An Alternative Bootstrap for Proxy Vector Autoregressions

Martin Bruns[1] · Helmut Lütkepohl[2]

Abstract

We propose a new bootstrap algorithm for inference for impulse responses in structural vector autoregressive models identified with an external proxy variable. Simulations show that the new bootstrap algorithm provides confidence intervals for impulse responses which often have more precise coverage than and similar length to the competing moving-block bootstrap intervals. An empirical example shows how the new bootstrap algorithm can be applied in the context of identifying monetary policy shocks.

Keywords Bootstrap inference · Structural vector autoregression · Impulse responses · Instrumental variable

JEL Classification C32

1 Introduction

In structural vector autoregressive (VAR) analysis one strand of the literature uses external instruments, also called proxies, to identify shocks of interest (e.g., Stock & Watson, 2012; Mertens & Ravn, 2013; Piffer & Podstawski, 2018; Kilian & Lütkepohl, 2017,Chapter 15). The related models and methods are often labelled proxy VARs. In this context, frequentist inference for impulse responses is typically based on bootstrap methods. In some of the literature, the wild bootstrap (WB) is

Martin Bruns
martin.bruns@uea.ac.uk

Helmut Lütkepohl
hluetkepohl@diw.de

[1] School of Economics, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK

[2] DIW Berlin and Freie Universität Berlin, Mohrenstr. 58, 10117 Berlin, Germany

used (e.g., Mertens & Ravn, 2013; Gertler & Karadi, 2015; Carriero et al., 2015). However, work by Brüggemann et al. (2016) and Jentsch and Lunsford (2019, 2021) shows that wild bootstrap methods are not asymptotically valid in this context and they propose a moving-block bootstrap (MBB) which provides asymptotically correct confidence intervals for impulse responses under very general conditions. It can cope, for example, with conditionally heteroskedastic (GARCH) VAR errors which is an advantage in many applied studies where financial data are of interest. On the other hand, (Lütkepohl and Schlaak, 2019) demonstrate by simulations that the MBB can result in confidence intervals with low coverage rates in small samples.

In this study, we propose an alternative bootstrap method for proxy VARs which is based on resampling not only the VAR residuals but also the residuals of a model for the proxy and is therefore signified as PRBB (proxy residual-based bootstrap). We show by simulation that it leads to quite precise confidence intervals for impulse responses in small samples. This makes it attractive for macroeconomic analysis where often smaller samples with less than 200 observations are available. A major advantage of the MBB is that it remains asymptotically valid even if the data exhibit conditional heteroskedasticity. Although the PRBB does not explicitly account for GARCH, we show by simulation that in small samples it may even outperform the MBB if the VAR errors are driven by a GARCH process.

The remainder of the paper is structured as follows. The proxy VAR model is presented in the next section. Estimation of proxy VAR models is considered in Sect. 3. The alternative bootstrap methods considered in this study are presented in Sect. 4 and a small sample Monte Carlo comparison of the bootstrap methods is discussed in Sect. 5. An illustrative example is presented in Sect. 6, Sect. 7 concludes and Sect. 8 discusses possible extensions.

2 The Proxy VAR Model

A K-dimensional reduced-form VAR process,

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \tag{2.1}$$

is considered. Here v is a $(K \times 1)$ constant term and the A_i, $i = 1, \ldots, p$, are $(K \times K)$ slope coefficient matrices. The reduced-form error, u_t, is a zero mean white noise process with covariance matrix Σ_u, i.e., $u_t \sim (0, \Sigma_u)$. The vector of structural errors, $w_t = (w_{1t}, \ldots, w_{Kt})'$, is such that $u_t = Bw_t$, where B is the nonsinguar $(K \times K)$ matrix of impact effects of the shocks on the observed variables y_t. Thus, $w_t \sim (0, \Sigma_w = B^{-1} \Sigma_u B^{-1'})$, where Σ_w is a diagonal matrix.

If the first column, say b, of B is known, the structural impulse responses of the first shock, $\boldsymbol{\theta}_i = (\theta_{11,i}, \ldots, \theta_{K1,i})'$, can be computed as

$$\boldsymbol{\theta}_i = (\theta_{11,i}, \ldots, \theta_{K1,i})' = \Phi_i b, \quad i = 1, \ldots, H,$$

where the $(K \times K)$ matrices $\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j$ can be obtained recursively from the VAR slope coefficients using $\Phi_0 = I_K$ (e.g.,(Lütkepohl, 2005, Chapter 2). In the following, the $(K \times (H + 1))$ matrix of impulse responses,

$$\Theta(H) = [\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_H] = [b, \Phi_1 b, \ldots, \Phi_H b], \tag{2.2}$$

is of interest. It is assumed that the first shock increases the first variable by one unit on impact. In other words, the first component of $b = \boldsymbol{\theta}_0$ is assumed to be 1.

If b, Σ_u and the reduced-form errors are given, the first structural shock can be obtained as

$$w_{1t} = b'\Sigma_u^{-1}u_t / b'\Sigma_u^{-1}b \tag{2.3}$$

(see Stock & Watson, 2018, Footnote 6, p. 933) or (Bruns & Lütkepohl, 2021, Appendix A.1).

Suppose there is an instrumental variable z_t satisfying

$$\mathbb{E}(w_{1t}z_t) = c \neq 0 \quad \text{(relevance)}, \tag{2.4}$$

$$\mathbb{E}(w_{kt}z_t) = 0, \quad k = 2, \ldots, K, \quad \text{(exogeneity)}. \tag{2.5}$$

These conditions imply that

$$\mathbb{E}(z_t u_t) = B\mathbb{E}(z_t w_t) = cb.$$

In other words, the proxy z_t identifies a multiple of b.

In line with some of the proxy VAR literature (e.g., Jentsch & Lunsford, 2019 or Bruns & Lütkepohl, 2021), the proxy z_t is assumed to be generated as

$$z_t = D_t(\phi w_{1t} + \eta_t), \tag{2.6}$$

where D_t is a random 0-1 variable which determines the number of nonzero values of the proxy. It is assumed to have a Bernoulli distribution, $B(d)$, with parameter d, $0 < d \leq 1$, and captures the fact that many proxies are measured only at certain announcement days or when special events occur. The D_t are assumed to be stochastically independent of w_{1t} and the error term η_t which is thought of as representing measurement error. This error term is assumed to have mean zero and variance σ_η^2, i.e., $\eta_t \sim (0, \sigma_\eta^2)$, and it is distributed independently of w_{1t}. The parameter ϕ, the error η_t and the distribution of the Bernoulli random variable D_t (i. e., the parameter d) determine the strength of the correlation between z_t and w_{1t} and, hence, the strength of the proxy as an instrument.

The variance of z_t is $\text{Var}(z_t) = d(\phi^2 \text{Var}(w_{1t}) + \sigma_\eta^2)$ for $0 < d \leq 1$. Moreover, the covariance between w_{1t} and z_t is

$$\mathbb{E}(w_{1t}z_t) = \mathbb{E}(w_{1t}D_t(\phi w_{1t} + \eta_t)) = \phi d \text{Var}(w_{1t})$$

so that the correlation between w_{1t} and z_t is

$$\text{Corr}(w_{1t}, z_t) = \phi\sqrt{d}\sqrt{\text{Var}(w_{1t})} \Big/ \sqrt{\phi^2 \text{Var}(w_{1t}) + \sigma_\eta^2}.$$

Thus, the correlation between the proxy and the first shock declines with declining d and increasing σ_η^2.

3 Estimation

Suppose an effective sample $y_1, \ldots, y_T$ of size T is available for the model variables, plus all required presample values, $y_{-p+1}, \ldots, y_0$. Moreover, a corresponding sample $z_1, \ldots, z_T$ is available for the proxy.

Then the VAR(p) is estimated by bias-adjusted least squares (LS) giving estimates $\hat{v}, \hat{A}_1, \ldots, \hat{A}_p$, residuals $\hat{u}_1, \ldots, \hat{u}_T$ and an error covariance matrix estimator

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

based on mean-adjusted residuals. Kilian (1998) shows that employing bias-adjusted LS estimators improves inference for impulse responses. Therefore we use the bias-adjustment based on Pope (1990), as proposed by Kilian (1998), throughout the paper.

The first column b of B is estimated using the proxy z_t,

$$\hat{b} = \sum_{t=1}^{T} \hat{u}_t z_t \Big/ \sum_{t=1}^{T} \hat{u}_{1t} z_t, \tag{3.1}$$

where $\hat{u}_t$ are the residuals corresponding to bias-adjusted LS estimation and $\hat{u}_{1t}$ is their first entry. The impulse response matrix $\Theta(H)$ is estimated as

$$\widehat{\Theta}(H) = [\hat{b}, \hat{\Phi}_1 \hat{b}, \ldots, \hat{\Phi}_H \hat{b}],$$

where

$$\hat{\Phi}_i = \sum_{j=1}^{i} \hat{\Phi}_{i-j} \hat{A}_j, \quad i = 1, \ldots, H.$$

Moreover, the first shock is estimated as

$$\hat{w}_{1t} = \hat{b}' \widehat{\Sigma}_u^{-1} \hat{u}_t / \hat{b}' \widehat{\Sigma}_u^{-1} \hat{b}, \quad t = 1, \ldots, T, \tag{3.2}$$

and ϕ is estimated by LS from

$$z_t = \mu + \phi \hat{w}_{1t} + \eta_t$$

for all $t \in \mathcal{T}_D$, where $\mathcal{T}_D = \{t | D_t = 1\}$. The estimate of ϕ is denoted by $\hat{\phi}$ and the residuals are $\hat{\eta}_t$ for $t \in \mathcal{T}_D$ and $\hat{\eta}_t = 0$ for $t \notin \mathcal{T}_D$.

4 Bootstraps

As mentioned in the introduction, the WB and the MBB are the bootstrap methods most frequently used in the proxy VAR literature for frequentist inference for impulse responses. The WB generates asymptotically invalid confidence intervals while the MBB yields confidence intervals with the correct coverage level asymptotically

under quite general conditions (Jentsch & Lunsford, 2019, 2021). It may be imprecise in small samples, however, and therefore we propose the PRBB which turns out to have better properties in small samples. The three bootstrap versions differ in the way they generate bootstrap samples of y_t and z_t. Based on N bootstrap samples $y_{-p+1}^{(n)}, \ldots, y_0^{(n)}, y_1^{(n)}, \ldots, y_T^{(n)}$ and $z_1^{(n)}, \ldots, z_T^{(n)}$, $n = 1, \ldots, N$, they all use the following steps to determine bootstrap impulse responses and confidence intervals:

1. A VAR(p) model is fitted to the sample by LS and bias-adjusted, giving bootstrap estimates $\hat{A}^{(n)}$,

$$\hat{\Phi}_i^{(n)} = \sum_{j=1}^{i} \hat{\Phi}_{i-j}^{(n)} \hat{A}_j^{(n)}, \quad i = 1, \ldots, H, \quad \text{with } \hat{\Phi}_0^{(n)} = I_K,$$

and residuals $\hat{u}_t^{(n)}$.

2. Then bootstrap estimates

$$\hat{b}^{(n)} = \sum_{t=1}^{T} \hat{u}_t^{(n)} z_t^{(n)} \Big/ \sum_{t=1}^{T} \hat{u}_{1t}^{(n)} z_t^{(n)}.$$

of the structural parameters are determined.

3. Finally bootstrap estimates of the impulse responses of interest are computed as

$$\widehat{\Theta}(H)^{(n)} = [\hat{b}^{(n)}, \hat{\Phi}_1^{(n)} \hat{b}^{(n)}, \ldots, \hat{\Phi}_H^{(n)} \hat{b}^{(n)}]$$

and stored.

The N bootstrap estimates $\widehat{\Theta}(H)^{(1)}, \ldots, \widehat{\Theta}(H)^{(N)}$ are used to construct pointwise confidence intervals based on the relevant quantiles of the bootstrap distributions. Alternatively, percentile-t or Hall intervals could be used (see (Kilian & Lütkepohl, 2017, Section 12.2). However, the intervals based on quantiles are quite common in practice and the relative performance of the alternative bootstrap versions is not expected to depend on the type of interval used.

The samples are generated by one of the three alternative bootstrap methods, WB, MBB and PRBB, as follows:

WB: For $t = 1, \ldots, T$, independent standard normal variates ψ_t, $\psi_t \sim \mathcal{N}(0, 1)$, are drawn and bootstrap residuals and proxy variables are generated as

$$\begin{pmatrix} u_t^{WB} \\ z_t^{WB} \end{pmatrix} = \psi_t \begin{pmatrix} \hat{u}_t \\ z_t \end{pmatrix}.$$

The u_t^{WB} are de-meaned and multiplied by $\sqrt{T/(T - Kp - 1)}$, as in (Davidson and MacKinnon (2004, p. 597), and they are used to generate $y_t^{WB} = \hat{v} + \hat{A}_1 y_{t-1}^{WB} + \cdots + \hat{A}_p y_{t-p}^{WB} + u_t^{WB}$, $t = 1, \ldots, T$, starting from $y_{-p+1}^{WB}, \ldots, y_0^{WB}$, which are obtained as a random draw of p consecutive values from the original sample.

MBB: A block length $\ell < T$ has to be chosen for the MBB. The blocks of length ℓ of the estimated residuals and proxies are arranged in the form of the matrix

$$\begin{bmatrix} \begin{pmatrix} \hat{u}_1 \\ z_1 \end{pmatrix} & \begin{pmatrix} \hat{u}_2 \\ z_2 \end{pmatrix} & \cdots & \begin{pmatrix} \hat{u}_\ell \\ z_\ell \end{pmatrix} \\ \begin{pmatrix} \hat{u}_2 \\ z_2 \end{pmatrix} & \begin{pmatrix} \hat{u}_3 \\ z_3 \end{pmatrix} & \cdots & \begin{pmatrix} \hat{u}_{1+\ell} \\ z_{1+\ell} \end{pmatrix} \\ \vdots & \vdots & & \vdots \\ \begin{pmatrix} \hat{u}_{T-\ell+1} \\ z_{T-\ell+1} \end{pmatrix} & \begin{pmatrix} \hat{u}_{T-\ell+2} \\ z_{T-\ell+2} \end{pmatrix} & \cdots & \begin{pmatrix} \hat{u}_T \\ z_T \end{pmatrix} \end{bmatrix}.$$

The bootstrap residuals and proxy are re-centered columnwise by constructing

$$\tilde{u}_{j\ell+i} = \hat{u}_{j\ell+i} - \frac{1}{T-\ell+1} \sum_{r=0}^{T-\ell} \hat{u}_{i+r}$$

and

$$\tilde{z}_{j\ell+i} = z_{j\ell+i} - \frac{1}{T-\ell+1} \sum_{r=0}^{T-\ell} z_{i+r}$$

for $i = 1, 2, \ldots, \ell$ and $j = 0, 1, \ldots, s - 1$. Then $s = [T/\ell]$ of the re-centered rows of the matrix are drawn with replacement, where $[\cdot]$ denotes the smallest number greater than or equal to the argument such that $\ell s \geq T$. These randomly drawn blocks are joined end-to-end and the first T bootstrap residuals and proxies are retained,

$$\begin{pmatrix} u_t^{MBB} \\ z_t^{MBB} \end{pmatrix}, \quad t = 1, \ldots, T.$$

Finally, the u_t^{MBB} are de-meaned, multiplied by $\sqrt{T/(T - Kp - 1)}$, and used to generate $y_t^{MBB} = \hat{v} + \hat{A}_1 y_{t-1}^{MBB} + \cdots + \hat{A}_p y_{t-p}^{MBB} + u_t^{MBB}$, $t = 1, \ldots, T$, starting from $y_{-p+1}^{MBB}, \ldots, y_0^{MBB}$, which are obtained as a random draw of p consecutive values from the original sample.

PRBB: Samples

$$\begin{pmatrix} u_t^{PRBB} \\ \eta_t^{PRBB} \\ w_{1t}^{PRBB} \end{pmatrix}, \ t = 1, \ldots, T, \ \text{are drawn from} \ \begin{pmatrix} \hat{u}_1 \\ \hat{\eta}_1 \\ \hat{w}_{11} \end{pmatrix}, \ldots, \begin{pmatrix} \hat{u}_T \\ \hat{\eta}_T \\ \hat{w}_{1T} \end{pmatrix},$$

with replacement. Bootstrap samples $y_t^{PRBB} = \hat{v} + \hat{A}_1 y_{t-1}^{PRBB} + \cdots + \hat{A}_p y_{t-p}^{PRBB} + u_t^{PRBB}$, $t = 1, \ldots, T$, are generated, starting from $y_{-p+1}^{PRBB}, \ldots, y_0^{PRBB}$, which are obtained as a random draw of p consecutive values from the original sample. Samples of the proxy are generated as

$$z_t^{PRBB} = \tilde{D}_t(\hat{\phi} w_{1t}^{PRBB} + \eta_t^{PRBB}),$$

where $\tilde{D}_t$ is a random 0-1 variable following a Bernoulli distribution, $B(\hat{d})$, with $\hat{d}$ being the share of non-zero observations of the proxy in the original sample.

We emphasize again that the WB does not result in asymptotically valid confidence intervals but is presented here and included in the simulation comparison in Sect. 5 because it has been used in the proxy VAR literature. The WB and the MBB draw the proxies directly from observed values and, hence, do not make assumptions on the exact DGP of z_t. In contrast, the PRBB samples from the residuals of the assumed DGP for z_t in (2.6) and constructs new proxy values in each bootstrap replication. In addition, while the WB by construction sets the share of non-zero observations of the bootstrap proxy equal to the share in the original sample, this is not generally the case in the MBB and the PRBB design. Apart from that, all three bootstraps are recursive-design residual based bootstraps for generating the y_t samples.

In all three bootstrap algorithms, the initial values $y_{-p+1}^{(n)}, \ldots, y_0^{(n)}$ are a random draw of p consecutive values from the original y_t sample. Alternatively, the original initial values $y_{-p+1}, \ldots, y_0$ could have been used as initial values for each bootstrap sample. If the y_t are mean-adjusted, one could even simply use zero initial values if stationary models are under consideration. For example, Jentsch & Lunsford (2019) used zero initial values, generated more than T sample values and then dropped some burn-in values.

For the MBB a decision on the block length ℓ is needed. To make the asymptotic theory work, it has to be chosen such that $\ell \to \infty$ and $\ell^3/T \to 0$ as $T \to \infty$ (see Jentsch & Lunsford, 2019, 2021). The choice is less clear in small samples. Choosing ℓ too small, the blocks may not capture the data features well and may result in poor confidence intervals. A small block length may not be a big problem if the VAR errors and the proxy are iid (independently, identically distributed) and, hence, no higher order moments and dependencies have to be captured within the blocks but a small ℓ may be a problem if there are higher order features and dependencies. On the other hand, choosing ℓ large undermines inference precision because there are too few blocks to choose from. Note that the number of available blocks is $T - \ell + 1$ and, hence, depends on the block length. Jentsch and Lunsford (2019) mention a block length of $\ell = 5.03T^{1/4}$ as a rule of thumb and we use this rule of thumb in our simulations in Sect. 5 and in the empirical example in Sect. 6.

5 Small Sample Comparison of Bootstraps

In this section a small sample simulation comparison of the three bootstrap methods is presented. The simulation design is considered first and then the simulation results are discussed.

5.1 Monte Carlo Design

5.1.1 DGP 1

The first data generating process (DGP1) is similar to a DGP that has been used frequently in related work on comparing inference methods for impulse responses (e. g., Kilian, 1998; Kilian & Kim, 2011; Lütkepohl et al., 2015a, b). It is a two-dimensional VAR(1) of the form:

$$y_t = \begin{bmatrix} a_{11} & 0 \\ 0.5 & 0.5 \end{bmatrix} y_{t-1} + u_t, \tag{5.1}$$

where $0 < a_{11} < 1$. The process is stable with more persistence for a_{11} closer to one.

The structural errors, w_t, are normally distributed with mean zero and variances 4 and 1 such that $w_t \sim \mathcal{N}(0, \text{diag}(4, 1))$, and $u_t = Bw_t$ with

$$B = \begin{bmatrix} 1 & 0 \\ 0.5 & 3 \end{bmatrix},$$

such that $b = (1, 0.5)'$. These u_t errors are used to generate the y_t as in Eq. (5.1), starting from a standard normal y_0, i.e., $y_0 \sim \mathcal{N}(0, I_2)$. In the simulations, we fit VAR models of order $p = 1$ and $p = 12$, without constant term, to de-meaned data.

In line with the related literature (e.g., Jentsch & Lunsford, 2019), the proxy z_t is generated as in Eq. (2.6), i.e., $z_t = D_t(\phi w_{1t} + \eta_t)$, where D_t, ϕ and the error η_t determine the strength of the correlation between z_t and w_{1t} and, hence, the strength of the proxy which is important for how well the impact effects of the shock can be estimated and these estimates are of central importance for estimating the impulse responses. The error term η_t is generated independently of w_{1t} as $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$, with different values of σ_η^2. The random variable D_t has a Bernoulli distribution with parameter d, $B(d)$, which specifies the average proportion of nonzero z_t variables. D_t is stochastically independent of η_t and w_{1t}. For $d = 1$, the proxy variable is nonzero with probability one for all sample periods $t = 1, \ldots, T$.

The parameter values used in our simulations are summarized in Table 1. We use propagation horizons up to $H = 20$ to capture not only the short-term effects of a shock but also the longer-term effects which may still be a bit away from zero for the more persistent processes. Sample sizes $T = 100, 250$, and 500 are considered. The

Table 1 Design Parameters for DGP1

T	p	a_{11}	B	Σ_w	d	ϕ	σ_η^2	Corr(w_{1t}, z_t)	H
100	1	0.5	$\begin{bmatrix} 1 & 0 \\ 0.5 & 3 \end{bmatrix}$	$\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$	1	1	0.2346	0.9	20
250	12	0.9			0.4		3	0.5	
500		0.95						0.5692	
								0.3162	

number of replications for each Monte Carlo design is $R = 1000$ and we use $N = 2000$ bootstrap repetitions within each replication.

Because the MBB is constructed so as to account for GARCH errors while this is not the case for the PRBB, we also generate DGP1 with GARCH errors to explore the performance of the bootstraps under conditions which may be unfavorable for the PRBB but may still be present in practice. The way the GARCH errors are generated is described in detail in Appendix A. Here we just mention that we use a bivariate GARCH process with high persistence in each component as it is often observed for financial data.

Our criteria for evaluating the bootstrap methods are the coverage precision and the lengths of the confidence intervals obtained from the bootstraps. These criteria capture main features of interest in related empirical studies and they have also been used in related small sample comparisons of bootstrap inference (e.g., Kilian & Kim, 2011; Lütkepohl & Schlaak, 2019).

5.1.2 DGP 2

Our second DGP (DGP2) mimics a VAR model from a study of Gertler and Karadi (2015). It is based on parameters estimated from their dataset. One of the models used by Gertler and Karadi is a four-dimensional US monthly model. We use their data from 1990M1 to 2016M6 and fit a VAR(1) model with constant term to the data. Using bias-adjusted estimates, the reduced-form parameters of DGP2 are $v = 0$,

$$A_1 = \begin{bmatrix} 0.97 & 0.00 & 0.00 & -0.13 \\ 0.01 & 1.00 & 0.00 & -0.09 \\ -0.03 & 0.00 & 1.00 & -0.53 \\ 0.02 & 0.00 & 0.00 & 0.91 \end{bmatrix} \text{ and } \Sigma_u = \begin{bmatrix} 0.04 & 0.00 & 0.01 & -0.01 \\ 0.00 & 0.11 & 0.03 & 0.00 \\ 0.01 & 0.03 & 0.37 & -0.01 \\ -0.01 & 0.00 & -0.01 & 0.07 \end{bmatrix}.$$

The maximum eigenvalue of A_1 has modulus 0.9997. Thus, DGP2 is stable but very persistent. These parameters are used to generate the y_t based on $u_t \sim \mathcal{N}(0, \Sigma_u)$ and starting from $y_0 = 0$, the unconditional mean of the y_t.

We also use a proxy with similar properties as the proxy for monetary policy shocks constructed by Gertler and Karadi (2015). More precisely, we estimate the b vector of impact effects of the first shock, giving a vector $b = (1, -0.14, 0.70, 0.24)'$, and estimate the parameters ϕ and σ_η^2 of the model (2.6) as described in Sect. 3 from the Gertler/Karadi data with nonzero z_t values and the first shock obtained from Eq. (2.3). This yields values $\phi = 0.1019$ and $\sigma_\eta^2 = 0.0020$ that are used for generating z_t as in Eq. (2.6) with D_t having a $B(0.82)$ distribution. The parameter, $d = 0.82$, of the Bernoulli distribution is chosen because the Gertler-Karadi proxy has nonzero values for 82% of the sample periods. The implied correlation between proxy and shock is 0.36 and, hence, it is rather low.

Note that the generation mechanism for DGP2 differs from that of DGP1, where the structural shocks are generated directly and the reduced-form data as well as the proxy are computed from the generated structural shocks and the generated η_t series. In contrast, we generate the reduced-form errors for DGP2, construct the first structural shock from the structural parameters b and the error covariance matrix Σ_u

as in equation (3.2) and then generate z_t as in Eq. (2.6) with an additionally generated $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$.

The rational behind using DGP2 is that we will also use the Gertler/Karadi data for an illustrative example in Sect. 6 and, hence, the simulation results for DGP2 may be indicative of what to expect in the example. Moreover, it is, of course, of interest to see whether the results for our small bivariate process underlying DGP1 carry over to a higher-dimensional DGP.

For DGP2, we fit VAR models of orders $p = 1$ and $p = 12$, including a constant term, to samples of size $T = 200$ and 500. The smallest sample size considered is a bit larger than for DGP1 to account for the larger model dimension. It is not far from the sample size used in the example in Sect. 6. The number of bootstrap replications is again $N = 2000$ and the number of Monte Carlo repetitions is $R = 1000$, as for DGP1.

5.2 Small Sample Results

5.2.1 Results for DGP1

Some key findings from simulating DGP1 are presented in Figs. 1 and 2. Specifically, in Fig. 1 the implications of changing the VAR order p, the persistence of the VAR process (a_{11}) and the strength of the proxy reflected in the Bernoulli parameter d on the coverage and average length of the bootstrap confidence intervals can be seen for relatively short samples of size $T = 100$. In Fig. 2, the impact of increasing the sample size is presented.

A main observation from Fig. 1 is that, for some designs and propagation horizons, there are clear differences in the coverage of the confidence intervals of the three bootstrap variants. The PRBB yields overall the coverage results closest to the desired 90% while the MBB tends to yield coverage rates a bit smaller and, in particular for short propagation horizons, the WB yields conservative intervals with coverage rates often larger than 90% and greater length than MBB and PRBB. Interestingly, the average lengths of the intervals of all three bootstrap methods are often very similar despite differences in coverage. For example, the PRBB intervals for short propagation horizons are in most cases very similar to the MBB intervals although the latter have a coverage which is below the PRBB coverage and it is also lower than the nominal 90%. Thus, Fig. 1 clearly shows that the PRBB tends to be more precise in terms of coverage and often it does so without sacrificing much interval length. Thus, under these two criteria it is preferable to the MBB and the WB. The latter bootstrap is often conservative and also yields larger confidence intervals than the other two bootstrap methods.

There are also some more specific results related to the VAR order and the proxy strength that can be seen in Fig. 1.

- The VAR order p has an important impact on both the coverage and average lengths of the confidence intervals. In particular, considering the order $p = 12$ of a short-memory process with $a_{11} = 0.5$ results in substantial over-coverage,
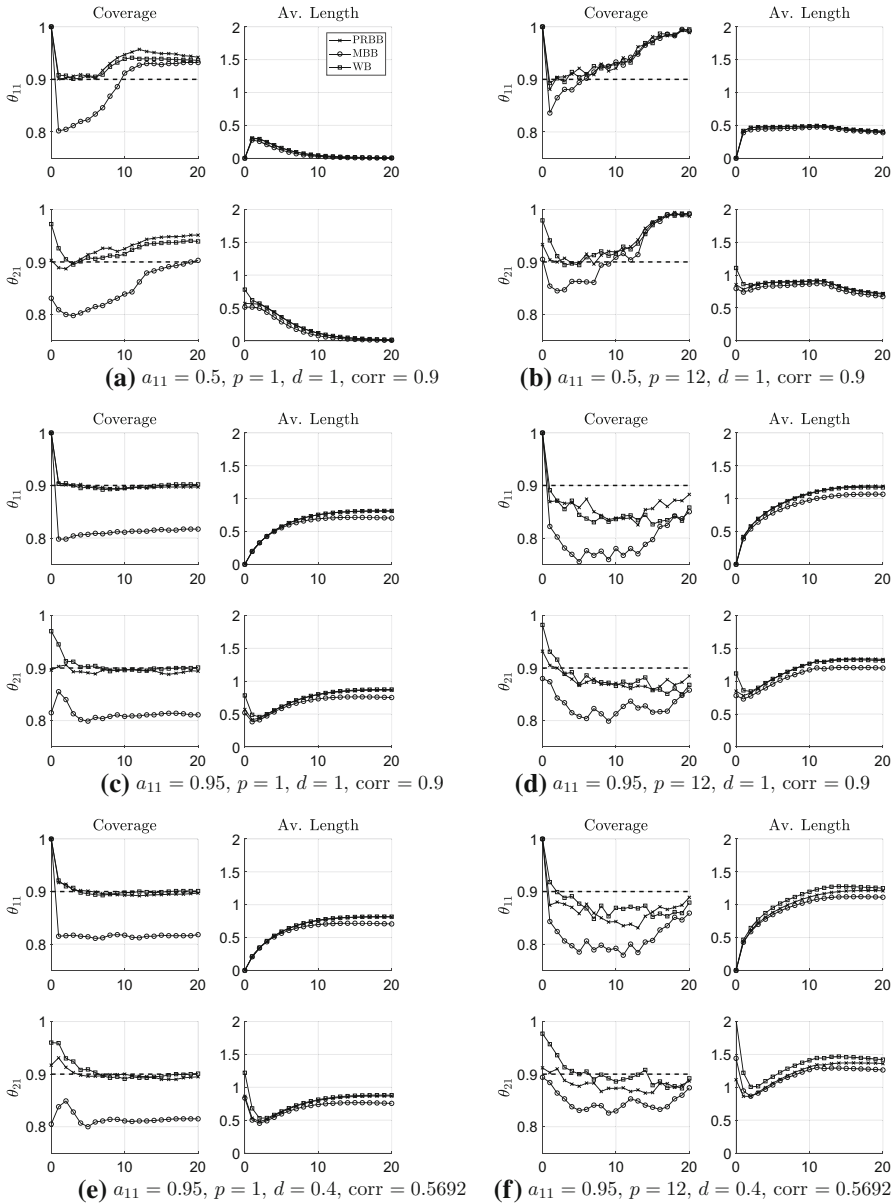
Fig. 1 Coverage and average lengths of alternative pointwise bootstrap 90% confidence intervals for DGP1 with iid errors and $T = 100$

especially for longer propagation horizons, for all three bootstrap variants. The interval lengths tend to increase for all propagation horizons and substantially so for the longer propagation horizons for all three bootstraps if the VAR order
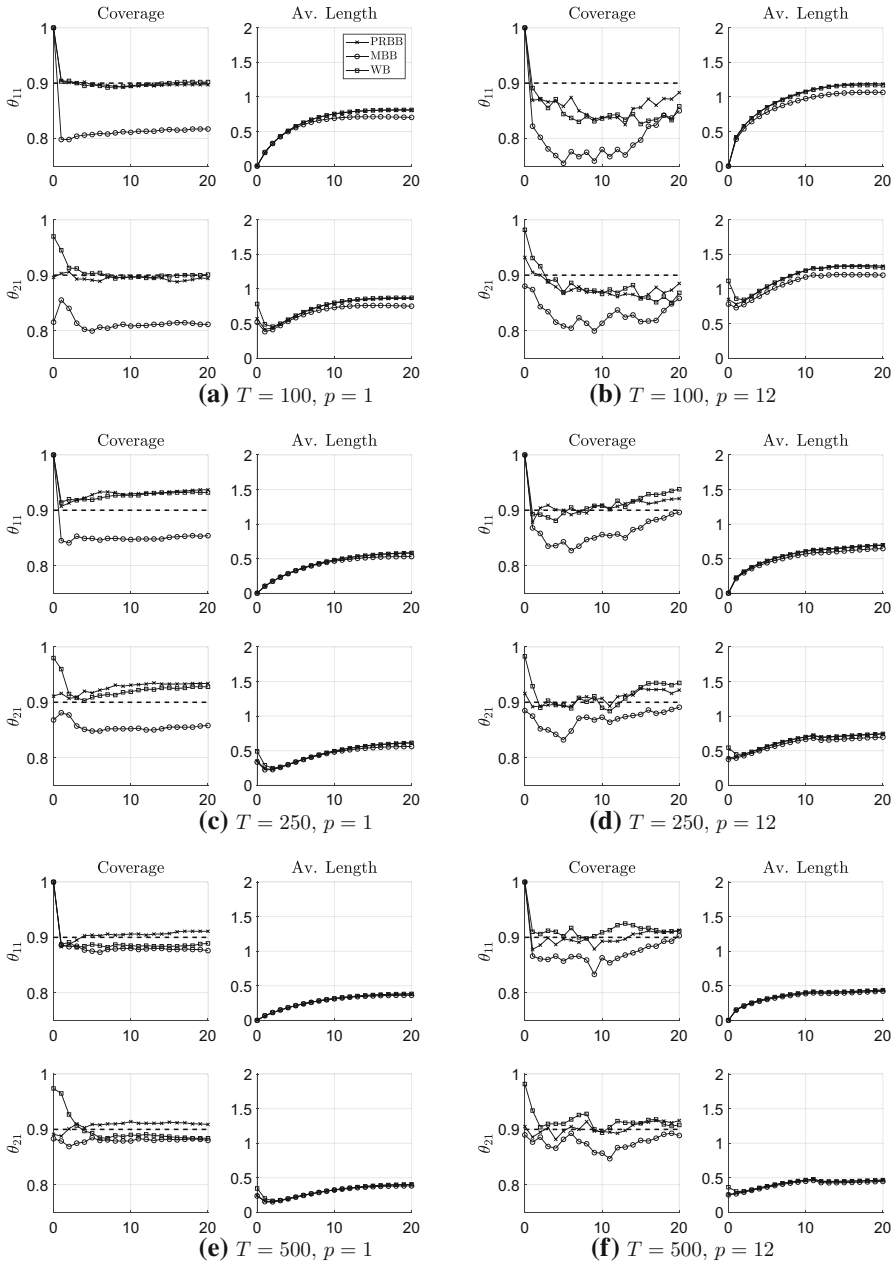
Fig. 2 Coverage and average lengths of alternative pointwise bootstrap 90% confidence intervals for DGP1 with iid errors and $d = 1$, $a_{11} = 0.95$, corr $= 0.9$

increases from $p = 1$ to $p = 12$ (compare the second and fourth columns of Fig. 1).

- Comparing panels (c) and (e) as well as (d) and (f) in Fig. 1, it is apparent that the proxy strength does not have much of an effect on the coverage but partly leads to larger intervals (see in particular the average lengths of θ_{21} intervals for short horizons). In panels (e) and (f) in Fig. 1, the proxy has a lower correlation with the structural shock of interest due to the reduced number of event dates, d, for which the proxy is constructed. A similar result is obtained, however, if $d = 1$ is maintained but the correlation between proxy and shock is reduced due to a larger variance σ_η^2 of the error term in Eq. (2.6), as can be seen in Fig. 6 in the Appendix.

- The impact of higher persistence (a larger a_{11} parameter) of the process can be seen by comparing panels (a) and (c) as well as (b) and (d) in Fig. 1. Generally the coverage is reduced and the intervals become larger, especially for longer propagation horizons, if a_{11} increases from 0.5 to 0.95. The reduction in coverage is most severe for the MBB, while the PRBB continues to have acceptable coverage for persistent processes. In Fig. 7 of the Appendix, additional results for $a_{11} = 0.9$ are presented and it can be seen that the results for $a_{11} = 0.9$ are similar to those of $a_{11} = 0.95$.

In Fig. 2, the impact of the sample size on the confidence intervals is exhibited for the case of a persistent process with $a_{11} = 0.95$ and a relatively strong proxy with correlation 0.9 with the shock and $d = 1$. As we saw in Fig. 1 already, in this situation the MBB has a coverage clearly smaller than the nominal 90% for $p = 1$ and all three bootstrap methods tend to yield under-coverage for $T = 100$. In Fig. 2 it can be seen that the coverage clearly improves for $T = 250$ already and the coverage deficiencies largely disappear for $T = 500$. Also, the interval lengths for all three methods become very similar and are reduced for larger sample sizes, as one would expect. Only the WB intervals for some short propagation horizons remain wider and less precise for larger samples. This result may be a reflection of the asymptotic inadmissibility of the WB.

As the PRBB does not explicitly account for GARCH in the VAR residuals, we have also applied the three bootstraps to processes with GARCH errors to see how such features impact on their properties. Results corresponding to Figs. 1 and 2 are presented in Figs. 8 and 9 in the Apprendix. It turns out that for small samples of $T = 100$ the relative performance of the three bootstraps in terms of coverage and interval length is not much affected. Despite the fact that the MBB is the only asymptotically valid procedure for this case, comparing its confidence intervals to the PRBB intervals, they have lower coverage and similar length as in the case of iid residuals. In Fig. 9 it can be seen that even for larger samples with $T = 500$, MBB is not clearly superior to the PRBB. Thus, at least for our DGP1 the MBB does not have an advantage over PRBB even if data features such as GARCH are present that are not accounted for explicitly by the PRBB. These results suggest that for macroeconomic studies where rarely samples larger than $T = 500$ are available, the PRBB may lead to superior inference as compared to MBB and WB.

5.2.2 Results for DGP2

Coverage and average interval lengths for DGP2 are depicted in Fig. 3 for sample size $T = 200$ and in Fig. 4 for $T = 500$. Even for the smaller sample size $T = 200$, all coverage rates of the nominal 90% confidence intervals of WB and PRBB are between 80% and 100%, except for the long-run response of the third variable. In other words, the two bootstrap methods yield rather precise confidence intervals for three out of four variables across our Monte Carlo designs. Given the asymptotic inadmissibility of the WB, this result may, of course, not be generalizable to other simulation designs.

Even the MBB has coverage rates above 80% for variables 1, 2 and 4 and propagation horizons up to 30 periods when $T = 200$. Thus, even the MBB is relatively precise in terms of coverage for three out of four variables. There are, however, differences in interval lengths among the three bootstraps. Typically, the WB intervals are a bit longer than the MBB and PRBB intervals which are often close together on average. Overall the performance of the WB is inferior to MBB and PRBB. Thus, although there is often not much to choose between MBB and PRBB in terms of coverage and interval length, it is remarkable that the PRBB tends to have typically coverage rates closer to 90% than the MBB. Thus, even for the higher-dimensional DGP2, PRBB performs well relative to its competitors, at least for three of the four variables.

For the third variable, VAR order $p = 1$ and $T = 200$ all three bootstrap methods yield coverage rates below 80% for a propagation horizon of 48 periods. For $T = 500$ and $p = 1$, only the MBB still has a coverage below 80% for long horizons (see Fig. 4). The coverage rates for long horizons are actually a bit closer to 90% for $p = 12$, although one might expect lower precision for the larger VAR order as the larger VAR order implies a model with substantially more parameters. Even for $p = 12$, the PRBB outperforms the MBB in terms of coverage and is almost as good in terms of interval length.

In summary, our simulations show that the WB is often conservative and yields more than the nominal coverage. In turn, its confidence intervals are often considerably larger than those of the MBB and PRBB. Thus, the WB is overall inferior to the MBB and the PRBB. Between the latter two, the PRBB is preferable because it yields typically coverage rates closer to the nominal rate than the MBB. Moreover, the PRBB confidence intervals are often about as long on average as those of the MBB. Hence, our simulations show that the PRBB has merit. In the next section, it will be applied to an illustrative example model from the literature.

6 Empirical Example

We consider an example based on the study of Gertler and Karadi (2015) mentioned earlier to illustrate the differences between the three bootstrap methods. One of the models used by Gertler and Karadi is a four-dimensional US monthly model for the variables (1) one-year government bond rate, (2) log consumer price index (CPI), (3) log industrial production (IP) and (4) excess bond premium. They employ the three
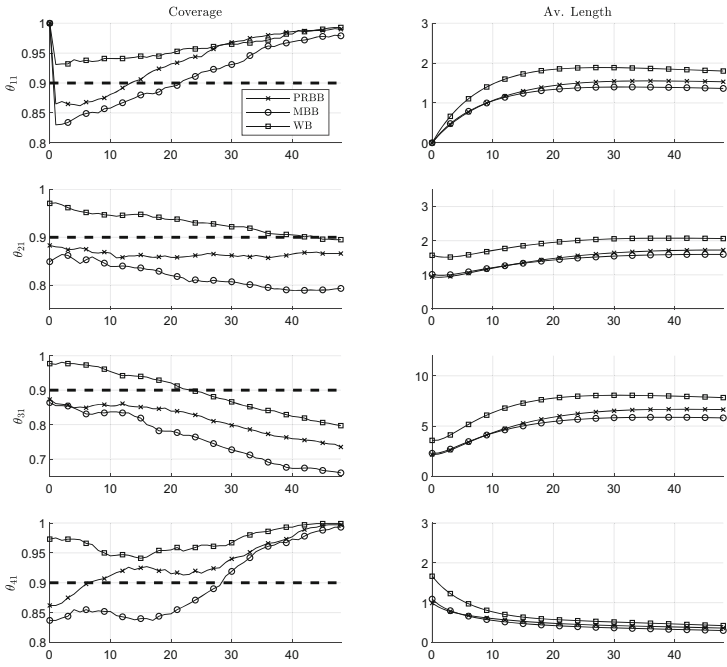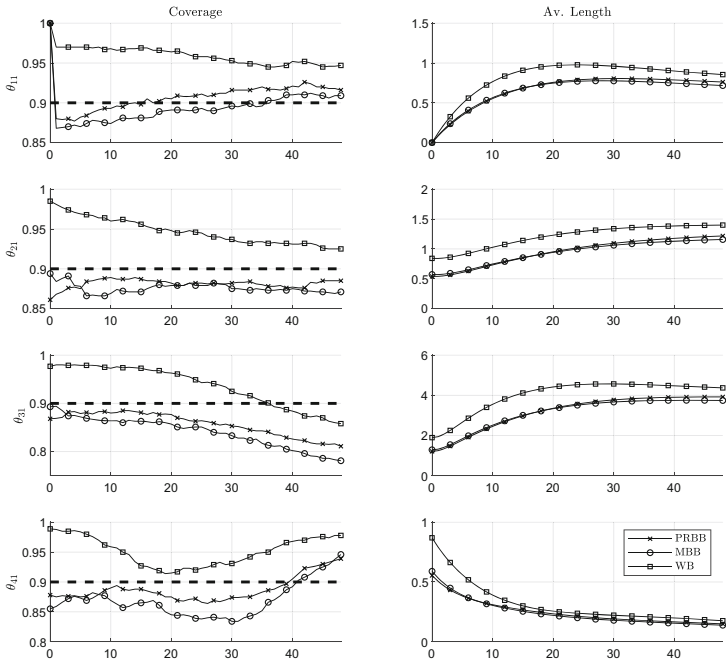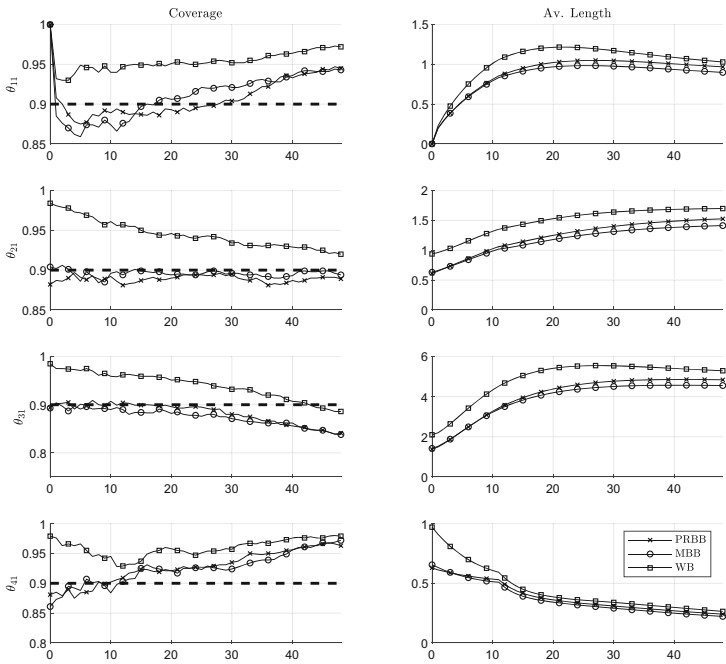
(a) $T = 200$, $p = 1$
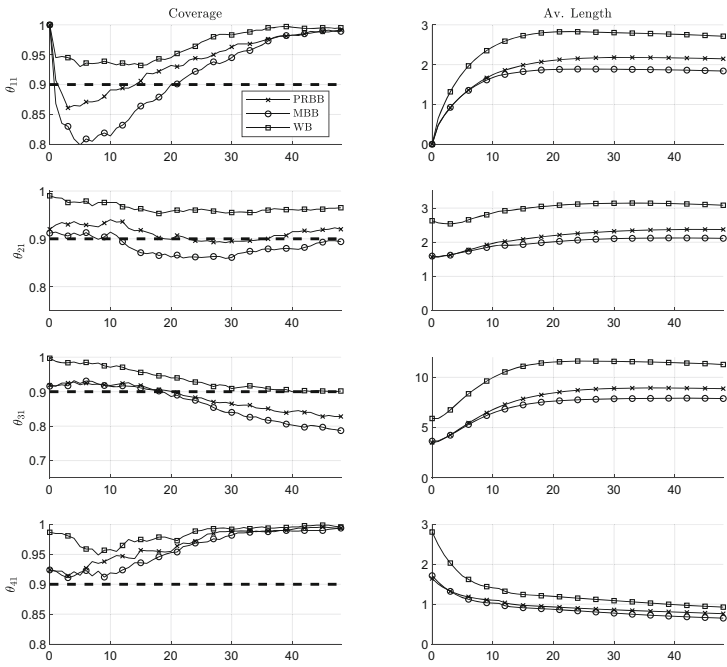
(b) $T = 200$, $p = 12$

Fig. 3 Coverage and average lengths of alternative pointwise bootstrap 90% confidence intervals for DGP2

(a) $T = 500$, $p = 1$

(b) $T = 500$, $p = 12$

Fig. 4 Coverage and average lengths of alternative pointwise bootstrap 90% confidence intervals for DGP2

months ahead federal funds rate future surprises as the baseline proxy to identify a monetary policy shock and they find their proxy to be a strong instrument. We re-estimate their model, shortening the sample to include only periods for which all four variables and the proxy are available. This leaves us with a sample running from 1990M1 through 2016M6, i.e., the sample size is $T = 270$. As the Gertler-Karadi proxy is autocorrelated and predictable, we pre-whiten it by regressing it on its own lags and lags of the endogenous variables and use the residuals as our proxy.[1] The proxy is available in $d = 82\%$ of the sample periods. Following the Gertler-Karadi baseline model specification we include a constant and 12 lags in the VAR.

Figure 5 shows the pointwise 90% confidence bands of the impulse responses to a monetary policy shock that increases the one-year-rate by 25 basis points on impact. Such a shock corresponds roughly to a one standard deviation shock in Gertler and Karadi (2015). The point estimates of the impulse responses are qualitatively in line with the findings by Gertler and Karadi (2015).[2] A monetary tightening induces declining point estimates of the response of industrial production and consumer prices and an increase in the excess bond premium by slightly more than 10 basis points on impact. However, the bootstrap confidence intervals indicate that the responses of industrial production and the CPI may not be significant. Clearly, the choice of the bootstrap procedure affects the widths of the confidence intervals, in line with the simulation results reported in Sect. 4.

From Fig. 5a, it is apparent that the bands estimated via PRBB and MBB tend to be either very similar or the PRBB intervals are slightly larger than the MBB intervals. This outcome is consistent with the simulation evidence, see for example Fig. 3, panel (b). Recall, however, that the shorter MBB intervals in the simulations come at the price of a lower coverage rate which may be below the nominal 90% rate. Although the interpretation of the impulse responses does not depend on the choice of bootstrap in this case, it is, of course, desirable to employ the most reliable inference procedure. Figure 5b compares the PRBB intervals to the WB intervals and shows that the intervals estimated via WB tend to be larger than the PRBB intervals which is again in line with our simulations.

7 Conclusions

In proxy VAR models, an external proxy variable that is correlated with a structural shock of interest and uncorrelated with all other shocks, is used for inference for the impulse responses. In this study, we have proposed a new bootstrap algorithm for such inference. So far frequentist inference in this context is typically based on the WB or the MBB. The former is not valid asymptotically and often yields rather wide

---

[1] Miranda-Agrippino & Ricco, (2021) find that the proxy by Gertler and Karadi (2015) is explained by up to 4 of its own lags, 10 lags of state variables (dynamic factors extracted from the set of monthly variables in McCracken and Ng (2015)), and a constant term. We regress the proxy by Gertler and Karadi (2015) on these variables and use the residuals as our proxy.

[2] Differences compared to Gertler and Karadi (2015) are the result of a shorter sample, the adjustment of the proxy, and differences in the bootstrap procedures. Note that Gertler & Karadi (2015) use a WB with ψ_t from a Rademacher distribution.

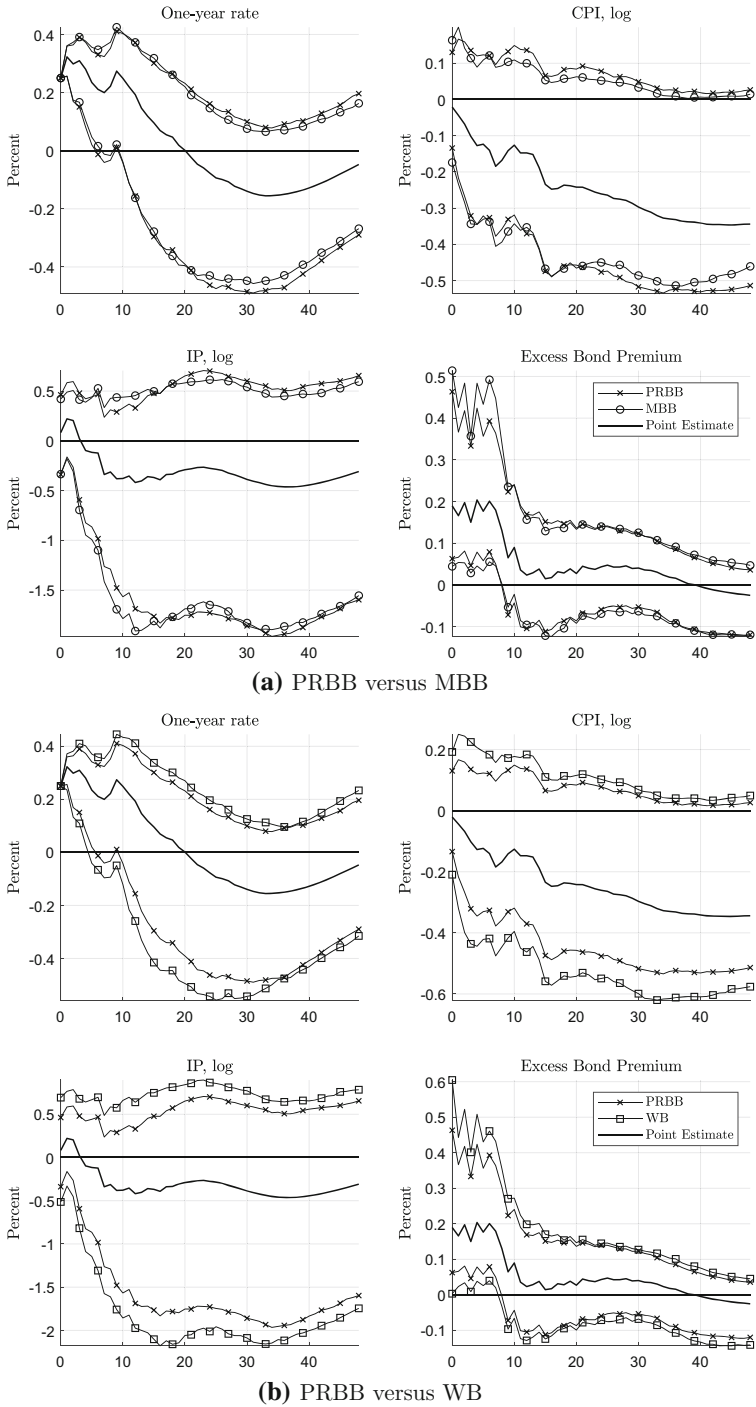**(a)** PRBB versus MBB

(b) PRBB versus WB

Fig. 5 Pointwise bootstrap 90% confidence intervals for the empirical example

confidence intervals, whereas the latter has poor coverage properties in small samples as they are often encountered in macroeconomic studies. We have proposed an alternative bootstrap method which assumes a specific model for the DGP of the proxy variable and samples from the estimated reduced-form errors and the residuals of the proxy model to generate bootstrap samples.

We have shown by simulation that our new PRBB method works well in relatively small samples. Specifically, it yields bootstrap confidence intervals for impulse responses with more accurate coverage than and similar length for comparable coverage to the WB and the MBB. Thus, it has merit for empirical studies for which only relatively small samples are available.

One advantage of the MBB is that it also works asymptotically for conditionally heteroskedastic model errors while the PRBB is not designed for such data features. In our simulations we have found, however, that the PRBB also outperforms the MBB in small samples if such data properties are present. The price paid for the additional generality of the MBB is its reduced accuracy in small samples.

8 Discussion and Extensions

We have proposed a new bootstrap algorithm for inference for proxy VAR models which differs from its main competitors by utilizing an explicit model for the DGP of the proxy. Thereby we achieve more accurate small sample properties of confidence intervals for impulse responses than the WB and the MBB. Clearly, for using the bootstrap algorithm for applied research, the superior small sample accuracy is the key advantage of the new bootstrap algorithm, while the need for modelling the DGP of the proxy is a limitation. So far, the setup of the new bootstrap algorithm does not account for heteroskedastic or conditionally heteroskedastic VAR processes. As the MBB remains asymptotically valid even under changing volatility, it is more general than the new bootstrap algorithm with respect to asymptotic properties. As we show by simulation, the new bootstrap algorithm may still yield more precise confidence intervals than the other two bootstrap algorithms if the data exhibit changing volatility. Thus, our new bootstrap algorithm can be recommended, in particular, for macroeconometric studies where only small or medium sample sizes are available, for example, if only 20 years of quarterly or monthly data are available. In that situation, its superior accuracy makes the new bootstrap algorithm an attractive alternative to the WB and MBB.

To account for more general data features, it may be of interest to explore more general models for the proxy variable in future research. In particular, allowing for serially correlated proxies may be of interest. Moreover, accounting explicitly for changing volatility by modelling such features and designing a bootstrap algorithm accordingly may be an interesting topic for future research. Given the variety of potential deviations of real economic data from the standard model setup, it may also be a fruitful topic for future research to investigate the accuracy of the simple setup of the new bootstrap algorithm considered in this study if more elaborate data features are ignored.

Appendix

A Generating DGP1 with GARCH Residuals

If, instead of drawing u_t from $\mathcal{N}(0, \Sigma_u)$, we aim to generate data with DGP1 with GARCH residuals, we follow Lütkepohl & Schlaak (2018) and proceed in the following steps:

1. Generate $2T$ bivariate standard normal variates, $(e_{1t}, e_{2,t}) \sim \mathcal{N}(0, I_2)$.
2. Generate $\sigma_{k,t|t-1}^2 = (1 - \gamma_k - g_k) + \gamma_k \epsilon_{k,t-1}^2 + g_k \sigma_{k,t-1|t-2}^2$, where $\epsilon_{k,t} = e_{k,t}\sigma_{k,t|t-1}$ and $\sigma_{k,1-1|1-2}^2 = 1$, $k = 1, 2$, are chosen as initial values.
3. Generate $u_t = B \operatorname{diag}(2, 1)\Lambda_{t|t-1}^{0.5} e_t$, where $\Lambda_{t|t-1} = \operatorname{diag}(\sigma_{1,t|t-1}^2, \sigma_{2,t|t-1}^2)$.
4. Generate data y_t recursively from Eq. (5.1), the proxy from Eq. (2.6), and retain the last T observations.

The unconditional variance of u_t is $\Sigma_u = B\Sigma_w B'$, where $\Sigma_w = \operatorname{diag}(4, 1)$. We use GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)'$ and $(\gamma_2, g_2) = (0.05, 0.92)'$.

B Additional Simulation Results

See Figs. 6, 7, 8 and 9.
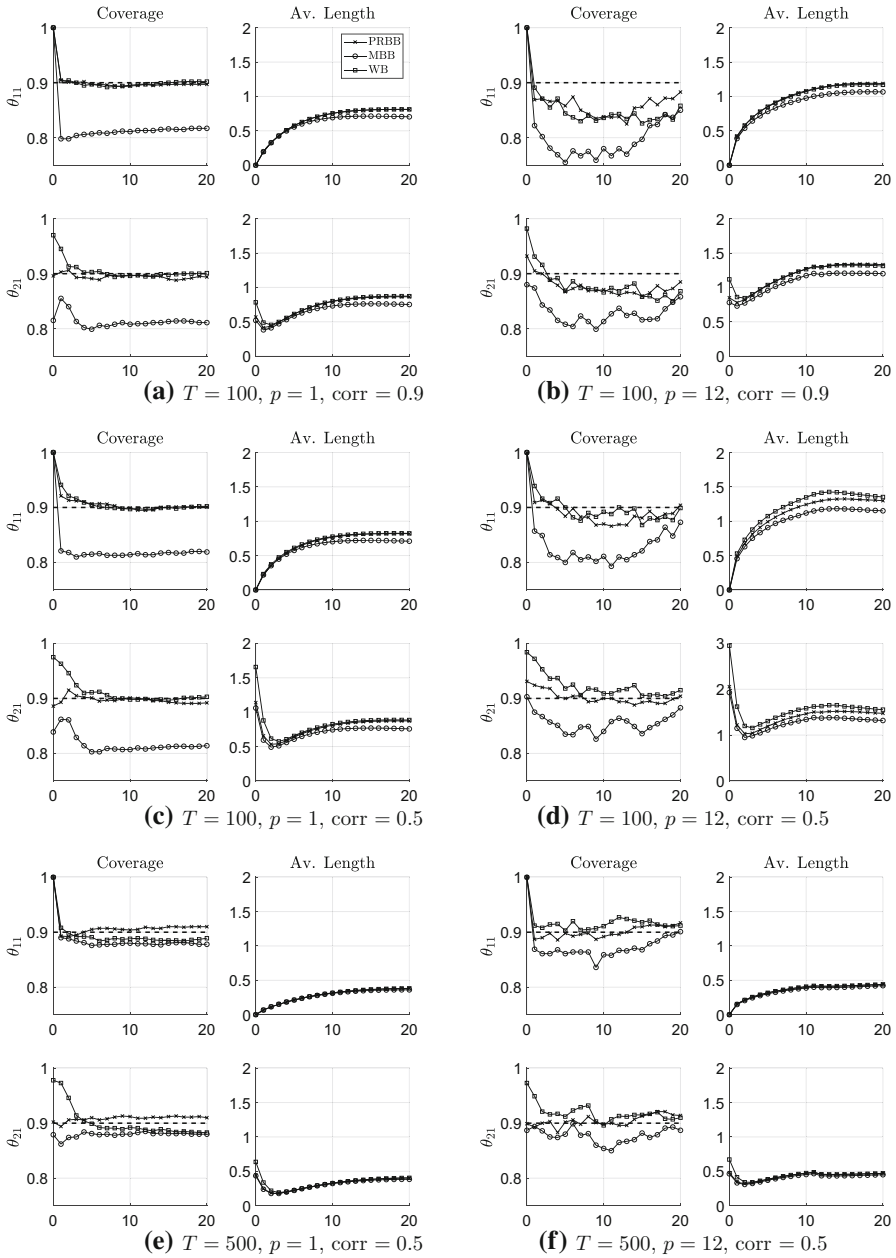
Fig. 6 Coverage and average lengths of alternative pointwise bootstrap 90% confidence intervals for DGP1 with iid errors and $d = 1$, $a_{11} = 0.95$
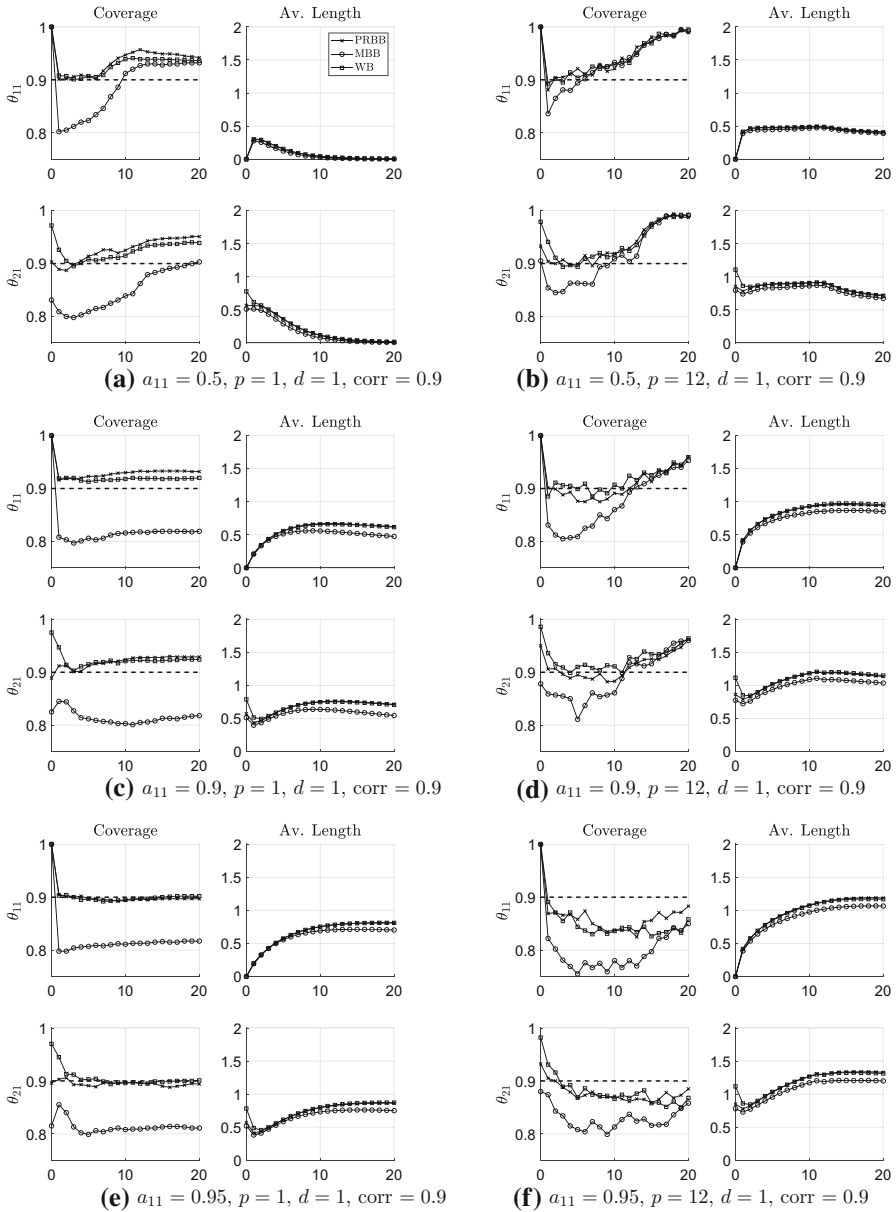
Fig. 7 Coverage and average lengths of alternative pointwise bootstrap 90% confidence intervals for DGP1 with iid errors and $T = 100$
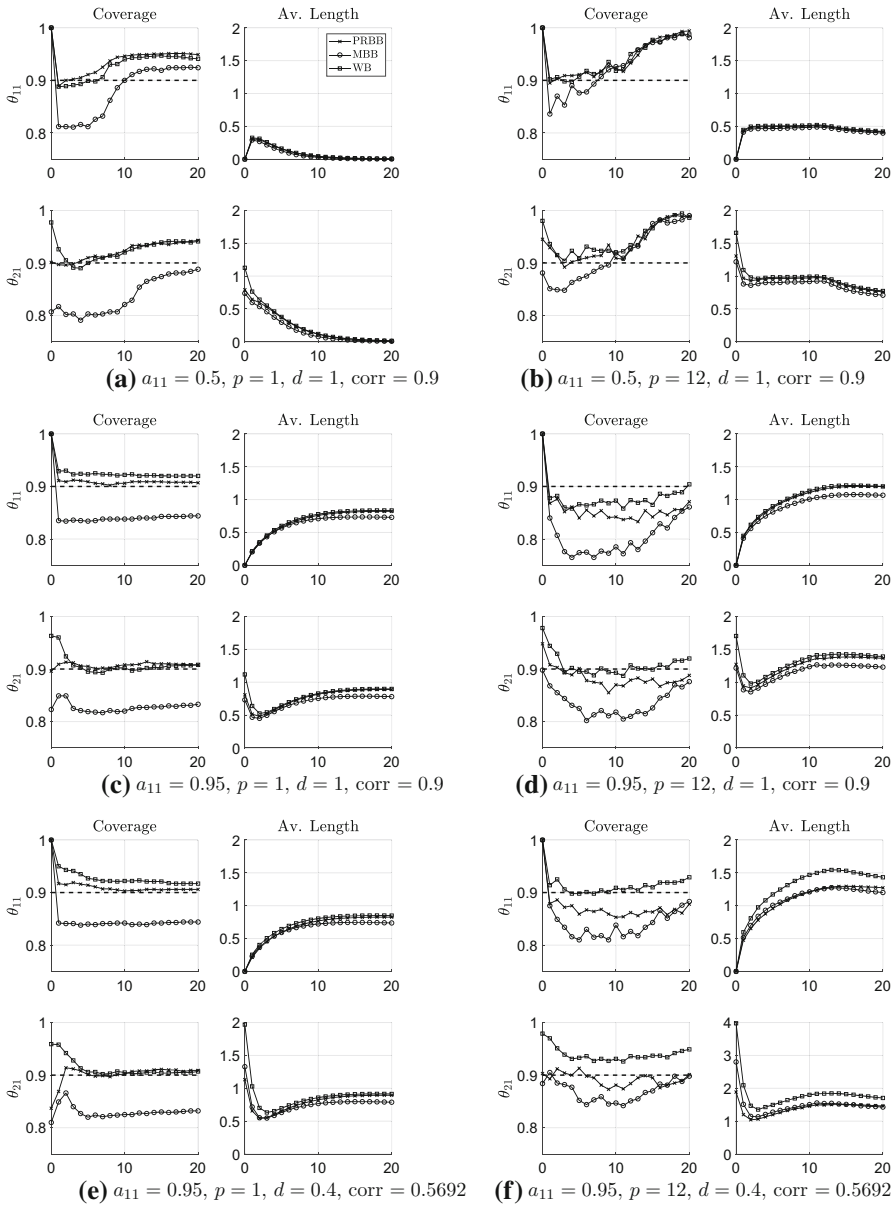
Fig. 8 Coverage and average lengths of alternative pointwise bootstrap 90% confidence intervals for DGP1 with GARCH errors and $T = 100$
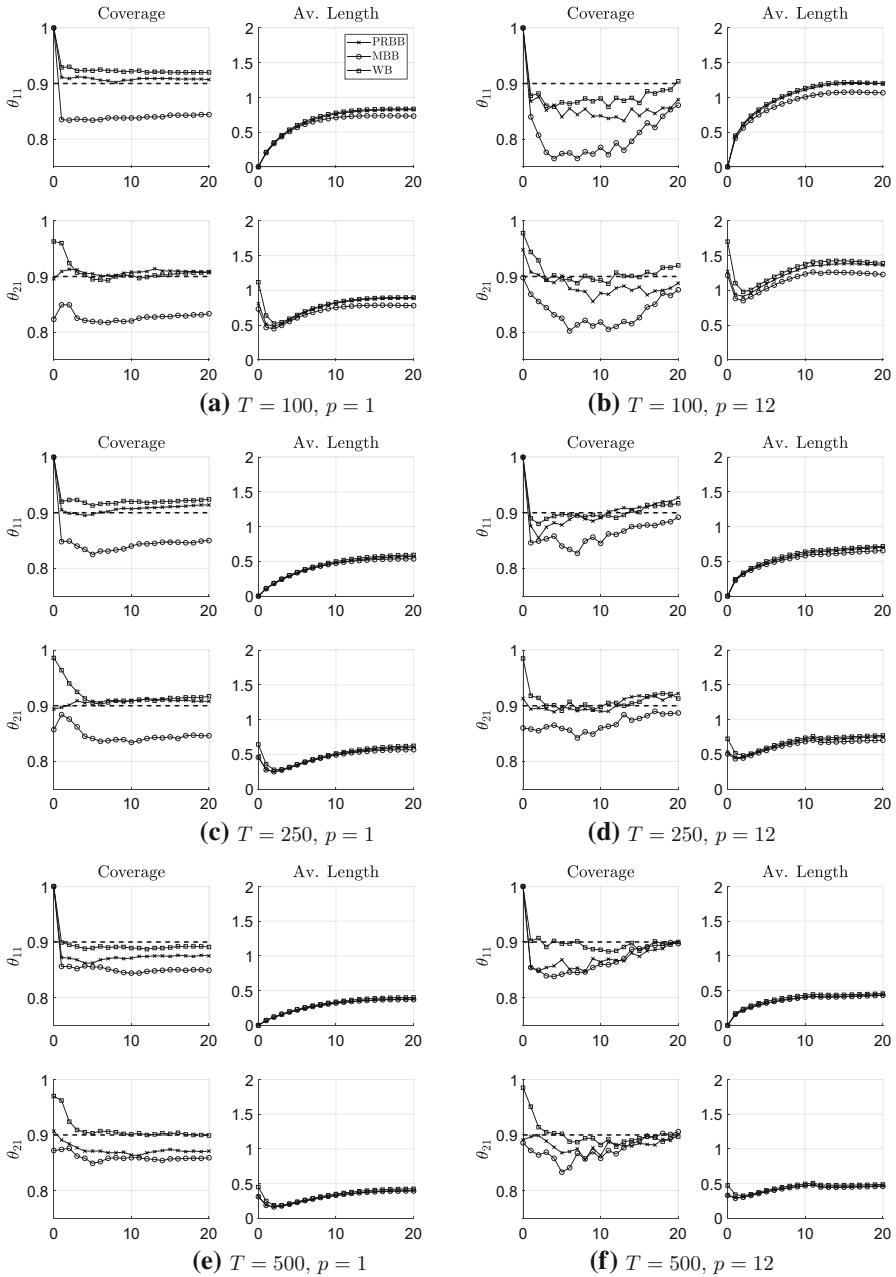
Fig. 9 Coverage and average lengths of alternative pointwise bootstrap 90% confidence intervals for DGP1 with GARCH errors and $d = 1$, $a_{11} = 0.95$, corr $= 0.9$

Declarations

References

Brüggemann, R., Jentsch, C., & Trenkler, C. (2016). Inference in VARs with conditional heteroskedasticity of unknown form. *Journal of Econometrics, 191,* 69–85.

Bruns, M., & Lütkepohl, H. (2022). Comparison of local projection estimators for proxy vector autoregressions. *Journal of Economic Dynamics and Control*, *134*, 104277.

Carriero, A., Mumtaz, H., Theodoridis, K., & Theophilopoulou, A. (2015). The impact of uncertainty shocks under measurement error: A proxy SVAR approach. *Journal of Money, Credit and Banking, 47*, 1223–1238.

Davidson, R., & MacKinnon, J. G. (2004). *Econometric theory and methods*. New York: Oxford University Press.

Gertler, M., & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics, 7*, 44–76.

Jentsch, C., & Lunsford, K. G. (2019). The dynamic effects of personal and corporate income tax changes in the United States: Comment. *American Economic Review, 109*, 2655–2678.

Jentsch, C., & Lunsford, K. G. (2021) . Asymptotically valid bootstrap inference for proxy SVARs, *Journal of Business & Economic Statistics*. https://doi.org/10.1080/07350015.2021.1990770

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics, 80*, 218–230.

Kilian, L., & Kim, Y. (2011). How reliable are local projection estimators of impulse responses? *Review of Economics and Statistics, 93*, 1460–1466.

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

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer-Verlag.

Lütkepohl, H., & Schlaak, T. (2018). Choosing between different time-varying volatility models for structural vector autoregressive analysis. *Oxford Bulletin of Economics and Statistics, 80*(4), 715–735.

Lütkepohl, H., & Schlaak, T. (2019). Bootstrapping impulse responses of structural vector autoregressive models identified through GARCH. *Journal of Economic Dynamics and Control, 101*, 41–61.

Lütkepohl, H., Staszewska-Bystrova, A., & Winker, P. (2015a). Comparison of methods for constructing joint confidence bands for impulse response functions. *International Journal of Forecasting, 31*, 782–798.

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

McCracken, M. W., & Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research, Technical report. Federal Reserve Bank of St. Louis.

Mertens, K., & Ravn, M. O. (2013). The dynamic effects of personal and corporate income tax changes in the United States. *American Economic Review, 103*, 1212–1247.

Miranda-Agrippino, S., & Ricco, G. (2021). The transmission of monetary policy shocks. *American Economic Journal: Macroeconomics, 13*(3), 74–107.

Piffer, M., & Podstawski, M. (2018). Identifying uncertainty shocks using the price of gold. *The Economic Journal, 128*(1549), 3266–3284.

Pope, A. L. (1990). Biases for estimators in multivariate non-Gaussian autoregressions. *Journal of Time Series Analysis, 11*, 249–258.

Stock, J. H., & Watson, M. W. (2012) . Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Activity* pp. 81–135.

Stock, J. H., & Watson, M. W. (2018). Identification and estimation of dynamic causal effects in macroeconomics using external instruments. *The Economic Journal, 128*, 917–948.