t follows Equation (2) where ϵ_t is the bivariate structural shock and

$$H = \begin{bmatrix} 0.592 & -0.806 \\ -0.592 & -0.806 \end{bmatrix}.$$

In the iid simulation, each element of ϵ_t is an independent standard normal random variable. In the GARCH(1,1) simulation, each element of ϵ_t is independent and follows

$$\epsilon_t^{(i)} = g_t^{(i)} w_t^{(i)},$$

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, $\gamma_1 = 0.05$, $\gamma_2 = 0.90$, and $\gamma_0 = 1 - \gamma_1 - \gamma_2$, which follows specification G2 in Brüggemann, Jentsch, and Trenkler (2016).

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

$$m_t = \Pi\epsilon_t^{(1)} + v_t,$$

where v_t is a standard normal random variable. This corresponds to model (23) without censoring, and we use $\Pi = 0.5$.

For each bootstrap method, we run the Monte Carlo simulations with effective sample sizes of 100, 250, 500 and 1000. 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 the iid DGP, we use $y_0 = 0$ and u_t to recursively generate $T + 1000$ observations of y_t. In the GARCH(1,1) DGP, we use $y_0 = 0$, $(g_0^{(i)})^2 = 1$ for $i = 1, 2$, and $(\epsilon_0^{(i)})^2 = 1$ for $i = 1, 2$ to recursively generate $T + 1000$ observations of ϵ_t and y_t. We then drop the first 999 observations of y_t to get a sample of length T plus one pre-sample value denoted by $y_0, y_1, \ldots, y_T$.

For the residual-based MBB, we use block lengths of 16, 20, 24, and 28 for the sample sizes 100, 250, 500, and 1000, respectively. These block lengths yield $N = [7, 13, 21, 36]$. From Theorem 3.2, these block lengths must satisfy $\ell \to \infty$ and $\ell^3/T \to 0$ as $T \to \infty$. Because of this, we follow the rule $\ell = \kappa T^{1/4}$ where κ is normalized so that a sample size of $T = 250$ corresponds exactly to $\ell = 20$. This yields $\kappa = 5.03$.

For each Monte Carlo trial, we use 1000 simulations and 2000 bootstrap replications. Then, we compute the coverage rate of a confidence interval to be the fraction of simulations where the true IRF lies within the confidence interval. Tables 1 and 2 show the coverage rates for the 68% and 95% confidence intervals, respectively, for each bootstrap method and all four sample sizes. In both tables, we display the coverage rates for the first 6 impulse responses, and the IRFs are produced with a one standard deviation change in $\epsilon_t^{(1)}$.[7] Finally, we use $y_t^{(1)}$ and $y_t^{(2)}$ to denote the first and second elements of y_t, respectively.

[7] We show the coverage rates of IRFs where the shock has been normalized so that the response of $y_t^{(1)}$ is -1 on impact in Appendix C. Those coverage rates are very similar to Tables 1 and 2 except that the coverage rate of the initial response of $y_t^{(1)}$ is 1 by construction.

Table 1: Coverage Rates of the 68% Confidence Intervals

Moving Block Bootstrap (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.58	0.58	0.63	0.62	0.66	0.66	0.66	0.66
1	0.61	0.61	0.65	0.62	0.64	0.67	0.67	0.66
2	0.70	0.62	0.69	0.61	0.67	0.66	0.68	0.68
3	0.74	0.63	0.71	0.61	0.71	0.65	0.71	0.68
4	0.74	0.61	0.73	0.64	0.73	0.66	0.70	0.67
5	0.78	0.58	0.76	0.62	0.75	0.67	0.72	0.67

Moving Block Bootstrap (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.58	0.55	0.61	0.60	0.63	0.64	0.63	0.65
1	0.60	0.60	0.64	0.61	0.64	0.66	0.66	0.67
2	0.69	0.63	0.69	0.62	0.67	0.66	0.68	0.67
3	0.73	0.63	0.72	0.62	0.72	0.66	0.70	0.67
4	0.74	0.62	0.73	0.64	0.73	0.67	0.70	0.68
5	0.78	0.58	0.76	0.62	0.75	0.66	0.73	0.67

Wild Bootstrap - Rademacher (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.11	0.13	0.08	0.06	0.05	0.06	0.04	0.03
1	0.67	0.37	0.64	0.31	0.63	0.37	0.65	0.34
2	0.70	0.48	0.70	0.43	0.68	0.45	0.69	0.46
3	0.70	0.56	0.68	0.48	0.67	0.49	0.67	0.50
4	0.71	0.57	0.66	0.51	0.65	0.52	0.65	0.51
5	0.74	0.53	0.69	0.52	0.68	0.52	0.68	0.52

Wild Bootstrap - Rademacher (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.11	0.12	0.07	0.05	0.05	0.06	0.05	0.04
1	0.66	0.38	0.65	0.31	0.64	0.37	0.63	0.35
2	0.70	0.48	0.70	0.43	0.68	0.46	0.68	0.46
3	0.70	0.56	0.68	0.49	0.68	0.50	0.67	0.50
4	0.70	0.56	0.67	0.52	0.65	0.53	0.66	0.52
5	0.75	0.52	0.69	0.51	0.67	0.53	0.67	0.51

Wild Bootstrap - Normal (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.82	0.84	0.83	0.83	0.84	0.84	0.83	0.84
1	0.73	0.81	0.72	0.81	0.71	0.81	0.72	0.81
2	0.85	0.78	0.77	0.78	0.73	0.79	0.73	0.81
3	0.86	0.79	0.84	0.77	0.80	0.76	0.78	0.81
4	0.87	0.78	0.85	0.77	0.81	0.76	0.78	0.79
5	0.89	0.72	0.85	0.74	0.83	0.76	0.78	0.78

Wild Bootstrap - Normal (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.80	0.80	0.82	0.80	0.82	0.81	0.80	0.81
1	0.72	0.81	0.72	0.81	0.71	0.81	0.71	0.82
2	0.84	0.79	0.77	0.77	0.73	0.77	0.72	0.80
3	0.87	0.80	0.85	0.77	0.78	0.76	0.76	0.79
4	0.88	0.76	0.86	0.78	0.82	0.76	0.78	0.78
5	0.90	0.71	0.85	0.74	0.82	0.76	0.78	0.78

Table 2: Coverage Rates of the 95% Confidence Intervals

Moving Block Bootstrap (i.i.d. DGP):

	$T = 100$		$T = 250$		$T = 500$		$T = 1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.88	0.90	0.92	0.92	0.93	0.94	0.93	0.93
1	0.90	0.90	0.92	0.93	0.94	0.92	0.94	0.94
2	0.97	0.90	0.95	0.92	0.94	0.92	0.93	0.94
3	0.97	0.92	0.98	0.93	0.97	0.92	0.96	0.93
4	0.97	0.91	0.97	0.93	0.97	0.92	0.97	0.94
5	0.98	0.86	0.97	0.92	0.96	0.93	0.96	0.94

Moving Block Bootstrap (GARCH DGP):

	$T = 100$		$T = 250$		$T = 500$		$T = 1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.87	0.87	0.91	0.91	0.92	0.92	0.92	0.92
1	0.90	0.90	0.92	0.93	0.93	0.93	0.93	0.94
2	0.97	0.91	0.95	0.92	0.95	0.92	0.94	0.94
3	0.96	0.92	0.97	0.93	0.97	0.93	0.96	0.94
4	0.97	0.91	0.97	0.93	0.97	0.93	0.97	0.94
5	0.98	0.86	0.97	0.92	0.96	0.93	0.96	0.93

Wild Bootstrap - Rademacher (i.i.d. DGP):

	$T = 100$		$T = 250$		$T = 500$		$T = 1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.29	0.28	0.18	0.16	0.12	0.13	0.10	0.07
1	0.92	0.66	0.92	0.59	0.93	0.62	0.92	0.60
2	0.97	0.79	0.95	0.75	0.94	0.76	0.94	0.76
3	0.95	0.85	0.96	0.80	0.94	0.80	0.94	0.81
4	0.97	0.87	0.96	0.84	0.94	0.82	0.94	0.81
5	0.99	0.83	0.96	0.84	0.95	0.84	0.94	0.83

Wild Bootstrap - Rademacher (GARCH DGP):

	$T = 100$		$T = 250$		$T = 500$		$T = 1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.28	0.28	0.17	0.16	0.12	0.14	0.09	0.09
1	0.91	0.66	0.92	0.60	0.92	0.62	0.92	0.61
2	0.97	0.79	0.95	0.75	0.94	0.74	0.94	0.76
3	0.94	0.84	0.96	0.80	0.95	0.80	0.95	0.81
4	0.97	0.86	0.96	0.85	0.95	0.83	0.93	0.82
5	0.99	0.83	0.96	0.86	0.95	0.85	0.95	0.83

Wild Bootstrap - Normal (i.i.d. DGP):

	$T = 100$		$T = 250$		$T = 500$		$T = 1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.99	1.00	1.00	0.99	0.99	1.00	0.99	0.99
1	0.97	0.99	0.95	0.99	0.96	0.99	0.96	0.99
2	1.00	0.99	0.98	0.99	0.97	0.99	0.96	0.99
3	1.00	0.99	1.00	0.99	1.00	0.99	0.99	0.99
4	1.00	0.98	0.99	0.99	0.99	0.99	0.99	0.98
5	1.00	0.97	0.99	0.98	0.98	0.98	0.98	0.98

Wild Bootstrap - Normal (GARCH DGP):

	$T = 100$		$T = 250$		$T = 500$		$T = 1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	0.98	0.99	0.99	0.99	0.99	0.99	0.99	0.99
1	0.96	0.99	0.95	0.99	0.96	0.99	0.95	0.99
2	1.00	0.99	0.98	0.99	0.97	0.99	0.95	0.98
3	1.00	0.98	1.00	0.99	1.00	0.99	0.99	0.98
4	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.98
5	1.00	0.97	0.99	0.98	0.98	0.98	0.98	0.98

Both tables show that the Rademacher wild bootstrap produces coverage rates that can be very small, particularly at impact. This is true for both DGPs and both the 68% and 95% confidence intervals. For $T = 250$, which most closely corresponds to MR's effective sample size of 224, the coverage rates for the initial response for the 68% confidence interval are only between 5% and 8%. Further, the 95% confidence intervals only have coverage rates of 16% and 18% at a sample size of 250. In addition, the coverage of the initial impulse responses gets worse as the sample size increases, indicating that more data worsens the Rademacher wild bootstrap.

As discussed in Section 3.2, the Rademacher wild bootstrap ignores the uncertainty in estimating Equations (3), (7) and (8). Hence, it ignores the uncertainty in estimating $H^{(1)}$, yielding the very small coverage for the initial responses. With this in mind, two additional results are worth noting. First, after the initial response, the coverage rates from the Rademacher wild bootstrap quickly approach the target coverage rates for $y_t^{(1)}$. This is because $y_t^{(1)}$ is essentially an AR(1) with a coefficient near zero leading to corresponding entries in the VMA coefficient matrices Φ_j in (16) that are quickly vanishing. For the DGP in (28), the first four VMA coefficient matrices are

$$\Phi_1 = \begin{bmatrix} 0.20 & 0.00 \\ 0.50 & 0.50 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} 0.04 & 0.00 \\ 0.35 & 0.25 \end{bmatrix}, \quad \Phi_3 = \begin{bmatrix} 0.01 & 0.00 \\ 0.20 & 0.13 \end{bmatrix}, \quad \Phi_4 = \begin{bmatrix} 0.00 & 0.00 \\ 0.10 & 0.06 \end{bmatrix},$$

and the wild bootstrap correctly measures the uncertainty around these coefficients in the first row of Φ_j (Gonçalves and Kilian, 2004). Thus, $H^{(1)}$ has little influence on $y_t^{(1)}$'s multi-step ahead IRFs. Second, the Rademacher wild bootstrap's small coverage rates for $y_t^{(2)}$ are more persistent. This is because $y_t^{(2)}$ has persistently larger VMA coefficients than $y_t^{(1)}$, implying that $H^{(1)}$ has more persistent influence on the IRFs of $y_t^{(2)}$ than $y_t^{(1)}$. Hence, the Rademacher wild bootstrap has the potential to produce persistently small coverage rates when the VMA coefficients are persistently large.

In contrast to the Rademacher wild bootstrap, the normal wild bootstrap produces coverage rates that are too large. For $T = 250$, the coverage rates for the initial response for the 68% confidence interval are between 80% and 83%. Further, the 95% confidence intervals are over 99% at a sample of 250. In addition, these initial coverage rates stay too large as the sample size increases and are generally too large for the multi-step ahead IRFs. In addition to the asymptotic invalidity of wild bootstraps established above, these results show that wild bootstraps can have huge variation in coverage rates in practice depending on the distribution of the bootstrap multipliers. Hence, wild bootstraps should not be used for inference on structural IRFs.

In contrast to both wild bootstraps, the MBB is proven to be asymptotically valid for inference on the structural IRFs. Tables 1 and 2 indicate that this bootstrap produces modestly undersized coverage rates for the initial impulse responses. However, unlike the wild bootstraps, these coverage rates generally improve as the sample size increases. Further, for the sample size 250, the MBB's coverage rates are generally better than those from the normal wild bootstrap and much better than those from the Rademacher wild bootstrap.

For the iid DGP, the MBB produces undersized coverage rates for the initial response because the block-wise resampling produces less variation than in a standard iid bootstrap, yielding somewhat smaller confidence intervals. However, the requirement that $\ell^3/T \to 0$ as $T \to \infty$ implies that this is not a problem asymptotically and explains the MBB's improving coverage rates with the sample size. Further, an iid bootstrap would not capture the nonlinear dependence structure in the GARCH DGP and produce even smaller coverage rates than the MBB.[8] In practice, this leads to a trade-off for the choice of block length, ℓ. Blocks that are too small will not be able to effectively account for dependence of the VAR innovations or proxy variables, but blocks that are too large will not produce enough bootstrap variation. For example, a block length of 1 will produce the appropriate coverage rates for iid innovations, but will produce coverage rates that are too small for GARCH innovations. As noted above, we set a block length of 20 for a sample size of 250, striking a balance between getting good coverage in the presence of conditional heteroskedasticity without producing coverage rates that are too low in the iid case.

4 The Effects of Tax Changes in the United States

As an application of the residual-based MBB, we recreate several figures from MR using the MBB instead of their Rademacher wild bootstrap. MR study the dynamic effects of two types of tax changes on the U.S. economy: average personal income tax rates (APITRs) and average corporate income tax rates (ACITRs). To do this, they construct narrative accounts of both shocks by decomposing Romer and Romer's (2009) narrative account of postwar tax changes (see Figure 1 of MR). Then, they use these narrative accounts as proxy variables for the tax shocks in a SVAR as described in Section 2. For these recreations, we use MR's replication files from the American Economic Association's website at `https://www.aeaweb.org/articles.php?doi=10.1257/aer.103.4.1212`.

We begin with MR's baseline specification, which includes the APITR, the ACITR, the logarithm of the personal income tax base (PITB), the logarithm of the corporate income tax base (CITB), the logarithm of government spending, the logarithm of GDP divided by population, and the logarithm of government debt held by the public divided by the GDP deflator and population, giving $K = 7$.[9] The data are quarterly from 1950:Q1 to 2006:Q4. We include a constant in the VAR, and we estimate the VAR with $p = 4$, giving an effective

[8]For example, see Figure 2 of Brüggemann, Jentsch, and Trenkler (2016).

[9]In their replication files, MR define the APITR as federal personal income tax revenues including contributions to government social insurance divided by personal income tax base, the ACITR as federal corporate income tax revenues divided by corporate income tax base, the PITB as personal income less government transfers plus contributions to government social insurance divided by GDP deflator and by population, the CITB as corporate profits less Federal Reserve Bank profits divided by GDP deflator and by population, and government spending as real Federal government consumption and investment expenditures divided by population.

sample size of 224. Given this sample size, we use a block length of $\ell = 19$ based on the rule used for the Monte Carlo simulations above.[10] Because we have two proxy variables and are estimating two structural shocks, $r = 2$.

Following MR, we consider two orderings for the variables. In the first ordering, the APITR comes before the ACITR, and in the second ordering, the ACITR comes before the APITR. As discussed at the end of Section 2.1, this ordering influences how the shocks will impact one another. Figure 1 displays the IRFs of a shock to personal taxes that has been normalized to a 1% cut in the APITR for both orderings along with the 68% confidence intervals from the MBB with 10,000 replications. Blue solid lines are the point estimate when ordering the APITR first, and the blue dashed lines are the corresponding confidence intervals. Red diamonds are the point estimate when ordering the ACITR first, and the red dashed lines are the corresponding confidence intervals. This figure parallels Figure 2 in MR but with two important differences. First, the confidence intervals here are produced with the MBB. Second, the confidence intervals presented by MR are 95% intervals – not the 68% intervals presented here. We present the 68% intervals here because the 95% intervals with the MBB are huge, and no inferences can be drawn from them. We show the 95% confidence intervals in Appendix C.

When using the wild bootstrap, MR find that a cut to the APITR causes an increase in output, an increase in the PITB, and a decrease in personal income tax revenues that are statistically significant with 95% confidence intervals. With the MBB, despite the smaller confidence level, the confidence intervals here are larger than those in MR. Because of this, Figure 1 indicates that no inference can be made about the effects of an APITR cut on output or the PITB, even at a 68% level. However, personal income tax revenues do fall with statistical significance at impact.[11] Further, unlike in MR where the confidence intervals were similar for both orderings, Figure 1 indicates that the confidence intervals can be quite different depending on the ordering. For example, when the ACITR is ordered first, the confidence interval for output is much narrower. At the third step in the IRF, which gives the peak point estimate, this confidence interval is [-0.3%, 2.8%]. In contrast, the confidence interval when the APITR is ordered first is [-2.9%, 4.0%].

Figure 2 displays the IRFs of a shock to corporate taxes that has been normalized to a 1% cut in the ACITR for both orderings along with the 68% confidence intervals from the MBB with 10,000 replications. Blue solid lines are the point estimate when the APITR is ordered first, and the blue dashed lines are the corresponding confidence intervals. Red diamonds are the point estimate when the ACITR is ordered first, and the red dashed lines are the corresponding confidence intervals. This figure parallels Figure 3 in MR but with

[10] Setting $\ell = 1$ yields an iid bootstrap. This produces modestly smaller confidence intervals than the MBB with $\ell = 19$ when applied to MR. However, autocorrelations of $|\widehat{u}_t^{(j)}|$ and $(\widehat{u}_t^{(j)})^2$ provided in a table in Appendix C yield several values that are statistically distinct from zero. This indicates some (non-linear) dependence is present in the residuals the requires the use of the MBB.

[11] Personal income tax revenues are not a variable included in the VAR. Rather, following MR, we compute them from the other IRFs as APITR/0.1667 + PITB.

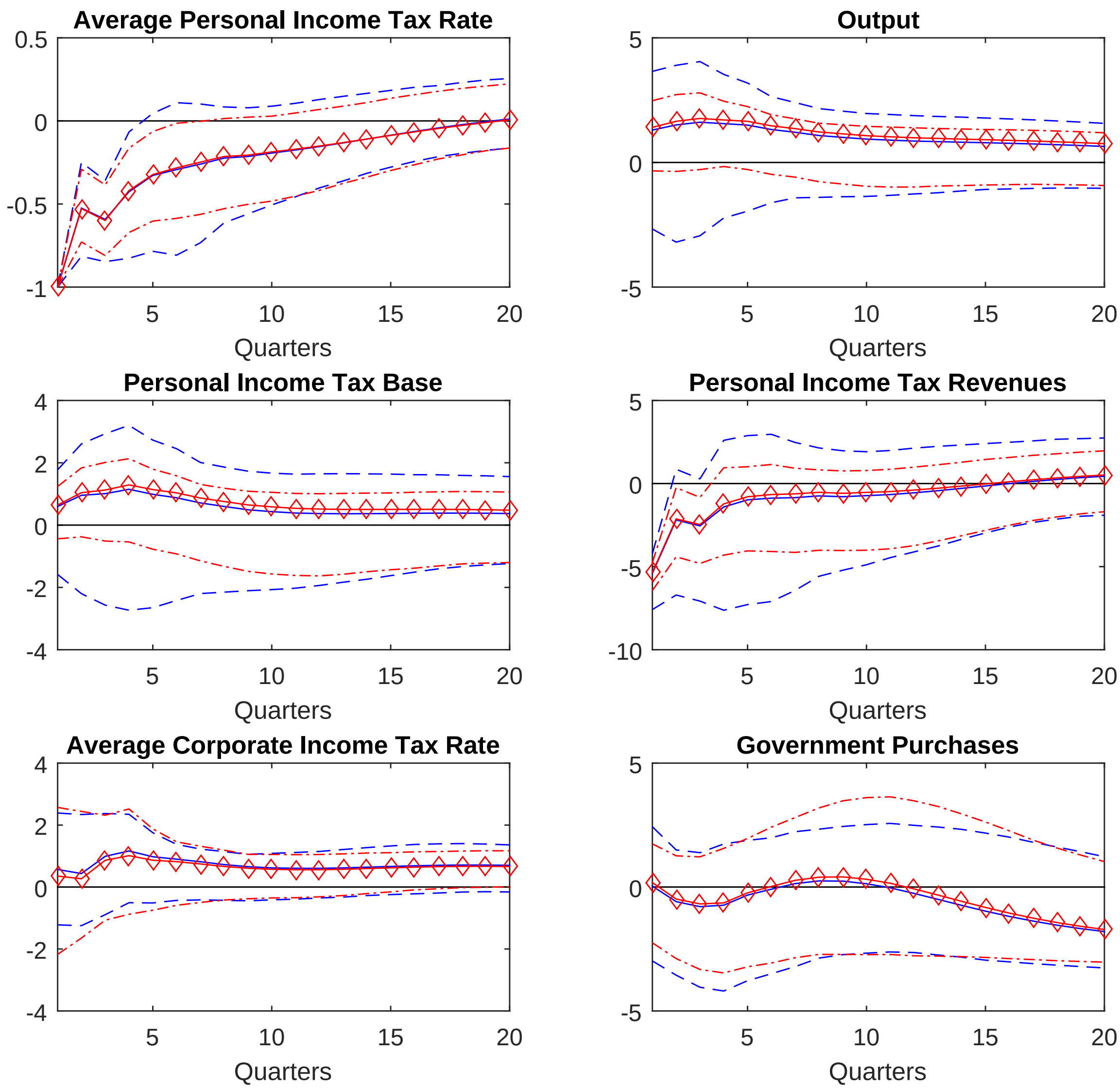

Figure 1: IRFs of a 1% cut in the APITR. Blue lines show the model with the APITR ordered first, and red diamonds show the model with the ACITR ordered first. Dashed lines are 68% confidence intervals from the MBB.

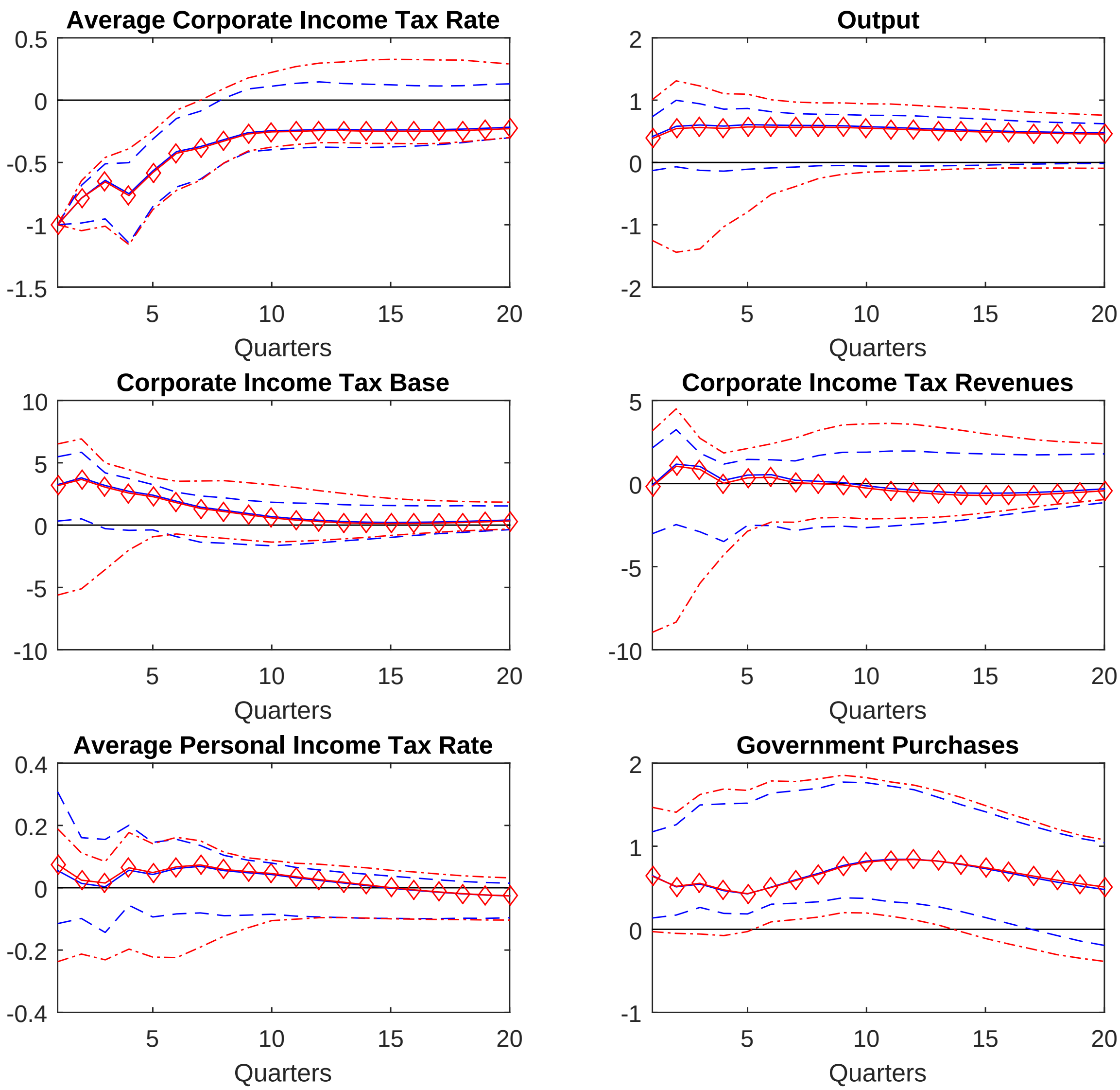

Figure 2: IRFs of a 1% cut in the ACITR. Blue lines show the model with the APITR ordered first, and red diamonds show the model with the ACITR ordered first. Dashed lines are 68% confidence intervals from the MBB.

the same differences as Figure 1 above. The same figure with 95% confidence intervals can be found in Appendix C.[12]

When using the wild bootstrap, MR find that a cut to the ACITR causes increases in output and the CITB that are statistically significant with 95% confidence intervals. However, with the MBB, no inference can be made about the effects of an ACITR cut on output. Further, as with the APITR cut, the confidence intervals surrounding output in Figure 2 are noticeably different depending on the ordering of the variables. With the MBB, inference about the effect of an ACITR cut on the CITB is ambiguous. When the APITR is ordered first, it appears that the CITB increases in the first two quarters after the shock by a statistically significant amount. However, this result disappears when the ACITR is ordered first. Further, as shown in Appendix C, this result is insignificant for both orderings at the 95% level. Finally, Figure 2 suggests that a cut to the ACITR will increase government purchases by a statistically significant amount after a year and a half. However, as in MR, this result disappears at the 95% level.

As noted in Section 3, a potential drawback of the MBB is that the possibility exists for all of the bootstrapped proxy variable observations to be censored to zero. This problem arises because many of the observations in MR's narrative account are censored to zero. Of the 224 observed proxy APITR shocks, 211 (94%) are zero. For the ACITR, this number is 208 (93%). Further, there can be many periods between non-censored proxies. The largest gap between non-censored proxies is 39 quarters for the APITR and 49 quarters for the ACITR. Because of this, a block length of 19 does not guarantee that a non-censored proxy will be observed in every block. However, given this block length, the MBB algorithm draws 12 blocks and the probability of all 12 blocks having zeros is very small. For the APITR, 153 of the 206 blocks contain at least one non-censored proxy, implying that the probability of drawing a block of zeros is 0.257. Then, the probability of drawing 12 blocks of zeros will be $0.257^{12} \approx 8 \times 10^{-8}$. For the ACITR, 162 of the 206 blocks contain at least one non-censored proxy, yielding a probability of about 9×10^{-9} of drawing 12 blocks of zeros. Hence, even with 10,000 bootstrap replications, the probability of drawing one replication of all zeros is very small, and this is not a relevant drawback in practice. In producing Figures 1 and 2, all 10,000 bootstrap samples contained at least 3 non-zero observations for both the APITR and the ACITR proxies.

Following MR, we also consider the effects of APITR and ACITR cuts on labor market variables, consumption and investment. For the labor market, we follow MR and estimate a VAR with $K = 8$ that includes the APITR, the ACITR, the logarithm of government spending, the logarithm of GDP divided by population, the logarithm of government debt held by the public divided by the GDP deflator and population, the logarithm of total economy employment divided by population, the logarithm of total economy hours worked divided by total economy employment, and the logarithm of labor force divided by population.

The left hand panels of Figure 3 display the IRFs of the labor market variables to a

[12] Note that corporate income tax revenues are not included in the VAR. Following MR, they are computed from the other IRFs as ACITR/0.2996 + CITB.

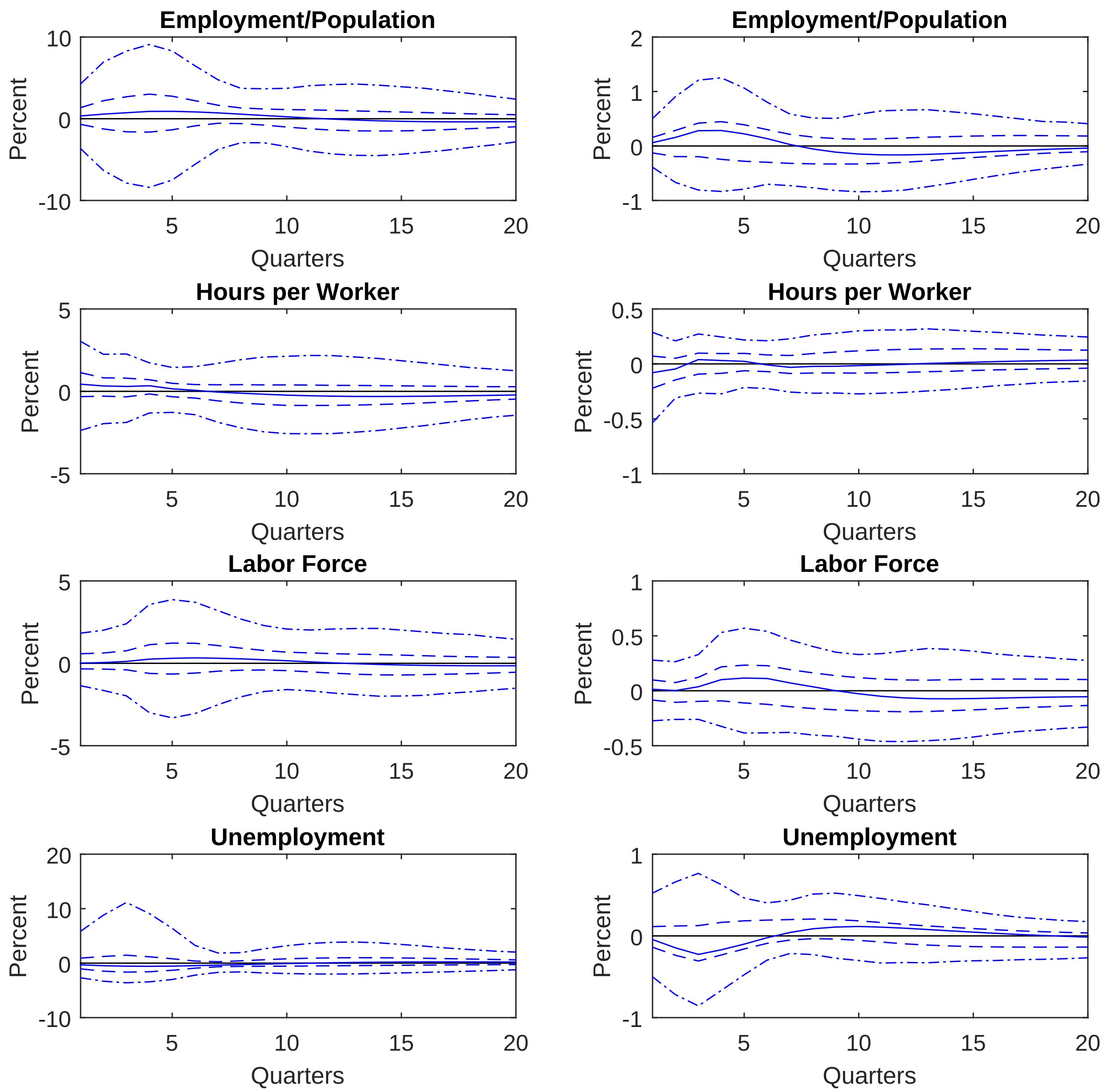

Figure 3: The left hand panels display IRFs of a 1% cut in the APITR. The right hand panels display IRFs of a 1% cut in the ACITR. Blue lines show the point estimates and the dashed lines show the 68% and 90% confidence intervals from the MBB.

1% APITR cut, and the right hand panels display the corresponding IRFs to a 1% ACITR cut. Solid blue lines are the point estimates, and the dashed lines show the 68% and 90% confidence intervals. This figure corresponds to Figure 9 in MR.[13] With the wild bootstrap, MR find statistically significant increases in the employment to population ratio and hours per worker along with a statistically significant decrease in the unemployment rate to an APITR cut. However, Figure 3 shows that no such inferences can be made, even at the 68% level. Rather, as with output, cuts to both the APITR and the ACITR have no inferable impact on the labor market in the United States.

To study the effects of APITR and ACITR cuts on consumption, we estimate a VAR with $K = 8$ that includes the APITR, the ACITR, the logarithm of the PITB, the logarithm of government spending, the logarithm of GDP divided by population, the logarithm of government debt held by the public divided by the GDP deflator and population, the logarithm of chain-aggregated nondurable consumption and service goods consumption divided by population, and logarithm of real durable consumption goods expenditures divided by population. The left hand panels of Figure 4 display the IRFs of consumption and durable goods purchases to a 1% APITR cut, and the right hand panels display the corresponding IRFs to a 1% ACITR cut. Solid blue lines are the point estimates, and the dashed lines show the 68% and 90% confidence intervals. This figure corresponds to the top panels in Figure 10 in MR.[14] With the wild bootstrap, MR find a statistically significant increase in durable goods purchases at both the 90% and 95% levels from an APITR cut. However, Figure 4 shows that no such inferences can be made at the 90% level. However, at the 68% level, a statistically significant positive response begins in quarter 5.

Finally, to study the effects of APITR and ACITR cuts on investment, we estimate a VAR with $K = 8$ that includes the APITR, the ACITR, the logarithm of the CITB, the logarithm of government spending, the logarithm of GDP divided by population, the logarithm of government debt held by the public divided by the GDP deflator and population, the logarithm of real non-residential fixed investment divided by population, and the logarithm of real residential fixed investment divided by population. The left hand panels of Figure 5 display the IRFs of nonresidential and residential investment to a 1% APITR cut, and the right hand panels display the corresponding IRFs to a 1% ACITR cut. Solid blue lines are the point estimates, and the dashed lines show the 68% and 90% confidence intervals. This figure corresponds to the bottom panels in Figure 10 in MR.[15] With the wild bootstrap, MR find a statistically significant increase in nonresidential investment at both the 90% and 95% levels from APITR and ACITR cuts. Further, they find a statistically significant increase in residential investment at both the 90% and 95% levels from an ACITR cut. In contrast, Figure 5 shows no statistically significant response of residential or nonresidential investment

[13] The APITR is ordered first for all IRFs in this figure. Also, the unemployment rate is not included in the VAR. Following MR, we compute it from the other IRFs as $5.25\{\exp[-0.9475(\text{Employment/Pop} - \text{Labor Force/Pop})/5.25] - 1\}$.

[14] The APITR is ordered first for all IRFs in this figure.

[15] The APITR is ordered first for all IRFs in this figure.

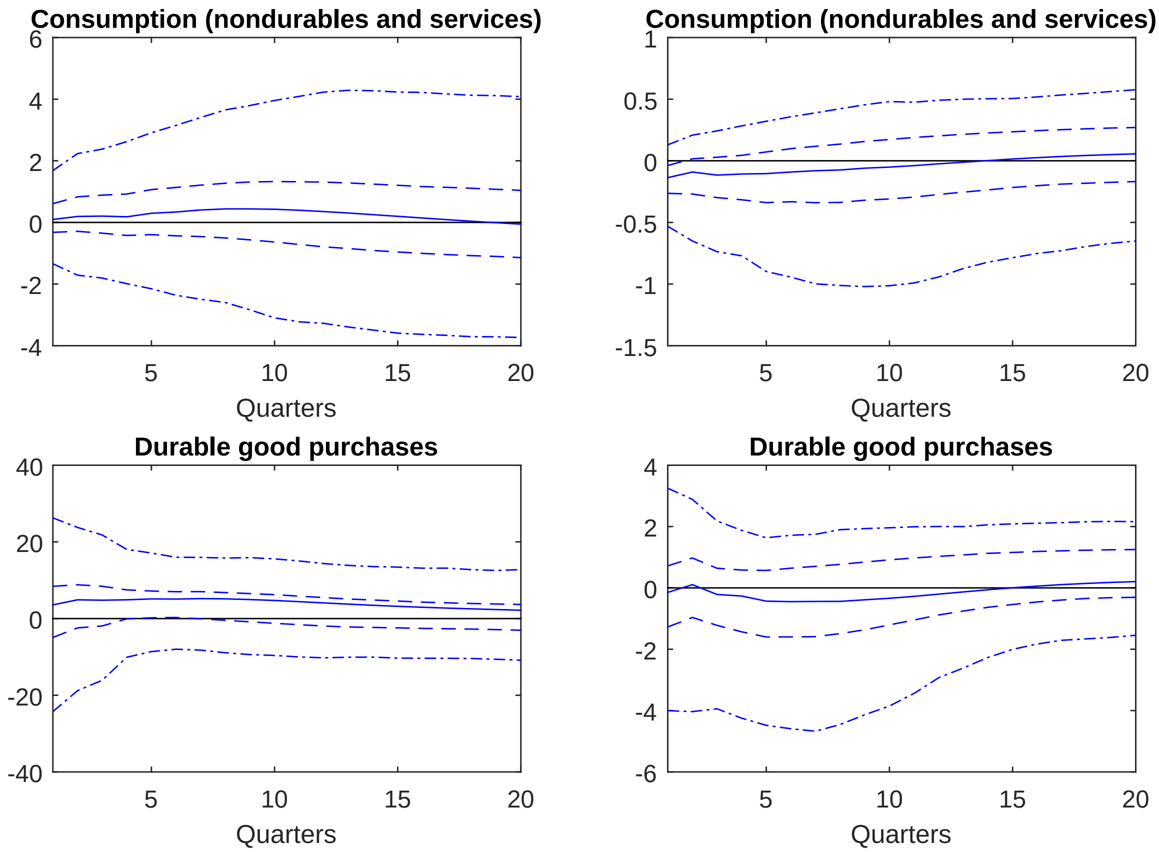

Figure 4: The left hand panels display IRFs of a 1% cut in the APITR. The right hand panels display IRFs of a 1% cut in the ACITR. Blue lines show the point estimates and the dashed lines show the 68% and 90% confidence intervals from the MBB.

to either tax cut.

5 Conclusions

Estimating the dynamic effects of structural shocks from SVARs is important for macroeconomic research. Recently, Stock and Watson (2008, 2012), Montiel Olea, Stock, and Watson (2012), and Mertens and Ravn (2013) developed a method for estimating SVARs that uses an external proxy variable that is correlated with the structural shocks of interest but uncorrelated with the other structural shocks. This paper studies methods for inference when using this proxy SVAR method. First, we provide a joint central limit theorem for the VAR coefficients, the variance matrix of the VAR innovations, and the covariance matrix of the VAR innovations with the proxy variables under mild α-mixing conditions. Second, we prove that a residual-based moving block bootstrap is asymptotically valid for inference on statistics that are smooth functions of the VAR coefficients, the variance matrix of the VAR

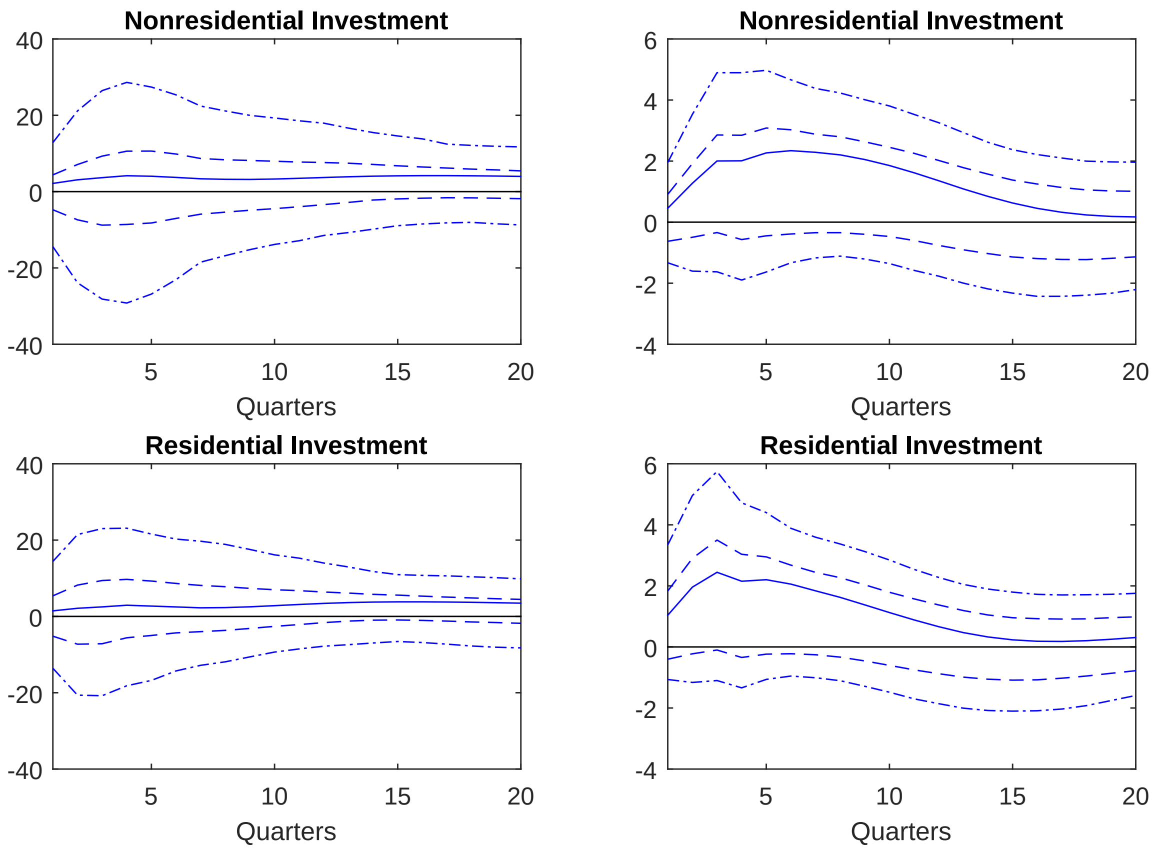

Figure 5: The left hand panels display IRFs of a 1% cut in the APITR. The right hand panels display IRFs of a 1% cut in the APITR. Blue lines show the point estimates and the dashed lines show the 68% and 90% confidence intervals from the MBB.

innovations, and the covariance matrix of the VAR innovations with the proxy variables. In contrast, wild bootstraps are not asymptotically valid for these statistics.

When the moving block bootstrap is applied to Mertens and Ravn (2013), we find that many of their results are no longer statistically significant. Specifically, cuts to both personal and corporate tax rates have no inferable effect on output, investment, employment, hours worked per worker or the unemployment rate. These results suggest that the narrative proxy variables used by MR are not informative enough to discern the dynamic effects of tax changes on economic activity in the United States. However, these results do not imply that proxy SVARs will always be uninformative. Lunsford (2015b) shows that inferences can be made at the 90% level with the moving block bootstrap when using Fernald's (2014) utilization-adjusted total factor productivities as proxy variables. Thus, proxy SVARs and the moving block bootstrap are useful tools for inferring the dynamic effects of structural shocks.

A Derivation of the Estimators

For the purposes of notation, define $\mathbb{E}(m_t u_t^{(1)\prime}) = \Sigma_{mu}^{(1)}$, $\mathbb{E}(m_t u_t^{(2)\prime}) = \Sigma_{mu}^{(2)}$, $\mathbb{E}(u_t^{(1)} u_t^{(1)\prime}) = \Sigma_u^{(1,1)}$, $\mathbb{E}(u_t^{(2)} u_t^{(1)\prime}) = \Sigma_u^{(2,1)}$, and $\mathbb{E}(u_t^{(2)} u_t^{(2)\prime}) = \Sigma_u^{(2,2)}$ to be the moments that can be estimated from the data. Then, Equation (9) can be re-written as

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

Next, Equations (3) and (6) imply

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

$$\Sigma_u^{(2,1)} = H^{(2,1)}H^{(1,1)\prime} + H^{(2,2)}H^{(1,2)\prime}, \tag{A.3}$$

and

$$\Sigma_u^{(2,2)} = H^{(2,1)}H^{(2,1)\prime} + H^{(2,2)}H^{(2,2)\prime}. \tag{A.4}$$

Using Equations (A.2) through (A.4), it is the case that

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

Next, define

$$Z = (H^{(2,2)} - H^{(2,1)}H^{(1,1)-1}H^{(1,2)})(H^{(2,2)} - H^{(2,1)}H^{(1,1)-1}H^{(1,2)})'. \tag{A.5}$$

Then, Z can also be written as

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

That is, Z is also given by

$$Z = \Sigma_u^{(2,2)} - H^{(2,1)} H^{(1,1)-1} \Sigma_u^{(2,1)\prime} - \Sigma_u^{(2,1)} (H^{(2,1)} H^{(1,1)-1})' \\ + H^{(2,1)} H^{(1,1)-1} \Sigma_u^{(1,1)} (H^{(2,1)} H^{(1,1)-1})', \tag{A.6}$$

which allows us to estimate Z from the data. Next, Equation (A.5) implies

$$\begin{aligned} H^{(1,2)} H^{(1,2)\prime} &= H^{(1,2)} [(H^{(2,2)} - H^{(2,1)} H^{(1,1)-1} H^{(1,2)})^{-1} (H^{(2,2)} - H^{(2,1)} H^{(1,1)-1} H^{(1,2)})]' \\ &\quad \times [(H^{(2,2)} - H^{(2,1)} H^{(1,1)-1} H^{(1,2)})(H^{(2,2)} - H^{(2,1)} H^{(1,1)-1} H^{(1,2)})^{-1}] H^{(1,2)\prime} \\ &= (\Sigma_u^{(2,1)} - H^{(2,1)} H^{(1,1)-1} \Sigma_u^{(1,1)})' Z^{-1} (\Sigma_u^{(2,1)} - H^{(2,1)} H^{(1,1)-1} \Sigma_u^{(1,1)}). \end{aligned}$$

That is,

$$H^{(1,2)} H^{(1,2)\prime} = (\Sigma_u^{(2,1)} - H^{(2,1)} H^{(1,1)-1} \Sigma_u^{(1,1)})' Z^{-1} (\Sigma_u^{(2,1)} - H^{(2,1)} H^{(1,1)-1} \Sigma_u^{(1,1)}). \tag{A.7}$$

Then, estimation of the model occurs as follows. First, estimate Equation (A.1). This then allows for the estimation of Equation (A.6), which then allows for the estimation of Equation (A.7). After this, we can estimate

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

which follows from Equation (A.2),

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

which follows from Equation (A.4), and

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

which follows from Equation (A.3).

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

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

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

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

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

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

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

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

□

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*

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

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*

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

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

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

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

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

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 Theorem 3.1

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

$$\mathbb{E}^*\left(\text{vec}(u_t^* u_{t-a}^{*\prime})\text{vec}(u_{t-b}^* u_{t-c}^{*\prime})'\right) = \text{vec}(\widehat{u}_t \widehat{u}_{t-a}')\text{vec}(\widehat{u}_{t-b}\widehat{u}_{t-c}')'\mathbb{E}^*\left(\eta_t \eta_{t-a}\eta_{t-b}\eta_{t-c}\right), \quad \text{(B.10)}$$

where

$$\mathbb{E}^*\left(\eta_t \eta_{t-a}\eta_{t-b}\eta_{t-c}\right) = \begin{cases} \mathbb{E}(\eta_t^4), & a = b = c = 0 \\ 1, & a = 0 \neq b = c \text{ or } b = 0 \neq a = c \text{ or } c = 0 \neq a = b \\ 0, & \text{otherwise} \end{cases}. \quad \text{(B.11)}$$

Note that analogous representations also hold for $\mathbb{E}^*\left(\text{vec}(m_t^* u_{t-a}^{*\prime})\text{vec}(u_{t-b} u_{t-c}^{*\prime})'\right)$ as well as $\mathbb{E}^*\left(\text{vec}(m_t^* u_{t-a}^{*\prime})\text{vec}(m_{t-b} u_{t-c}^{*\prime})'\right)$. Now, by using similar arguments as used in the proof of Theorem 2.1 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.3 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 $\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}} := ((\widetilde{Z}^* \widetilde{Z}^{*\prime})^{-1}\widetilde{Z}^* \otimes I_K)\widetilde{\mathbf{u}}^*$, $\widetilde{\boldsymbol{\sigma}}^* = \text{vech}(\widetilde{\Sigma}_u^*)$ with $\widetilde{\Sigma}_u^* = \frac{1}{T}\sum_{t=1}^T \widetilde{u}_t^* \widetilde{u}_t^{*\prime}$, $\widetilde{\boldsymbol{\sigma}} = \text{vech}(\widetilde{\Sigma}_u)$ with $\widetilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^T u_t u_t'$, $\widetilde{\boldsymbol{\varphi}}^* = \text{vec}(\widetilde{\Psi H^{(1)*\prime}})$ with $\widetilde{\Psi H^{(1)*\prime}} = \frac{1}{T}\sum_{t=1}^T m_t^* \widetilde{u}_t^{*\prime}$ and $\widetilde{\boldsymbol{\varphi}} = \text{vec}(\widetilde{\Psi H^{(1)\prime}})$ 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

$$\begin{aligned} \widetilde{Z}_t^* &= \text{vec}(\widetilde{y}_t^*, \ldots, \widetilde{y}_{t-p+1}^*) \quad (Kp \times 1) \\ \widetilde{Z}^* &= (\widetilde{Z}_0^*, \ldots, \widetilde{Z}_{T-1}^*) \quad (Kp \times T) \\ \widetilde{\mathbf{u}}^* &= \text{vec}(\widetilde{u}_1^*, \ldots, \widetilde{u}_T^*) \quad (KT \times 1). \end{aligned}$$

Similarly to Equation (B.2), we get the representation

$$\sqrt{T}\begin{pmatrix}\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\sigma}}^* - \widetilde{\boldsymbol{\sigma}} \\ \widetilde{\boldsymbol{\varphi}}^* - \widetilde{\boldsymbol{\varphi}}\end{pmatrix} = \begin{pmatrix}\left\{(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\otimes I_K\right\}\frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}(C_j\otimes I_K)\sum_{t=j+1}^{T}\left\{\operatorname{vec}(\widetilde{u}_t^*\widetilde{u}_{t-j}^{*\prime})\right\} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T} L_K\left\{\operatorname{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}) - \operatorname{vec}(u_t u_t^\prime)\right\} \\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{\operatorname{vec}(m_t^*\widetilde{u}_t^{*\prime}) - \operatorname{vec}(m_t u_t^\prime)\right\}\end{pmatrix} \tag{B.12}$$
$$= A_m^* + (A^* - A_m^*),$$

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

$$\begin{aligned}A_m^* &= \begin{pmatrix}(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\otimes I_K & O_{K^2p\times\widetilde{K}} & O_{K^2p\times Kr} \\ O_{\widetilde{K}\ timesK^2p} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr} \\ O_{Kr\times K^2p} & O_{Kr\times\widetilde{K}} & I_{Kr}\end{pmatrix}\begin{pmatrix}C_1\otimes I_K & \cdots & C_m\otimes I_K & O_{K^2p\times\widetilde{K}} & O_{K^2p\times Kr} \\ O_{\widetilde{K}\times K^2} & \cdots & O_{\widetilde{K}\times K^2} & I_{\widetilde{K}} & O_{\widetilde{K}\times Kr} \\ O_{Kr\times K^2} & \cdots & O_{Kr\times K^2} & O_{Kr\times\widetilde{K}} & I_{Kr}\end{pmatrix} \\ &\quad\times\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\begin{pmatrix}\operatorname{vec}(\widetilde{u}_t^*\widetilde{u}_{t-1}^{*\prime}) \\ \vdots \\ \operatorname{vec}(\widetilde{u}_t^*\widetilde{u}_{t-m}^{*\prime}) \\ L_K\left\{\operatorname{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}) - \operatorname{vec}(u_t u_t^\prime)\right\} \\ \operatorname{vec}(m_t^*\widetilde{u}_t^{*\prime}) - \operatorname{vec}(m_t u_t^\prime)\end{pmatrix} \\ &= \widetilde{Q}_T^* R_m \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{W}_{t,m}^*\end{aligned}$$

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 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. □

C Supplemental Figures and Tables

Figures 6 and 7 display the baseline specification used to produce Figures 1 and 2 above, but with 95% confidence intervals from the residual-based moving block bootstrap.

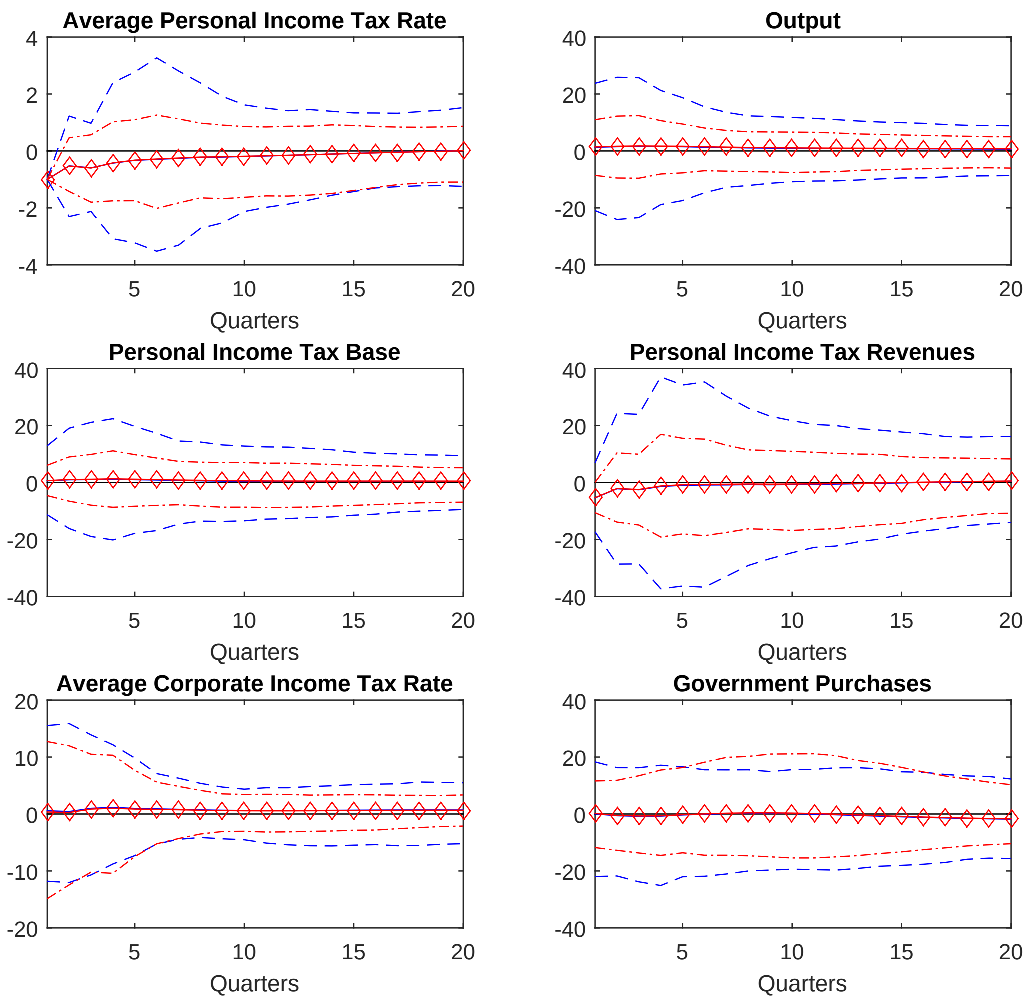

Figure 6: IRFs of a 1% cut in the APITR. Blue lines show the model with the APITR ordered first, and red diamonds show the model with the ACITR ordered first. Dashed lines are 95% confidence intervals from the MBB.

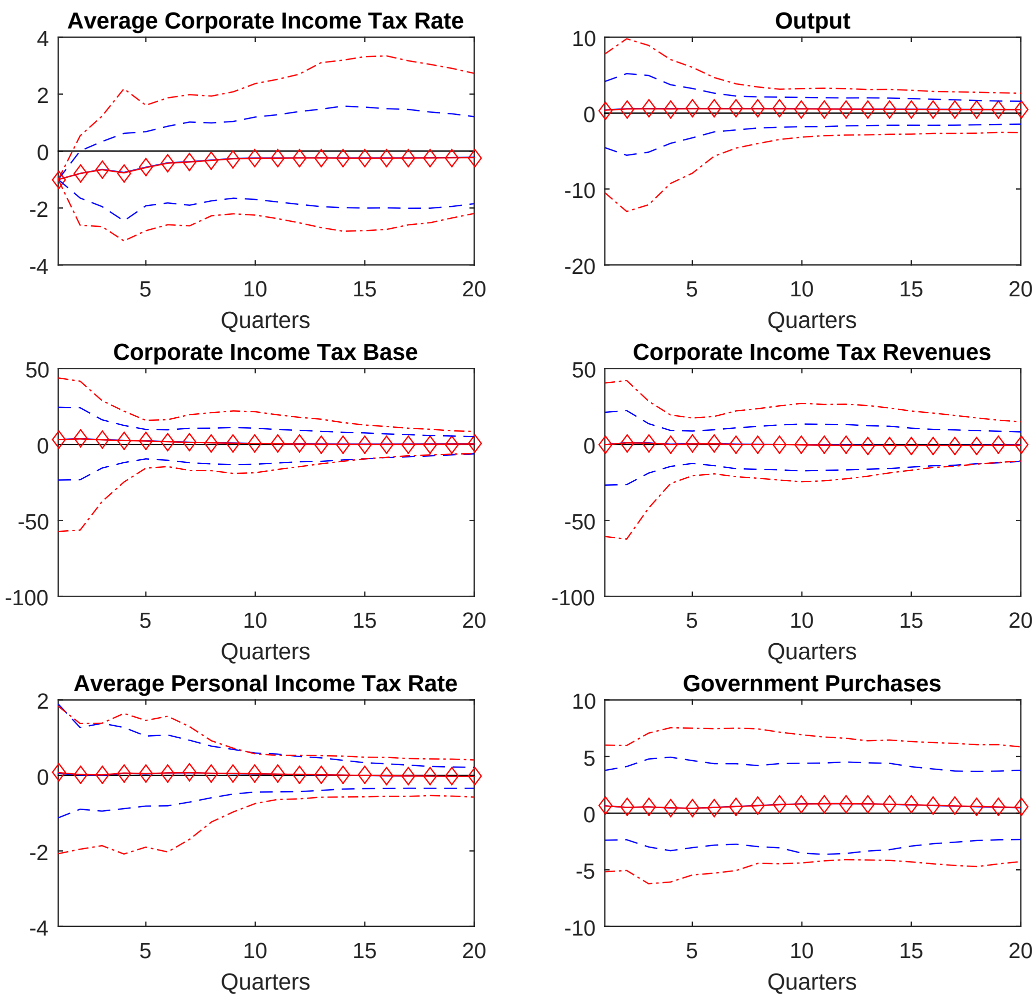

Figure 7: IRFs of a 1% cut in the ACITR. Blue lines show the model with the APITR ordered first, and red diamonds show the model with the ACITR ordered first. Dashed lines are 95% confidence intervals from the MBB.

Table 3: Coverage Rates of the 68% Confidence Intervals (with the initial response of $y_t^{(1)}$ normalized to -1)

Moving Block Bootstrap (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.65	1.00	0.64	1.00	0.65	1.00	0.66
1	0.64	0.67	0.65	0.62	0.67	0.67	0.68	0.66
2	0.73	0.66	0.68	0.61	0.69	0.65	0.69	0.67
3	0.78	0.68	0.72	0.64	0.73	0.65	0.72	0.68
4	0.78	0.68	0.75	0.65	0.75	0.67	0.72	0.69
5	0.80	0.65	0.77	0.65	0.77	0.67	0.73	0.67

Moving Block Bootstrap (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.66	1.00	0.64	1.00	0.66	1.00	0.65
1	0.65	0.67	0.67	0.62	0.66	0.66	0.67	0.67
2	0.72	0.68	0.70	0.62	0.69	0.66	0.68	0.67
3	0.76	0.69	0.73	0.63	0.72	0.66	0.71	0.68
4	0.77	0.68	0.76	0.65	0.76	0.67	0.73	0.67
5	0.80	0.66	0.78	0.64	0.76	0.66	0.73	0.67

Wild Bootstrap - Rademacher (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.13	1.00	0.07	1.00	0.07	1.00	0.04
1	0.68	0.37	0.69	0.31	0.68	0.37	0.70	0.34
2	0.69	0.53	0.70	0.48	0.70	0.49	0.68	0.50
3	0.72	0.61	0.67	0.55	0.69	0.55	0.69	0.55
4	0.73	0.63	0.68	0.58	0.67	0.57	0.68	0.57
5	0.77	0.58	0.71	0.58	0.69	0.58	0.68	0.56

Wild Bootstrap - Rademacher (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.13	1.00	0.07	1.00	0.08	1.00	0.05
1	0.67	0.38	0.69	0.31	0.68	0.37	0.69	0.35
2	0.70	0.53	0.70	0.49	0.70	0.50	0.68	0.51
3	0.71	0.61	0.69	0.56	0.68	0.56	0.69	0.56
4	0.73	0.63	0.68	0.58	0.67	0.58	0.68	0.57
5	0.77	0.59	0.70	0.58	0.69	0.58	0.69	0.58

Wild Bootstrap - Normal (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.90	1.00	0.85	1.00	0.84	1.00	0.84
1	0.78	0.85	0.70	0.83	0.69	0.81	0.69	0.81
2	0.89	0.83	0.78	0.77	0.73	0.77	0.71	0.80
3	0.89	0.84	0.86	0.75	0.80	0.73	0.79	0.78
4	0.88	0.85	0.87	0.77	0.84	0.74	0.79	0.77
5	0.91	0.81	0.86	0.75	0.84	0.75	0.79	0.76

Wild Bootstrap - Normal (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.89	1.00	0.85	1.00	0.84	1.00	0.83
1	0.78	0.85	0.71	0.83	0.68	0.81	0.69	0.82
2	0.89	0.81	0.78	0.77	0.73	0.76	0.69	0.78
3	0.89	0.83	0.87	0.76	0.80	0.74	0.78	0.77
4	0.89	0.85	0.87	0.77	0.84	0.75	0.80	0.76
5	0.92	0.80	0.86	0.75	0.84	0.74	0.80	0.75

Table 4: Coverage Rates of the 95% Confidence Intervals (with the initial response of $y_t^{(1)}$ normalized to -1)

Moving Block Bootstrap (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.93	1.00	0.93	1.00	0.93	1.00	0.94
1	0.93	0.92	0.94	0.94	0.94	0.92	0.94	0.94
2	0.98	0.93	0.96	0.93	0.94	0.92	0.95	0.94
3	0.97	0.95	0.98	0.94	0.97	0.93	0.97	0.94
4	0.98	0.94	0.97	0.95	0.97	0.93	0.98	0.93
5	0.99	0.91	0.97	0.94	0.97	0.94	0.97	0.94

Moving Block Bootstrap (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.93	1.00	0.93	1.00	0.93	1.00	0.93
1	0.93	0.92	0.93	0.93	0.93	0.92	0.94	0.94
2	0.98	0.94	0.96	0.94	0.94	0.92	0.95	0.94
3	0.97	0.95	0.98	0.94	0.97	0.92	0.97	0.94
4	0.98	0.95	0.97	0.95	0.97	0.93	0.97	0.93
5	0.99	0.92	0.97	0.94	0.96	0.94	0.96	0.94

Wild Bootstrap - Rademacher (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.29	1.00	0.18	1.00	0.15	1.00	0.10
1	0.94	0.67	0.95	0.59	0.95	0.62	0.95	0.60
2	0.98	0.85	0.95	0.80	0.94	0.80	0.95	0.82
3	0.95	0.90	0.96	0.86	0.94	0.86	0.95	0.86
4	0.97	0.91	0.96	0.89	0.95	0.89	0.94	0.87
5	0.99	0.89	0.96	0.89	0.95	0.89	0.95	0.88

Wild Bootstrap - Rademacher (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	0.31	1.00	0.16	1.00	0.16	1.00	0.10
1	0.93	0.68	0.95	0.60	0.95	0.62	0.95	0.62
2	0.97	0.85	0.96	0.80	0.95	0.79	0.95	0.82
3	0.95	0.90	0.96	0.86	0.95	0.86	0.95	0.86
4	0.97	0.91	0.96	0.90	0.96	0.89	0.94	0.87
5	0.99	0.89	0.97	0.90	0.95	0.89	0.95	0.88

Wild Bootstrap - Normal (i.i.d. DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99
1	0.99	0.99	0.97	0.99	0.95	0.99	0.95	0.99
2	1.00	0.99	0.99	0.99	0.97	0.99	0.96	0.99
3	1.00	0.99	1.00	0.99	1.00	0.98	0.99	0.98
4	1.00	1.00	0.99	0.99	0.99	0.98	0.99	0.98
5	1.00	0.98	0.99	0.99	0.99	0.98	0.99	0.98

Wild Bootstrap - Normal (GARCH DGP):

	$T=100$		$T=250$		$T=500$		$T=1000$	
t	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$	$y_t^{(1)}$	$y_t^{(2)}$
0	1.00	1.00	1.00	1.00	1.00	0.99	1.00	0.99
1	0.98	0.99	0.97	0.99	0.95	0.99	0.95	0.99
2	1.00	0.99	0.99	0.99	0.97	0.98	0.96	0.99
3	1.00	0.99	1.00	0.99	1.00	0.98	0.99	0.98
4	1.00	0.99	1.00	0.99	0.99	0.98	0.99	0.98
5	1.00	0.98	0.99	0.99	0.98	0.98	0.99	0.98

Tables 5 gives the autocorrelation functions for $\widehat{u}_t^{(j)}$, $|\widehat{u}_t^{(j)}|$ and $(\widehat{u}_t^{(j)})^2$, for $j = 1, \dots, K$ where $K = 7$ in the baseline model. The ordering of the variables in this baseline model are the APITR, the ACITR, the log of the PITB, the log of the CITB, the log of government spending, the log of GDP divided population, and the log of government debt divided by the GDP deflator and population.

Table 5: Autocorrelations

Autocorrelations of $u_t^{(j)}$:

h	$u_t^{(1)}$	$u_t^{(2)}$	$u_t^{(3)}$	$u_t^{(4)}$	$u_t^{(5)}$	$u_t^{(6)}$	$u_t^{(7)}$
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	-0.01	0.02	0.01	0.00	-0.06	-0.01	-0.01
2	-0.01	-0.05	0.02	0.00	0.01	0.01	-0.08
3	0.02	-0.02	0.00	0.02	-0.03	0.02	-0.03
4	0.06	-0.02	-0.02	-0.16	0.16	-0.08	-0.04
5	-0.08	0.03	-0.16	-0.07	-0.01	-0.08	-0.03
6	-0.02	0.02	0.04	0.02	0.13	0.03	-0.02

Autocorrelations of $|u_t^{(j)}|$:

h	$\|u_t^{(1)}\|$	$\|u_t^{(2)}\|$	$\|u_t^{(3)}\|$	$\|u_t^{(4)}\|$	$\|u_t^{(5)}\|$	$\|u_t^{(6)}\|$	$\|u_t^{(7)}\|$
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.26	0.29	0.11	0.06	0.07	0.12	0.30
2	0.06	0.13	0.00	0.00	0.04	0.19	0.08
3	-0.06	0.00	-0.02	0.01	-0.06	0.13	0.09
4	-0.06	0.04	0.20	0.12	0.08	0.12	0.09
5	0.02	-0.03	-0.01	-0.02	-0.10	0.07	0.01
6	0.03	-0.06	0.03	0.14	0.08	0.04	0.09

Autocorrelations of $(u_t^{(j)})^2$:

h	$(u_t^{(1)})^2$	$(u_t^{(2)})^2$	$(u_t^{(3)})^2$	$(u_t^{(4)})^2$	$(u_t^{(5)})^2$	$(u_t^{(6)})^2$	$(u_t^{(7)})^2$
0	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1	0.19	0.22	0.09	0.05	0.05	0.08	0.18
2	0.01	0.10	0.02	-0.04	0.00	0.10	0.03
3	-0.04	-0.01	0.03	-0.03	-0.05	0.04	0.07
4	-0.05	0.00	0.15	0.10	0.09	0.06	0.09
5	-0.02	-0.01	-0.02	-0.01	-0.10	0.03	-0.02
6	-0.02	-0.02	-0.01	0.10	0.01	0.05	0.00

Notes: h is the autocorrelation horizon: $\text{corr}(x_t, x_{t-h})$ for a given variable x_t. The 95% confidence interval is given by $(-0.13, 0.13)$.

References

Blanchard, Olivier and Roberto Perotti. 2002. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *Quarterly Journal of Economics* 117 (4):1329–1368.

Bloom, Nicholas. 2009. "The Impact of Uncertainty Shocks." *Econometrica* 77 (3):623–685.

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.

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.

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.

Drautzburg, Thorsten. 2015. "A Narrative Approach to a Fiscal DSGE Model." Working Paper, Federal Reserve Bank of Philadelphia.

Fernald, John. 2014. "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity." Working Paper, Federal Reserve Bank of San Francisco.

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

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

———. 2007. "Asymptotic and Bootstrap Inference for AR(∞) Processes with Conditional Heteroskedasticity." *Econometric Reviews* 26 (6):609–641.

Gürkaynak, Refet S, Brian Sack, and Eric T. Swanson. 2005. "Do Actions Speak Louder Than Words? The Response of Asset Prices to Monetary Policy Actions and Statements." *International Journal of Central Banking* 1 (1):55–93.

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

Kliem, Martin and Alexander Kriwoluzky. 2013. "Reconciling Narrative Monetary Policy Disturbances with Structural VAR Model Shocks?" *Economics Letters* 121 (2):247–251.

Künsch, Hans R. 1989. "The Jackknife and the Bootstrap for General Stationary Observations." *Annals of Statistics* 17 (3):1217–1241.

Lunsford, Kurt G. 2015a. "Housing Search and Fluctuations in Residential Construction." Job Market Paper, University of Wisconsin – Madison.

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

———. 2016. "Monetary Policy, Residential Investment, and Search Frictions: An Empirical and Theoretical Synthesis." Federal Reserve Bank of Cleveland Working Paper no. 16-07.

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

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

———. 2014. "A Reconciliation of SVAR and Narrative Estimates of Tax Multipliers." *Journal of Monetary Economics* 68 (Supplement):S1–S19.

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

———. 2016. "Uniform Inference in SVARs with External Instruments." Working Paper, New York University.

Mountford, Andrew and Harald Uhlig. 2009. "What Are the Effects of Fiscal Policy Shocks?" *Journal of Applied Econometrics* 24 (6):960–992.

Mumtaz, Haroon, Gabor Pinter, and Konstantinos Theodoridis. 2015. "What Do VARs Tell Us About the Impact of a Credit Supply Shock?" Working Paper No. 739 Queen Mary University of London.

Nakamura, Emi and Jón Steinsson. 2015. "High Frequency Identification of Monetary Non-Neutrality." Working Paper, Columbia University.

Ramey, Valerie A. 2016. "Macroeconomic Shocks and Their Propagation." NBER Working Paper No. 21978.

Romer, Christina D. and David H. Romer. 2009. "A Narrative Analysis of Postwar Tax Changes." Unpublished manuscript, University of California, Berkeley.

———. 2010. "The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks." *American Economic Review* 100 (3):763–801.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48 (1):1–48.

Staiger, Douglas and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3):557–586.

Stock, James H. and Mark W. Watson. 2008. "Lecture 7 - Recent Developments in Structural VAR Modeling." Presentd 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.