DIW BERLIN

1762

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2018

Constructing Joint Confidence Bands for Impulse Response Functions of VAR Models

A Review

Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker

IMPRESSUM

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
http://www.diw.de

ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
http://www.diw.de/discussionpapers

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
http://ideas.repec.org/s/diw/diwwpp.html
http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html

Constructing Joint Confidence Bands for Impulse Response Functions of VAR Models – A Review

Helmut Lütkepohl
DIW Berlin and Freie Universität Berlin
Mohrenstr. 58
10177 Berlin, Germany
email: hluetkepohl@diw.de

Anna Staszewska-Bystrova
University of Lodz
Rewolucji 1905r. 41
90-214 Lodz, Poland
email: emfans@uni.lodz.pl

Peter Winker
University of Giessen
Licher Str. 64
35394 Giessen, Germany
email: Peter.Winker@wirtschaft.uni-giessen.de

September 26, 2018

Abstract
Methods for constructing joint confidence bands for impulse response functions which are commonly used in vector autoregressive analysis are reviewed. While considering separate intervals for each horizon individually still seems to be the most common approach, a substantial number of methods have been proposed for making joint inferences about the complete impulse response paths up to a given horizon. A structured presentation of these methods is provided. Furthermore, existing evidence on the small-sample performance of the methods is gathered. The collected information can help practitioners to decide on a suitable confidence band for a structural VAR analysis.

Key Words: Impulse responses, vector autoregressive model, joint confidence bands

JEL classification: C32

1 Introduction

Since the seminal paper by Sims (1980), vector autoregressive (VAR) models became a standard work horse in applied economic analysis. Fields of applications include, e.g., transmission of monetary policy shocks (Bagliano & Favero, 1998), international business cycle linkages (Stock & Watson, 2005) or real effects of oil price shocks (Hamilton, 2009; Kilian, 2009).

In such applications the interest is in the dynamic reaction of variables to specific shocks hitting the system. Impulse responses are one of the most commonly applied tools for describing these dynamic reactions in structural vector autoregressive analysis (for some recent applications and reviews see, e.g., Kilian & Murphy (2012), Kilian (2013), Gertler & Karadi (2015), Fisher & Huh (2016), Kilian & Lütkepohl (2017), and Kapetanios, Price & Young (2018)). A typical set of results consists of a number of estimated response paths of endogenous variables to selected structural shocks. Usually, a standard graph presenting these outcomes, still offered, e.g., by a number of computer packages, includes the estimated response function constructed over the propagation periods $h = 0, \ldots, H$ and a collection of $H + 1$ individual confidence intervals. Each of these intervals could be used for making marginal inferences concerning the response of a particular variable to a specific shock for a given horizon h.

However, in the graphical output of impulse response analysis the lower and upper end points of the individual confidence intervals are typically connected providing the impression of a joint band for the complete estimated response function up to horizon H. This graphical representation suggests, as discussed in a number of papers lately, that the impulse response functions are interpreted as a whole. If the focus of the analysis is on the whole path, appropriate bands are required which correspond to this interpretation. What is more, sometimes the interest might lie in discussing the shape of several impulse response functions jointly. In such cases, the uncertainty associated with the estimates should be assessed using joint confidence bands.

The aim of this paper is to review the methods proposed in the literature for constructing such joint bands. We focus on frequentist approaches which include asymptotic and bootstrap methods. Bayesian procedures which could be alternatively used include the methods of Sims & Zha (1999), Inoue & Kilian (2013) and Montiel Olea & Plagborg-Møller (2018) and are not covered by this review.

Related to the problem of computing confidence bands for impulse responses is the issue of constructing prediction bands for path forecasts, i.e., forecasts calculated for a number of consecutive periods, from VARs. Methods of calculating such prediction bands were being developed in the literature in parallel to procedures for obtaining confidence bands for impulse responses. Although a detailed presentation of these methods is beyond the scope of the present paper, we provide some comments on the approaches used in forecasting in the concluding section of the article.

Impulse response analysis and path forecasting are also relevant for analyses based on non-linear models and some of the methods considered here

might also be applied in a non-linear setting, e.g., the method proposed by Wolf & Wunderli (2015). Since the number of contributions concerning construction of confidence or prediction bands for such models is very limited, we do not include these methods in this review (see, however, Grabowski, Staszewska-Bystrova & Winker (2017) who discuss bootstrap prediction bands for SETAR models).

In the following Section 2, we introduce the notation used for VAR models, the resulting impulse response functions and the corresponding confidence intervals and bands. Section 2 also describes the classification applied to the methods used for constructing joint bands, which will be presented in Sections 3 and 4. Some existing Monte Carlo based evidence on the performance of several methods is summarized in Section 5, while Section 6 provides concluding remarks.

2 Impulse response functions and confidence bands

2.1 Structural vector autoregressive analysis

The presentation of the methods will be based on a standard reduced-form VAR(p) model with p lags, although some of the methods have a broader scope of application. Let $y_t = (y_{1t}, \ldots, y_{Kt})'$ denote the vector of endogenous variables of dimension K at time t. Then, the VAR model is given by

$$y_t = \nu + \sum_{i=1}^{p} A_i y_{t-i} + u_t, \tag{2.1}$$

where the A_i, $i = 1, 2, \ldots$, are $K \times K$ slope coefficient matrices, ν is a fixed $K \times 1$ intercept term and $u_t = (u_{1t}, \ldots, u_{Kt})'$ is a zero mean white noise error process such that $u_t \sim (0, \Sigma_u)$. For simplicity we assume that the u_t are independently, identically distributed although some results discussed in the following hold under more general conditions. The covariance matrix Σ_u is assumed to be positive definite. Using the lag operator L defined such that $Ly_t = y_{t-1}$ and defining $A(L) = I_K - A_1 L - \cdots - A_p L^p$, the model can also be expressed as

$$A(L)y_t = \nu + u_t. \tag{2.2}$$

We do not consider additional exogenous regressors. Furthermore, the processes are assumed to be stable and stationary satisfying

$$\det A(z) = \det\left(I_K - \sum_{i=1}^{p} A_i z^i\right) \neq 0 \text{ for } z \in \mathbb{C}, |z| \leq 1. \tag{2.3}$$

Consequently, the roots are bounded away from the unit circle, and the process can be expressed in its moving average (MA) representation

$$y_t = A(1)^{-1}\nu + A(L)^{-1}u_t = \mu + \sum_{h=0}^{\infty} \Phi_h u_{t-h}, \tag{2.4}$$

where $\mu = A(1)^{-1}\nu$, $\Phi_0 = I_K$ and $\sum_{h=0}^{\infty} \Phi_h L^h = A(L)^{-1}$. Thus, the MA coefficient matrices Φ_h are functions of the A_i parameter matrices.

Structural shocks, ε_t, are obtained by a linear transformation of the u_t, i.e., $\varepsilon_t = \boldsymbol{B}^{-1} u_t$. A classical choice of $\boldsymbol{B}$ is a lower-triangular matrix obtained by a Cholesky decomposition of Σ_u. Many alternative ways to specify and identify the structural matrix $\boldsymbol{B}$ have been proposed in the literature. In particular, more general exclusion restrictions on the impact effects or the long-run effects of the shocks, sign restrictions on the impulse reponses, external instruments or specific features of the distribution of the DGP have been used to identify the $\boldsymbol{B}$ matrix and hence the structural shocks (for an extensive review of the related literature see Kilian & Lütkepohl (2017)). In the following the specific shape of $\boldsymbol{B}$ is not essential. We assume, however, that it is point-identified and a consistent, asymptotically Gaussian estimator is available. Replacing u_t in (2.4) by $\boldsymbol{B}\varepsilon_t$ provides the responses of the system to structural shocks as $\Theta_h = \Phi_h \boldsymbol{B}$, $h = 0, 1, \ldots$. These functions of the VAR parameters are the impulse response functions (IRFs) of interest, which are considered up to a fixed propagation horizon H.

In applications, most of the time, the interest is in a specific IRF, i.e., the response of some variable i to the j^{th} shock. We denote this response at horizon h by $\theta_h \, (= \theta_{ij,h})$. In order to ease notation, we will skip the subscript ij in the following and consider a generic IRF up to a fixed horizon H given by $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_H)$. Then, the goal consists in finding a joint confidence band, i.e., a hyper-rectangle $\times_{h=0}^{H} [l_h, u_h]$ such that the band contains $\boldsymbol{\theta}$ with a predefined nominal coverage probability of $1 - \alpha$, $0 < \alpha < 1$. The constraint of using a hyper-rectangle rather than a more general form, e.g., hyper-ellipses, follows from the usual graphical representation of IRFs by plotting (estimates of) θ_h against h as shown in Figure 1, which exhibits a stylized impulse response function with a corresponding confidence band.

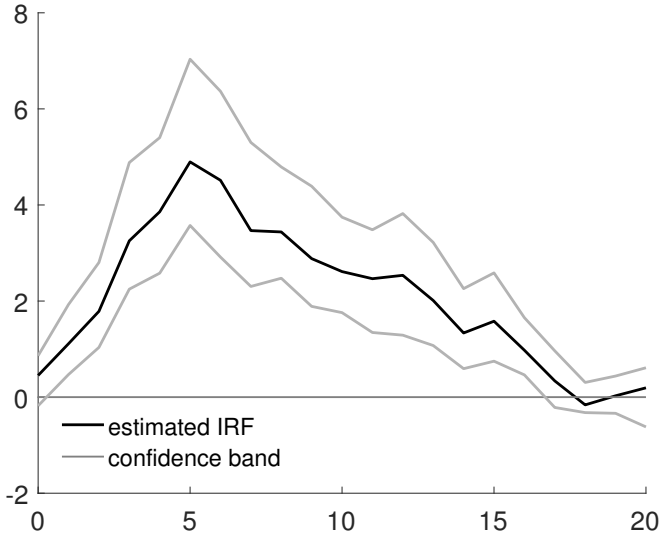

Figure 1: Stylized impulse response function for $H = 20$ with a confidence band.

If a lower-triangular matrix $\boldsymbol{B}$ is used to identify structural shocks, some

impulse response coefficients for $h = 0$ are zero by construction. Then, the band is constructed only for $h = 1, \ldots, H$. This applies analogously if some other parameter of an IRF is constrained to zero. In the following, we consider IRFs without zero restrictions. If such constraints are imposed the necessary adjustments to the methods are straightforward.

To enable a meaningful interpretation of IRFs, it appears preferable to have narrow bands for a given nominal coverage probability. The most commonly used measure of band width corresponds closely to the optical perception of figures like Figure 1. It consists in summing up the widths of the individual intervals at each horizon h for $h = 0, \ldots, H$. Given the equidistance of the horizons, this measure is approximately proportional to the area between the interpolated upper and lower bounds exhibited in Figure 1. This might be the reason why, e.g., Bruder & Wolf (2018) use the term `'volume'` for this measure of band width. By contrast, Bruder (2014) proposed using the geometric-average widths, which corresponds more closely to the idea of volume in the $(H + 1)$-dimensional space, where each IRF is represented by a point and a band by a hyper-rectangle. Given that this concept of volume is not common in the literature, we stick to the standard measure of band width.

Several of the methods presented in the following section allow constructing simultaneous confidence bands for several IRFs at a time with straightforward adjustments. Given that the joint consideration of several IRFs is not very common in empirical applications, we do not comment further on this possibility.

2.2 Classification of methods for constructing bands

Given a substantial and still growing number of methods proposed for constructing joint confidence bands for impulse response functions of VAR models, the presentation of these methods has to follow some order. We decided to group the methods in two broad classes.

The first class comprises methods which can be thought of as constructing the joint bands based on some estimate of the variances of the individual elements of the IRFs using one parameter for scaling. Then, for each horizon h, the band is given by the estimate $\hat{\theta}_h$ plus/minus a fixed multiple c of its estimated standard deviation $\hat{\sigma}_h$. Given this structure, these methods are classified as members of a one-parameter family. For these methods, typically, substantial knowledge is available regarding their asymptotic behavior. We consider their implementation based both on asymptotic and bootstrap approaches in Section 3.

The second group of methods comprises all remaining approaches. Some of these methods impose a specific structure on the bands, others use sequential or global optimization algorithms in order to obtain an adequate finite sample coverage, and also a semi-parametric concept is included in the survey provided in Section 4.

3 One-parameter family methods

3.1 Asymptotic methods

3.1.1 Naïve bands

The classical approach to constructing confidence bands for IRFs is based on standard asymptotic theory for individual estimators $\hat{\theta}_h$. Given that these estimators are non-linear (polynomial) transformations of the estimators of the VAR coefficient matrices A_i, asymptotic normality of the latter translates to asymptotic normality of $\hat{\theta}_h$. Consequently, the asymptotic variance-covariance matrix of the estimators $\hat{\theta}_h$ can be obtained by the Delta-method (for details and necessary assumptions see, e.g., Kilian & Lütkepohl (2017, pp. 335ff)). Let $\hat{\sigma}_h$ denote the estimated standard error of $\hat{\theta}_h$. Then, an asymptotic confidence interval for θ_h at confidence level $1-\alpha$ has the form

$$[\hat{\theta}_h - c_{1-\alpha/2}\hat{\sigma}_h, \hat{\theta}_h + c_{1-\alpha/2}\hat{\sigma}_h], \tag{3.1}$$

where $c_{1-\alpha/2}$ is the quantile of the order $1-\alpha/2$ of the standard normal distribution.

If the necessary assumptions for asymptotic normality of the parameter estimators are satisfied, this confidence interval will include the actual impulse response at horizon h, θ_h, with a probability of $1-\alpha$ asymptotically. However, for practical applications, two caveats are relevant. First, the asymptotic result does not imply correct coverage for finite (small) sample sizes. Therefore, we will consider alternative implementations based on bootstrap estimates in Subsection 3.2. Second, even asymptotically, the correct coverage at a given horizon h does not imply that the full IRF for $h = 0, \ldots, H$ falls into the hyper-rectangle

$$\times_{h=0}^{H}[\hat{\theta}_h - c_{1-\alpha/2}\hat{\sigma}_h, \hat{\theta}_h + c_{1-\alpha/2}\hat{\sigma}_h] \tag{3.2}$$

with a probability of $1-\alpha$. In fact, only for very specific settings, e.g., a perfect correlation of the $\hat{\theta}_h$ over h, the coverage of these naïve bands will be close to the nominal level. Even asymptotically, the actual coverage might be substantially smaller than $1-\alpha$, which might not come as a surprise given that the method considers each horizon separately. In the sequel, we present some adjustments to the quantile $c_{1-\alpha/2}$ in order to avoid or at least reduce this deviance.

3.1.2 Bonferroni bands

Exploiting the correspondence between hypothesis testing and confidence intervals, it appears straightforward to consider the problem of constructing a joint band for a complete IRF up to horizon H as a multiple testing issue. Assuming the worst case of complete independence between the $\hat{\theta}_h$ over h, according to Bonferroni's principle the quantile c has to be chosen corresponding to the level $1-\alpha/2(H+1)$ in order to guarantee a coverage

probability for the complete IRF of at least $1-\alpha$. Thus, the Bonferroni band is given by

$$\times_{h=0}^{H}\left[\hat{\theta}_h - c_{1-\alpha/2(H+1)}\hat{\sigma}_h, \hat{\theta}_h + c_{1-\alpha/2(H+1)}\hat{\sigma}_h\right]. \tag{3.3}$$

Obviously, if all assumptions regarding the asymptotic normality are satisfied, this band is conservative in the sense that the actual coverage rate will be at least $1-\alpha$, but tends to be substantially larger in real applications as the elements of an IRF will typically exhibit some correlation. In Section 4, adjustments to the Bonferroni bands are suggested which reduce the width of the band and, consequently, its (over-)coverage.

3.1.3 Šidák bands

Alternatively, joint confidence bands can also be constructed using Šidák's principle (Šidák, 1967). Similarly to the Bonferroni approach, these bands are designed to achieve the nominal coverage probability even if the $\hat{\theta}_h$'s are independent over h. However, in contrast to Bonferroni bands, the Šidák bands rest on the assumption of a multivariate Gaussian distribution of the impulse response estimators, which can be justified by asymptotic considerations. In fact, the approach is justified for a larger set of distributions, due to a result in Royen (2014). However, the underlying probability inequality is less general than the Bonferroni inequality. While the Bonferroni band is constructed from intervals with individual coverage of $(1-\alpha/(H+1))$, the Šidák band is computed using a marginal coverage level of $(1-\alpha)^{1/(H+1)}$. Consequently, the relevant factor is obtained as the quantile of the standard normal distribution of the order

$$1 - \frac{1-(1-\alpha)^{1/(H+1)}}{2}$$

and will be denoted as c_S. Then, the band is given by

$$\times_{h=0}^{H}\left[\hat{\theta}_h - c_S\hat{\sigma}_h, \hat{\theta}_h + c_S\hat{\sigma}_h\right]. \tag{3.4}$$

Since $1-\alpha/(H+1) \approx (1-\alpha)^{1/(H+1)}$ for values of H and α as they are typically used in econometric applications, there is almost no difference or very little difference between Bonferroni and Šidák bands for practical purposes, in particular for small α, e.g., $0 < \alpha \leq 0.1$ (see also Montiel Olea & Plagborg-Møller (2017, Figure 2)).

3.1.4 Wald bands

The methods described so far either assume a very specific and strong dependence between impulse responses (naïve bands) or provide conservative bands with coverage at least $1-\alpha$ asymptotically independently of the specific dependence structure. The Wald bands represent a method which takes this dependence structure into account explicitly. There exist two versions

of Wald bands, one based on the joint distribution of the $\hat{\theta}_h$ and one based directly on the joint distribution of the parameter estimators of the VAR model.

We start with the method based on the joint distribution of the $\hat{\theta}_h$ for $h = 0, \ldots, H$ considered in the context of joint confidence bands by Inoue & Kilian (2016) and, consequently, labelled as $\mathcal{W}_{IK}$. If the estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is asymptotically normally distributed, i.e.,

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{\theta}}),$$

and the asymptotic covariance matrix $\Sigma_{\boldsymbol{\theta}}$ of $\hat{\boldsymbol{\theta}}$ is nonsingular, the band is obtained based on the Wald confidence ellipse

$$\mathcal{W}(1-\alpha) = \{\boldsymbol{\theta} \mid W = T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\hat{\Sigma}_{\boldsymbol{\theta}}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq \chi^2_{H+1,1-\alpha}\},$$

where $\hat{\Sigma}_{\boldsymbol{\theta}}$ is a consistent estimator of $\Sigma_{\boldsymbol{\theta}}$ and $\chi^2_{H+1,1-\alpha}$ is the relevant quantile from a χ^2 distribution with $H+1$ degrees of freedom. The confidence band is obtained by the projections of the set $\mathcal{W}(1-\alpha)$, i.e.,

$$\mathcal{W}_{IK}(1-\alpha) = \times_{h=0}^{H} [\min\{\theta_h \mid \boldsymbol{\theta} \in \mathcal{W}(1-\alpha)\}, \max\{\theta_h \mid \boldsymbol{\theta} \in \mathcal{W}(1-\alpha)\}] .$$

Figure 2 provides a stylized representation of the different bands described so far. In particular, it becomes obvious, that exploiting the covariance between elements by means of the Wald ellipse leads to a smaller confidence set, while the encompassing rectangle used for the bands turns out to be even larger than the Bonferroni bands in this case. Lütkepohl, Staszewska-Bystrova & Winker (2015b) show that this outcome might be rather considered as a rule than an exception.

Since $\mathcal{W}_{IK}(1-\alpha)$ includes the Wald ellipse, it is conservative, i.e., the probability of including the actual IRF $\boldsymbol{\theta}$ is at least $1-\alpha$. Furthermore, as shown in Montiel Olea & Plagborg-Møller (2018), when $\Sigma_{\boldsymbol{\theta}}$ is non singular, it can be represented as a member of the one parameter family using $c_{IK} = \sqrt{\chi^2_{H+1,1-\alpha}}$.

It is well-known that the Wald statistic underlying the $\mathcal{W}_{IK}(1-\alpha)$ band in general does not have the assumed asymptotic χ^2_{H+1} distribution if the covariance matrix $\Sigma_{\boldsymbol{\theta}}$ is singular (e.g., Andrews (1987)). This situation occurs in the present context if the number of elements in the IRFs to be estimated (i.e., $H+1$ if a single IRF is considered) exceeds the number of slope parameters in the estimated VAR model. In that case the asymptotic distribution of the IRF is degenerate. Inoue & Kilian (2016) resolve the degeneracy of the asymptotic distribution by transforming the estimator. The corresponding Wald statistic then has a nonstandard asymptotic distribution which can, however, be approximated by a suitable bootstrap method proposed by Inoue & Kilian (2016).

An alternative method for constructing Wald confidence bands which avoids the problem of a degenerate asymptotic distribution of the estimator underlying the Wald statistic for impulse responses is based directly on the

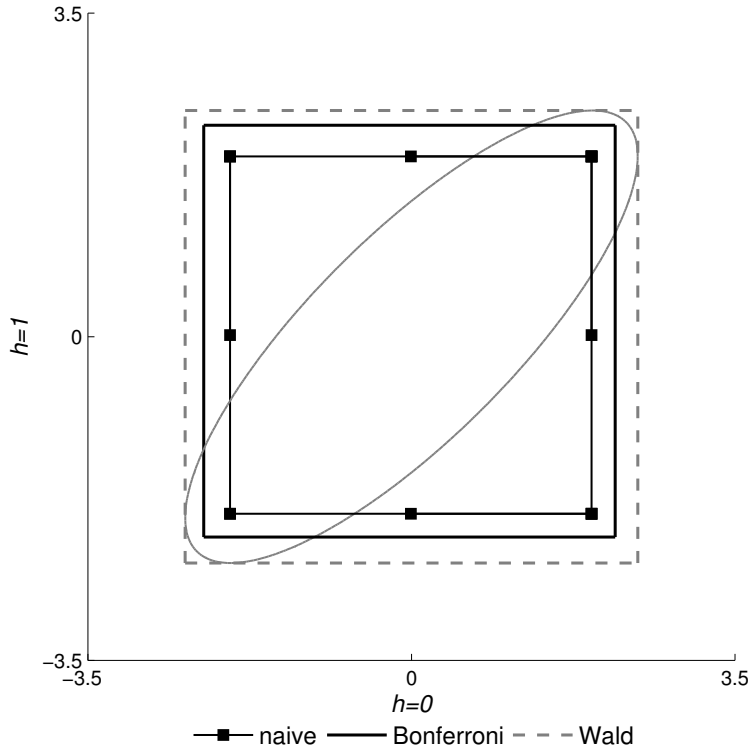

Figure 2: Stylized representation of alternative (naïve, Bonferroni and Wald) confidence sets for a 2-dimensional normal distribution

estimators of the VAR parameters and was considered by Lütkepohl et al. (2015b). Therefore, we will label this band as $\mathcal{W}_{LSW}$. We assume that all parameters of the VAR process required for calculating the IRFs of interest are collected in the M-dimensional vector η such that the IRF, $\boldsymbol{\theta} = \boldsymbol{\theta}(\eta)$, is a differentiable function of η.

Assuming that the estimator $\hat{\eta}$ of η has an asymptotic normal distribution such that

$$\sqrt{T}(\hat{\eta} - \eta) \xrightarrow{d} \mathcal{N}(0, \Sigma_\eta),$$

where Σ_η is nonsingular, the Wald confidence region with asymptotic coverage $1 - \alpha$ is defined as

$$\mathcal{W}^\eta(1-\alpha) = \{\eta \mid W = T(\hat{\eta} - \eta)'\hat{\Sigma}_\eta^{-1}(\hat{\eta} - \eta) \leq \chi^2_{M,1-\alpha}\},$$

where $\hat{\Sigma}_\eta$ is a consistent estimator of Σ_η. From this confidence set for η, we can construct an asymptotic confidence set for $\boldsymbol{\theta}$ with a coverage of at least $1 - \alpha$:

$$\mathcal{W}^{\boldsymbol{\theta}}(1-\alpha) = \{\boldsymbol{\theta}(\eta) \mid \eta \in \mathcal{W}^\eta(1-\alpha)\}.$$

Finally, the confidence band is obtained as for $\mathcal{W}_{IK}$:

$$\mathcal{W}_{LSW}(1-\alpha) = \times_{h=0}^{H} \left[\min\{\theta_h \mid \boldsymbol{\theta} \in \mathcal{W}^{\boldsymbol{\theta}}(1-\alpha)\}, \max\{\theta_h \mid \boldsymbol{\theta} \in \mathcal{W}^{\boldsymbol{\theta}}(1-\alpha)\}\right].$$

According to Montiel Olea & Plagborg-Møller (2018), this projection can also be expressed in the form typical for the one-parameter family class of bands up to terms of order $o_p(T^{-1/2})$.

3.1.5 sup-t bands

Given the shortcomings in guaranteeing the correct coverage asymptotically of the methods presented so far, Montiel Olea & Plagborg-Møller (2018) propose a different choice of c. They start with the observation that the bands from the one-parameter family can also be represented by the set of all IRFs $\tilde{\boldsymbol{\theta}} = (\tilde{\theta}_0, \ldots, \tilde{\theta}_H)$ satisfying

$$\max_{h=0,\ldots,H} \frac{|\tilde{\theta}_h - \hat{\theta}_h|}{\hat{\sigma}_h} \leq c\,. \tag{3.5}$$

Consequently, according to their Lemma 1, the asymptotic coverage probability is given by the cumulative distribution function of the maximum of the absolute values of $H+1$ standard normal variables, which are typically correlated. Let Q denote this distribution, which has to be approximated by means of Monte Carlo simulation for a given estimate of the asymptotic variance-covariance matrix. Then, the obvious choice for c is the $1-\alpha$ quantile $q_{1-\alpha}$ of Q which provides the sup-t bands:

$$\bigtimes_{h=0}^{H} [\hat{\theta}_h - q_{1-\alpha}\hat{\sigma}_h, \hat{\theta}_h + q_{1-\alpha}\hat{\sigma}_h]. \tag{3.6}$$

The sup-t bands are asymptotically balanced, i.e., the pointwise coverage probabilities for each horizon h have the same asymptotic limit.

For a given estimate of $\hat{\Sigma}_{\boldsymbol{\theta}}$ the widths of the bands described in this subsection are proportional to the factor c. Montiel Olea & Plagborg-Møller (2017) list the following relations:

c**(naïve)** $\leq c$**(sup-t)** $\leq c$**(Šidák):** This relation always holds but it should be noted that the naïve band has coverage at most $1-\alpha$, but typically substantially smaller than $1-\alpha$ and is hence not a serious competitor. It is listed here for completeness.

c**(Šidák)** $\leq c$**(Bonferroni):** Formally, the Bonferroni bands are larger than the Šidák bands. However, the difference is very small for all relevant significance levels. Also, the Šidák band is based on a normal distribution assumption for $\hat{\boldsymbol{\theta}}$. This assumption is no problem here since all the bands are justified with asymptotic arguments only, using the asymptotic normal distributions of the estimators. On the other hand, the actual distribution of the impulse responses are non-Gaussian, which may make the use of the more generally valid Bonferroni band more plausible.

c**(Bonferroni)** $\leq c$**(Wald**$_{IK}$**)** $\leq c$**(Wald**$_{LSW}$**):** This relation holds if $\alpha < 0.5$ and $2 \leq H+1 \leq M$, where M is the total number of underlying VAR parameters including the relevant residual covariance parameters.

Montiel Olea & Plagborg-Møller (2017) attribute this result to Alt & Spruill (1977).

It may be worth emphasizing that the confidence intervals and, hence, the confidence bands rely on nonzero asymptotic variances σ_h^2. That condition is satisfied under common assumptions, even if the asymptotic distribution of the IRF is degenerate, for example, because the propagation horizon is chosen larger than the number of VAR slope parameters. However, Benkwitz, Lütkepohl & Neumann (2000) discuss cases where standard estimators of impulse responses have higher convergence rates than $\sqrt{T}$ such that the term $\sqrt{T}(\hat{\theta}_h - \theta_h)$ converges to zero and, hence, has a zero asymptotic variance even under usual assumptions.

In real applications, when IRFs are considered, their small-sample distributions are not known. They are not Gaussian in general. Therefore, the confidence bands as described above are typically based on bootstrap approximations of the sampling distributions of the quantities of interest as discussed in more detail in the next subsection.

3.2 Bootstrap methods

Both in the case when small-sample distributions are not known and cannot be assumed to be Gaussian and when there is not a closed-form solution for the asymptotic standard error as in the case of the sup-t statistic, bootstrap methods might be the appropriate approach when dealing with confidence bands for IRFs. For this reason, most applications are based on bootstrap methods. These methods proceed by generating samples of estimates of the quantities of interest, in our case the IRFs. Thus, they are based on a sample of estimates $\hat{\boldsymbol{\theta}}^b$, $b = 1, \ldots, N$. This sample can be used to estimate the standard deviations σ_h, $h = 0, 1, \ldots, H$, used to construct the confidence intervals and bands in the previous subsection. Alternatively, they could be used to estimate the quantiles of the t ratios underlying some of the intervals or they can be used to directly estimate the quantiles of the distributions of the estimated impulse responses $\hat{\theta}_h$. The latter approach is in fact rather common in the structural VAR literature. We will therefore primarily focus on that approach in the following. We are presenting the associated confidence bands under the heading of the one-parameter family because the basic construction principles are very similar to those presented in the previous subsection.

Bootstrap methods are often applied in the hope to obtain a better small-sample coverage than with their asymptotic counterparts. We emphasize, however, that just like asymptotic properties, bootstrap methods in general only approximate the true finite-sample properties of estimators and test statistics because they are based on empirical distributions and not on true distributions. The validity of bootstrap procedures is typically justified by asymptotic arguments assuming that both the sample size T and the number of bootstrap replications N go to infinity. In some cases bootstrap approximations of the distributions of certain statistics converge at a faster rate

than approximations based on conventional asymptotic arguments. This result has stimulated hopes that the bootstrap can also provide more accurate confidence intervals and bands. However, better theoretical convergence rates have not been obtained for the bootstrap confidence bands discussed in the following. Thus, their accuracy in small samples may not be better than that obtained for asymptotic theory based confidence bands. In practice, the relative accuracy of asymptotic and bootstrap based confidence intervals and bands for IRFs is typically investigated by Monte Carlo simulation techniques and may depend on the specific data generation processes (DGPs) used in such simulations. We will review some small-sample evidence in Section 5.

Despite the similarity of the ideas underlying the asymptotic and bootstrap confidence bands, the actual appearance of the two types of bands can differ substantially in small samples. For example, it is obvious that the impulse response estimate $\hat{\theta}_h$ always lies in the center of a corresponding interval such as (3.1). Similarly, the estimated IRF lies in the center of the band (3.2). In contrast, the point estimate of the IRF may be at the edge or even outside a bootstrap confidence band.

There exists a plethora of different bootstrap implementations, which differ in their suitability for different settings (see, e.g., Kilian & Lütkepohl (2017, pp. 340ff)). Therefore, we restrain ourselves from providing an overview and present one standard implementation, i.e., the residual-based bootstrap as suggested by Kilian (1998b, 1999), which is used for generating bootstrapped IRFs. In the following subsections, we describe how these bootstrapped IRFs are used in the bootstrap versions of the methods introduced above in order to obtain joint confidence bands.

Assuming that a sample $y_1, \ldots, y_T$ and presample values $y_{-p+1}, \ldots, y_0$ are available, the steps of the bootstrap algorithm are as follows:

1. The parameters of (2.1) are estimated, resulting in $\hat{\nu}, \hat{A}_1, \ldots, \hat{A}_p, \hat{\Sigma}_u$ and the corresponding residuals are computed as

   $$\hat{u}_t = y_t - \hat{\nu} - \hat{A}_1 y_{t-1} - \cdots - \hat{A}_p y_{t-p}$$

   for $t = 1, \ldots, T$. Unless the $\hat{u}_t$ have mean zero by construction as in the case of ordinary least squares estimation, they should be re-centered.

2. The bootstrap DGP is defined by

   $$y_t^* = \hat{\nu} + \sum_{i=1}^{p} \hat{A}_i y_{t-i}^* + u_t^*,$$

   where u_t^* follows the empirical distribution of the (re-centered) $\hat{u}_t$. This DGP is used to generate N bootstrap samples y_t^* of size T, which are conditional on the p initial observations $y_{-p+1}, \ldots, y_0$ of y_t. Realizations of u_t^* are obtained by random drawing with replacement from $\{\hat{u}_t\}_{t=1}^T$.

3. Each of the N samples is used to estimate the parameters of (2.1) by the same estimation technique as applied in step 1. This leads to obtaining N different sets of estimates $\hat{\nu}^b, \hat{A}_1^b, \ldots, \hat{A}_p^b, \hat{\Sigma}_u^b$ for $b = 1, \ldots, N$.

4. Based on $\{\hat{\nu}^b, \hat{A}_1^b, \ldots, \hat{A}_p^b, \hat{\Sigma}_u^b\}_{b=1}^N$, N bootstrap replicates of impulse response functions $\hat{\theta}_0^b, \ldots, \hat{\theta}_H^b$, $b = 1, \ldots, N$, are computed.

The estimation of the VAR model in step 1 may be performed using least squares (see, e.g., Lütkepohl (2005)) or some other suitable estimation method. In practice, bias-corrected least squares is used in many applications. The bias correction can be based either on the asymptotic mean bias formula presented by Nicholls & Pope (1988) and Pope (1990) or the, more generally applicable, bootstrap estimator described by Kilian (1998b). The second approach is less often included in Monte Carlo comparisons of methods for constructing confidence bands due to the high computational complexity related to the implementation of the double bootstrap.

If the bias adjustment changes the dynamic properties of the estimated system by introducing non-stationarity, it is advisable to apply the stationarity correction of Kilian (1998b), consisting in down-scaling bias estimates in such a way that the corrected bias-adjusted values describe a stationary system.

In step 2, there are also alternative approaches to choose the initial observations. One may either use the original presample observations $y_{-p+1}, \ldots, y_0$ or one may randomly sample p consecutive values from the sample values $y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T$.

Generally, the above algorithm may need some modifications for specific empirical applications. For example, an adjustment might be required because the lag order p is rarely known and has to be estimated. The lag order selection may then be repeated in step 3 giving rise to the endogenous lag order algorithm of Kilian (1998a). Further changes would be necessary if the errors u_t were not independent but just serially uncorrelated and possibly conditionally heteroskedastic. In such cases, the residual-based bootstrap method described above, should be substituted with a suitable wild bootstrap or block-bootstrap procedure as proposed by Brüggemann, Jentsch & Trenkler (2016) (for an overview see, e.g., Kilian & Lütkepohl (2017)).

3.2.1 Naïve bands

The construction of naïve bands based on N bootstrap realisations is straightforward. For each $h = 0, \ldots, H$ the $1-\alpha$ interval is obtained by determining the $\alpha/2$ and the $1-\alpha/2$ percentiles of the bootstrapped $\hat{\theta}_h^b$.

If the bootstrap distribution corresponds well to the finite sample distribution of the impulse responses, at each horizon h, a share of $1-\alpha$ confidence intervals will cover the true θ_h. However, this does not imply that a share of $1-\alpha$ of the bands covers the true IRF $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_H)'$. Figure 3 illustrates the point by showing a random sample from a bivariate normal distribution and confidence sets corresponding to naive (left) and Bonferroni bands (right) for $\alpha = 0.05$. While in the left-hand panel exactly 95% of the observations are in each of the two intervals extending to infinity in the second dimension (dotted lines), the joint coverage probability of the intersection amounts to 92.3%, falling short of the nominal level.

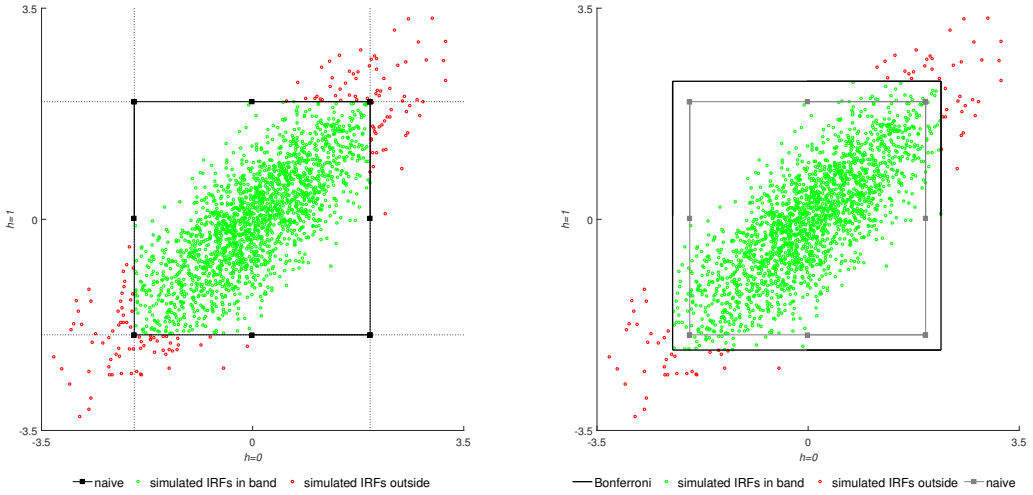

Figure 3: Comparison of naïve (left) and Bonferroni (right) confidence sets for simulated 2-dimensional IRFs.

3.2.2 Bonferroni bands

The bootstrap based construction of Bonferroni bands follows exactly the procedure for the naïve bands with the sole difference that for each $h = 0, \ldots, H$ the $1-\alpha/(H+1)$ interval is obtained by determining the $\alpha/2(H+1)$ and the $1-\alpha/2(H+1)$ percentiles of the bootstrapped $\hat{\theta}_h^b$.

The Bonferroni bands are illustrated in the right panel of Figure 3. A direct comparison with the naïve bands indicates that the joint coverage increases. Now, the individual coverage of the sample points in each dimension amounts to 97.5%, while, as might have been expected given the theoretical properties of Bonferroni bands, the joint coverage exceeds the nominal level and is equal to 96.15%.

3.2.3 Šidák bands

Šidák bands are quite similar to Bonferroni bands and so is the construction based on the bootstrapped IRFs by just replacing the bootstrapped quantiles of orders $\alpha/2(H+1)$ and $1-\alpha/2(H+1)$ by the percentiles

$$\frac{1-(1-\alpha)^{1/(H+1)}}{2} \quad \text{and} \quad 1-\frac{1-(1-\alpha)^{1/(H+1)}}{2},$$

respectively. However, given that the construction depends on the joint normal assumption, these bands cannot be guaranteed to have at least a coverage of $1-\alpha$ even if the bootstrap procedure provides the correct finite sample distribution without approximation error. For most practical purposes, however, the difference between Bonferroni bands and Šidák bands will be negligible also for their bootstrap based versions.

3.2.4 Wald bands

The bootstrap approaches for both types of Wald bands differ from the previous ones as they are based on a confidence ellipse instead of univariate intervals. They differ insofar as the bootstrapped version of $\mathcal{W}_{IK}(1-\alpha)$ is based on the $\hat{\boldsymbol{\theta}}^b$, while $\mathcal{W}_{LSW}$ is constructed from the bootstrapped parameter estimates $\{\hat{\nu}^b, \hat{A}_1^b, \ldots, \hat{A}_p^b, \hat{\Sigma}_u^b\}_{b=1}^N$.

For $\mathcal{W}_{IK}(1-\alpha)$, in a first step the Wald statistics

$$W^b = T(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}})'\hat{\Sigma}(N)_{\boldsymbol{\theta}}^{-1}(\hat{\boldsymbol{\theta}}^b - \hat{\boldsymbol{\theta}})$$

have to be calculated, where $\hat{\Sigma}(N)_{\boldsymbol{\theta}}$ denotes an estimator of the covariance matrix of the $\hat{\boldsymbol{\theta}}^b$. Given that the asymptotic covariance matrix is singular if more impulse responses are considered than there are VAR parameters, it is important to use a suitable estimator $\hat{\Sigma}(N)_{\boldsymbol{\theta}}$ as proposed by Inoue & Kilian (2016) to ensure a valid bootstrap approximation of the potentially nonstandard asymptotic distribution of the Wald statistic.

Once the W^b are obtained, the $1-\alpha$ quantile $c_{1-\alpha}$ can be determined. It is not necessary to determine the corresponding Wald confidence ellipse as only its projections are required. Let $\mathcal{W}^{boot}(1-\alpha)$ denote the set of bootstrapped IRFs corresponding to the $(1-\alpha)N$ smallest values of W^b. Then, the bootstrapped confidence band is given by

$$\mathcal{W}_{IK}^{boot}(1-\alpha) = \times_{h=0}^{H} \left[\min\{\theta_h \mid \boldsymbol{\theta} \in \mathcal{W}^{boot}(1-\alpha)\}, \max\{\theta_h \mid \boldsymbol{\theta} \in \mathcal{W}^{boot}(1-\alpha)\}\right].$$

The procedure is illustrated in Figure 4. The left panel shows a random sample from a bivariate normal distribution and the corresponding Wald confidence ellipse. The resulting band is provided by the dashed rectangle in the right panel. The empirical coverage of this band is larger than or equal to that of the confidence ellipse.

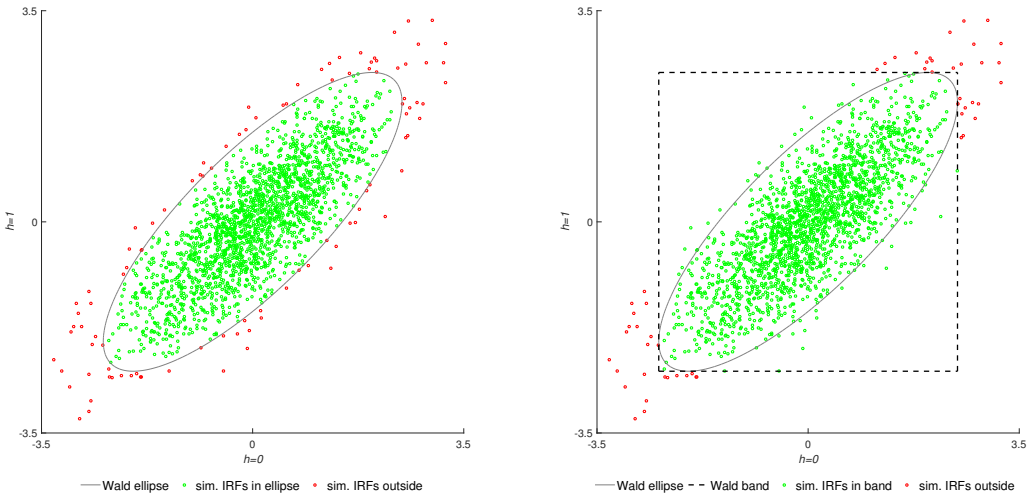

Figure 4: Comparison of Wald confidence sets for simulated 2-dimensional standard normal distribution.

The alternative method for constructing Wald confidence bands proposed by Lütkepohl et al. (2015b) constructs a Wald confidence set for the estimators of the VAR parameters. This avoids the issues with a potentially (near) singular covariance matrix. Otherwise, the procedure is quite similar. For each bootstrapped parameter vector η^b, the Wald statistic is calculated. The confidence set is given by the $(1-\alpha)N$ smallest values. For the elements in this set, the corresponding IRFs are calculated, and the resulting set of IRFs is projected as above, providing $\mathcal{W}_{LSW}^{boot}(1-\alpha)$.

3.2.5 sup-t bands

Montiel Olea & Plagborg-Møller (2017) propose two bootstrap methods for generating sup-t bands. The first one is based on the empirical percentiles of the bootstrapped $\hat{\theta}_h^b$, while the second one involves estimation of the bootstrap distribution of the maximum of the absolute values of the standardized estimates, which requires estimators of the empirical standard deviations.

Starting with the bootstrap realisations $\hat{\theta}_0^b, \ldots, \hat{\theta}_H^b$, $b = 1, \ldots, N$, for each $h = 0, \ldots, H$, the empirical ζ quantile of $\hat{\theta}_h^b$ can be defined and is denoted as $q_{h,\zeta}$. The goal is to find a value of $\hat{\zeta}$ such that the rectangle

$$\times_{h=0}^{H} [q_{h,\hat{\zeta}}, q_{h,1-\hat{\zeta}}] \tag{3.7}$$

covers at least $(1-\alpha)N$ of the bootstrap realisations. Montiel Olea & Plagborg-Møller (2017) propose to solve numerically for $\hat{\zeta}$ on the interval $[\alpha/2(H+1), \alpha/2]$, where the lower bound is motivated by the Bonferroni bounds.

In the second approach, the bootstrap realisations are used in a first step to obtain estimators of the empirical standard deviation $\hat{\sigma}_h$ of $\hat{\theta}_h$, $h = 0, \ldots, H$. Then, for each bootstrap realisation b, the maximum

$$\hat{m}^b = \max_{h=0,\ldots,H} \frac{|\hat{\theta}_h^b - \hat{\theta}_h|}{\hat{\sigma}_h} \tag{3.8}$$

is calculated. The quantity $\hat{q}_{1-\alpha}$ is obtained as the empirical $(1-\alpha)$ quantile of $\hat{m}^1, \ldots, \hat{m}^N$. Finally, the band is given by

$$\times_{h=0}^{H} [\hat{\theta}_h - \hat{\sigma}_h \hat{q}_{1-\alpha}, \hat{\theta}_h + \hat{\sigma}_h \hat{q}_{1-\alpha}] \,. \tag{3.9}$$

Montiel Olea & Plagborg-Møller (2018) mention that their implementation of the bootstrap version of the sup-t bands is closely related to the bootstrap based adjusted Bonferroni and Wald bands of Lütkepohl, Staszewska-Bystrova & Winker (2015a, 2015b) which will be described in more detail in Section 4.3. The latter authors present simulation evidence on the small sample performance of these methods and find that they can lead to rather wide confidence bands in practice which led to further research on improving the methods and motivated some of the proposals presented in the following section. The small sample properties of the methods will be further discussed in Section 5.

4 Other methods

In this section, we review those methods of constructing confidence bands which do not fall into the category of one-parameter family procedures. The approaches are quite diverse. We start with presenting the bands proposed by Bruder & Wolf (2018) in Subsection 4.1, which are constructed using a bootstrap. Subsection 4.2 presents the method proposed by Jordà (2009), which is based on asymptotic considerations. The remaining proposals either use or can be applied using the bootstrap as introduced in Subsection 3.2. Despite the differences, the last group of methods introduced in Subsection 4.3 shares the feature that bands are formed as envelopes of selected $(1-\alpha)N$ bootstrap IRFs. The methods differ in the way the relevant impulse responses are chosen.

4.1 Bands of Bruder and Wolf

Bruder & Wolf (2018) propose to construct confidence bands using a method that was originally developed by Romano & Wolf (2010) to control the joint size of a sequential testing procedure. The algorithm uses the bootstrap to approximate the distributions of the terms $\max \sqrt{T}|\hat{\theta}_h - \theta_h|$. In the first step, bootstrap distributions $\{\sqrt{T}|\hat{\theta}_h^b - \hat{\theta}_h|\}_{b=1}^N$ for propagation horizons $h \in \{0, \dots, H\}$ are obtained, which allows to compute $H+1$ empirical distribution functions $G_h(s)$ of $\sqrt{T}|\hat{\theta}_h^b - \hat{\theta}_h|$ of the form

$$G_h(s) = \frac{1}{N}\sum_{b=1}^{N} \mathbb{1}\left(\sqrt{T}|\hat{\theta}_h^b - \hat{\theta}_h| \leq s\right), \quad s \in \mathbb{R},$$

where $\mathbb{1}(\cdot)$ denotes an indicator function which is 1 if the condition in the argument is satisfied and 0 otherwise. The corresponding empirical quantile functions $G_h^{-1}(q)$ of $\sqrt{T}|\hat{\theta}_h^b - \hat{\theta}_h|$ are given by $G_h^{-1}(q) = \inf\{s \mid G_h(s) \geq q\}$.

Denoting the set of propagation horizons, where $\sqrt{T}|\hat{\theta}_h^b - \hat{\theta}_h|$ have nondegenerate distributions by S such that $S \subseteq \{0, \dots, H\}$, the empirical distribution functions $D_h(s)$ of $\max_{h\in S} G_h(\sqrt{T}|\hat{\theta}_h^b - \hat{\theta}_h|)$, are defined as

$$D_h(s) = \frac{1}{N}\sum_{b=1}^{N} \mathbb{1}\left(\max_{h\in S} G_h\left(\sqrt{T}|\hat{\theta}_h^b - \hat{\theta}_h|\right) \leq s\right), \quad s \in \mathbb{R},$$

with corresponding empirical quantile functions $D_h^{-1}(q) = \inf\{s \mid D_h(s) \geq q\}$. The bands for the coverage level of $1-\alpha$ are then computed as:

$$\times_{h=0}^{H}\left[\hat{\theta}_h - \frac{1}{\sqrt{T}}G_h^{-1}\left(D_h^{-1}(1-\alpha)\right), \hat{\theta}_h + \frac{1}{\sqrt{T}}G_h^{-1}\left(D_h^{-1}(1-\alpha)\right)\right]. \quad (4.1)$$

Bruder & Wolf (2018) show that these bands have asymptotically correct coverage for points in the parameter space where the usual maximum likelihood estimators of the impulse responses have nondegenerate asymptotic distributions. Moreover, for those points they are asymptotically balanced,

i.e., the pointwise coverage probabilities for each horizon h have the same asymptotic limit. Bruder & Wolf (2018) consider bands with this property because it is rather common for IRFs and is shared, for example, by the naïve, Bonferroni, Šidák and sup-t bands presented earlier. It is not obvious by which alternative property it should be replaced in this context. Note, however, that allowing for unbalanced bands may lead to overall smaller bands.

4.2 Asymptotic bands of Jordà

Another procedure for constructing bands for structural impulse response functions was considered by Jordà (2009) who based his approach on Scheffé's S-method of simultaneous inference (Scheffé, 1953) and asymptotic normality of the joint distribution of impulse response estimators. The band is computed as

$$\begin{pmatrix} \hat{\theta}_0 \\ \vdots \\ \hat{\theta}_H \end{pmatrix} \pm P \begin{pmatrix} \sqrt{\frac{\chi^2_{H+1,1-\alpha}}{H+1}} \\ \vdots \\ \sqrt{\frac{\chi^2_{H+1,1-\alpha}}{H+1}} \end{pmatrix} \tag{4.2}$$

where P stands for the Cholesky factor obtained from decomposing an estimate of the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}} = (\hat{\theta}_0, \ldots, \hat{\theta}_H)'$ as $\hat{\Sigma}_{\boldsymbol{\theta}}/T = PP'$, and $\chi^2_{H+1,1-\alpha}$ represents the $1-\alpha$ percentile of the χ^2 distribution with $H+1$ degrees of freedom.

The formula (4.2) was later modified by Jordà & Marcellino (2010) who used the bands as prediction bands when forecasting from VARs. Their version has the form:

$$\begin{pmatrix} \hat{\theta}_0 \\ \vdots \\ \hat{\theta}_H \end{pmatrix} \pm P \begin{pmatrix} \sqrt{\frac{\chi^2_{1,1-\alpha}}{1}} \\ \vdots \\ \sqrt{\frac{\chi^2_{H+1,1-\alpha}}{H+1}} \end{pmatrix} \tag{4.3}$$

The methods (4.2) and (4.3) were subsequently criticized by Staszewska-Bystrova (2013) and Wolf & Wunderli (2015) who pointed out some inconsistencies in deriving these formulae. Staszewska-Bystrova (2013) suggested to substitute all negative entries of P by their absolute values, which improves the coverage frequencies of the bands. Wolf & Wunderli (2015) objected, among other things, to the reasoning behind division with $H+1$ in (4.2) or with $h+1$ in (4.3). They also point out some theoretical deficiencies in the derivation of the bands, which also concern the method of Staszewska-Bystrova (2013).

4.3 Bands constructed as envelopes of sets of IRFs

The group of methods presented in this subsection is based on selecting, from N replicates, $(1-\alpha)N$ sample IRFs with certain properties. These

sample IRFs are usually generated by bootstrapping, even though some of the papers used functions obtained from other types of simulations to describe the procedures. In what follows we focus on bootstrap based applications of the methods. Once the set $\mathcal{N}^*$ of $(1-\alpha)N$ IRFs is defined, the bands are obtained as:

$$\times_{h=0}^{H} \left[\min_{b \in \mathcal{N}^*} (\hat{\theta}_h^b), \max_{b \in \mathcal{N}^*} (\hat{\theta}_h^b) \right]. \tag{4.4}$$

Below we review the criteria used by alternative methods in the pre-selection process in order to obtain the set $\mathcal{N}^*$.

Staszewska (2007) proposes to obtain confidence regions for vector error correction (VEC) models by finding the $1-\alpha$ proportion of the bootstrapped impulse responses whose envelope, given by (4.4), provides the narrowest band. Since full enumeration of all possible sets of paths, which would require to consider $\binom{N}{(1-\alpha)N}$ options, or even some equivalent simplified enumerations (Staszewska-Bystrova & Winker, 2013), would be too time consuming for typically used values of N and α, she resorts to optimization using a genetic algorithm (GA). The minimization procedure is applied after simplifying the problem by indicating those bootstrapped impulse responses which, by construction, must be covered by the band. These paths are found by sorting the bootstrapped values for all horizons and selecting those with no elements belonging to the $\alpha N + 1$ largest or smallest values. Schüssler & Trede (2016) follow a similar approach replacing the optimization heuristic by an exact mixed-integer optimization algorithm to find minimum-width confidence bands. It involves a branch-and-bound algorithm improved through the use of cuts and heuristics. Thus, the method becomes feasible with current computational resources for problem instances as they are typically used in macroeconometric applications.

Apart from the methods using global optimization, some other heuristic approaches were proposed, which can be viewed as sequential optimization procedures. Staszewska (2007) describes three such algorithms based on iterative rejection from the initial set of size N of those bootstrapped IRFs whose removal shrinks the envelope of the paths in each of αN steps. One of these procedures, labelled as the neighboring path (NP) method, consists in rejecting those paths which are most distant from the estimated impulse response function. The distance is either defined as Euclidean distance or computed using absolute deviations. The band is obtained as the envelope (see equation (4.4)) of the remaining $(1-\alpha)N$ bootstrap IRFs.

Lütkepohl et al. (2015a, 2015b) suggest to adjust the classical approaches to forming bands given by bootstrap Bonferroni and Wald methods. The need for an adjustment arises since both types of bands are conservative and may be substantially too wide. As explained in Section 3.2.2, Bonferroni bands tend to have a coverage larger than the nominal confidence level. The adjusted procedures aim at reducing the coverage level to $1-\alpha$. An adjusted Bonferroni band is obtained by first identifying those bootstrap impulse responses which are completely covered by the Bonferroni band, then reducing this set until $(1-\alpha)N$ paths remain and obtaining the band according to

equation (4.4). The excessive paths are eliminated one at a time. In each iteration, an IRF is rejected whose removal reduces the current envelope by the largest amount.

An adjusted Wald band of Lütkepohl et al. (2015b) recognizes that $\mathcal{W}_{LSW}^{boot}(1-\alpha)$ may contain more than $(1-\alpha)N$ bootstrap IRFs. Therefore it removes those bootstrapped IRFs which, after sorting, provide the $(1-\alpha)N$, $(1-\alpha)N-1$, $(1-\alpha)N-2, \ldots$ values of the Wald statistic. The process continues until the removal of the next IRF would violate the constraint that at least $(1-\alpha)N$ IRFs have to be covered by the envelope of the retained bootstrapped impulse responses.

Lütkepohl, Staszewska-Bystrova & Winker (2017) construct confidence bands by considering a version of the highest density region (HDR) approach introduced by Hyndman (1995, 1996). The basic HDR method is applied to a collection of IRFs obtained from N bootstrap replications, by first estimating their individual densities, second selecting $(1-\alpha)N$ IRFs with highest density values and then finding the smallest rectangular box which covers the chosen replicates (for $\mathcal{N}^*$ indicating the set of bootstrap drawings corresponding to the $(1-\alpha)N$ largest density values, this box is given by (4.4)). Density estimates are calculated using a normal kernel estimator. To take into account different variances of response parameter estimators for alternative periods $h = 0, \ldots, H$ or even their whole covariance structure, two additional variants of the procedure are introduced. They consist in computing density estimates for appropriately transformed bootstrap IRFs and constructing the band on the basis of the original (pre-transformation) values of the $(1-\alpha)N$ IRFs with highest densities. The transformations involve respectively, weighing the bootstrapped values period-wise by the estimated standard deviations or whitening the bootstrap replicates.

5 Asymptotic and finite sample comparison

Given a substantial number of methods based on asymptotic arguments which could be used to compute confidence bands, it is useful to summarize their small-sample properties which might help to select the best method for a particular application. Some evidence on the small-sample performance can be gathered from the results of Monte Carlo simulations reported in the literature. The most important criteria used to assess the bands are their coverage, width and, in some cases, also their balance. Before we review some small-sample results for the confidence bands for IRFs it is useful to compare them on the basis of their asymptotic properties.

If the asymptotic distribution of the impulse responses can be assumed to be multivariate normal, large sample features of the procedures for constructing bands for a single impulse response function from the one-parameter family are easily compared. Sup-t bands are the most attractive bands, as only for this method asymptotic coverage is equal to the nominal rate of $1-\alpha$. In addition, these bands are balanced in large samples and they are narrower than the remaining bands designed for making joint inferences. Bonferroni,

Šidák and Wald bands are conservative with a large sample coverage of at least $1-\alpha$ and usually exceeding the nominal coverage probability. Although Šidák and Bonferroni bands are usually almost identical, under some conditions listed in Subsection 3.1.5, Šidák bands are slightly narrower and hence better than Bonferroni bands, which have in turn smaller width than the Wald bands. Additionally, Šidák and Bonferroni bands are asymptotically balanced. As mentioned before, despite being the smallest, the naïve bands are not a proper tool for building confidence bands as, in general, their asymptotic coverage is unknown and smaller than $1-\alpha$.

In the group of the remaining procedures, the bands of Bruder & Wolf (2018) and also the regions constructed as envelopes have asymptotically correct coverage (see, Lütkepohl et al. (2015a, 2017) and Bruder & Wolf (2018)). Moreover, the former bands are asymptotically balanced. All these procedures are preferred to the method of Jordà which may not have the desired coverage even in large samples (see e.g. Staszewska-Bystrova (2013), Wolf & Wunderli (2015) or Kilian & Lütkepohl (2017)).

For smaller sample sizes, as argued earlier, it may be beneficial to substitute the asymptotic methods with their bootstrap counterparts or alternatives. For this reason, most of the simulation evidence on the performance of the procedures for constructing bands involves the bootstrapped versions. While the small-sample properties have been studied for a variety of data generating processes (DGPs), a common DGP, originally considered by Kilian (1998b) in the context of investigating the properties of confidence intervals for impulse responses, has the form of a bivariate VAR(1):

$$y_t = \begin{bmatrix} \varphi & 0 \\ 0.5 & 0.5 \end{bmatrix} y_{t-1} + u_t, \quad u_t \sim \text{iid}\, \mathcal{N}\left(0, \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}\right). \tag{5.1}$$

This DGP allows to study the behaviour of alternative procedures for constructing bands in different scenarios, e.g., for stationary processes (for $|\varphi| < 1$), stationary persistent processes (e.g., for $|\varphi| = 0.95$) and unit root, cointegrated processes (assuming $\varphi = 1$). Other settings concern the length of the propagation horizon H, the sample size T and the nominal coverage rate $1-\alpha$.

Using this DGP with $\varphi \in \{-0.95, -0.9, -0.5, 0, 0.5, 0.9, 0.95, 1\}$, Lütkepohl et al. (2015a) compare the features of the bootstrapped naïve, Bonferroni, adjusted Bonferroni and NP methods with the properties of the Jordà (2009) procedure, for $1-\alpha = 0.9$. Using $H \in \{10, 20\}$ and $T \in \{50, 100, 200\}$, it is shown that the naïve bands and the bands described by Jordà (2009) may undercover considerably in small samples and are inferior to the competing methods. The Bonferroni band is, as expected, conservative as it often leads to estimated coverage rates higher than the nominal value. At the same time, it is competitive with the best methods given by the adjusted Bonferroni followed by the NP procedure. The latter two exhibit the smallest deviations from the nominal coverage level and are attractive in terms of their moderate width. These conclusions are confirmed by results obtained for a different DGP in the form of a three-dimensional VAR(3) with param-

eter values based on estimation results using the crude oil market dataset of Kilian (2009), $H \in \{18, 36\}$ and $T \in \{100, 200, 400\}$.

Lütkepohl et al. (2015b) also report simulation results based on the DGP (5.1) with the same parameter choices as in Lütkepohl et al. (2015a). These results show that $\mathcal{W}_{LSW}^{boot}$ bands typically have larger coverage and width than the bootstrapped Bonferroni bands for small and large sample sizes. Thus, they are considerably conservative. It is also concluded that the adjusted Wald bands have very similar features to the adjusted Bonferroni bands, i.e., their actual coverage approximates the nominal value reasonably well and they are much more narrow than the unadjusted Bonferroni and Wald regions. The adjusted methods might be preferred, unless the sample size is very small, the propagation horizon H is large or the data are generated by a process with a unit root, as these circumstances may lead to their undercoverage. In these cases, conservative methods, especially the Bonferroni procedure, might be a safer choice.

Bruder & Wolf (2018) perform simulations to compare the small-sample performance of their balanced bands to selected competing approaches. One of the DGPs follows (5.1) with $\varphi \in \{-0.95, -0.9, -0.5, 0, 0.5, 0.9, 0.95\}$, $H \in \{10, 20\}$ and $T \in \{100, 400\}$, while the remaining experiments use a trivariate VAR(4) process. It is found that the coverage bias of the balanced bands is similar to that of the adjusted Wald bands for the smaller DGPs and that the former bands have advantages for the larger DGP. Both approaches, as well as the bootstrapped Bonferroni method, may however exhibit coverage distortions if the sample size is very small. In such cases, $\mathcal{W}_{LSW}^{boot}$ bands could be used. Bruder & Wolf (2018) examine also the empirical balance of their bands and find that it may be quite distorted in small samples.

DGP (5.1) is also used by Lütkepohl et al. (2017) who investigate the small-sample properties of alternative variants of HDR bands. An additional DGP is based on a corporate bond spreads model with six variables and four lags suggested by Deutsche Bundesbank (2005) and used for forecasting by Staszewska-Bystrova & Winker (2014). Lütkepohl et al. (2017) find that one variant of the HDR bands outperforms the $\mathcal{W}_{LSW}^{boot}$ procedure in terms of smaller width and more accurate coverage, as long as the sample size is not very small. The bands are also competitive with bootstrap Bonferroni bands, especially for the larger DGP and for some scenarios involving large propagation horizons.

Montiel Olea & Plagborg-Møller (2017) present some small-sample simulation results for DGP (5.1) with $\varphi \in \{0, 0.5, 0.9, 1\}$ which indicate that their bootstrap implementation of sup-t bands may have lower coverage than the nominal level for persistent processes, similar to what was found by Lütkepohl et al. (2015a, 2015b) for other methods. The bands are much more narrow than the conservative Bonferroni, Šidák and Wald regions. At the same time, it is shown that, in practice, Bonferroni and Šidák bands have almost identical width and coverage.

Some additional Monte Carlo evidence, using different DGPs than (5.1), is presented by Inoue & Kilian (2016) who study the features of bootstrapped

$\mathcal{W}_{IK}(1-\alpha)$ sets considered for a number of impulse responses jointly. It is shown that these confidence sets maintain the nominal coverage rates to a satisfactory degree, even for highly persistent processes with large dimensions. They exhibit similar coverage accuracy to bootstrapped $\mathcal{W}_{LSW}(1-\alpha)$ regions, but may perform in a more stable way, i.e., have smaller coverage errors in some problematic cases. The results also indicate that the bootstrapped Bonferroni sets may become more conservative than the Wald sets, if the number of statistics analyzed jointly becomes very large as in the case where a large number of impulse responses is considered simultaneously for a substantial propagation horizon.

6 Conclusions

The literature on computing joint confidence bands for impulse response functions from vector autoregressive models is still growing. The purpose of this review was to describe and classify frequentist methods which have been proposed and tested by various authors so far.

We have grouped the procedures by considering differences in the way the alternative bands are calculated. This led to distinguishing asymptotic and bootstrapped one-parameter family methods and the remaining approaches. Another possibility would be to classify the procedures according to their properties. If, e.g., asymptotic coverage was considered, it would be possible to distinguish methods with unknown coverage properties (naïve and Jordà's bands), exact asymptotic coverage (sup-t, Bruder and Wolf, NP, band obtained through global optimization, HDR, adjusted Bonferroni and adjusted Wald bands) and excessively large sample coverage (Bonferroni, Šidák and Wald bands). This could help to select an approach to constructing bands in empirical work for a given sample size, where for large samples, procedures with exact coverage would be preferred, while for very small sample sizes, a conservative method could be chosen. Classifying the procedures according to their coverage properties would be similar to differentiating the bands with respect to their width, as regions with larger coverage are usually also wider.

As indicated in the introduction, additional insights into the problem of constructing confidence regions could be gained by investigating the literature on building prediction bands for VARs. Such bands are designed to include future trajectories of predicted variables with a pre-defined probability of $1-\alpha$ and the principles for deriving them are very similar to the ideas behind constructing confidence bands for IRFs. For this reason, many approaches to building bands described in this review were also applied to the problem of forecasting. For example, the method of Jordà (2009) was later applied by Jordà & Marcellino (2010) to build prediction bands for VAR models and an improved version of this procedure was presented by Staszewska-Bystrova (2013). The bootstrap version of a sup-t method was used by Wolf & Wunderli (2015) in the context of predicting from VARs, but could be applied to prediction in non-linear models as well. A number of papers dealt also with the application of global and sequential optimization

methods to the problem of building prediction bands. Kolsrud (2007, 2015) proposed procedures labeled as the adjusted interval method, the Chebyshev method and the minimal content method to, respectively, univariate and multivariate time series forecasting. The neighbouring paths method of Staszewska (2007) was used by Staszewska-Bystrova (2011) to predict from VARs and the procedure based on a global optimization heuristic was considered by Staszewska-Bystrova & Winker (2013) who optimized the bands with the threshold accepting (TA) algorithm.

Directions of future research will be an extension of the available finite sample evidence based on Monte Carlo simulations. This will allow a better choice of the appropriate method for constructing joint confidence bands based on the properties of the data. Furthermore, the finite sample performance of the methods might benefit from refinements of the bootstrap procedure including the bias correction. Finally, given the growing interest in different types of non-linear multivariate time series models, the generation of joint confidence bands and the properties of methods in finite samples in such settings are research topics of potential interest.

Acknowledgements

Part of the work on this paper was conducted while the first author was a Fernand Braudel Fellow at the European University Institute in Florence. We are indebted to Daniel Grabowski, Lutz Kilian and two anonymous referees for helpful comments on a preliminary version of this paper. Financial support from the National Science Center (NCN) through MAESTRO 4: DEC-2013/08/A/HS4/00612 is gratefully acknowledged.

References

Alt, F. & Spruill, C. (1977). A comparison of confidence intervals generated by the Scheffé and Bonferroni methods, *Communications in Statistics – Theory and Methods* **A6(15)**: 1503–1510.

Andrews, D. W. K. (1987). Asymptotic results for generalized Wald tests, *Econometric Theory* **3**: 348–358.

Bagliano, F. C. & Favero, C. A. (1998). Measuring monetary policy with VAR models: An evaluation, *European Economic Review* **42**(6): 1069–1112.

Benkwitz, A., Lütkepohl, L. & Neumann, M. (2000). Problems related to bootstrapping impulse responses of autoregressive processes, *Econometric Reviews* **19**: 69–103.

Bruder, S. (2014). Comparing several methods to compute joint prediction regions for path forecasts generated by vector autoregressions, *ECON* –

Working Papers 181, University of Zurich, Department of Economics. revised 2015.

Bruder, S. & Wolf, M. (2018). Balanced bootstrap joint confidence bands for structural impulse response functions, *Journal of Time Series Analysis* (forthcoming).

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

Deutsche Bundesbank (2005). *Finanzmarktstabilitätsbericht 2005*, Deutsche Bundesbank, Frankfurt a.M.

Fisher, L. A. & Huh, H.-S. (2016). Monetary policy and exchange rates: Further evidence using a new method for implementing sign restrictions, *Journal of Macroeconomics* **49**: 177–191.

Gertler, M. & Karadi, P. (2015). Monetary policy surprises, credit costs, and economic activity, *American Economic Journal: Macroeconomics* **7(1)**: 44–76.

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

Hamilton, J. D. (2009). Causes and consequences of the oil shock of 2007-08, *Brookings Papers on Economic Activity* **40**: 215–283.

Hyndman, R. (1995). Highest-density forecast regions for nonlinear and non-normal time series models, *Journal of Forecasting* **14**(5): 431–441.

Hyndman, R. (1996). Computing and graphing highest density regions, *The American Statistician* **50**(2): 120–126.

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

Inoue, A. & Kilian, L. (2016). Joint confidence sets for structural impulse responses, *Journal of Econometrics* **192(2)**: 421–432.

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

Jordà, O. & Marcellino, M. (2010). Path forecast evaluation, *Journal of Applied Econometrics* **25**: 635–662.

Kapetanios, G., Price, S. & Young, G. (2018). A UK financial conditions index using targeted data reduction: Forecasting and structural identification, *Econometrics and Statistics* (forthcoming).

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

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

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

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.

Kilian, L. (2013). Structural vector autoregressions, *in* N. Hashimzade & M. Thornton (eds), *Handbook of Research Methods and Applications on Empirical Macroeconomics*, Edward Elger, Cheltenham, UK, pp. 515–554.

Kilian, L. & Lütkepohl, H. (2017). *Structural Vector Autoregressive Analysis*, Cambridge University Press, Cambridge.

Kilian, L. & Murphy, D. P. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models, *Journal of the European Economic Association* **10(5)**: 1166–1188.

Kolsrud, D. (2007). Time-simultaneous prediction band for a time series, *Journal of Forecasting* **26(3)**: 171–188.

Kolsrud, D. (2015). A time-simultaneous prediction box for a multivariate time series, *Journal of Forecasting* **34(8)**: 675–693.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

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

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

Lütkepohl, H., Staszewska-Bystrova, A. & Winker, P. (2017). Calculating joint confidence bands for impulse response functions using highest density regions, *Empirical Economics* (forthcoming).

Montiel Olea, J. & Plagborg-Møller, M. (2017). Simultaneous confidence bands: Theoretical comparisons and recommendations for practice, *Unpublished manuscript*, Harvard University.

Montiel Olea, J. & Plagborg-Møller, M. (2018). Simultaneous confidence bands: Theorym implementation, and an application to svars, *Journal of Applied Econometrics* (forthcoming).

Nicholls, D. F. & Pope, A. L. (1988). Bias in estimation of multivariate autoregression, *Australian Journal of Statistics* **30A**: 296–309.

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

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

Royen, T. (2014). A simple proof of the Gaussian correlation conjecture extended to some multivariate Gamma distributions, *Far East Journal of Theoretical Statistics* **48**: 139–145.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance, *Biometrika* **49**: 87–104.

Schüssler, R. & Trede, M. (2016). Constructing minimum-width confidence bands, *Economics Letters* **145**: 182–185.

Sims, C. A. (1980). Macroeconomics and reality, *Econometrica* **48**: 1–48.

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

Staszewska, A. (2007). Representing uncertainty about impulse response paths: The use of heuristic optimization methods, *Computational Statistics & Data Analysis* **52**: 121–132.

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

Staszewska-Bystrova, A. (2013). Modified Scheffé's prediction bands, *Journal of Economics and Statistics* **233(5-6)**: 680–690.

Staszewska-Bystrova, A. & Winker, P. (2013). Constructing narrowest pathwise bootstrap prediction bands using threshold accepting, *International Journal of Forecasting* **29**: 221–233.

Staszewska-Bystrova, A. & Winker, P. (2014). Measuring forecats uncertainty of corporate bond spreads by Bonferroni-type prediction bands, *Central European Journal of Economic Modeling and Econometrics* **6**(2): 89–104.

Stock, J. & Watson, M. (2005). Understanding changes in international business cycle dynamics, *Journal of the European Economic Association* **3**(5): 968–1006.

Šidák, Z. (1967). Rectangular confidence regions for the means of multivariate normal distributions, *Journal of the American Statistical Association* **62**: 626–633.

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