JOURNAL OF TIME SERIES ANALYSIS
*J. Time Ser. Anal.* **39**: 641–664 (2018)
Published online 9 May 2018 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12289

ORIGINAL ARTICLE

# BALANCED BOOTSTRAP JOINT CONFIDENCE BANDS FOR STRUCTURAL IMPULSE RESPONSE FUNCTIONS

STEFAN BRUDER AND MICHAEL WOLF*
*Department of Economics, University of Zurich, Zurich, Switzerland*

Constructing joint confidence bands for structural impulse response functions based on a VAR model is a difficult task because of the non-linear nature of such functions. We propose new joint confidence bands that cover the entire true structural impulse response function up to a chosen maximum horizon with a pre-specified probability $(1-\alpha)$, at least asymptotically. Such bands are based on a certain bootstrap procedure from the multiple testing literature. We compare the finite-sample properties of our method with those of existing methods via extensive Monte Carlo simulations. We also investigate the effect of endogenizing the lag order in our bootstrap procedure on the finite-sample properties. Furthermore, an empirical application to a real dataset is provided.

*Received 10 March 2017; Accepted 15 January 2018*

Keywords: Bootstrap; impulse response functions; joint confidence bands; vector autoregressive process

JEL classification: C12; C32

MOS subject classification: 62G09, 62G15, 62M10, 62M20.

## 1. INTRODUCTION

Impulse response analysis based on low-dimensional structural vector autoregressions (VARs) is still a popular tool in applied work; for example, see Barsky and Sims (2011), Kurmann and Otrok (2013), and Bian and Gete (2015). In practice, the impulse response functions have to be estimated from the data, and it is standard in the literature to report the corresponding estimation uncertainty in the form of confidence bands.

It is by now a well-known fact that simply connecting individual marginal confidence intervals with nominal confidence level $(1-\alpha)$ does not result in confidence bands that cover the *entire* true impulse response function with the pre-specified confidence level $(1-\alpha)$. Instead, such a procedure results in joint confidence bands that are too narrow and hence cover the true impulse responses with probability less than the desired level.[1] Consequently, the literature has proposed a number of methods to construct 'proper' joint confidence bands that are designed to actually cover the entire true impulse response function with a pre-specified probability; for example, see Staszewska (2007), Jordà (2009), and Lütkepohl *et al.* (2015a,b).

The finite-sample properties of the existing methods are compared in Lütkepohl *et al.* (2015a,b) via extensive Monte Carlo experiments. They find that the traditional Bonferroni bands and the Wald bands of Lütkepohl *et al.* (2015b) mostly exhibit empirical coverage rates close to or above the nominal level but that the bands can be excessively wide. In contrast, the bands of the other competing methods, namely the bands of Staszewska (2007) and Jordà (2009) as well as the size-adjusted Bonferroni and Wald bands of Lütkepohl *et al.* (2015a,b), are narrower but suffer from finite-sample coverage rates below the nominal level in certain scenarios. Consequently, there is no method so far that produces joint confidence bands for impulse response functions that enjoy both (i)

* Correspondence to: Michael Wolf, Department of Economics, University of Zurich, Zürichbergstrasse 14, 8032 Zurich, Switzerland.
Email: michael.wolf@econ.uzh.ch

[1] This property is obvious from a theoretical point of view; in addition, for example, see Lütkepohl *et al.* (2015a) for Monte Carlo evidence.

robust empirical coverage rates close to the nominal confidence level and (ii) moderate volumes compared to the Bonferroni and Wald bands.

We propose new joint confidence bands for impulse response functions that are based on the methodology of Romano and Wolf (2010), who provide a bootstrap-based method to construct rectangular joint confidence bands for a generic parameter $\theta \in \mathbb{R}^d$. Furthermore, they prove that, under weak regularity conditions, their proposed joint confidence bands have asymptotically the correct coverage probability and are also asymptotically balanced. The resulting joint confidence bands are subsequently labeled as balanced bootstrap (BB) bands.

In addition, we conduct a Monte Carlo experiment to compare the finite-sample properties of the proposed BB bands with those of a set of competing methods. We find that the BB bands are smaller than the Bonferroni and the Wald bands. Furthermore, the BB confidence bands seem to work reliably in scenarios where the ratio of the sample size to the number of coefficients is not small (that is, in scenarios with medium to high degrees of freedom), even when the maximum propagation horizon is large.

The remainder of the paper is organized as follows. Section 2 reviews impulse response functions of structural vector autoregressions. Section 3 presents the new confidence bands. Section 4 briefly describes the competing methods to construct confidence bands. Section 5 describes the Monte Carlo experiment and presents the empirical findings. Section 6 presents an empirical application. Section 7 concludes. The Appendix contains additional details about the estimation of impulse response functions, an algorithm to construct the BB bands, and boxplots describing the finite-sample properties of the various methods. The Supporting information contains detailed tables with the simulation results and figures corresponding to the Monte Carlo simulations and the empirical application.

## 2. STRUCTURAL IMPULSE RESPONSE FUNCTIONS

Consider an $m$-dimensional reduced-form VAR($p$) process of the form

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

where $y_t$ is an $m$-dimensional random vector, the $A_i$ are $m \times m$ coefficient matrices, $\nu$ is an $m$-dimensional intercept vector, and $\{u_t\}$ is an $m$-dimensional independent and identically distributed (i.i.d.) process with $\mathbb{E}[u_t] = 0$ and positive-definite covariance matrix $\Sigma_u := \mathbb{E}[u_t u_t']$. The process in (1) is stable and stationary if and only if

$$\det\left(I_m - A_1 z^1 - \cdots - A_p z^p\right) \neq 0 \quad \text{for all } z \in \mathbb{C} \text{ with } |z| \leq 1 \,.$$

A stationary VAR($p$) process admits a Wold vector moving average (VMA) representation of the form

$$y_t = \mu + \sum_{i=0}^{+\infty} \phi_i u_{t-i} \tag{2}$$

where $\mu := \mathbb{E}[y_t] = \left(I_m - A_1 - \cdots - A_p\right)^{-1} \nu$, and, the $\phi_i$ are fixed $m \times m$ VMA-coefficient matrices that satisfy $\phi_0 = I_m$, and $\phi_s = \sum_{j=1}^{s} \phi_{s-j} A_j$, for $s \in \mathbb{N}_+$.

The structural representation of (1) is given by

$$B_0^{-1} y_t = B_0^{-1}\nu + B_0^{-1} A_1 y_{t-1} + \cdots + B_0^{-1} A_p y_{t-p} + B_0^{-1} u_t \,, \tag{3}$$

where $B_0^{-1} \in \mathbb{R}^{m \times m}$ is a non-singular linear mapping that transforms the reduced-form errors $u_t$ into the structural shocks $\varepsilon_t$, that is, $\varepsilon_t := B_0^{-1} u_t$. The key restriction on $B_0^{-1}$ (or equivalently on $B_0$) emerges from imposing that the structural shocks be instantaneously uncorrelated and have unit variance.[2] Thus, $B_0$ needs to satisfy the following equation:

$$\Sigma_u = B_0 B_0' \,. \tag{4}$$

[2] This implies that the covariance matrix of the $\varepsilon_t$ is equal to the $m$-dimensional identity matrix, that is, $\mathbb{E}[\varepsilon_t \varepsilon_t'] = I_m$.

  *J. Time Ser. Anal.* **39**: 641–664 (2018)
DOI: 10.1111/jtsa.12289

Simple accounting reveals that there are $m(m-1)/2$ degrees of freedom in specifying $B_0$, and hence further restrictions are needed to achieve identification.[3] The literature offers a wide variety of different identification strategies; for example, see Lütkepohl (2005) for a brief overview.

However, we will be agnostic about the particular identification procedure, as our goal is to provide new joint confidence bands with good finite-sample properties rather than to propose a new identification procedure. Thus, at this point, we only assume that the structural VAR is exactly identified via an arbitrary identification procedure.

The identification of the impact matrix $B_0$ allows one to exactly express the reduced-form VAR($p$) process $\{y_t\}$ as a structural vector moving average (VMA) process:

$$y_t = \mu + \sum_{i=0}^{+\infty} \Theta_i \varepsilon_{t-i}\,, \tag{5}$$

where $\Theta_h := \phi_h B_0$. The $(i,j)$th structural impulse response function with a maximum propagation horizon $H \in \mathbb{N}$, denoted by $\mathbf{\Theta}_{ij,H}$, measures the partial effect of a one-standard-deviation shock[4] in the $j$th variable on the $i$th variable over $H+1$ periods and is given by the vector that collects the $(i,j)$-th element of the corresponding structural VMA coefficient matrices, that is,

$$\mathbf{\Theta}_{ij,H} := \begin{pmatrix} \frac{\partial y_{t,i}}{\partial \varepsilon_{t,j}} \\ \vdots \\ \frac{\partial y_{t+H,i}}{\partial \varepsilon_{t,j}} \end{pmatrix} = \begin{pmatrix} \Theta_{ij,0} \\ \vdots \\ \Theta_{ij,H} \end{pmatrix} \qquad \text{for } i,j = 1,\dots,m\,. \tag{6}$$

The structural VMA coefficient matrices at propagation horizon $h \in \{0,\dots,H\}$ can be obtained as $\Theta_h = (JA^hJ')B_0$, where $J := \begin{bmatrix} I_m : 0 : \dots : 0 \end{bmatrix} \in \mathbb{R}^{m \times mp}$ is a selector matrix, and $A$ denotes the reduced-form coefficient matrix of the $mp$-dimensional companion form of a VAR($p$) process. Thus, the structural impulse response function $\mathbf{\Theta}_{ij,H}$ is a non-linear function of the reduced-form model coefficients $(A_1,\dots,A_p)$ and the impact matrix $B_0$, that is,

$$\mathbf{\Theta}_{ij,H} = \mathbf{\Theta}_{ij,H}\left(A_1,\dots,A_p,B_0\right). \tag{7}$$

The reduced-form coefficient matrices $(A_1,\dots,A_p)$ are usually estimated by a standard procedure such as least squares (LS). The impact matrix $B_0$ is in general a function of the reduced-form coefficient matrices $(A_1,\dots,A_p,\Sigma_u)$ and a set of identifying restrictions. An estimator for $B_0$, denoted by $\hat{B}_0$, is found by replacing the true coefficient matrices by the corresponding estimators (and by imposing the identifying restrictions). Thus, a plug-in estimator of the impulse response function is obtained as

$$\widehat{\mathbf{\Theta}}_{ij,H} := \mathbf{\Theta}_{ij,H}\big(\hat{A}_1,\dots,\hat{A}_p,\hat{B}_0\big). \tag{8}$$

The estimator in (8) is consistent if the estimators $(\hat{A}_1,\dots,\hat{A}_p,\hat{B}_0)$ are consistent because $\mathbf{\Theta}_{ij,H}(\cdot)$ is a continuous function.

## 3. NEW JOINT CONFIDENCE BANDS

### 3.1. Motivation and Notation

Romano and Wolf (2010) propose a method to construct joint confidence bands for a generic parameter $\theta \in \mathbb{R}^d$. For them, this method is just a means to an end, where the end is a step-wise multiple testing procedure that controls the family-wise error rate. But we can adapt this method to our 'direct' end of constructing joint confidence bands

[3] Rubio-Ramirez *et al.* (2010) provide a necessary and sufficient condition for global (exact) identification of structural VARs; in particular, the necessary condition of Rubio-Ramirez *et al.* (2010) is equivalent to the widely used (necessary) rank condition of Rothenberg (1971).

[4] Given the normalization $\mathbb{E}[\varepsilon_t\varepsilon_t'] = I_m$, a one-standard-deviation shock is equivalent to a unit shock.

  
DOI: 10.1111/jtsa.12289

for impulse response functions. The method of Romano and Wolf (2010) is based on the availability of a consistent estimator of the parameter of interest and a bootstrap procedure that estimates the sampling distribution of the aforementioned estimator. Therefore, it can be used in a one-to-one manner to construct joint confidence bands for impulse response functions of structural vector autoregressions because both a consistent estimator for $\mathbf{\Theta}_{ij,H}$ and such a bootstrap procedure are available; for example, see Kilian (1998b).

The asymptotic properties of the generic bands hinge on a set of regularity conditions about the asymptotic distribution of the estimator (of the parameter of interest) and the bootstrap; see Romano and Wolf (2010). A discussion of the validity of the regularity conditions in the present context, that is, the construction of joint confidence bands for impulse response functions of structural vector autoregressions, can be found in Section 3.3.

The bands of Romano and Wolf (2010) are rectangular by construction, in contrast to methods that produce joint confidence sets of a non-rectangular shape in first place, from which then rectangular joint confidence bands are obtained by projection on the axes; an example of the latter approach is the Wald bands of Lütkepohl *et al.* (2015b). Such projection methods usually result in conservative joint confidence bands, even asymptotically, which are excessive in volume and thus lead to a loss of information.

Furthermore, the method of Romano and Wolf (2010) is attractive from a computational point of view, since it involves only the computation of the estimator of the impulse response function $\mathbf{\Theta}_{ij,H}$ and an estimator (via the bootstrap) of the sampling distribution of the statistic $\max \sqrt{T}|\widehat{\mathbf{\Theta}}_{ij,H} - \mathbf{\Theta}_{ij,H}|$, where $T$ denotes the sample size, and both the maximum and the absolute value of a vector are understood to be element-wise operators. Both quantities are straightforward to compute and do not contain any sort of potential numerical difficulties such as, for example, the inversion of a large-dimensional matrix.

In contrast, the construction of the size-adjusted Wald joint confidence bands of Lütkepohl *et al.* (2015b) requires on the one hand the computation and the inversion of the (potentially large-dimensional) asymptotic covariance matrix of the vectorized estimators of $(A_1, \dots, A_p, \Sigma_u)$, and on the other hand an iterative procedure to decrease the volume of the confidence bands; see Section 4.3.

Next, the following notation is introduced:

- Let $\left\{ \sqrt{T} \left| \widehat{\Theta}^*_{ij,h,b} - \widehat{\Theta}_{ij,h} \right| \right\}_{b=1}^{B}$ denote the marginal bootstrap distribution at propagation horizon $h \in \{0, \dots, H\}$ based on $B$ bootstrap replications.
- Let $\widehat{\mathrm{H}}^*_h(t)$, $h \in \{0, \dots, H\}$, denote the following empirical distribution function:

$$\forall t \in \mathbb{R}, \quad \widehat{\mathrm{H}}^*_h(t) := \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}_{\left\{ \sqrt{T} \left| \widehat{\Theta}^*_{ij,h,b} - \widehat{\Theta}_{ij,h} \right| \le t \right\}},$$

  and the corresponding empirical quantile function is then given by

$$\widehat{\mathrm{H}}^{*,-1}_h(q) := \inf \left\{ t : \widehat{H}^*_h(t) \ge q \right\}.$$

- Let $\widehat{\mathrm{L}}^*(t)$ denote the following empirical distribution function:

$$\forall t \in \mathbb{R}, \quad \widehat{\mathrm{L}}^*(t) := \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}_{\left\{ \max\limits_{h \in \tilde{S}_{ij}} \left\{ \widehat{H}^*_h \left( \sqrt{T} \left| \widehat{\Theta}^*_{ij,h,b} - \widehat{\Theta}_{ij,h} \right| \right) \right\} \le t \right\}},$$

  where $\tilde{S}_{ij} \subseteq \{0, \dots, H\}$ denotes the propagation horizons when $\sqrt{T} \left| \widehat{\Theta}_{ij,h} - \Theta_{ij,h} \right|$ exhibits a non-degenerate distribution; see Remark 1. The corresponding empirical quantile function is then given by

$$\widehat{\mathrm{L}}^{*,-1}(q) := \inf \left\{ t : \widehat{L}^*(t) \ge q \right\}.$$

  *J. Time Ser. Anal.* **39**: 641–664 (2018)
DOI: 10.1111/jtsa.12289

**Remark 1.** Identifying restrictions may pre-determine the response at one or multiple propagation horizons $\tilde{h}$, that is, for some known constant $c_{\tilde{h}} \in \mathbb{R}$, $\widehat{\Theta}_{ij,\tilde{h}} = c_{\tilde{h}}$ and also $\widehat{\Theta}^*_{ij,\tilde{h},b} = c_{\tilde{h}}$ for all $b$. Consequently, $\widehat{\mathrm{H}}^*_{\tilde{h}}(t) = \mathbb{1}_{[0,\infty)}(t)$ with probability one, and hence the empirical distribution of $\max_{h\in\{0,\dots,H\}} \left\{ \widehat{H}^*_h \left( \sqrt{T} \left| \widehat{\Theta}^*_{ij,h,b} - \widehat{\Theta}_{ij,h} \right| \right) \right\}$ is degenerate at one (with probability one). Defining $\widehat{\mathrm{L}}^*(t)$ as the empirical distribution function of the aforementioned distribution would result in joint confidence bands that are excessively wide because $\widehat{\mathrm{L}}^{*,-1}(1-\alpha) = 1$ for all $\alpha \in [0,1)$; see formula (9). Hence, $\widehat{\mathrm{L}}^*(t)$ is defined as the empirical distribution function of $\max_{h\in\hat{S}_{ij}} \left\{ \widehat{H}^*_h \left( \sqrt{T} \left| \widehat{\Theta}^*_{ij,h,b} - \widehat{\Theta}_{ij,h} \right| \right) \right\}$.

### 3.2. Balanced Bootstrap Joint Confidence Bands

Based on equation (3.7) in Romano and Wolf (2010), we define the BB joint confidence bands for $\mathbf{\Theta}_{ij,H}$ with nominal coverage probability $(1-\alpha)$ as the Cartesian product of the following $(H+1)$ marginal intervals:

$$\left[ \widehat{\Theta}_{ij,h} - \frac{1}{\sqrt{T}} \widehat{\mathrm{H}}^{*,-1}_h \left( \widehat{\mathrm{L}}^{*,-1}(1-\alpha) \right), \widehat{\Theta}_{ij,h} + \frac{1}{\sqrt{T}} \widehat{\mathrm{H}}^{*,-1}_h \left( \widehat{\mathrm{L}}^{*,-1}(1-\alpha) \right) \right] \quad \text{for } h = 0, \dots, H\,. \tag{9}$$

A detailed algorithm for the construction of the BB bands is given in Appendix C. In the following, the BB bands for $\mathbf{\Theta}_{ij,H}$ with a nominal coverage of $(1-\alpha)$ are denoted by $\mathrm{CB}^{(1-\alpha)}_{\mathrm{BB},ij}$. It is worth providing some further discussion about the BB joint confidence bands.

As is evident from (9), the BB bands are based on the estimated sampling distributions of the non-studentized roots $\sqrt{T}|\widehat{\Theta}_{ij,h} - \Theta_{ij,h}|$. Often, confidence intervals based on studentized roots are preferred from a higher-order asymptotic point of view; for example, see Hinkley and Wei (1984). However, under the assumption of stationarity of $\{y_t\}$, the standard deviations of the scaled estimator $\sqrt{T}\widehat{\Theta}_{ij,h}$, denoted by $\sigma_h$, are decreasing in the propagation horizon (for fixed $T$), that is, $\sigma_h \to 0$, and the same is true for the standard errors $\hat{\sigma}_h$; for example, see Lütkepohl (1990). As a consequence, using the estimated sampling distributions of the studentized roots $\sqrt{T}|\widehat{\Theta}_{ij,h} - \Theta_{ij,h}|/\hat{\sigma}_h$ results in joint confidence bands that can have excessively large volume, as pointed out by Lütkepohl *et al.* (2015a).

The construction of the BB bands involves the pre-pivoting transformation of Beran (1987); that is, the roots that underlie the joint confidence bands are monotonically transformed by their estimated empirical distribution function $\widehat{H}^*_h$. Beran (1987) argues that the pre-pivoting transformation reduces the coverage bias of marginal confidence intervals and also results in improved higher-order properties, similar to studentized roots. Consequently, using the pre-pivoting transformation results in BB joint confidence bands with good coverage properties but without excessive volume; see the Monte Carlo simulations in Section 5.

The BB bands are symmetric around the estimated impulse response function. The methodology of Romano and Wolf (2010) also allows the construction of asymmetric, 'equal-tailed' joint confidence bands based on the the estimated distribution of the one-sided roots $\sqrt{T}(\widehat{\Theta}_{ij,h} - \Theta_{ij,h})$. But simulation results (not reported here) suggest that the symmetric bands are superior to the asymmetric bands in terms of finite-sample coverage properties.

**Remark 2.** In the absence of any 'favouritism' of certain propagation horizons, the property of balance is a desirable one, as has previously been argued by Beran (1987,1988) and Romano and Wolf (2010) in more general contexts: Balanced confidence bands spread out the probability of missing at least one element of the impulse response function evenly over the individual propagation horizons (up to the maximum propagation horizon $H$ considered).

Another way to look at this issue is the following: If the property of balance were considered completely irrelevant, it would be easy to construct joint confidence bands with coverage $(1-\alpha)$: Construct a marginal confidence interval for the impulse response function at propagation horizon zero with coverage $(1-\alpha)$ and take the Cartesian product of it with the Cartesian product of $H$ times the real line. The resulting Cartesian product then trivially results in valid joint confidence bands with maximum propagation horizon $H$, for any $H$. Such joint confidence bands are extremely unbalanced and are of no use in practice.

  
DOI: 10.1111/jtsa.12289

Of course, this example is perverse, since all intervals but the first are unbounded. However, it can be considered as a limiting case for a non-perverse example where all but the first interval have individual coverage probabilities that are close to 1 (but less than 1) and where the first interval has individual coverage probability close to $(1-\alpha)$ (but greater than $1-\alpha$), in a way such that the coverage probability of the confidence bands is equal to $(1-\alpha)$. Clearly, such imbalanced bands are also not desirable from a practical point of view.

Last but not least, it can be expected that imposing the property of balance, at least asymptotically, will result in joint confidence bands with small volume; though it may well be possible to find joint confidence bands with even smaller volume if the property of balance is abandoned, a topic which is left to future research.

If certain propagation horizons are 'favoured' over others, then it is desirable to construct imbalanced joint confidence bands such that the marginal coverage probabilities at the favoured propagation horizons are suitably higher than the other propagation horizons. An explicit construction of this sort is beyond the scope of this paper. However, the solution in such a case can certainly not employ joint confidence bands whose balance properties are unknown and which do not adapt to any 'favoured' propagation horizons, either, such as the Wald-type bands of Lütkepohl *et al.* (2015b). ■

**Remark 3.** Cao and Sun (2011) derive the asymptotic distribution of structural impulse response functions of short panel vector autoregressions. Furthermore, Cao and Sun (2011) compare the finite-sample coverage properties of marginal confidence for individual responses based on the asymptotic distribution with the properties of various bootstrap intervals, but joint confidence bands for the entire impulse response function are not considered in their study. We expect that our proposed method can also be applied to construct joint confidence bands for impulse response functions of short panel vector autoregressions. However, a detailed analysis of this topic is beyond the scope of this paper. ■

### 3.3. Asymptotic Properties

The regularity conditions underlying Theorem 3.1 of Romano and Wolf (2010), which states the asymptotic properties of the generic bootstrap joint confidence bands, involve the asymptotic distribution of the estimator of the impulse response function and the asymptotic validity of the bootstrap. Thus, for the sake of completeness, both assumptions are subsequently reviewed.

Under standard assumptions and when LS is used for the estimation of $\left(A_1, \ldots, A_p, \Sigma_u\right)$, the asymptotic distribution of $\sqrt{T}(\widehat{\boldsymbol{\Theta}}_{ij,H} - \boldsymbol{\Theta}_{ij,H})$ is generally derived via an application of the delta method; for example, see Lütkepohl (1990). Thus, the asymptotic distribution of the (standardized) estimator of the impulse response function is typically normal because the (vectorized) LS estimator of $\left(A_1, \ldots, A_p, \Sigma_u\right)$ is asymptotically normal under weak high-level assumptions: for example, see Lütkepohl (2005). However, Lütkepohl (1989) and Benkwitz *et al.* (2000) note that the asymptotic covariance matrix, denoted by $\Sigma_{\widehat{\Theta}}$, is singular in certain scenarios; hence, in such scenarios, the limiting distribution of $\sqrt{T}(\widehat{\boldsymbol{\Theta}}_{ij,H} - \boldsymbol{\Theta}_{ij,H})$ is not normal, but degenerate normal instead.

This characteristic of the asymptotic distribution of the estimated impulse response function (that is, normal versus degenerate normal) has an impact on the consistency of the bootstrap for the joint sampling distribution of $\sqrt{T}(\widehat{\boldsymbol{\Theta}}_{ij,H} - \boldsymbol{\Theta}_{ij,H})$. More specifically, in case the asymptotic distribution is non-degenerate normal, the bootstrap is consistent because the usual smoothness conditions underlying the bootstrap are satisfied; for example, see Horowitz (2001). However, the bootstrap may not be consistent when the asymptotic distribution is degenerate normal; for example, see Benkwitz *et al.* (2000).

It is evident that assumptions B1–B4 of Romano and Wolf (2010) are satisfied if the asymptotic distribution of $\widehat{\boldsymbol{\Theta}}_{ij,H}$ is non-degenerate normal, which is the case if $\Sigma_{\widehat{\Theta}}$ is positive definite. Thus, the asymptotic properties of $\mathrm{CB}_{\mathrm{BB},ij}^{(1-\alpha)}$ can then be deduced from Theorem 3.1 of Romano and Wolf (2010). More specifically, it holds that

$$\lim_{T\to\infty} \mathbb{P}\left(\boldsymbol{\Theta}_{ij,H} \in \mathrm{CB}_{\mathrm{BB},ij}^{(1-\alpha)}\right) = (1-\alpha)\,, \tag{10}$$

  
DOI: 10.1111/jtsa.12289

if $\Sigma_{\widehat{\Theta}}$ is positive definite. Furthermore, let $\mathrm{CB}^{(1-\alpha)}_{\mathrm{BB},ij,h}$ denote the $h$th marginal confidence interval for $\Theta_{ij,h}$; then it holds that

$$\lim_{T\to\infty} \mathbb{P}\left(\Theta_{ij,h} \in \mathrm{CB}^{(1-\alpha)}_{\mathrm{BB},ij,h}\right) = \rho \in (0,1) \quad \forall h \in \{0,\dots,H\} \ , \tag{11}$$

if $\Sigma_{\widehat{\Theta}}$ is positive definite. Summarizing, under the condition of a positive definite covariance matrix $\Sigma_{\widehat{\Theta}}$, the BB joint confidence bands have asymptotically the correct coverage rate and are asymptotically balanced in the sense that the coverage rates of the marginal intervals $\mathrm{CB}^{(1-\alpha)}_{\mathrm{BB},ij,h}$ are asymptotically independent of $h \in \{0,\dots,H\}$.

Data generating processes that give rise to a singular asymptotic covariance matrix $\Sigma_{\widehat{\Theta}}$ are included in the Monte Carlo experiment in Section 5 to gain some simulation-based insights about the properties of the BB bands in scenarios with an asymptotic degenerate normal distribution.

**Remark 4.** The traditional Bonferroni bands and the size-adjusted Bonferroni bands of Lütkepohl *et al.* (2015a) have a similar handicap: These joint confidence bands are also proven to work only if the bootstrap is consistent for all marginal distributions $\sqrt{T}(\widehat{\Theta}_{ij,h} - \Theta_{ij,h})$, which is the case if $\sqrt{T}(\widehat{\Theta}_{ij,h} - \Theta_{ij,h})$ converges to a non-degenerate normal distribution for all $h \in \{0,\dots,H\}$; for more details, see Lütkepohl et al. (2015b). ■

The simultaneous test of $H_{0,h} : \Theta_{ij,h} = 0, h \in \{0,\dots,H\}$, is of great interest in applied work. Such a test can be carried out by 'inverting' the joint confidence bands for $\mathbf{\Theta}_{ij,H}$. In particular, any $H_{0,h}$ is rejected for which zero is not contained in the marginal confidence interval $\mathrm{CB}^{(1-\alpha)}_{\mathrm{BB},ij,h}$. It follows from Corollary 3.1 of Romano and Wolf (2010) that such a testing procedure asymptotically controls the probability of falsely rejecting at least one true hypothesis $H_{0,h}$, that is,

$$\limsup_{T\to\infty} \mathbb{P}(\text{reject at least one true hypothesis } H_{0,h}) \leq \alpha \ , \tag{12}$$

at least as long as $\Sigma_{\widehat{\Theta}}$ is positive definite. In other words, for all $h \in \{0,\dots,H\}$ for which zero is not contained in the corresponding marginal confidence interval, one can be jointly confident that the true impulse response $\Theta_{ij,h}$ is non-zero; that is, the confidence holds jointly for all such $h$ and not just individually (for a given such $h$).

## 4. COMPETING METHODS

In order to assess the finite-sample performance of our proposed method, we compare its finite-sample properties with those of relevant competing methods in the literature. More specifically, the list of the competing bands consists of the Naïve bands, the traditional Bonferroni bands, and the recently proposed Wald and Adjusted-Wald bands of Lütkepohl *et al.* (2015b). In the following, each of the four competing methods is briefly outlined; more details are found in the corresponding references.

### 4.1. Naïve Confidence Bands

The Naïve confidence bands for $\mathbf{\Theta}_{ij,H}$, as defined in Lütkepohl *et al.* (2015a), are given by the collection of the $(H+1)$ marginal confidence intervals with individual confidence level $(1-\alpha)$, that is

$$\mathrm{CB}^{(1-\alpha)}_{\mathrm{Naïve},ij} := \left[q^{*,ij}_{0,\frac{\alpha}{2}}, q^{*,ij}_{0,(1-\frac{\alpha}{2})}\right] \times \cdots \times \left[q^{*,ij}_{H,\frac{\alpha}{2}}, q^{*,ij}_{H,(1-\frac{\alpha}{2})}\right] \ , \tag{13}$$

where $q^{*,ij}_{h,\frac{\alpha}{2}}$ and $q^{*,ij}_{h,(1-\frac{\alpha}{2})}$ denote the $\frac{\alpha}{2}$ and $1-\frac{\alpha}{2}$ quantiles of the bootstrap distribution of the estimated impulse response coefficient at horizon $h \in \{0,\dots,H\}$.

  
DOI: 10.1111/jtsa.12289

### 4.2. Bonferroni Joint Confidence Bands

The Bonferroni joint confidence bands for $\mathbf{\Theta}_{ij,H}$ consist of the Cartesian product of $(H+1)$ marginal confidence intervals for the individual responses $\Theta_{ij,h}, h \in \{0,\dots,H\}$, where the nominal confidence level of the marginal intervals is adjusted via Bonferroni's inequality in order to ensure that the joint coverage probability is, asymptotically, at least $(1-\alpha)$. (Of course, this can be guaranteed only if the underlying bootstrap method is consistent.) More specifically, the adjusted marginal nominal confidence level is equal to $(1-\beta)$, where $\beta := \alpha/(H+1)$.[5] The rectangular Bonferroni joint confidence bands for $\mathbf{\Theta}_{ij,H}$, as defined in Lütkepohl *et al.* (2015a), are given by

$$\mathrm{CB}^{(1-\alpha)}_{\mathrm{B},ij} := \left[q^{*,ij}_{0,\frac{\beta}{2}}, q^{*,ij}_{0,(1-\frac{\beta}{2})}\right] \times \cdots \times \left[q^{*,ij}_{H,\frac{\beta}{2}}, q^{*,ij}_{H,(1-\frac{\beta}{2})}\right] , \tag{14}$$

where $q^{*,ij}_{h,\frac{\beta}{2}}$ and $q^{*,ij}_{h,(1-\frac{\beta}{2})}$ denote the $\frac{\beta}{2}$ and $1-\frac{\beta}{2}$ quantiles of the bootstrap distribution of the estimated impulse response coefficient at horizon $h \in \{0,\dots,H\}$.

### 4.3. Wald and Adjusted-Wald Joint Confidence Bands

The Wald joint confidence bands for $\mathbf{\Theta}_{ij,H}$ of Lütkepohl *et al.* (2015b) are constructed in two steps. First, the bootstrap Wald test for the relevant reduced-form parameters of the underlying VAR is inverted to obtain a joint confidence ellipse. Second, the ellipse is projected onto the axes of the impulse response space, resulting in rectangular joint confidence bands for $\mathbf{\Theta}_{ij,H}$.

More specifically, the Wald joint confidence ellipse for the relevant reduced-form coefficients, denoted by $\theta$, with a nominal confidence level of $(1-\alpha)$ is given by

$$\mathcal{W}^{\theta}_{(1-\alpha)} = \left\{ \theta : T\left(\hat{\theta}-\theta\right)' \left(\widehat{\Sigma}_{\theta}\right)^{-1} \left(\hat{\theta}-\theta\right) \leq w^{*}_{(1-\alpha)} \right\} ,$$

where $\hat{\theta}$ denotes a consistent estimator for $\theta$, $\widehat{\Sigma}_{\theta}$ denotes a consistent estimator of the asymptotic variance of $\hat{\theta}$, and $w^{*}_{(1-\alpha)}$ denotes the $(1-\alpha)$ quantile of the bootstrap distribution of the Wald statistic.[6] The bootstrap impulse responses are ordered according to the set of increasing bootstrap Wald statistics $\left\{w^{*}_{1} \leq \cdots \leq w^{*}_{B}\right\}$ in the sense that $\widehat{\mathbf{\Theta}}^{*}_{ij,H,n}$ corresponds to $w^{*}_{n}$. Finally, the rectangular Wald joint confidence bands for $\mathbf{\Theta}_{ij,H}$ with nominal confidence level $(1-\alpha)$ are given as the envelope of the ordered set of bootstrap impulse response functions $\left\{\widehat{\mathbf{\Theta}}^{*}_{ij,H,1},\dots,\widehat{\mathbf{\Theta}}^{*}_{ij,H,(1-\alpha)\times B}\right\}$, that is,

$$\mathrm{CB}^{(1-\alpha)}_{\mathrm{Wald},ij} := \left[l^{*}_{ij,0}, u^{*}_{ij,0}\right] \times \cdots \times \left[l^{*}_{ij,H}, u^{*}_{ij,H}\right] , \tag{15}$$

where $l^{*}_{ij,s} := \min\left\{\widehat{\Theta}^{*}_{ij,b,n} : n = 1,\dots,(1-\alpha)\times B\right\}$, and $u^{*}_{ij,s}$ is defined as the corresponding upper bound; we assume here tacitly that $(1-\alpha)\times B$ is an integer; otherwise, we take the smallest integer larger than $(1-\alpha)\times B$.

Lütkepohl *et al.* (2015b) point out that the Wald bands are conservative by construction, and hence usually cover more than $(1-\alpha)\times B$ of the bootstrap impulse response functions. Thus, a volume adjustment of the Wald bands can be considered, and Lütkepohl *et al.* (2015b) propose the following iterative adjustment:[7] Remove iteratively

[5] In case the initial response is zero by construction, $\beta$ is equal to $\alpha/H$.

[6] This is the empirical distribution of $w^{*}_{b} := T\left(\hat{\theta}^{*}_{b}-\hat{\theta}\right)' \left(\widehat{\Sigma}^{*}_{\theta}\right)^{-1} \left(\hat{\theta}^{*}_{b}-\hat{\theta}\right), b = 1,\dots,B$, where $\hat{\theta}^{*}_{b}$ and $\widehat{\Sigma}^{*}_{\theta}$ are estimators based on bootstrap data $\left\{y^{*}_{1},\dots,y^{*}_{T}\right\}$.

[7] Lütkepohl *et al.* (2015b) also propose another adjustment procedure and the resulting joint confidence bands are called Bonferroni-adjusted Wald bands, see Lütkepohl et al. (2015b).

  
DOI: 10.1111/jtsa.12289

the last element in the set of ordered bootstrap impulse responses until the bootstrap coverage of the envelope of the remaining functions is greater than or equal to $(1-\alpha)$. The resulting rectangular Adjusted-Wald joint confidence bands for $\mathbf{\Theta}_{ij,H}$ with a nominal confidence level $(1-\alpha)$ are given as the envelope of the remaining bootstrap impulse responses, that is,

$$\mathrm{CB}^{(1-\alpha)}_{\text{Adj-W},ij} := \left[\tilde{l}^{*}_{ij,0}, \tilde{u}^{*}_{ij,0}\right] \times \cdots \times \left[\tilde{l}^{*}_{ij,H}, \tilde{u}^{*}_{ij,H}\right] , \tag{16}$$

where $\tilde{l}^{*}_{ij,h} := \min\left\{\widehat{\Theta}^{*}_{ij,b,h} : b = 1,\ldots,\tilde{B}\right\}$, and $\tilde{B}$ denotes index of the first bootstrap impulse response that is not removed, and $\tilde{u}^{*}_{ij,H}$ is defined as the corresponding upper bound.

## 5. MONTE CARLO SIMULATION

### 5.1. Lag Selection and Estimation of Impulse Responses

The lag order is selected using the Akaike information criterion (AIC), as Kilian (2001) provides simulation evidence that confidence intervals (for individual responses) based on the AIC exhibit superior finite-sample coverage properties compared to confidence intervals based on the Schwarz information criterion and the Hannan–Quinn criterion. The maximum lag order $p_{\max}$ is determined endogenously using the rule of thumb proposed in Schwert (1989). According to this rule, the maximum lag order is given by

$$p_{\max} := \left\lfloor 12(T/100)^{0.25} \right\rfloor , \tag{17}$$

where $\lfloor\cdot\rfloor$ denotes the integer part of a real number, and $T$ denotes the sample size. Thus, the maximum lag order is given by 12, 14, and 16 for sample sizes of 100, 200, and 400, respectively. In the following, $\hat{p}$ denotes the lag order selected by the AIC, that is, $\hat{p} := \hat{p}_{\mathrm{AIC}}$.

We estimate the reduced-form coefficients $(\nu, A_1, \ldots, A_{\hat{p}})$ of the VAR model by least squares; see Lütkepohl (2005) for more details. It is well known that the LS estimator is biased in finite samples because of the presence of lagged endogenous variables. Thus, we correct for the finite-sample bias using the closed-form bias estimator of Pope (1990); see Appendix A for details. The corresponding bias-corrected estimators of the slope coefficients are given by

$$\hat{A}^{\mathrm{BC}}_{i} := \hat{A}_{\mathrm{LS},i} - \widehat{\mathrm{Bias}}(\hat{A}_{\mathrm{LS},i}) , \quad \text{for } i = 1,\ldots,\hat{p} ,$$

where $\hat{A}_{\mathrm{LS},i}$ denotes the LS estimator of $A_i$, and $\widehat{\mathrm{Bias}}(\hat{A}_{\mathrm{LS},i})$ denotes Pope's corresponding bias estimator. Furthermore, in scenarios where the bias correction causes nonstationarity, that is, where the process corresponding to $(\nu, \hat{A}_{\mathrm{LS},1}, \ldots, \hat{A}_{\mathrm{LS},\hat{p}})$ is stationary but the process corresponding to $(\nu^{\mathrm{BC}}, \hat{A}^{\mathrm{BC}}_{1}, \ldots, \hat{A}^{\mathrm{BC}}_{\hat{p}})$ is non-stationary, the stationarity correction of Kilian (1998b) is applied instead; see Appendix B for details.

We assume a recursive structure of the structural VAR model, that is, the impact matrix $B_0$ is given by the lower triangular Cholesky decomposition of $\Sigma_u$. The corresponding estimator is naturally given by

$$\widehat{B}_0 := \mathrm{chol}\left(\widehat{\Sigma}^{\mathrm{BC}}_{u}\right) ,$$

where $\widehat{\Sigma}^{\mathrm{BC}}_{u}$ denotes the estimated residual covariance matrix based on the bias-corrected VAR coefficient estimators. Summarizing, the estimated structural impulse response functions $\widehat{\mathbf{\Theta}}_{ij,H}$ are obtained as

$$\widehat{\mathbf{\Theta}}_{ij,H} := \mathbf{\Theta}_{ij,H}\left(\hat{A}^{\mathrm{BC}}_{1}, \ldots, \hat{A}^{\mathrm{BC}}_{\hat{p}}, \mathrm{chol}\left(\widehat{\Sigma}^{\mathrm{BC}}_{u}\right)\right) . \tag{18}$$

  
DOI: 10.1111/jtsa.12289

### 5.2. Bootstrap Details

The bootstrap distribution of the estimator of the structural impulse response functions is generated by the following nonparametric bootstrap procedure of Kilian (1998b):

(a) Given $\{y_t\}_{t=1}^{T}, \hat{p}, \left(\hat{\nu}^{\mathrm{BC}}, \hat{A}_1^{\mathrm{BC}}, \ldots, \hat{A}_{\hat{p}}^{\mathrm{BC}}\right)$, and the corresponding series of residuals $\{\hat{u}_t\}_{t=\hat{p}+1}^{T}$, generate a bootstrap sample $\{y_1^*, \ldots, y_T^*\}$ via the following recursion:

$$y_t^* := \begin{cases} y_t\,, & t = 1, \ldots, \hat{p}, \\ \hat{\nu}^{\mathrm{BC}} + \hat{A}_1^{\mathrm{BC}} y_{t-1}^* + \cdots + \hat{A}_{\hat{p}}^{\mathrm{BC}} y_{t-\hat{p}}^* + e_t^*\,, & t = \hat{p}+1, \ldots, T, \end{cases} \tag{19}$$

where $e_t^*$ is a random draw with replacement from the empirical distribution of the residuals that are rescaled and centered to have mean zero.[8]

(b) Obtain $\left(\hat{A}_1^{*,\mathrm{BC}}, \ldots, \hat{A}_{\hat{p}}^{*,\mathrm{BC}}\right)$ and $\widehat{\Sigma}_u^{*,\mathrm{BC}}$ by fitting a VAR($\hat{p}$) model to $\{y_t^*\}_{t=1}^{T}$.

(c) Obtain $\widehat{\boldsymbol{\Theta}}_{ij,H}^{*} = \boldsymbol{\Theta}_{ij,H}\left(\hat{A}_1^{*,\mathrm{BC}}, \ldots, \hat{A}_{\hat{p}}^{*,\mathrm{BC}}, \mathrm{chol}\left(\widehat{\Sigma}_u^{*,\mathrm{BC}}\right)\right)$.

(d) Repeat steps (a)–(c) $B$ times resulting in the bootstrap sample $\left\{\widehat{\boldsymbol{\Theta}}_{ij,H,b}^{*}\right\}_{b:=1}^{B}$.

The previously outlined bootstrap algorithm is subsequently referred to as the *exogenous bootstrap* because the lag order is *not* re-estimated based on the bootstrap sample $\{y_1^*, \ldots, y_T^*\}$.

The Wald-type joint confidence bands of Lütkepohl *et al.* (2015b) can be constructed only by using the exogenous bootstrap because the computation of the bootstrap Wald statistic requires that the vector of the estimator of the slope coefficients $(\hat{A}_1, \ldots, \hat{A}_{\hat{p}})$ and the bootstrap analogues $(\hat{A}_1^*, \ldots, \hat{A}_{\hat{p}^*}^*)$ have the same dimension and hence the same lag order, that is, $\hat{p} := \hat{p}^*$. Therefore, a fair comparison of the finite-sample performance of the BB bands with the competing methods should be based on the exogenous bootstrap.

**Remark 5.** The exogenous bootstrap algorithm differs from the original algorithm in Kilian (1998b) in one minor aspect. The LS parameter estimates of the reduced-form coefficients are corrected for their finite-sample bias using the closed-form bias formula of Pope (1990) instead of a bootstrap-based bias correction as in Kilian (1998b). This modification can be justified, on one hand, since both procedures remove only the first-order bias and, on the other hand, since both procedure exhibit a similar finite-sample performance, as is shown in the Monte Carlo study by Engsted and Pedersen (2014). ■

Kilian (1998a) provides simulation-based evidence that endogenizing the lag order selection in the bootstrap procedure results in an improved coverage accuracy of marginal bootstrap intervals for impulse responses of structural vector autoregressions. In order to investigate whether a similar effect can be observed for joint confidence bands, the BB bands, the Naïve bands, and the Bonferroni bands will be additionally constructed based on the endogenous bootstrap procedure of Kilian (1998a).

**Remark 6.** Our suggested bootstrap procedure is an extension of previous proposals for univariate finite-order ARMA models to multivariate finite-order VAR models, where the order is determined in a data-dependent fashion as opposed to being assumed known; in particular, we follow Kilian (2001) in using the AIC to select the order. As stated before, the maximum order considered is allowed to tend to infinity together with the sample size $T$; see (17). Therefore, our bootstrap procedure can also be considered a sieve bootstrap, whose validity in more general models, that is, in models more general than a finite-order VAR model, is studied in Meyer and Kreiss (2015).

[8] The centering and rescaling is carried out as suggested in Stine (1987).

  
DOI: 10.1111/jtsa.12289

There are recent bootstrap procedures, such as the hybrid bootstrap of Jentsch and Kreiss (2010) and the linear process bootstrap of Jentsch and Politis (2015), which can be applied in the present context to generate the bootstrap data $\{y_1^*, \dots, y_T^*\}$. However, an in-depth analysis of the effect of employing different bootstrap procedures on the finite-sample performances of confidence bands for structural impulse response functions is beyond the scope of this paper. ■

### 5.3. Data Generating Processes

*5.3.1. Bivariate VAR(1) Models*

We first consider the bivariate data generating processes from Kilian (1998b), which were previously considered in Lütkepohl *et al.* (2015a,b) in the context of joint confidence bands for structural impulse response functions, that is

$$\textbf{DGP-1} \quad y_t = \begin{pmatrix} \rho & 0.0 \\ 0.5 & 0.5 \end{pmatrix} y_{t-1} + u_t \,, \tag{20}$$

with $\rho \in \{0.95, 0.9, 0.5, 0, -0.5, -0.9, -0.95\}$. The specific variants of DGP-1 will be denoted by DGP-1$i$, $i \in \{a, \dots, g\}$, depending on the specific value of $\rho$. The characteristic roots of the processes are presented in Table I.

Table I. Characteristic roots of the various variants of DGP-1

| DGP | Roots |
|---|---|
| 1a | (2.000, 1.053) |
| 1b | (2.000, 1.111) |
| 1c | (2.000, 2.000) |
| 1d | (2.000) |
| 1e | (2.000, −2.000) |
| 1f | (2.000, −1.111) |
| 1g | (2.000, −1.053) |

Some properties of DGP-1 are worth mentioning: First, all processes are stationary but some are persistent (DGP-1a, DGP-1g). Second, independent of $\rho$, the true response of the first variable to a shock in the second variable is zero at all propagation horizons, that is, $\boldsymbol{\Theta}_{12,H} = 0 \in \mathbb{R}^{H+1}$, and hence the asymptotic distribution of the estimator is degenerate normal as noted in Benkwitz *et al.* (2000). Third, for $\rho = 0$ (DGP-1d), the true response of the first variable to a shock in the first variable is also zero at all propagation horizons, that is, $\boldsymbol{\Theta}_{11,H} = 0$, and hence the estimator is also asymptotically degenerate normal.

Furthermore, we use $u_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma_u^1)$, where the population covariance matrix is given by

$$\Sigma_u^1 = \begin{pmatrix} 1.00 & 0.30 \\ 0.30 & 1.00 \end{pmatrix} .$$

In the Monte Carlo simulation, data samples of length $T \in \{100, 400\}$ are generated for each variant of DGP-1 and the maximum propagation horizon is $H \in \{10, 20\}$; Lütkepohl *et al.* (2015a,b) use the same choices of $T$ and $H$.

*5.3.2. Trivariate VAR(4) Model*

DGP-2 is a trivariate VAR(4) model previously considered in Staszewska-Bystrova (2011). More specifically, the population parameters of DGP-2 are the estimates of a model of the inflation rate, the unemployment rate, and the

  
DOI: 10.1111/jtsa.12289

federal fund rate using U.S. quarterly data from 1960-Q1 through 2004-Q1; for more details about the dataset, see Stock and Watson (2001) or Staszewska-Bystrova (2011). The DGP is given by

$$\textbf{DGP-2} \quad y_t = \nu + A_1 y_{t-1} + A_2 y_{t-2} + A_3 y_{t-3} + A_4 y_{t-4} + u_t \,, \tag{21}$$

where

$$A_1 := \begin{pmatrix} 0.549 & -0.965 & 0.164 \\ 0.029 & 1.480 & 0.003 \\ 0.084 & -1.567 & 0.962 \end{pmatrix}, \quad A_2 := \begin{pmatrix} 0.118 & 1.506 & -0.128 \\ -0.013 & -0.494 & 0.043 \\ 0.197 & 1.763 & -0.364 \end{pmatrix},$$

$$A_3 := \begin{pmatrix} 0.060 & -0.954 & 0.054 \\ 0.002 & -0.029 & -0.024 \\ -0.070 & -0.848 & 0.333 \end{pmatrix}, \quad A_4 := \begin{pmatrix} 0.261 & 0.250 & -0.098 \\ -0.012 & -0.014 & 0.008 \\ -0.046 & 0.563 & -0.010 \end{pmatrix},$$

and $\nu := (1.076, 0.125, 0.347)'$. Furthermore, we use $u_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma_u^2)$, where the population covariance matrix of DGP-2 is the covariance estimate based on the same data as the intercept and the slope coefficients and is given by

$$\Sigma_u^2 := \begin{pmatrix} 0.962 & -0.018 & 0.116 \\ -0.018 & 0.049 & -0.087 \\ 0.116 & -0.087 & 0.693 \end{pmatrix} .$$

In the Monte Carlo study, data samples of length $T \in \{100, 400\}$ are generated, and the maximum propagation horizon is $H \in \{4, 8, 12, 16, 20, 24, 28\}$. The choices of $H$ reflect the fact that DGP-2 is an empirical DGP based on quarterly data and, hence, the considered values of the maximum propagation horizon $H$ correspond to impulse responses over one up to seven years.

### 5.4. Simulation Parameters and Performance Evaluation

The nominal confidence level of the various joint confidence bands is 90%. The number of bootstrap replications is $B = 2000$ throughout, and the number of Monte Carlo replications is also 2000. The finite-sample performance of the various bands is evaluated via the empirical volume and the empirical coverage rate. More specifically, the empirical coverage rate is calculated in the usual way, that is,

$$EC := \frac{1}{2000} \sum_{m=1}^{2000} \mathbb{1}_{\{\boldsymbol{\Theta}_{ij,H} \in \mathrm{CB}_{m,ij}\}} \,,$$

where $\mathrm{CB}_{m,ij}$ denotes particular joint confidence bands for $\boldsymbol{\Theta}_{ij,H}$, and $\mathbb{1}_{\{A\}}$ denotes the indicator function of an event $A$. The empirical volume of particular joint confidence bands for $\boldsymbol{\Theta}_{ij,H}$ is computed as the average of the sum of the lengths of the corresponding marginal intervals, that is,

$$V := \frac{1}{2000} \sum_{m=1}^{2000} \sum_{h=0}^{H} \left( u_{m,h} - l_{m,h} \right) \,,$$

where $u_{m,h}$ denotes the upper bound of the $h$th marginal interval of the bands in the $m$th Monte Carlo repetition, and $l_{m,h}$ denotes the corresponding lower bound.

  
DOI: 10.1111/jtsa.12289

### 5.5. Results

#### *5.5.1. DGP-1: Bivariate VAR(1) Models*

The tables with the simulation results are relegated to the Supporting information for the sake of clarity. Boxplots summarizing the performance of the various methods across different scenarios are given in Appendices D and E. The focus is on the finite-sample performance of the BB bands because the performance of the competing methods have already been investigated individually for this specific DGP in Lütkepohl *et al.* (2015a,b). The main conclusions are as follows:

- For $T = 100$ and $H = 10$, the BB bands exhibit coverages rates close to or slightly below the nominal level of 90% except in scenarios where the asymptotic distribution of the estimator of the impulse response function is degenerate normal (that is, joint confidence bands for $\Theta_{1,1}$ of DGP-1d and $\Theta_{1,2}$ of all processes) and the scenario where joint confidence bands are constructed for $\Theta_{1,1}$ of DGP-1g. The bands of the former scenarios exhibit coverage rates above the nominal level, whereas the bands for the latter scenario exhibit substantial undercoverage. Furthermore, the BB bands are robust with respect to the propagation horizon. Overall, the coverage bias of the BB bands is substantially reduced for the large sample size of $T = 400$; see Figures D1 and D2.
- In general, the coverage bias of the BB bands is comparable to that of the Adjusted-Wald bands but smaller than the coverage bias of the Naïve bands, the Bonferroni bands, and the Wald bands.
- In all scenarios, the BB bands are smaller than the Bonferroni bands, where the excess volume (vis-à-vis the volume of the BB bands) of the Bonferroni bands ranges from 0.7% to 55%. Overall, the excess volume tends to increase with the sample size $T$ and the maximum propagation horizon $H$. However, there does not seem to be a clear pattern between the stationary characteristics of the processes and the excess volume of the Bonferroni bands.
- In 108 out of the 112 scenarios, the BB bands exhibit a smaller volume than the conservative Wald bands. In these scenarios, the volume of the Wald bands is substantially larger, and the excess volume (vis-à-vis the volume of the BB bands) ranges from 24.5% to 82.9%. Furthermore, the excess volume decreases with the maximum propagation horizon $H$ and the sample size $T$ in almost all scenarios.
- In 98 out of the 112 scenarios, the volume of the Adjusted-Wald bands is smaller than the volume of the BB bands. However, the difference is usually small. As expected, the BB bands are larger than the Naïve bands, which completely ignore the inherent simultaneity in the construction of the bands.

#### *5.5.2. DGP-2: Trivariate VAR(4) Model*

The complexity of DGP-2 is substantially larger than that of the bivariate VAR(1) models of DGP-1. The true impulse response functions of DGP-2 depend on 42 population reduced-from coefficients (compared to 7 population coefficients in DGP-1). Thus, the results for $T = 100$ give some indication about the performance of the joint confidence bands in scenarios where the ratio of the sample size to the number of coefficients is small (that is, low degrees of freedom).[9] The tables with the simulation results are relegated to the Supporting information. Boxplots summarizing the performance of the various methods across different scenarios are given in Appendices F and G. The main conclusions are as follows:

- For $T = 100$, there are systematic differences in the coverage rates of the BB bands, that is, the bands for $\Theta_{1,1}, \Theta_{1,3}, \Theta_{3,3}$ perform worse than the bands for the other impulse responses; see Figure N.1 in the Supporting information. More specifically, the bands for $\Theta_{1,1}, \Theta_{1,3}, \Theta_{3,3}$ exhibit substantial undercoverage, whereas the bands for the remaining impulse responses exhibit only mild under- and overcoverage, even for large $H$. Overall, the coverage distortion as well as the variation in coverage rates is substantially reduced for a sample size of $T = 200$, except the coverage rates of the bands for $\Theta_{1,1}$, which are still much below 90%. For $T = 400$,

[9] For $T = 100$, there are 300 individual data points to estimate the 42 coefficients.

  
DOI: 10.1111/jtsa.12289

the coverage rates of the BB bands are consistently close to the nominal coverage of 90%, even for large $H$; see Figure N.2 in the Supporting information.

- In principle, the BB bands are smaller than the two conservative bands (Bonferroni and Wald) and larger than the Naïve bands. Interestingly, the BB bands are even smaller than the Adjusted-Wald bands in 172 out of 189 scenarios. In general, the volume of the BB bands is strictly decreasing in the sample size, but strictly increasing in the propagation horizon.
- For $T = 100$, there are systematic differences in the coverage rates of the Bonferroni bands. The Bonferroni bands for $\Theta_{1,1}, \Theta_{1,3}, \Theta_{3,3}$ exhibit massive to substantial undercoverage in some scenarios, whereas the bands for the other impulse responses exhibit only mild under- and overcoverage. For a larger sample size of $T = 200$, the coverage rates are close to or above the nominal level of 90% except the coverage rates of the bands for $\Theta_{1,1}$, which are still much below 90%. For $T = 400$, all coverage rates are above the nominal level.
- The coverage rates of the Wald bands are markedly above the nominal level of 90% in 173 out of 189 scenarios. Overall, the positive coverage bias is enhanced with the sample size; see Figures N.1 and N.2 in the Supporting information. In all scenarios, the Wald bands exhibit the largest volume and are substantially larger than the Bonferroni bands.
- For $T = 100$, the Adjusted-Wald bands exhibit coverage rates that are substantially distorted and systematically differ among the different impulse response functions; see Figure N.1 in the Supporting information. Nevertheless, the coverage distortions and the systematic variation in coverage rates are mitigated as the sample size increases; see Figure N.2 in the Supporting information.
- The Naïve bands massively under-represent the joint estimation uncertainty in all scenarios. The empirical coverage rate falls below 40% in some scenarios with $T = 100$. Thus, these results provide additional evidence that the Naïve bands should not be used in practice, at least not when joint confidence bands are desired.

#### *5.5.3. Empirical Balance*

From a theoretical point of view, the BB bands, the Naïve bands, and the Bonferroni bands are asymptotically balanced, that is, the marginal coverage probability is (asymptotically) independent of the propagation horizon $h \in \{0, \dots, H\}$. The Wald-type bands of Lütkepohl *et al.* (2015b) have not been theoretically investigated in terms of asymptotic balance.

Table II presents the mean absolute deviations (MAD) from the mean of the marginal empirical coverage rates for DGP-2, $H = 20$, and $T \in \{100, 400\}$. For the small sample size ($T = 100$), the (unadjusted) Wald bands exhibit the smallest deviations from balance, followed closely by the Bonferroni bands. The third place is shared by the BB bands and the Adjusted-Wald bands, and the last place goes to the Naïve bands. Increasing the sample size to $T = 400$ reduces the MAD of all methods in all scenarios. The first two places are again awarded to the Wald and the Bonferroni bands respectively. However, the BB bands exhibit smaller MAD in six out of nine scenarios when compared to the Adjusted-Wald bands and in seven out of nine scenarios when compared to the Naïve bands respectively, resulting in the third place. The last place is shared by the Adjusted-Wald and the Naïve bands.

#### *5.5.4. Exogenous vs. Endogenous Bootstrap*

The tables with the simulation results for the Naïve bands, the Bonferroni bands, and the BB bands based on the endogenous bootstrap are given in the Supporting information. The main conclusions are as follows:

- The simulation results for the bivariate VAR(1) models show that endogenizing the lag uncertainty in general results in an upward shift of the coverage rates of the BB bands for both $T = 100$ and $T = 400$; see Figure N.3 in the Supporting information. The effect on the coverage rates of endogenizing the lag uncertainty is ambiguous, as there are scenarios where the coverage bias is reduced (for example, bands for $\Theta_{1,1}$ of DGP-1g), but also scenarios where the opposite is true (for example, bands for $\Theta_{1,1}$ of DGP-1e). However, the BB bands based on the endogenous bootstrap are larger than those based on the exogenous bootstrap.
- The results for the trivariate VAR(4) model show that the coverage rates of the BB bands based on the endogenous bootstrap are surprisingly inferior to those of the BB bands based on exogenous bootstrap for $T = 100$;

  
DOI: 10.1111/jtsa.12289

Table II. MAD of the empirical marginal coverage rates (from their mean) of nominal 90% confidence bands with $H = 20$, normal errors, and AIC lag selection

| DGP2 | Method | $\Theta_{1,1}$ | $\Theta_{1,2}$ | $\Theta_{1,3}$ | $\Theta_{2,1}$ | $\Theta_{2,2}$ | $\Theta_{2,3}$ | $\Theta_{3,1}$ | $\Theta_{3,2}$ | $\Theta_{3,3}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $T = 100$ | Naïve | 7.97 | 1.63 | 5.61 | 2.46 | 1.44 | 0.51 | 3.41 | 3.44 | 2.91 |
| | BB | 1.95 | 0.95 | 4.35 | 0.73 | 1.02 | 0.47 | 1.39 | 2.20 | 0.64 |
| | Bonferroni | 2.26 | 0.44 | 3.73 | 0.47 | 0.57 | 0.31 | 0.47 | 1.32 | 0.75 |
| | Wald | 1.32 | 0.35 | 2.58 | 0.28 | 0.43 | 0.30 | 0.35 | 0.72 | 0.53 |
| | Adj-Wald | 3.15 | 0.70 | 5.35 | 1.64 | 0.61 | 0.74 | 1.77 | 1.27 | 1.35 |
| $T = 400$ | Naïve | 0.37 | 0.38 | 0.54 | 0.70 | 0.58 | 0.65 | 0.95 | 0.69 | 0.72 |
| | BB | 0.51 | 0.43 | 0.32 | 0.13 | 0.27 | 0.22 | 0.66 | 0.52 | 0.27 |
| | Bonferroni | 0.20 | 0.20 | 0.13 | 0.19 | 0.29 | 0.15 | 0.41 | 0.14 | 0.30 |
| | Wald | 0.11 | 0.08 | 0.07 | 0.16 | 0.15 | 0.08 | 0.22 | 0.13 | 0.12 |
| | Adj-Wald | 0.82 | 0.36 | 1.04 | 0.21 | 0.25 | 0.42 | 1.28 | 0.32 | 0.80 |

see Figure N.4 in the Supporting information. For $T = 200, 400$, there are scenarios where the BB bands based on the endogenous bootstrap are superior, but in the majority of the scenarios, the bands based on the exogenous bootstrap are superior. For $T = 100$, endogenizing the lag uncertainty results in bands that are smaller than those based on the exogenous bootstrap. The same is true for the majority of the scenarios with $T = 200, 400$, although the differences are decreasing in the sample size.

- The results for the the bivariate VAR(1) models show that the coverage rates of the Bonferroni bands based on the endogenous bootstrap are larger than those of the Bonferroni bands based on the exogenous bootstrap in the majority of the scenarios, although the differences tend to decrease with the sample size. Hence, endogenizing the lag uncertainty increases the coverage bias of the Bonferroni bands, as the Bonferroni bands based on the exogenous bands exhibit coverage rates above the nominal level.
- The simulation results for the trivariate VAR(4) model show that the effect of endogenizing the lag uncertainty on the Bonferroni bands is ambiguous (for all sample sizes); there are scenarios where the bands based on the endogenous bootstrap are superior, but there are also scenarios where the opposite is true. The Bonferroni bands based on the endogenous bootstrap are smaller than the Bonferroni bands based on the exogenous bootstrap in almost all scenarios.

### 5.6. Summary of Simulation Evidence

We have conducted a Monte Carlo simulation to compare the finite-sample properties of the BB bands with a number of competing methods. We have included several bivariate VAR(1) models and an empirical trivariate VAR(4) model in the set of data generating processes.

The simulation results of the bivariate VAR(1) models show that the BB bands and the Adjusted-Wald bands exhibit both a smaller coverage bias and a smaller volume than the Bonferroni and the Wald bands. Both methods are robust with respect to the maximum propagation horizon and also produce reasonable joint confidence bands when the true impulse response is zero and hence the asymptotic distribution of the estimator is degenerate.

The simulation results of the trivariate VAR(4) model show that the coverage rates of the BB bands, the Bonferroni bands, and the Adjusted-Wald bands are potentially downward biased in scenarios with low degrees of freedom. In such scenarios, the Wald bands may be preferred to avoid the use of bands that underestimate the estimation uncertainty, but the price to pay is the large volume of the Wald bands, especially for large maximum propagation horizons. For $T = 400$, the BB bands exhibit the smallest coverage bias and at the same time the smallest volume among all joint confidence bands (except for the Naïve bands).

Overall, both the Adjusted-Wald bands and the BB bands work reliably in small models. However, the simulation results of the trivariate VAR(4) model show that the BB bands generally outperform the Adjusted-Wald bands in more complex models.




DOI: 10.1111/jtsa.12289

Endogenizing the lag uncertainty reduces the coverage bias of the BB bands only in particular scenarios. However, in the majority of the scenarios, the BB bands based on the exogenous bootstrap are superior in that regard. The effect on the volume is ambiguous and depends on the data generating process; the same holds for the Bonferroni bands. Based on these findings, we do not promote the endogenous bootstrap.

Furthermore, the results of the trivariate VAR(4) model confirm two of the main empirical findings in Lütkepohl *et al.* (2015b). First, the Wald bands tend to exhibit a larger volume than the Bonferroni bands. Second, the volume adjustment of the Wald bands can result in coverage rates markedly below the nominal level in scenarios with low degrees of freedom.

## 6. EMPIRICAL APPLICATION

We illustrate the BB joint confidence bands and the competing methods using the structural VAR model of Kilian (2009). The three-dimensional VAR model of Kilian includes the following three variables:

- $\Delta prod_t$: the percentage change in global crude oil production
- $real_t$: a business cycle index of global real activity (expressed in logs)
- $rpo_t$: the real price of oil (expressed in logs)

The monthly dataset from 1973–01 through 2007–12 was downloaded from the homepage of the *American Economic Review*.[10] We estimate the parameters of the reduced-form VAR(3) model (as suggested by the AIC) by the same methodology as in the Monte Carlo simulation. Following Kilian (2009), we use a recursive identification Scheme, where $B_0$ is the lower triangular Cholesky decomposition of the reduced-form residual covariance matrix, and the maximum propagation horizon is $H = 18$. The 90% joint confidence bands are constructed based on the bootstrap procedure of Kilian (1998b) with $B = 2000$ replications.

Table III. Volume of joint confidence bands for $H = 18$

| | BB | Bon | Wald | A-Wald | Naïve |
|---|---|---|---|---|---|
| $\Theta_{11}$ | 31.77 | 35.69 | 42.63 | 34.50 | 20.25 |
| $\Theta_{21}$ | 23.06 | 27.77 | 33.72 | 24.09 | 16.24 |
| $\Theta_{31}$ | 21.25 | 25.27 | 32.03 | 21.82 | 14.48 |
| $\Theta_{12}$ | 41.73 | 54.35 | 62.56 | 44.54 | 31.99 |
| $\Theta_{22}$ | 59.99 | 73.66 | 94.45 | 64.41 | 44.67 |
| $\Theta_{32}$ | 45.72 | 62.34 | 75.75 | 44.79 | 34.25 |
| $\Theta_{13}$ | 65.61 | 88.09 | 107.61 | 65.31 | 51.12 |
| $\Theta_{23}$ | 89.19 | 114.63 | 128.32 | 95.58 | 67.83 |
| $\Theta_{33}$ | 85.05 | 115.42 | 133.36 | 87.68 | 64.98 |

Table III presents the volumes of the 90% joint confidence bands of all impulse responses with a maximum propagation horizon of $H = 18$. The Wald and the Bonferroni bands exhibit the largest volumes in all scenarios. The BB bands have the smallest volume (of the proper joint confidence bands) in seven out of nine scenarios. These findings are completely in line with the simulation-based findings about the volumes of the various joint confidence bands.

Figure 1 displays the estimated structural response of $\Delta prod_t$ to a one-standard-deviation shock in $\varepsilon_{t,2}$ over a maximum propagation horizon of $H = 18$ (that is, $\widehat{\boldsymbol{\Theta}}_{12,18}$) and the corresponding nominal 90% joint confidence bands.[11] The figure illustrates the different shapes and volumes of the joint confidence bands. Inference based on the BB bands results in rejecting the null hypothesis of a zero response for propagation horizons $h = 4$–18 (as the marginal intervals do not cover zero). Inverting the other confidence bands results in different conclusions about the simultaneous test of the $H + 1$ hypotheses. More specifically, inference based on the Bonferroni bands results

[10] https://www.aeaweb.org/articles?id=10.1257/aer.99.3.1053
[11] The figures of the remaining eight impulse responses are relegated to the Supporting information for the sake of clarity.

  
DOI: 10.1111/jtsa.12289

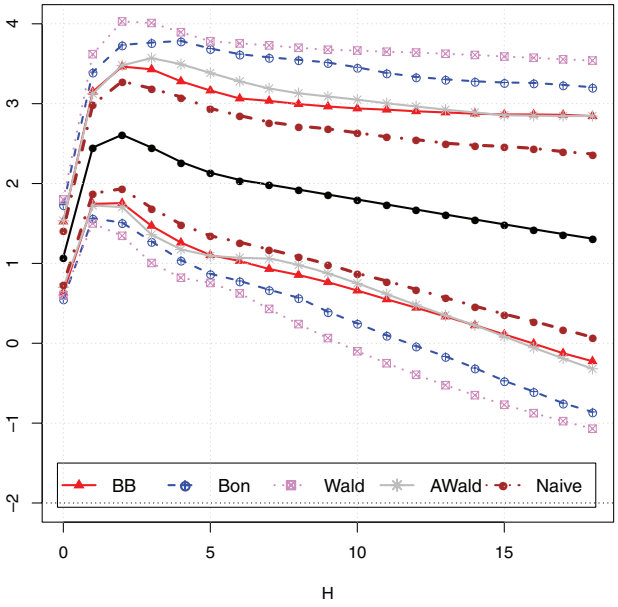


Figure 1. Estimated impulse response of $\Delta prod_t$ to a 1 standard deviation shock in $\epsilon_{t,2}$ over a maximum propagation horizon of $H = 18$ (solid line with circles) and the corresponding nominal 90% joint confidence bands [Color figure can be viewed at wileyonlinelibrary.com]

in rejecting the null hypothesis of a zero response for only two propagation horizons ($h = 3, 4$), inference based the Wald bands results in accepting the null hypothesis of a zero response at all propagation horizons $h \in \{0, \dots, 18\}$, and inference based on the Adjusted-Wald bands results in rejecting the null hypothesis of a zero response for propagation horizons $h = 3$–$18$.

## 7. CONCLUSION

Impulse response analysis based on low-dimensional structural vector autoregressions is still a popular tool in applied work. It is standard in applications to equip the estimated impulse response function with confidence bands that indicate the underlying estimation uncertainty. The literature has proposed several methods to construct joint confidence bands designed to cover the entire true impulse response function with a pre-specified probability. The existing methods suffer from deficiencies: they can exhibit empirical coverage rates substantially below the desired nominal level in certain scenarios, or they can be excessively large in terms of the aggregate volume.

We have proposed new joint confidence bands for impulse response functions of structural vector autoregressions based on multiple testing methodology of Romano and Wolf (2010). Under weak regularity conditions, these BB bands have asymptotically the desired coverage probability and are also asymptotically balanced.

We have compared the finite-sample properties of the BB bands with those of existing bands by means of a Monte Carlo simulation. The BB bands (i) have smaller volume than the two conservative bands—the traditional Bonferroni bands and the Wald bands of Lütkepohl *et al.* (2015b)—and (ii) have similar volume compared to the Adjusted-Wald bands of Lütkepohl *et al.* (2015b). In terms of coverage probability, the performance of the BB bands is overall the best.

Nevertheless, the BB bands—just like the Adjusted-Wald bands —can suffer from undercoverage for small sample sizes. It stands to reason that this problem can be fixed, or at least mitigated, by a double-bootstrap approach. However, studying the finite-sample properties of such an approach via Monte Carlo simulations does not seem feasible given the currently available computing power.

## ACKNOWLEDGMENT

We thank a co-editor and two anonymous reviewers for helpful comments which have improved the exposition of the paper.




DOI: 10.1111/jtsa.12289

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

## REFERENCES


Barsky RB, Sims ER. 2011. News shocks and business cycles. *Journal of Monetary Economics* **58**(3): 273–289.

Benkwitz A, Neumann MH, Lütekpohl H. 2000. Problems related to confidence intervals for impulse responses of autoregressive processes. *Econometric Reviews* **19**(1): 69–103.

Beran R. 1987. Prepivoting to reduce level error of confidence sets. *Biometrika* **74**(3): 457–468.

Beran R. 1988. Balanced simultaneous confidence sets. *Journal of the American Statistical Association* **83**(403): 679–686.

Bian TY, Gete P. 2015. What drives housing dynamics in China? A sign restrictions VAR approach. *Journal of Macroeconomics* **46**: 96–112.

Cao B, Sun Y. 2011. Asymptotic distributions of impulse response functions in short panel vector autoregressions. *Journal of Econometrics* **163**(2): 127– 143.

Engsted T, Pedersen TQ. 2014. Bias-correction in vector autoregressive models: a simulation study. *Econometrics* **2**(1): 45–71.

Hinkley D, Wei B-C. 1984. Improvements of jackknife confidence limit methods. *Biometrika* **71**(2): 331–339.

Horowitz JL. 2001. The bootstrap. In Handbook of Econometrics, Vol. 5, Chap. 52, 1st ed. Elsevier: Amsterdam; 3159–3228.

Jentsch C, Kreiss J-P. 2010. The multiple hybrid bootstrap: resampling multivariate linear processes. *Journal of Multivariate Analysis* **101**(10): 2320 – 2345.

Jentsch C, Politis DN. 2015. Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *Annals of Statistics* **43**(3): 1117–1140.

Jordà O. 2009. Simultaneous confidence regions for impulse responses. *Review of Economics and Statistics* **91**(3): 629–647.

Kilian L. 1998a. Accounting for lag order uncertainty in autoregressions: the endogenous lag order bootstrap algorithm. *Journal of Time Series Analysis* **19**(5): 531–548.

Kilian L. 1998b. Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics* **80**(2): 218–230.

Kilian L. 2001. Impulse response analysis in vector autoregressions with unknown lag order. *Journal of Forecasting* **20**(3): 161–179.

Kilian L. 2009. Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. *American Economic Review* **99**(3): 1053–1069.

Kurmann A, Otrok C. 2013. News shocks and the slope of the term structure of interest rates. *American Economic Review* **103**(6): 2612–2632.

Lütkepohl H. 1989. A note on the asymptotic distribution of impulse response functions of estimated var models with orthogonal residuals. *Journal of Econometrics* **42**(3): 371–376.

Lütkepohl H. 1990. Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. *Review of Economics and Statistics* **72**(1): 116–125.

Lütkepohl H. 2005. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

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**(3): 782–798.

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

Meyer M, Kreiss J-P. 2015. On the vector autoregressive sieve bootstrap. *Journal of Time Series Analysis* **36**(3): 377–397.

Pope AL. 1990. Biases of estimators in multivariate non-Gaussian autoregressions. *Journal of Time Series Analysis* **11**(3): 249–258.

Romano JP, Wolf M. 2010. Balanced control of generalized error rates. *Annals of Statistics* **38**(1): 598–633.

Rothenberg TJ. 1971. Identification in parametric models. *Econometrica* **39**(3): 577–591.

Rubio-Ramirez JF, Waggoner DF, Zha T. 2010. Structural vector autoregressions: theory of identification and algorithms for inference. *Review of Economic Studies* **77**(2): 665–696.

Schwert GW. 1989. Tests for unit roots: a Monte Carlo investigation. *Journal of Business & Economic Statistics* **7**(2): 147–159.

Staszewska A. 2007. Representing uncertainty about response paths: the use of heuristic optimisation methods. *Computational Statistics & Data Analysis* **52**(1): 121–132.

Staszewska-Bystrova A. 2011. Bootstrap prediction bands for forecast paths from vector autoregressive models. *Journal of Forecasting* **30**(8): 721–735.

Stine RA. 1987. Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association* **82**(400): 1072–1078.

Stock JH, Watson MW. 2001. Vector autoregressions. *Journal of Economic Perspectives* **15**(4): 101–115.


  
DOI: 10.1111/jtsa.12289

## APPENDIX A. BIAS CORRECTION

Let $A$ denote the matrix of the true slope coefficients of the VAR(1) representation of a general VAR($p$) process. Under some regularity conditions, Pope (1990) derives the following approximation for the finite-sample bias of the least squares (LS) estimator of $A$:

$$\text{Bias}(\hat{A}) = -\frac{b}{T} + \mathcal{O}(T^{-\frac{3}{2}}) ,$$

where

$$b := \Sigma_U \left[ \left(I_{kp} - A'\right)^{-1} + A' \left(I_{kp} - \left(A'\right)^2\right)^{-1} + \sum_{i=1}^{k} \lambda_i \left(I_{kp} - \lambda_i A'\right)^{-1} \right] \Sigma_Y^{-1} .$$

Here, $I_{kp}$ denotes the $kp \times kp$ identity matrix, $\lambda_i$ denotes the $i$th eigenvalue of $A$, $\Sigma_Y$ denotes the covariance matrix of $Y_t := \left(y_t', y_{t-1}', \ldots, y_{t-p+1}'\right)'$, and $\Sigma_U$ denotes the covariance matrix of $U_t := \left(u_t, 0, \ldots, 0\right)'$. Neglecting higher-order terms and replacing true parameters by their LS estimators yields the following estimator for the finite-sample bias of $\hat{A}$:

$$\widehat{\text{Bias}}(\hat{A}) := -\frac{1}{T}\hat{\Sigma}_U \left[ \left(I_{kp} - \hat{A}'\right)^{-1} + \hat{A}' \left(I_{kp} - \left(\hat{A}'\right)^2\right)^{-1} + \sum_{i=1}^{k} \hat{\lambda}_i \left(I_{kp} - \hat{\lambda}_i \hat{A}'\right)^{-1} \right] \hat{\Sigma}_Y^{-1} .$$

Thus, the bias-corrected LS estimator is given by

$$\hat{A}^{\text{BC}} := \hat{A} - \widehat{\text{Bias}}(\hat{A}) .$$

## APPENDIX B. STATIONARITY CORRECTION

In order to prevent that stationary parameter estimates may be pushed outside the stationary region by the bias correction, Kilian (1998b) proposes the following adjustment procedure:

(a) Calculate the modulus of the largest root of the (uncorrected) LS estimate $\hat{A}$ and denote this quantity by $r(\hat{A})$. If $r(\hat{A}) \geq 1$, set $\hat{\tilde{A}} := \hat{A}$. If $r(\hat{A}) < 1$, construct the bias-corrected estimator $\hat{A}^{\text{BC}} := \hat{A} - \widehat{\text{Bias}}(\hat{A})$.
(b) If $r(\hat{A}^{\text{BC}}) \geq 1$, obtain $\tilde{A}_i := \hat{A} - \delta_i \widehat{\text{Bias}}(\hat{A})$ with $\delta_1 = 1$ and $\delta_i = \delta_{i-1} - 0.01$.
(c) Repeat step (b) until $r(\tilde{A}_i) < 1$.
(d) Set $\hat{\tilde{A}} := \tilde{A}_i$.

## APPENDIX C. CONSTRUCTION OF BB BANDS

The following algorithm provides a step-by-step instruction for the construction of the BB bands with nominal confidence level $(1 - \alpha)$ for an arbitrary identification procedure.

(i) Fit a VAR($\hat{p}$) model to the the observed time series $\{y_1, \ldots, y_T\}$, where $\hat{p}$ denotes the estimated lag order.
(ii) Compute the impact matrix $\hat{B}_0$ according to the chosen approach of identification.
(iii) Estimate the structural impulse response function of interest with maximum propagation horizon $H$, that is,

$$\widehat{\boldsymbol{\Theta}}_{ij,H} := \boldsymbol{\Theta}_{ij,H}\left(\hat{A}_1, \ldots, \hat{A}_{\hat{p}}, \hat{B}_0\right) ,$$

where the $\hat{A}_1, \ldots, \hat{A}_{\hat{p}}$ are the estimated reduced-form coefficients.

(iv) Generate a bootstrap sample $\left\{ \sqrt{T} \left| \widehat{\boldsymbol{\Theta}}^*_{ij,H,b} - \widehat{\boldsymbol{\Theta}}_{ij,H} \right| \right\}_{b=1}^{B}$. The number $B$ of bootstrap replications should be at least 2000, if feasible.




DOI: 10.1111/jtsa.12289

(v) Compute the following $(H+1)$ empirical distribution functions:

$$\widehat{H}_h^*(t) := \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\left\{\sqrt{T}\left|\widehat{\Theta}^*_{ij,h,b}-\widehat{\Theta}_{ij,h}\right| \le t\right\}} \quad \text{for } h = 0, \dots, H\,.$$

Statistical software packages usually provide a built-in function for computing the empirical distribution function; for example, the function `ecdf` in the software package R.

(vi) Compute the following bootstrap sample $\left\{\max_{h\in\tilde{S}}\left\{\widehat{H}_h^*\left(\sqrt{T}\left|\widehat{\Theta}^*_{ij,h,b}-\widehat{\Theta}_{ij,h}\right|\right)\right\}\right\}_{b=1}^{B}$ and the corresponding $(1-\alpha)$ quantile. Statistical software packages usually provide a built-in function for computing empirical quantiles; for example, the function `quantile` in the software package R.

(vii) Construct the BB confidence bands for $\boldsymbol{\Theta}_{ij,H}$ with nominal confidence level $(1-\alpha)$ by computing $(H+1)$ marginal intervals

$$\left[\widehat{\Theta}_{ij,h} \pm \frac{1}{\sqrt{T}}\widehat{H}_h^{*,-1}\left(\widehat{L}^{*,-1}(1-\alpha)\right)\right] \quad \text{for } h = 0, \dots, H\,,$$

where $\widehat{H}_h^{*,-1}(q) := \inf\left\{t : \widehat{H}_h^*(t) \ge q\right\}$, and $\widehat{L}^{*,-1}(1-\alpha)$ denotes the $(1-\alpha)$ quantile from step 6.

**Remark 7.** The methodology of Romano and Wolf (2010) allows one to construct joint confidence bands that do not cover all but 'only' at least $H-k+2$, $k \ge 2$, of the elements of $\boldsymbol{\Theta}_{ij,H}$ with nominal confidence level $1-\alpha$. Step 6 in the previous algorithm has to be modified to construct such bands. More specifically, one has to compute the empirical $(1-\alpha)$ quantile of

$$\left\{k\text{-}\max_{h\in\tilde{S}}\left\{\widehat{H}_h\left(\sqrt{T}\left|\widehat{\Theta}^*_{ij,h,b}-\widehat{\Theta}_{ij,h}\right|\right)\right\}\right\}_{b=1}^{B},$$

where $k$-max is the function that returns the $k$th largest element of a vector.

## APPENDIX D. DGP-1: EMPIRICAL COVERAGES

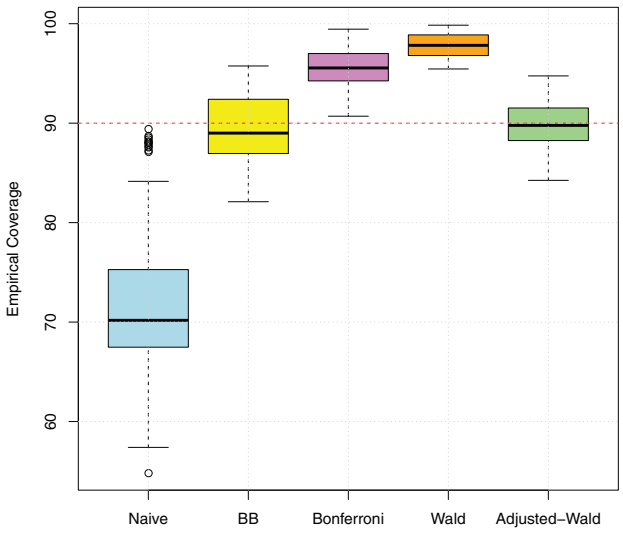


Figure D1. DGP-1: Boxplots of the empirical coverages across all variants of DGP-1, all impulse responses and all maximum propagation horizons (56 parameter constellations in total) of nominal 90% joint confidence bands for $T = 100$ [Color figure can be viewed at wileyonlinelibrary.com]

  
DOI: 10.1111/jtsa.12289

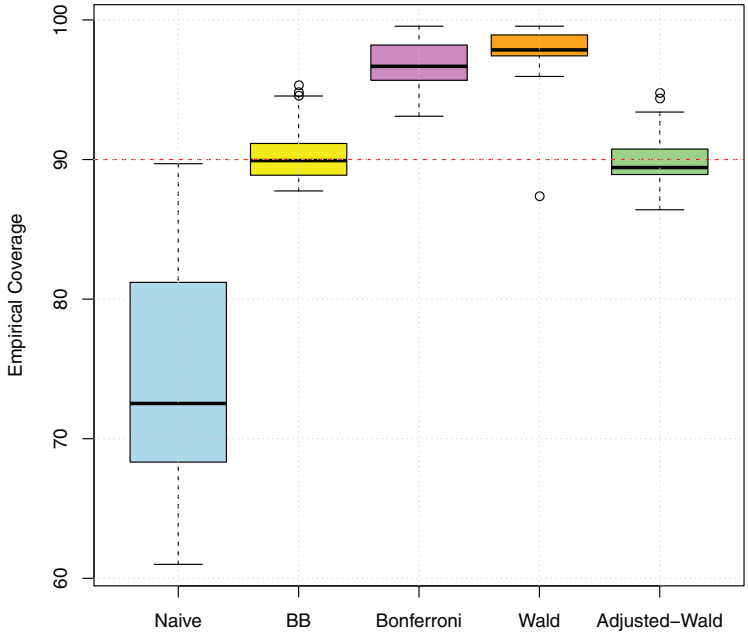


Figure D2. DGP-1: Boxplots of the empirical coverages across all variants of DGP-1, all impulse responses and all maximum propagation horizons (56 parameter constellations in total) of nominal 90% joint confidence bands for $T = 400$ [Color figure can be viewed at wileyonlinelibrary.com]

## APPENDIX E. DGP-1: VOLUMES

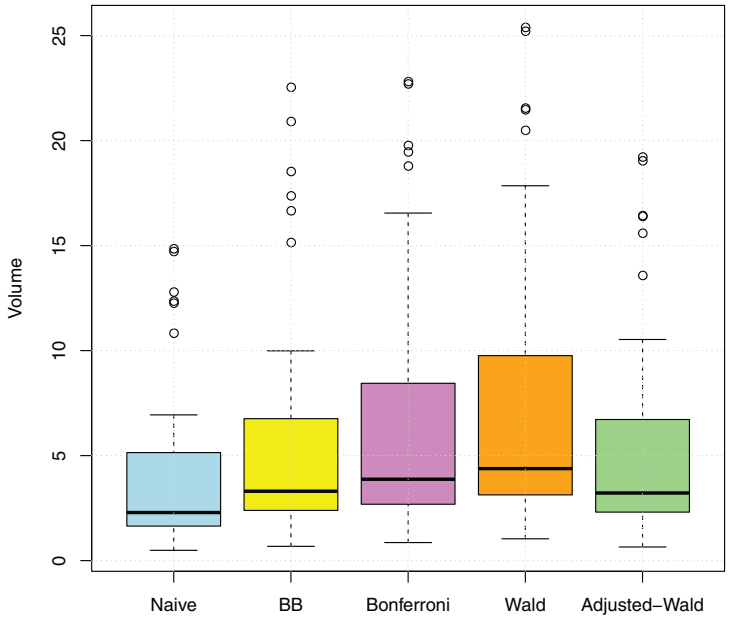


Figure E1. DGP-1: Boxplots of the volumes across all variants of DGP-1, all impulse responses and all maximum propagation horizons (56 parameter constellations in total) of nominal 90% joint confidence bands for $T = 100$ [Color figure can be viewed at wileyonlinelibrary.com]




DOI: 10.1111/jtsa.12289

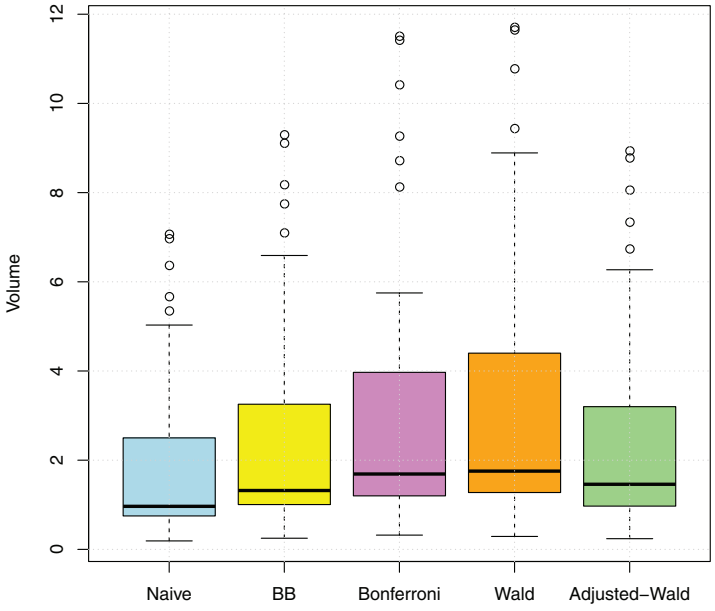


Figure E2. DGP-1: Boxplots of the volumes across all variants of DGP-1, all impulse responses and all maximum propagation horizons (56 parameter constellations in total) of nominal 90% joint confidence bands for $T = 400$ [Color figure can be viewed at wileyonlinelibrary.com]

## APPENDIX F. DGP-2: EMPIRICAL COVERAGES

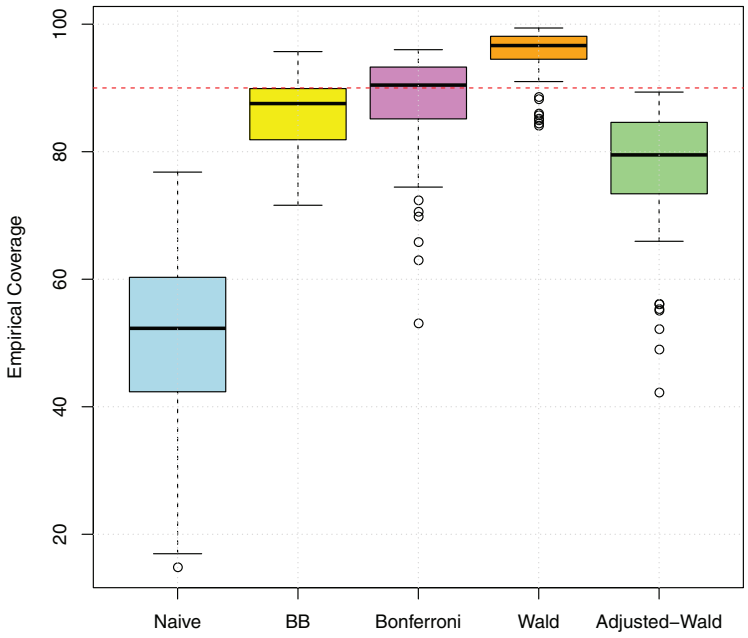


Figure F1. DGP-2: Boxplots of the empirical coverages across all impulse responses and all maximum propagation horizons (63 parameter constellations in total) of nominal 90% joint confidence bands for $T = 100$ [Color figure can be viewed at wileyonlinelibrary.com]




DOI: 10.1111/jtsa.12289

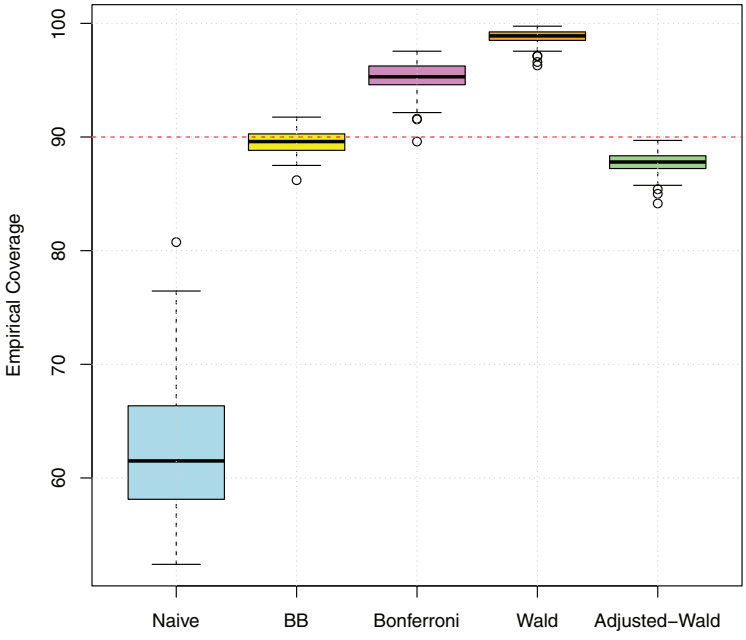


Figure F2. DGP-2: Boxplots of the empirical coverages across all impulse responses and all maximum propagation horizons (63 parameter constellations in total) of nominal 90% joint confidence bands for $T = 400$ [Color figure can be viewed at wileyonlinelibrary.com]

## APPENDIX G. DGP-2: VOLUMES

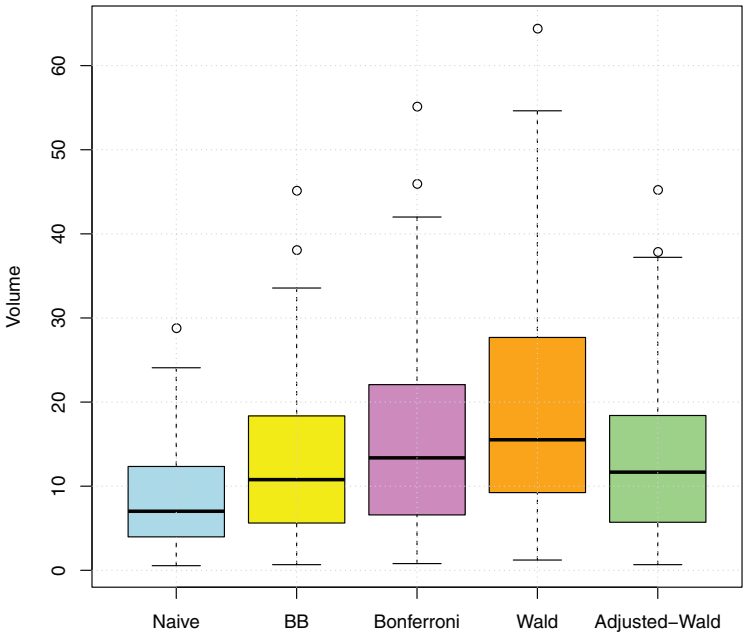


Figure G1. DGP-2: Boxplots of the volumes across all impulse responses and all maximum propagation horizons (63 parameter constellations in total) of nominal 90% joint confidence bands for $T = 100$ [Color figure can be viewed at wileyonlinelibrary.com]




DOI: 10.1111/jtsa.12289

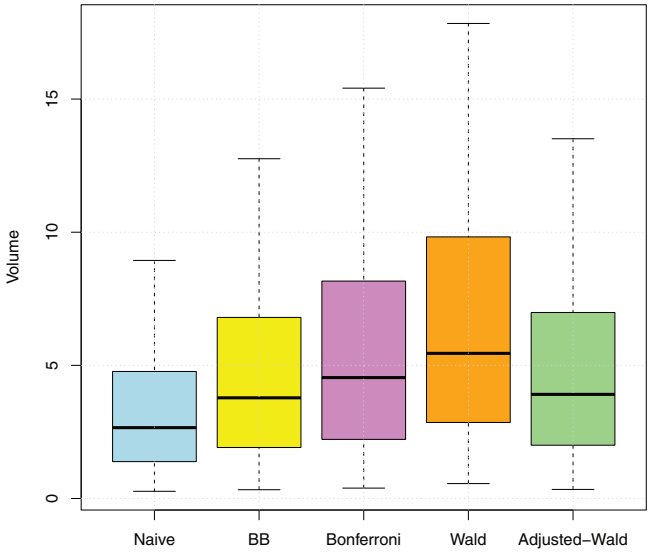


Figure G2. DGP-2: Boxplots of the volumes across all impulse responses and all maximum propagation horizons (63 parameter constellations in total) of nominal 90% joint confidence bands for $T = 400$ [Color figure can be viewed at wileyonlinelibrary.com]




DOI: 10.1111/jtsa.12289