View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

Lodz
Economics
Working
Papers

Skewness-Adjusted Bootstrap Confidence Intervals and Confidence Bands for Impulse Response Functions

Daniel Grabowski
Anna Staszewska-Bystrova
Peter Winker

1/2018

Faculty of Economics and Sociology
University of Lodz

Skewness-Adjusted Bootstrap Confidence Intervals and Confidence Bands for Impulse Response Functions☆

Daniel Grabowski[a,*], Anna Staszewska-Bystrova[b], Peter Winker[a]

[a] *University of Giessen, Department of Economics, Licher Str. 64, 35394 Giessen, Germany*
[b] *University of Lodz, 41 Rewolucji 1905r. Str., 90-214 Lodz, Poland*

Abstract

This article investigates the construction of skewness-adjusted confidence intervals and joint confidence bands for impulse response functions from vector autoregressive models. Three different implementations of the skewness adjustment are investigated. The methods are based on a bootstrap algorithm that adjusts mean and skewness of the bootstrap distribution of the autoregressive coefficients before the impulse response functions are computed. Using extensive Monte Carlo simulations, the methods are shown to improve the coverage accuracy in small and medium sized samples and for unit root processes for both known and unknown lag orders.

Keywords: Bootstrap, confidence intervals, joint confidence bands, vector autoregression
JEL Codes: C15, C32

☆Support from the National Science Center, Poland (NCN) through MAESTRO 4: DEC-2013/08/A/HS4/00612 is gratefully acknowledged.

*Corresponding author

Email addresses: Daniel.Grabowski@wirtschaft.uni-giessen.de (Daniel Grabowski), emfans@uni.lodz.pl (Anna Staszewska-Bystrova), Peter.Winker@wirtschaft.uni-giessen.de (Peter Winker)

1. Introduction

Time series with small sample sizes are common in econometrics and regularly analysed with vector autoregressive (VAR) models. In such cases, confidence intervals for impulse response functions (IRFs) for VAR models have to be constructed using bootstrapping. Confidence intervals for IRFs based on asymptotic approximation such as that of Lütkepohl (1990) are known to fall short of the nominal coverage level in small samples, see e.g. Kilian (1998a). Kilian proposed a bootstrap algorithm to construct bias corrected small sample confidence intervals. An extension allowing for an unknown lag order of the VAR process is given in Kilian (1998b). While the bootstrap intervals achieved higher coverage frequencies than the asymptotic intervals, they still deviated from the nominal coverage rate. The deviations in actual coverage from nominal coverage levels are particularly pronounced for small sample sizes. When the true lag order is no longer assumed to be known, the coverage accuracy declines further. For processes with unit roots or cointegrated processes the performance of asymptotic as well as bootstrap confidence intervals was also found to deteriorate in Kilian (1998a,b). This confirmed the result of Basawa et al. (1991), who show that the standard bootstrap algorithm is not valid for cointegrated VAR models in levels. However, unit roots may not always be detected and estimation of a vector error correction model (VECM) requires knowledge of the cointegration rank. Thus, misspecification can easily occur, as discussed in Kilian (1998a, p. 225) and Berkowitz and Kilian (2000, p. 30). Therefore, using a bootstrap method that maintains a high coverage accuracy even when the process is non-stationary appears desirable.

Two major issues in obtaining an adequate bootstrap distribution are the bias and the skewness of the least squares (LS) estimator in autoregressive models.[1] To the best of the authors' knowledge, there is no method for adjusting the skewness that is practical for IRF analyses. The percentile interval as given in Hall (1992), also called basic interval, can in principle deal with skewness, but does not perform well for IRF coefficients, as shown in Kilian (1999) and also reported in this article. We propose an alternative but related route for constructing skewness-adjusted confidence intervals. The approach is based on the idea of mirroring the bootstrap distribution of the least squares estimator of the autoregressive coefficients before computing the non-linear transformation that yields the impulse response functions. Thereby, the skewness of the bootstrap distribution is reversed. Three different implementations of this concept are investigated and compared to the standard approaches of Kilian (1998a,b) and of Hall (1992). The mirroring procedures are shown to improve coverage rates for unit-root processes and for stationary processes in small and medium sized samples.

[1] Berkowitz and Kilian (2000, p. 30) conjecture that the poor performance of the bootstrap confidence intervals in the presence of unit roots and roots close to unity is due to small sample bias. The evidence presented in this article suggests that skewness may also play an important role.

We also show that the proposed adjustments improve the coverage accuracy of joint confidence bands. Confidence bands are supposed to contain the true IRF throughout all considered periods with a given probability. Compared to confidence intervals, the actual coverage rate of joint confidence bands is more dependent on the tails of the bootstrap distribution of the VAR coefficients. Thus, adjusting for the skewness may be especially useful in their construction. Methods for constructing joint bands have been proposed by, among others, Sims and Zha (1999), Staszewska (2007), Jordà (2009), Jordà and Marcellino (2010), Staszewska-Bystrova (2011), Staszewska-Bystrova and Winker (2013), Inoue and Kilian (2013) and Wolf and Wunderli (2015). The performance of different methods is compared in Lütkepohl et al. (2015a,b). Therein it was shown that for samples of size 100 and larger and non-persistent processes, confidence bands constructed using bootstrapping and a method derived from the Bonferroni principle achieved good coverage rates. However, the coverage still falls short of the nominal level. We provide evidence that adjusting the skewness improves the coverage accuracy of joint confidence bands.

Section 2 discusses the problem with skewness and proposes ways to adjust the bootstrap distribution. Section 3 presents the bootstrap algorithm in detail and describes the way confidence intervals and bands are obtained. We conjecture that the proposed methods lead to improved coverage rates of confidence intervals and bands by adjusting the skewness in addition to correcting the bias. To confirm this, evidence from Monte Carlo simulations is provided in Section 4. We do not derive analytical results regarding the effect on impulse response functions or the asymptotic validity of the method. Section 5 concludes.

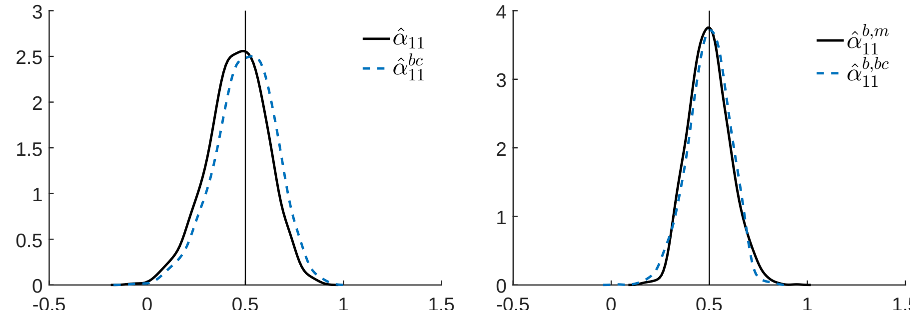
2. Skewness and Mirroring

In this section, we introduce and motivate adjustments to the bootstrap distribution used to compute confidence intervals and bands for impulse response functions. For explanatory purposes, consider the simple two-dimensional VAR(1) process

$$\begin{aligned} y_{1,t} &= \alpha_{11} y_{1,t-1} + \alpha_{12} y_{2,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= \alpha_{21} y_{1,t-1} + \alpha_{22} y_{2,t-1} + \varepsilon_{2,t}, \ t = 2, ..., T \end{aligned} \tag{1}$$

for the time series $y_{1,t}$ and $y_{2,t}$, where $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ denote error terms. For a sample of size T, the parameters α_{ij} can be estimated with least squares. Let $\hat{\alpha}_{ij}$ denote the LS estimators and $F_{\hat{\alpha}_{ij}}$ their unknown distribution. These estimators are biased in finite samples and the distribution $F_{\hat{\alpha}_{ij}}$ is skewed, as visualized in Figure 1a. Based on a Monte Carlo simulation, the figure shows an approximation of the distribution $F_{\hat{\alpha}_{11}}$. The data generating process (DGP) used in the simulation is the one given in (1) with α_{11}=0.5, α_{12}=0, α_{21}=0.5 and α_{22}=0.5, i.e.

$$\begin{aligned} y_{1,t} &= 0.5 y_{1,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= 0.5 y_{1,t-1} + 0.5 y_{2,t-1} + \varepsilon_{2,t}. \end{aligned} \tag{2}$$



a) Distribution of $\hat{\alpha}_{11}$ and $\hat{\alpha}_{11}^{bc}$. **b)** Bootstrap distribution for $\hat{\alpha}_{11}$.

Figure 1: a) Distribution of $\hat{\alpha}_{11}$ and $\hat{\alpha}_{11}^{bc}$ based on 2000 Monte Carlo simulations of a DGP given by equation (2). b) Distribution of $\hat{\alpha}_{11}^{b,bc}$ and $\hat{\alpha}_{11}^{b,m}$ based on 2000 bootstrap simulations using the mean values of $\hat{\alpha}_{ij}^{bc}$ and $\hat{\alpha}_{ij}$, respectively, from Figure 1a.

The error terms $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ follow a multivariate normal distribution with means of zero, variances of one, and a covariance of 0.3. The sample size is set to T=100 observations. For an unbiased estimator, the mean of $\hat{\alpha}_{11}$ should be close to 0.5. However, in this example, the mean LS estimate over 2000 simulated DGPs turns out to be 0.4589. Furthermore, the distribution is not symmetric. The skewness is -0.2344, i.e., $F_{\hat{\alpha}_{11}}$ is skewed left. When generating a bootstrap distribution, it is common to correct for the bias, but not for the skewness. Figure 1a also shows the distribution of the bias corrected estimator $\hat{\alpha}_{11}^{bc}$, obtained using the method of Pope (1990). Its mean is 0.4949, so the bias correction was successful. Unsurprisingly, the skewness remains at -0.2830.

In applications, $F_{\hat{\alpha}_{11}}$ is generally unknown and for constructing confidence intervals a bootstrap distribution that resembles—or is hoped to resemble—the true distribution $F_{\hat{\alpha}_{11}}$ is used. A bootstrap distribution of the bias corrected bootstrap coefficients $\hat{\alpha}_{ij}^{b,bc}$ is shown in Figure 1b for i, j=1. It approximately retains the properties of the distribution of $\hat{\alpha}_{11}^{bc}$. With a mean of 0.4975, the bias is successfully removed, while the skewness is still -0.2699. Thus, the bootstrap approximation to the true distribution was successful. Nonetheless, using this distribution to construct confidence intervals without taking into account the skewness can be inappropriate.

The problem resulting from the skewness is recognized in the literature (see, e.g., Hall (1992)), but is briefly restated here. A left-skewed distribution $F_{\hat{\alpha}_{ij}}$ implies a relatively higher probability of obtaining an estimate far to the left of the true parameter value of α_{ij}. When conducting inference, we start from the estimate $\hat{\alpha}_{ij}$ and want to infer from this the position of the true α_{ij}. Thus, if $F_{\hat{\alpha}_{ij}}$ has probability mass further to the left, we want confidence intervals extending further right to cover the true parameter with sufficient frequency. The standard approach for constructing percentile confidence intervals, used for example in Efron (1981) and Efron and Tibshirani (1986), does not accomplish this. Such a $1-a$ confidence interval for a parameter θ is obtained as

$$CI = [\hat{\theta}^b_{a/2}; \hat{\theta}^b_{1-a/2}], \tag{3}$$

where $\hat{\theta}^b_{a/2}$ and $\hat{\theta}^b_{1-a/2}$ are the $\frac{a}{2}$ and $1-\frac{a}{2}$ quantiles of the bootstrap distribution of $\hat{\theta}$. Hall (1992, p. 95) calls this 'looking up the wrong tail [...] of a distribution.' Hall suggests to compute percentile intervals instead as

$$CI^* = [2\hat{\theta} - \hat{\theta}^b_{1-a/2}; 2\hat{\theta} - \hat{\theta}^b_{a/2}]. \tag{4}$$

However, such intervals turned out to not work well if the statistic of interest, θ, is an impulse response coefficient, as shown in Kilian (1999).

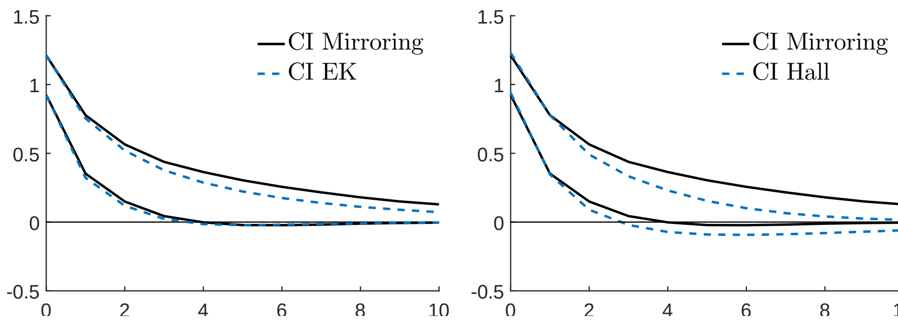
2.1. Mirroring the bootstrap distribution

As an alternative to (3) and (4), we propose to mirror the bootstrap distribution of the autoregressive coefficients before carrying out the non-linear transformation to the impulse response coefficients. The mirroring reverses the skewness of the bootstrap distribution and thus hopefully improves the coverage accuracy of confidence intervals. Similarly to the bias correction in Kilian (1998a), the skewness adjustment is done before computing the IRFs. Correcting for the skewness that is present in the distribution of the autoregressive coefficients after applying the non-linear IRF-transformation would be infeasible.

The mirroring algorithm is simple. Consider a model such as the VAR(1) in equation (1). First, we can obtain least squares estimates $\hat{\alpha}_{ij}$. Next, we generate B bootstrap time series $y^b_{i,t}$. This can be done by resampling residuals—details are given in the next section. Based on $y^b_{i,t}$, we obtain bootstrap estimates $\hat{\alpha}^b_{ij}$. The mirrored bootstrap estimates are then given by

$$\hat{\alpha}^{b,m}_{ij} = \hat{\alpha}_{ij} - (\hat{\alpha}^b_{ij} - \hat{\alpha}_{ij}) \; \forall \; i, j. \tag{5}$$

Note that, by flipping the bootstrap around the initial estimates, we automatically subtract the mean bias of the bootstrap distribution relative to the estimated coefficient two times and thus adjust for the bias that has occurred both in the initial estimation and in the estimation during the bootstrapping. Based on the mirrored bootstrap estimates, we can proceed to compute the impulse responses and subsequently the confidence intervals based on Efron's percentile interval given in (3). Details are given in Section 3. Using this approach, we obtain a distribution of the mirrored bootstrap estimator $\hat{\alpha}^{b,m}_{ij}$ that is plotted in Figure 1b for i, j=1. This achieves a bias correction as good as using the algorithm of Pope (1990). The mean of $\alpha^{b,m}_{11}$ is 0.4989. In addition, the sign of the skewness is now almost exactly reversed (from -0.2344 in the true distribution to 0.2391 in the mirrored bootstrap distribution), so we obtain a right skewed distribution as desired.



a) Efron-Kilian vs. Mirroring CI. **b)** Hall vs Mirroring CI.

Figure 2: a) Example of bias corrected percentile confidence intervals as in Kilian (1998a) compared to mirroring confidence intervals. b) Example of Hall percentile confidence intervals compared to mirroring confidence intervals.

If we are interested in confidence intervals for the AR coefficients, Hall's method and the mirroring method result in the same confidence intervals. The difference between the two comes into play when the quantity of interest is an impulse response parameter and not an autoregressive parameter. Because the former is a non-linear transformation of the latter, correcting for the bias and skewness that is present in the distribution of the AR coefficients after the IRF transformation is not straightforward. Hence, we intervene at an earlier stage and mirror the VAR coefficients instead of the IRF coefficients.

Figure 2a shows exemplarily confidence intervals for IRFs obtained using the ideas of Efron (1981) and Kilian (1998a) (called CI EK). The intervals are based on the bootstrap distributions shown in Figure 1b. This is compared to confidence intervals from the mirroring method (labelled M-method). The plot shows that, due to the right-skewness of the mirrored distribution, the mirroring confidence intervals collapse less quickly to zero and thus are wider in the later periods. Figure 2b compares confidence intervals obtained using the method based on Hall (1992) (called CI Hall) to the mirroring intervals. Hall's method also involves a sort of mirroring, but of the estimated IRF. These intervals appear to not correctly adjust for the skewness in the bootstrap distribution of the VAR coefficients. Hall's method moves the intervals further downwards, instead of extending them upwards, where we would rather expect the true IRF if the distribution of the estimators is skewed left.

2.2. Refinements to the mirroring

Mirroring the bootstrap coefficients around the LS estimates as described above is a relatively crude method, as it can completely change the dynamics of the VAR model. Because each coefficient is mirrored individually, the percentile ranks of bootstrap coefficients of the same model change in a non-consistent way. For example, consider some particular bootstrap values of $\hat{\alpha}_{11}^b$ and $\hat{\alpha}_{12}^b$, coming from the bth iteration. Let these values constitute the 10th and the 70th empirical percentile of their respective bootstrap distributions. Then, the values after applying the mirroring, i.e. $\hat{\alpha}_{11}^{b,m}$ and $\hat{\alpha}_{12}^{b,m}$, constitute the 90th and the 30th percentile of their respective distributions. Therefore, it seems that the dynamics of the VAR system resulting from the mirroring are possibly quite disconnected from the originally estimated structure. The least squares estimate represents a reasonable estimate because it minimizes a loss function. The mirrored estimates do nothing of the sort and have no particular justification in and on themselves, other than leading to a bootstrap distribution that has desirable properties in terms of mean and skewness.

For multidimensional data no clear ordering exists that would allow to jointly mirror the data. Therefore, we are left with adjusting individual coefficients instead of the entire system. However, there is a less disruptive way to adjust the distribution than simply mirroring the coefficients. We can mirror the distribution without mirroring the coefficients by adjusting the distance between the estimates and the percentiles of the bootstrap distribution. This percentile mirroring, labelled MP, is given by

$$\hat{\alpha}_{ij}^{b,mp} = \hat{\alpha}_{ij} - (\hat{\alpha}_{ij}^{b} - \hat{\alpha}_{ij})_{100-r(\hat{\alpha}_{ij}^{b})} \; \forall \, i, j, \tag{6}$$

where $r(\hat{\alpha}_{ij}^{b})$ is the percentile rank of $\hat{\alpha}_{ij}^{b}$ in its distribution and $(\hat{\alpha}_{ij}^{b} - \hat{\alpha}_{ij})_{100-r(\hat{\alpha}_{ij}^{b})}$ is the $(100 - r(\hat{\alpha}_{ij}^{b}))^{th}$ empirical percentile of the distribution of $(\hat{\alpha}_{ij}^{b} - \hat{\alpha}_{ij})$. To see the difference more clearly, the mirroring given in (5) can be restated as

$$\hat{\alpha}_{ij}^{b,m} = \hat{\alpha}_{ij} - (\hat{\alpha}_{ij}^{b} - \hat{\alpha}_{ij})_{r(\hat{\alpha}_{ij}^{b})} \; \forall \, i, j. \tag{7}$$

The mean and the skewness resulting from the two mirroring methods are identical. In fact, the distribution of $\hat{\alpha}_{ij}^{b,mp}$ is the same as that of $\hat{\alpha}_{ij}^{b,m}$, but the positions of individual coefficients are reversed. With MP, the bootstrap distribution is stretched in one tail and shrunk in the other instead of the coefficients being swapped within the distribution. When using MP, the coefficients will each retain their rank order in their respective distributions. This difference matters because the different VAR coefficients jointly determine the dynamics of the VAR system.

Even though the MP method also corrects for a bias, it only does so after bootstrapping. Thus, the bootstrap series are generated from a VAR model that has not been bias-corrected. To address this potential shortcoming, we investigate combining the skewness adjustment with a bias correction as a third option. As in Kilian (1998b) we use the bias correction from Pope (1990) to adjust the estimates of the VAR coefficients as well as the bootstrap coefficients. After this, the MP method from (6) is applied to the bias corrected coefficients. This combined approach is labeled the MPbc method.

Unknown lag orders represent a challenge for MP and MPbc. For their computation we need percentiles for each bootstrap coefficient. It appears reasonable to compute the percentiles using only bootstrap coefficients that come from a VAR with the same estimated lag order. However, some lag orders might be estimated only very few times during bootstrapping and thus there are very few matching bootstrap coefficients available. One could generate additional bootstrap draws until a satisfactory number of models is available for each lag length. We do not consider this approach here due to the high computational cost. Instead, we compute the percentiles based on however many bootstrap coefficients with the same lag order are available.

3. Bootstrap confidence intervals and bands

3.1. The model

Consider a VAR(p) model, p denoting the lag order, for t=p+1, ..., T, given by

$$y_t = A_0 + A_1 y_{t-1} + ... + A_p y_{t-p} + \varepsilon_t, \tag{8}$$

where $y_t = (y_{1t}, ..., y_{Kt})'$ is the vector of time series, A_0 is a $K \times 1$ vector of constants, A_1 to A_p are $K \times K$ parameter matrices and ε_t is a $K \times 1$ error term. The errors are assumed to be uncorrelated over time, having zero mean and covariance matrix Σ_ε. To estimate the VAR model, we first determine the lag order p. As discussed in Kilian (1998b) and Berkowitz and Kilian (2000), different information criteria may be used to accomplish this in the context of bootstrap confidence intervals. The criterion should, however, not be biased to underestimate the true lag order in small samples. This point is particularly important in the context of bootstrapping, where the lag order is estimated two times and a downward bias would thus be exacerbated. Kilian (1998b) suggests using the AIC (Akaike, 1974). Our simulations showed that this produces too wide confidence intervals and bands, in line with the results in Kilian (1998b).[2] This is probably due to the overestimation of the true lag order by the AIC and the subsequently higher estimation uncertainty of models with too many free parameters. We therefore use the corrected AIC (AICc) introduced by Sugiura (1978), applied to autoregressions in Hurvich and Tsai (1989) and to vector autoregressions in Hurvich and Tsai (1993). The AICc is given by

$$AICc = AIC + \frac{2\kappa(\kappa + 1)}{T - \kappa - 1} \tag{9}$$

with κ being the number of parameters per VAR equation. For a given lag order $\hat{p}$, we can estimate equation (8) using LS. Denote the estimators by $\hat{A}_0$ to $\hat{A}_{\hat{p}}$. When constructing confidence intervals it was found useful to correct for the bias in the estimated autoregressive coefficients, see e.g. Kilian (1998a)

[2]Simulations of DGPs with higher lag orders than considered here or by Kilian (1998b) indicate that the AICc might provide to low estimated lag orders and the AIC might be preferable. Research into the optimal information criterion for bootstrap confidence intervals and bands at different sample sizes may be helpful.

and Berkowitz and Kilian (2000). Two commonly used methods are the analytical method of Pope (1990) and the bootstrap method of Kilian (1998a). For a description of the method of Pope in a similar context to ours, the reader is referred to Kim (2004) and Staszewska-Bystrova and Winker (2013). Because of the lower computational cost we use the method of Pope (1990) to obtain bias corrected estimates $\hat{A}_0^{bc}$ to $\hat{A}_{\hat{p}}^{bc}$. The bias correction includes the stationarity correction suggested by Kilian (1998a). This means that the bias correction is not applied if a system is non-stationary. Further, if the estimates imply stationarity, but the bias corrected estimates correspond to a non-stationary model, only a fraction of the estimated bias is subtracted from the estimates. The fraction is gradually reduced until stationarity of the system is maintained throughout the bias correction. A process is considered stationary if the modulus of the largest eigenvalue of the companion matrix associated with the autoregressive coefficients is less than one.

Given estimated coefficients, we can compute residuals either as

$$\hat{\varepsilon}_t = y_t - \hat{A}_0 - \hat{A}_1 y_{t-1} - ... - \hat{A}_p y_{t-p} \tag{10}$$

or consistent with the bias corrected estimates as

$$\hat{\varepsilon}_t^{bc} = y_t - \hat{A}_0^{bc} - \hat{A}_1^{bc} y_{t-1} - ... - \hat{A}_p^{bc} y_{t-p}. \tag{11}$$

The residuals of (11) will no longer have a mean of zero, and should be recentered by subtracting the mean. Furthermore, the residuals of both (10) and (11) are rescaled by a factor of $\sqrt{(T-\hat{p})/(T-\hat{p}-K\hat{p}-1)}$ (see Stine, 1987). Let $\hat{\varepsilon}_t^*$ denote the recentered and rescaled residuals. Given $\hat{\varepsilon}_t^*$, we compute the least squares estimator $\hat{\Sigma}_\varepsilon$ of the covariance matrix Σ_ε. In a next step, the impulse response coefficients over H periods are obtained as

$$\hat{\Phi}_h = \sum_{i=1}^{h} \hat{\Phi}_{h-i} \hat{A}_i, \qquad h = 1, ..., H, \tag{12}$$

with $\hat{\Phi}_0 = I_K$ and $\hat{A}_i = 0$ for $i > \hat{p}$. Orthogonalized impulse response functions can be obtained as

$$\hat{\Theta}_h = \hat{\Phi}_h P, \qquad h = 0, ..., H, \tag{13}$$

where $PP' = \hat{\Sigma}_\varepsilon$ is the Cholesky decomposition of $\hat{\Sigma}_\varepsilon$. An element $\hat{\theta}_{k,j,h}$ of $\hat{\Theta}_h$ can be interpreted as the reaction of variable k to a shock in equation j, after h periods. In the case of a two dimensional VAR model, i.e. K=2, the recursive ordering of the Cholesky decomposition implies that the reaction of the first variable to the second shock is restricted to zero in the initial period h=0.

3.2. Bootstrap Algorithm

Our interest is in constructing confidence intervals and joint confidence bands for the orthogonalized impulse response functions. This section describes the algorithm to obtain bootstrap distributions of IRF coefficients. Based on these, confidence intervals and bands can be constructed, as detailed in Section 3.3. The bootstrap algorithm given below includes optional steps for conducting a bias correction or a skewness adjustment or both. Aside from the mirroring steps, the bootstrap procedure described here is in large parts similar to that in Kilian (1998a,b) as well as to that in Lütkepohl et al. (2015a). The bootstrap procedure is given in Algorithm 1. Based on the ideas of Efron (1981) and Kilian (1998a,b) we construct confidence intervals and bands, labelled EK, using Algorithm 1 with the bias correction options (steps 2 and 8). Hall-type intervals and bands are obtained using the same bootstrap algorithm, also including the bias correction. The difference between the two is whether we use the percentile intervals given in (3) or the interval given in (4). The M and the MP methods do not employ the bias correction options but use the mirroring and the percentile mirroring given in (5) and (6), respectively (step 13 in Algorithm 1). The MPbc method utilises the percentile mirroring as well as the bias correction steps of Algorithm 1.

Algorithm 1 Bootstrap Procedure

1: Estimate lag order and coefficients of VAR;
2: Optional: Bias-correct VAR coefficients;
3: Obtain, rescale, and re-center residuals;
4: Optional: When a mirroring method is used, estimate and store VAR parameters for all lag orders between 1 and p_{max}
5: **for** $b = 1$ to B **do**
6: Construct bootstrap time series $y_t^b = \hat{A}_0 + \hat{A}_1 y_{t-1}^b + ... + \hat{A}_{\hat{p}} y_{t-\hat{p}}^b + \varepsilon_t^{*b}$, $t = \hat{p}+1, ..., T$, where ε_t^{*b} is a random draw with replacement from the residuals. The initial values $(y_1^b, ..., y_{\hat{p}}^b)$ are set to a randomly chosen sequence $(y_\tau, ..., y_{\tau+\hat{p}-1})$ from the data $\{y_t\}$, $\tau \in \{1, ..., T-\hat{p}\}$. If the bias correction in step 2 is applied, use $\hat{A}_0^{bc}$ to $\hat{A}_p^{bc}$ instead of $\hat{A}_0$ to $\hat{A}_p$;
7: Estimate bootstrap lag order and bootstrap coefficients;
8: Optional: Bias-correct the bootstrap coefficients;
9: Obtain, rescale, and re-center bootstrap residuals;
10: Calculate the bootstrap covariance matrix;
11: **end for**
12: **for** $b = 1$ to B **do**
13: Optional: Apply the mirroring or percentile mirroring to the bootstrap coefficients;
14: Compute orthogonalized IRF coefficients;
15: **end for**

3.3. Construction of Confidence Intervals and Bands

For the construction of confidence intervals, the percentile interval of Efron (1981), given by (3), is used for all methods but Hall's. Hall's interval is based on the adjusted percentile interval given in (4). Thus, in contrast to the three mirroring methods, Hall's method 'mirrors' the distribution after computing the impulse response functions.

Furthermore, we construct five types of joint confidence bands. The bands are constructed based on the Bonferroni adjusted (`Ba`) method introduced in Lütkepohl et al. (2015a). Out of the different possibilities given in the literature, the `Ba` method is chosen here because it was shown to work well in previous studies. The method was successfully applied to VAR forecasting of corporate bond spreads in Staszewska-Bystrova and Winker (2014), to forecasting with SETAR models in Grabowski et al. (2017), and was shown to compare favourably for constructing confidence bands for impulse response functions in Lütkepohl et al. (2015a,b).

The Bonferroni adjusted bands are constructed as follows. First, select the $\frac{a}{2(H+1)}B$ smallest and largest bootstrap impulse response coefficients in each period and eliminate the corresponding impulse response functions from the set containing all B bootstrapped IRFs ($H+1$ is the number of periods over which the IRFs are investigated). This ensures that at most a fraction a of the bootstrap impulse response paths are eliminated, which is consistent with the Bonferroni principle. If the contour of the set of remaining IRFs is taken as confidence bands, it should ensure a coverage rate of at least $1-a$. In general, this procedure eliminates less than aB paths and can thus be considered conservative, exhibiting a large width. The Bonferroni adjusted method therefore proceeds to eliminate more bootstrap IRF paths. In each step, the path which contributes the most to the width of the current band is removed. The width is measured as the sum of the widths of the individual intervals. This continues until aB bootstrap impulse response functions have been eliminated. The envelope of the remaining $(1-a)B$ functions represents the `Ba` band. For a more detailed discussion of the `Ba` method see Staszewska-Bystrova and Winker (2014) and Lütkepohl et al. (2015a).

If the `Ba` bands are computed based on IRFs obtained with the bias corrected bootstrap, the bands are labelled Efron-Kilian (EK) bands. Based on bootstrap IRFs obtained with Algorithm 1 using the mirroring, the percentile mirroring or the percentile mirroring with bias correction, we obtain M bands, MP bands, and MPbc bands, respectively. To compute Hall-type bands, the bias corrected bootstrap is again used. Similarly to the construction of confidence intervals in (4), Hall bands reverse the `Ba` bands. For this, the bootstrap IRF coefficients of each period are mirrored at the estimated IRF coefficients, i.e.

$$\hat{\Theta}_h^{b,Hall} = 2\hat{\Theta}_h - \hat{\Theta}_h^b. \tag{14}$$

Based on the bootstrap distribution of $\hat{\Theta}_h^{b,Hall}$, Hall-type confidence bands are constructed using the `Ba` algorithm described above.

4. Monte Carlo Evaluation

4.1. Simulation Setup

To evaluate the performance of the proposed mirroring procedure, we simulate different DGPs that all follow a two-dimensional VAR(p) process as given in equation (8). Simulations are performed under the assumption that the true lag order p of the DGP is known and also under the assumption that it is not known and has to be estimated. If the true lag order is set to p=1 it is not possible to underestimate the lag length (if an order of zero is not considered). Underestimating the true lag order generally leads to substantial coverage errors of confidence intervals and bands. To critically evaluate the methods, we therefore consider the cases p=1 as well as p=2.

For the DGP with one lag we use a design common in the literature (e.g. Kilian (1998a)),

$$A_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; A_1 = \begin{pmatrix} \alpha_{11} & 0 \\ 0.5 & 0.5 \end{pmatrix}. \tag{15}$$

For the parameter α_{11} values in {-1, -0.8, -0.5, -0.3, 0, 0.1, 0.3, 0.5, 0.7, 0.9, 0.95, 0.99, 1} are used. Alternative values of α_{11} imply different persistences of the process. For $\alpha_{11} = \pm 1$ the process is non-stationary. In such a setting, estimation of a vector error correction model may be preferable—provided the unit root and the cointegration rank are both correctly tested for (Kilian, 1998a; Berkowitz and Kilian, 2000). Thus, it makes sense to distinguish the results for different cases. In what follows, we aggregate the results in two ways: over all values of α_{11} and separately over only the less persistent cases. As less persistent DGPs we consider the models with $\alpha_{11} \in$ {-0.8, -0.5, -0.3, 0, 0.1, 0.3, 0.5, 0.7, 0.9}.

For the DGP with two lags we use

$$A_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \; A_1 = \begin{pmatrix} \alpha_{11} & 0 \\ 0.4 & 0.4 \end{pmatrix}, \; A_2 = \begin{pmatrix} 0.3 & 0 \\ -0.2 & -0.1 \end{pmatrix}. \tag{16}$$

This setting uses $\alpha_{11} \in$ {-0.7 -0.6, -0.5, -0.3, 0, 0.1, 0.3, 0.5, 0.6, 0.65, 0.69, 0.7}. For $\alpha_{11} = \pm 0.7$ the process has a unit root. Hence, summary statistics for processes with $\alpha_{11} \in$ {-0.6, -0.5, -0.3, 0, 0.1, 0.3, 0.5, 0.6} are again reported separately.

The error term is assumed to follow a multivariate normal distribution with mean zero and covariance matrix

$$\Sigma_\varepsilon = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}. \tag{17}$$

The impulse responses are evaluated for horizons h=0,...,10. For a given DGP we simulate 2000 time series of size T. The sample size takes on values in {30, 50, 100, 200, 1000}. The bootstrap algorithm uses B=2000 iterations. The confidence level is set to $1-a$=95%. We present results for confidence intervals in section 4.2 and for joint confidence bands in section 4.3.

4.2. Results for Confidence Intervals

T	Across all α_{11} M	MP	MPbc	EK	Hall	$\alpha_{11} \in \{-0.8, ..., 0.9\}$ M	MP	MPbc	EK	Hall
Lag order p=1 known										
30	95.04	95.34	95.15	95.01	89.24	95.35	95.80	95.33	96.76	90.36
50	95.02	95.38	95.28	94.95	90.36	95.05	95.59	95.49	96.25	90.68
100	95.28	95.60	95.44	95.09	91.81	95.32	95.74	95.67	95.98	91.59
200	94.99	95.30	95.05	94.84	92.62	94.99	95.32	95.26	95.47	92.26
1000	95.14	95.23	95.09	95.05	93.66	95.20	95.32	95.29	95.33	93.18
Lag order p=2 known										
30	96.88	96.08	96.31	96.15	92.93	97.31	96.78	96.77	97.39	94.32
50	96.38	95.92	95.84	95.70	92.91	96.76	96.52	96.47	96.84	93.90
100	96.32	96.10	95.87	95.67	94.07	96.47	96.47	96.35	96.51	94.57
200	95.73	95.79	95.52	95.37	94.98	95.75	96.01	95.89	95.94	95.32
1000	95.38	95.43	95.26	95.15	95.47	95.36	95.40	95.36	95.35	95.70
Lag order p=1 unknown										
30	96.27	96.47	96.38	95.61	90.45	96.60	96.91	96.54	97.32	91.79
50	96.84	97.05	97.03	96.11	92.27	97.00	97.30	97.21	97.40	93.00
100	97.28	97.49	97.40	96.66	93.81	97.49	97.75	97.69	97.67	93.94
200	97.23	97.50	97.35	96.75	94.55	97.39	97.63	97.57	97.56	94.46
1000	97.34	97.57	97.47	97.27	96.14	97.49	97.65	97.59	97.74	96.22
Lag order p=2 unknown										
30	94.64	94.33	95.06	93.32	85.48	94.75	94.34	94.95	93.60	86.00
50	96.31	96.15	96.28	93.67	88.17	96.45	96.21	96.33	93.87	88.52
100	97.29	97.21	97.09	95.61	93.42	97.45	97.42	97.36	96.20	93.89
200	97.16	97.27	97.12	96.50	95.55	97.25	97.47	97.41	97.13	95.94
1000	97.04	97.20	97.11	96.71	96.29	97.15	97.23	97.20	97.05	96.48

Table 1: Mean coverage frequencies (in percent) for nominal 95% confidence intervals. Means of estimated coverage frequencies of intervals are computed over different parameter settings for α_{11}, over periods h=0, ..., 10 and over the four impulse responses in a two-dimensional VAR.

This section compares the performance of different confidence intervals for IRFs. We summarize results by averaging over the performances for all the different parameter choices for α_{11}, over horizons $h = 0, ..., 10$, and over the four IRFs ($y_1 \to y_1$, $y_1 \to y_2$, $y_2 \to y_1$, $y_2 \to y_2$). When p=1, each summary statistic is therefore based on 572 individual coverage frequencies when evaluating the entire range of α_{11}, and 396 values when looking only at the less persistent processes. For p=2, averages are obtained over 528 and 352 individual performance results. Table 1 shows the mean coverage frequencies for the different methods under investigation. Except for Hall's method in small samples, mean coverage rates approximate 95%. They tend to be larger when only less persistent processes are considered, when the lag order is endogenous, and when the sample size increases. Because the results of Table 1 are means that are computed across different periods, different settings for α_{11} and four impulse response

T	Across all α_{11}					$\alpha_{11} \in \{-0.5, ..., 0.6\}$				
	M	MP	MPbc	EK	Hall	M	MP	MPbc	EK	Hall
Lag order p=1 known										
30	2.78	2.62	2.40	5.33	11.81	2.84	2.71	2.54	3.35	10.70
50	2.13	1.92	1.79	3.94	10.45	2.18	2.02	1.86	2.35	10.69
100	1.60	1.55	1.44	2.79	8.78	1.67	1.62	1.54	1.73	9.71
200	1.25	1.21	1.05	2.09	7.29	1.24	1.17	1.09	1.20	8.53
1000	0.92	0.95	0.89	1.13	3.56	0.95	0.97	0.95	0.98	4.26
Lag order p=2 known										
30	3.64	3.46	3.02	4.47	9.05	3.87	3.61	3.27	3.75	8.01
50	2.84	2.64	2.47	3.69	8.01	3.11	2.87	2.74	3.00	8.76
100	2.27	2.07	2.06	2.79	6.74	2.48	2.35	2.31	2.39	6.76
200	1.81	1.60	1.50	1.95	4.45	1.97	1.84	1.78	1.79	4.77
1000	1.00	1.03	0.90	1.19	2.22	1.04	1.08	1.05	1.03	2.69
Lag order p=1 unknown										
30	3.11	3.08	2.75	5.42	10.76	3.27	3.28	2.90	3.77	9.29
50	2.78	2.83	2.72	4.32	8.35	2.95	3.06	2.90	3.41	7.97
100	2.89	2.96	2.88	3.67	6.50	3.12	3.21	3.15	3.36	6.89
200	2.73	2.86	2.73	3.16	5.36	2.93	3.01	2.95	3.16	6.21
1000	2.71	2.90	2.80	2.94	2.85	2.90	3.00	2.94	3.29	3.29
Lag order p=2 unknown										
30	4.27	4.49	3.96	6.21	19.75	4.51	4.70	4.35	5.89	20.14
50	2.92	2.90	2.90	5.29	14.50	3.10	3.05	3.12	5.05	15.40
100	2.97	2.86	2.73	3.18	6.65	3.17	3.08	3.00	2.93	6.90
200	2.80	2.77	2.64	2.69	3.93	2.98	3.00	2.95	2.83	4.14
1000	2.49	2.61	2.52	2.44	2.38	2.66	2.68	2.66	2.71	2.77

Table 2: Root mean squared coverage errors (RMSCEs) (in percentage points) for nominal 95% confidence intervals. Root means of the squared deviations of estimated coverage frequencies from the desired 95% nominal rate are computed over different parameter settings for α_{11}, over periods h=0, ..., 10 and over the four impulse responses in a two-dimensional VAR.

functions, they are not very informative with regard to the coverage accuracy. For some individual settings of α_{11}, horizon h, and IRF, we obtain coverage rates of around 99% while for others values drop below 80%. This might still result in mean coverages close to 95%, but cannot be considered accurate.

Table 2 presents root mean squared coverage errors (RMSCEs) to measure the percentage point deviations from the desired 95% coverage level. This number gives a good indication of how well each method is maintaining the nominal level across all settings for a given sample size.[3] The table shows that all three variations of the mirroring approach (M, MP, MPbc) substantially reduce the coverage errors in all four scenarios (known and unknown lag orders of p=1, 2) when highly persistent and non-stationary time series are allowed for. When

[3]This supposes that we equally dislike positive and negative deviations from the nominal coverage level. Arguably, too low coverages are a more severe violation of the idea underlying the construction of confidence intervals.

T	Across all α_{11} M	MP	MPbc	EK	Hall	$\alpha_{11} \in \{-0.8, ..., 0.9\}$ M	MP	MPbc	EK	Hall
Lag order p=1 known										
30	1.64	1.24	1.32	0.63	0.63	1.09	0.84	0.96	0.54	0.54
50	0.81	0.68	0.66	0.46	0.46	0.54	0.45	0.47	0.37	0.37
100	0.40	0.37	0.35	0.31	0.31	0.27	0.25	0.25	0.23	0.23
200	0.23	0.23	0.22	0.21	0.21	0.16	0.16	0.16	0.15	0.15
1000	0.09	0.09	0.09	0.09	0.09	0.07	0.06	0.06	0.06	0.06
Lag order p=2 known										
30	1.84	1.06	1.12	0.69	0.69	1.48	0.85	0.97	0.66	0.66
50	0.83	0.59	0.58	0.47	0.47	0.66	0.48	0.49	0.43	0.43
100	0.39	0.33	0.31	0.30	0.30	0.31	0.27	0.27	0.26	0.26
200	0.22	0.21	0.20	0.20	0.20	0.18	0.17	0.17	0.17	0.17
1000	0.08	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.07	0.07
Lag order p=1 unknown										
30	1.78	1.38	1.53	0.66	0.66	1.20	0.93	1.12	0.57	0.57
50	0.92	0.81	0.82	0.50	0.50	0.63	0.55	0.59	0.40	0.40
100	0.47	0.45	0.45	0.34	0.34	0.33	0.32	0.32	0.26	0.26
200	0.28	0.29	0.28	0.23	0.23	0.21	0.20	0.20	0.17	0.17
1000	0.11	0.12	0.11	0.10	0.10	0.08	0.09	0.08	0.07	0.07
Lag order p=2 unknown										
30	1.46	1.16	1.46	0.56	0.56	1.01	0.80	1.09	0.49	0.50
50	0.84	0.70	0.73	0.45	0.45	0.63	0.52	0.57	0.39	0.39
100	0.44	0.39	0.39	0.32	0.32	0.35	0.31	0.32	0.27	0.27
200	0.26	0.25	0.25	0.21	0.21	0.22	0.21	0.21	0.18	0.18
1000	0.10	0.10	0.10	0.09	0.09	0.09	0.09	0.09	0.08	0.08

Table 3: Mean widths for nominal 95% confidence intervals. The mean distance between the upper and lower bound of an interval is computed over different parameter settings for α_{11}, over periods h=0, ..., 10 and over the four impulse responses in a two-dimensional VAR.

only less persistent series are considered, the mirroring methods dominate the EK and Hall method in very small samples (T=30, 50). For medium and large sample sizes (T=100, 200), the performance of EK and the mirroring methods becomes comparable. Hall's method exhibits the largest deviations from the nominal coverage rate throughout all settings, but draws level with the other methods for T=1000. MPbc offers the smallest RMSCE out of the three newly proposed methods. Comparing the left and right side of Table 2 reveals an interesting difference between the mirroring methods and the two benchmark methods. The left blocks report summary statistics that include non-stationary processes while on the right only less persistent DGPs are considered. The mirroring methods generally do better for highly persistent and non-stationary processes than for less persistent processes. This is in contrast to the EK and Hall method, whose coverage properties are generally worse when highly persistent processes are allowed for. In conclusion, MPbc provides better or similar coverage accuracy as compared to the simpler mirroring schemes as well as to the EK and Hall intervals.

Besides the coverage frequencies, the size of confidence intervals is a relevant measure of how informative the intervals are. Table 3 gives the mean widths of the intervals. The EK and Hall's intervals have the lowest widths throughout all settings.[4] Out of the three mirroring methods, MP offers the smallest width. For small samples the mirroring methods produce particularly wide intervals. This might be justified as they also substantially improve coverage frequencies.

As discussed in Section 2, MP and MPbc are based on percentiles of a bootstrap distribution using only models with the same estimated lag length. This means that for scenarios where the true lag order is unknown, some percentiles may be based upon very few realizations. This might explain why MP and MPbc dominate the EK and Hall's intervals when the lag order is known, but cannot keep up their superiority when the lag order is endogenous. Nonetheless, both methods turn out to be relatively successful even for unknown lag orders.

In summary, mirroring the bootstrap distribution of the VAR coefficients seems to offer a path to achieving better coverage accuracy for confidence intervals in small samples, especially when unit roots may be present. However, the mirroring intervals are less informative due to their larger width.

4.3. Results for Joint Confidence Bands

A joint confidence band is considered to cover the true IRF only if the IRF is contained in the band at every horizon h. In presenting results for bands, we thus compute mean coverages and RMSCEs as averages across different settings of α_{11} and across the four impulse response functions. The width is still computed as the average width per period, however. Table 4 presents mean coverage frequencies for confidence bands. The results match those of Section 4.2. M, MP, MPbc, and EK bands all yield coverage rates around 95%. Coverage frequencies are lower for small samples and for known lag orders, while for larger samples and endogenous lag orders coverage rates are above 95%. Hall's bands again fall substantially short of the nominal level in samples of size 30, 50 and sometimes also 100. To measure the deviations of actual from nominal coverage rates for the different settings, we again turn to the RMSCEs presented in Table 5. The results show that the mirroring methods also work well for the construction of confidence bands. Coverage errors of M, MP, and MPbc bands are substantially lower than those of EK and Hall's bands when highly persistent and non-stationary processes are included in the evaluation. When looking only at the less persistent processes in the right columns of Table 5, the mirroring methods still outperform the benchmarks when the sample sizes are very small (T=30, 50). For medium (T=100, 200) and large samples (T=1000) the performance of the mirroring bands and the EK bands converge. The method providing the smallest RMSCE varies depending on sample size, lag length and between known and unknown lag orders.

We notice that again the comparatively strong performance of the percentile mirroring approaches (MP and MPbc) for known lag lengths appears to vanish for endogenous lag orders. Note also that the increases of the RMSCEs when

[4] By construction, the EK and Hall's intervals have the same width.

T	Across all α_{11} M	MP	MPbc	EK	Hall	$\alpha_{11} \in \{-0.8, ..., 0.9\}$ M	MP	MPbc	EK	Hall
	Lag order p=1 known									
30	93.20	93.31	93.67	92.48	82.99	93.31	93.62	93.85	94.22	85.20
50	94.35	94.40	94.55	93.47	84.13	94.28	94.63	94.78	94.85	85.12
100	94.93	94.97	94.85	94.19	86.32	94.82	95.09	95.16	95.08	85.89
200	94.80	94.86	94.63	94.26	87.53	94.78	94.90	94.92	94.94	86.32
1000	94.83	94.62	94.50	94.58	90.29	94.82	94.67	94.68	94.78	89.10
	Lag order p=2 known									
30	94.75	93.63	94.60	94.10	88.63	95.14	94.31	95.10	95.33	89.28
50	95.38	94.31	94.59	94.35	87.96	95.82	94.87	95.22	95.43	89.64
100	95.69	94.99	94.88	94.65	89.58	95.82	95.34	95.33	95.44	90.66
200	95.15	94.83	94.58	94.49	91.04	95.05	94.99	94.92	95.01	90.72
1000	94.41	94.35	94.14	94.12	92.91	94.33	94.29	94.23	94.25	92.10
	Lag order p=1 unknown									
30	94.67	94.64	95.31	93.45	86.34	94.82	94.93	95.47	94.98	88.95
50	96.33	96.37	96.65	94.94	89.67	96.40	96.57	96.84	96.19	91.90
100	97.18	97.30	97.30	96.12	93.08	97.27	97.53	97.57	97.05	94.37
200	97.30	97.57	97.50	96.59	94.19	97.37	97.69	97.67	97.41	94.63
1000	97.63	97.87	97.80	97.32	95.85	97.77	97.95	97.89	97.71	96.03
	Lag order p=2 unknown									
30	93.54	92.57	94.10	91.58	77.16	93.64	92.55	94.11	92.18	78.45
50	95.90	95.52	96.00	91.97	80.82	96.06	95.81	96.15	92.43	81.82
100	97.12	96.83	96.87	94.60	89.43	97.18	97.06	97.15	95.16	90.25
200	97.02	96.95	96.88	96.15	93.52	97.03	97.09	97.11	96.73	94.04
1000	96.95	97.11	97.07	96.54	95.19	96.92	97.11	97.12	96.76	95.22

Table 4: Mean coverage frequencies (in percent) for nominal 95% joint confidence bands. Means of estimated joint coverage frequencies of bands are computed over different parameter settings for α_{11} and over the four impulse responses in a two-dimensional VAR.

the sample size grows, which occur for all mirroring methods and also the EK method, are due to a shift from coverage rates being mostly below the nominal level in smaller samples to coverage rates mostly above the nominal level in larger samples.

Table 6 shows the mean widths per period for the different kinds of confidence bands. Again, EK's and Hall's method offer the lowest widths in all settings. Out of the three mirroring methods, MP has the lowest width while M bands might be considered excessively wide in small samples.

5. Conclusions

We investigate modifications of the bootstrap algorithm for the construction of confidence intervals and confidence bands for impulse response functions in vector autoregressive models. The simple mirroring method adjusts for bias and skewness of the bootstrap distribution of the coefficient estimators. This is achieved by mirroring the distribution of the bootstrap coefficients at the estimates, similarly to the percentile intervals suggested in Hall (1992). Mirroring

T	M	MP	MPbc	EK	Hall	M	MP	MPbc	EK	Hall
	Across all α_{11}					$\alpha_{11} \in \{-0.5, ..., 0.6\}$				
	Lag order p=1 known									
30	3.80	3.55	3.00	6.23	17.11	3.61	3.24	2.94	3.27	14.81
50	2.67	2.26	1.97	4.69	15.76	2.73	2.14	1.89	2.32	15.23
100	1.96	1.61	1.69	3.23	13.40	2.05	1.57	1.57	1.87	14.43
200	1.69	1.20	1.43	2.52	11.77	1.77	1.24	1.25	1.41	13.64
1000	1.38	1.14	1.29	1.63	6.97	1.40	1.12	1.16	1.37	8.26
	Lag order p=2 known									
30	2.84	3.32	2.22	3.81	11.95	2.91	2.73	1.98	2.16	9.74
50	2.09	2.36	1.89	3.44	12.86	2.11	2.00	1.61	1.93	11.06
100	1.77	1.68	1.68	2.63	10.24	1.67	1.60	1.58	1.80	7.85
200	1.23	0.97	1.17	1.70	7.02	1.41	0.93	0.97	1.04	6.48
1000	0.99	1.00	1.18	1.30	3.28	1.21	0.98	1.02	1.04	3.84
	Lag order p=1 unknown									
30	2.98	2.93	2.41	5.46	14.54	2.85	2.71	2.42	2.98	11.54
50	2.35	2.93	2.32	4.26	11.13	2.38	2.37	2.42	2.53	8.94
100	2.60	2.66	2.67	3.35	6.58	2.67	2.78	2.81	2.68	5.14
200	2.62	2.74	2.68	2.94	3.70	2.70	2.83	2.82	2.75	3.63
1000	2.83	2.97	2.90	2.76	1.82	2.96	3.05	3.00	2.98	2.07
	Lag order p=2 unknown									
30	4.35	5.31	3.54	6.92	27.67	4.30	5.34	3.58	6.07	26.53
50	2.52	2.56	2.31	6.34	21.46	2.57	2.48	2.40	5.59	21.06
100	2.49	2.38	2.33	3.21	9.85	2.57	2.46	2.49	2.65	9.43
200	2.25	2.14	2.12	2.04	4.44	2.28	2.21	2.25	2.03	4.00
1000	2.08	2.18	2.15	1.80	1.41	2.09	2.19	2.20	1.95	1.67

Table 5: Root mean squared coverage errors (RMSCEs) (in percentage points) for nominal 95% joint confidence bands. Root means of the squared deviations of estimated joint coverage frequencies from the desired 95% nominal rate are computed over different parameter settings for α_{11} and over the four impulse responses in a two-dimensional VAR.

individual coefficients moves them a considerable distance within the parameter space and does so for each coefficient individually. Because this might distort the dynamics of the estimated VAR systems, we explore a related but altered approach. The percentile mirroring equally mirrors the bootstrap distribution, but does so by squeezing and stretching the distribution rather than swapping coefficients. While both methods adjust the bias in the bootstrap distribution, they do so only after the bootstrapping. To address this problem, we suggest as a third approach to combine the percentile mirroring with a bias correction of the LS estimator. These three methods are compared to standard methods from the literature.

Monte Carlo evidence suggests that in samples with 50 or fewer observations, all three mirroring methods improve the coverage accuracy of confidence intervals and bands as compared to the benchmark methods. For samples of size 100 and larger, the different methods start to converge in terms of coverage rates. The best suited method might depend on the lag length of the true DGP and the kind of lag order selection. For unit root processes, the coverage rates of the

T	Across all α_{11} M	MP	MPbc	EK	Hall	$\alpha_{11} \in \{-0.8, ..., 0.9\}$ M	MP	MPbc	EK	Hall
					Lag order p=1 known					
30	1.70	1.32	1.38	0.74	0.74	1.17	0.92	1.03	0.64	0.64
50	0.87	0.75	0.73	0.53	0.53	0.60	0.52	0.53	0.44	0.44
100	0.44	0.42	0.40	0.35	0.35	0.31	0.29	0.29	0.27	0.27
200	0.26	0.26	0.25	0.24	0.24	0.19	0.19	0.18	0.18	0.18
1000	0.10	0.10	0.10	0.10	0.10	0.08	0.08	0.08	0.08	0.08
					Lag order p=2 known					
30	2.13	1.26	1.33	0.88	0.88	1.09	1.05	1.18	0.84	0.84
50	0.97	0.71	0.71	0.58	0.58	0.91	0.59	0.61	0.54	0.54
100	0.46	0.40	0.39	0.36	0.36	0.44	0.33	0.33	0.32	0.32
200	0.27	0.25	0.24	0.24	0.24	0.26	0.21	0.21	0.21	0.21
1000	0.10	0.10	0.10	0.10	0.10	0.10	0.09	0.09	0.09	0.09
					Lag order p=1 unknown					
30	2.00	1.54	1.67	0.80	0.80	1.42	1.09	1.27	0.70	0.70
50	1.10	0.96	0.97	0.61	0.61	0.81	0.69	0.74	0.51	0.51
100	0.58	0.56	0.55	0.42	0.42	0.44	0.41	0.42	0.33	0.33
200	0.35	0.35	0.35	0.29	0.29	0.27	0.26	0.27	0.23	0.23
1000	0.14	0.14	0.14	0.12	0.12	0.11	0.11	0.11	0.10	0.10
					Lag order p=2 unknown					
30	1.65	1.28	1.53	0.70	0.70	1.21	0.93	1.18	0.63	0.63
50	1.02	0.84	0.87	0.56	0.56	0.81	0.66	0.72	0.51	0.51
100	0.56	0.50	0.50	0.41	0.41	0.47	0.42	0.43	0.36	0.36
200	0.34	0.32	0.32	0.28	0.28	0.29	0.28	0.28	0.25	0.25
1000	0.13	0.13	0.13	0.12	0.12	0.11	0.11	0.11	0.10	0.10

Table 6: Mean widths for nominal 95% joint confidence bands. The mean distance between the upper and lower bound of a confidence band in one period is computed over different parameter settings for α_{11}, over the four impulse responses in a two-dimensional VAR and also over periods h=0, ..., 10.

mirroring methods clearly dominate the benchmarks. The mirroring methods maintain almost the same coverage accuracies whether or not non-stationary and highly persistent processes are allowed for.

The MPbc method offered the lowest squared coverage errors when the lag order was assumed to be known. The performance was less dominant for endogenously estimated lag lengths. As discussed in Sections 2 and 4, the implementation of MP and MPbc in this article can result in MP and MPbc intervals and bands that are based on very few bootstrap draws of VAR models for some lag orders. This might negatively affect their performance when the lag order has to be estimated. Finding a better implementation of this method is left to future research.

The article presents results for coverage frequencies as summary statistics due to the large number of simulation settings. When inspecting individual results for coverage rates, these reveal that coverage frequencies are usually below the nominal level in the initial period of a shock to the VAR system. At later horizons, the coverage rates are higher and usually above the nominal level. This indicates that the uncertainty about the covariance matrix of the

VAR system might be underestimated in the resampling procedure, while the bootstrap VAR coefficients exhibit too much variation. Future research might aim to reduce the variation of the bootstrap autoregressive coefficients while increasing the variation of the bootstrap covariance matrices.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.

Basawa, I. V., Mallik, A. K., Cormick, W. P., Reeves, J. H., Taylor, R. L., 1991. Bootstrapping unstable first-order autoregressive processes. Annals of Statistics 19, 1098–1101.

Berkowitz, J., Kilian, L., 2000. Recent developments in bootstrapping time series. Econometric Reviews 19 (1), 1–48.

Efron, B., 1981. Nonparametric standard errors and confidence intervals. Canadian Journal of Statistics 9, 139–172.

Efron, B., Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1 (1), 54–75.

Grabowski, D., Staszewska-Bystrova, A., Winker, P., 2017. Generating prediction bands for path forecasts from SETAR models. Studies in Nonlinear Dynamics & Econometrics 21 (5).

Hall, P., 1992. The bootstrap and Edgeworth expansion. Springer, New York.

Hurvich, C. M., Tsai, C. L., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.

Hurvich, C. M., Tsai, C. L., 1993. A corrected Akaike information criterion for vector autoregressive model selection. Journal of Time Series Analysis 14, 272–279.

Inoue, A., Kilian, L., 2013. Inference on impulse response functions in structural VAR models. Journal of Econometrics 177 (1), 1–13.

Jordà, Ò., 2009. Simultaneous confidence regions for impulse responses. The Review of Economics and Statistics 91 (3), 629–647.

Jordà, Ò., Marcellino, M., 2010. Path forecast evaluation. Journal of Applied Econometrics 25 (4), 635–662.

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

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

Kilian, L., 1999. Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses. Review of Economics and Statistics 81 (4), 652–660.

Kim, J. H., 2004. Bias-corrected bootstrap prediction regions for vector autoregression. Journal of Forecasting 23 (2), 141–154.

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

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.

Pope, A. L., 1990. Biases of estimators in multivariate non-Gaussian autoregressions. Journal of Time Series Analysis 11 (3), 249–258.

Sims, C. A., Zha, T., 1999. Error bands for impulse responses. Econometrica 67 (5), 1113–1155.

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.

Staszewska-Bystrova, A., Winker, P., 2013. Constructing narrowest pathwise bootstrap prediction bands using Threshold Accepting. International Journal of Forecasting 29 (2), 221–233.

Staszewska-Bystrova, A., Winker, P., 2014. Measuring forecast uncertainty of corporate bond spreads by Bonferroni-type prediction bands. Central European Journal of Economic Modelling and Econometrics 2, 89–104.

Stine, R. A., 1987. Estimating properties of autoregressive forecasts. Journal of the American Statistical Association 82 (400), 1072–1078.

Sugiura, N., 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. Communications in Statistics A7, 13–26.

Wolf, M., Wunderli, D., 2015. Bootstrap joint prediction regions. Journal of Time Series Analysis 36 (3), 352–376.