ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Lütkepohl, Helmut; Staszewska-Bystrova, Anna; Winker, Peter

Working Paper

Confidence bands for impulse responses: Bonferroni versus Wald

DIW Discussion Papers, No. 1354

Provided in Cooperation with:
German Institute for Economic Research (DIW Berlin)

Suggested Citation: Lütkepohl, Helmut; Staszewska-Bystrova, Anna; Winker, Peter (2014) : Confidence bands for impulse responses: Bonferroni versus Wald, DIW Discussion Papers, No. 1354, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:
https://hdl.handle.net/10419/90899

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

DIW BERLIN

1354

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2014

Confidence Bands for Impulse Responses

Bonferroni versus Wald

Helmut Lütkepohl, Anna Staszewska-Bystrova, Peter Winker

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2014

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 print edition 1433-0210
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

Confidence Bands for Impulse Responses: Bonferroni versus Wald[1]

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

January 9, 2014

Abstract. In impulse response analysis estimation uncertainty is typically displayed by constructing bands around estimated impulse response functions. These bands may be based on frequentist or Bayesian methods. If they are based on the joint distribution in the Bayesian framework or the joint asymptotic distribution possibly constructed with bootstrap methods in the frequentist framework often individual confidence intervals or credibility sets are simply connected to obtain the bands. Such bands are known to be too narrow and have a joint confidence content lower than the desired one. If instead the joint distribution of the impulse response coefficients is taken into account and mapped into the band it is shown that such a band is typically rather conservative. It is argued that a smaller band can often be obtained by using the Bonferroni method. While these considerations are equally important for constructing forecast bands, we focus on the case of impulse responses in this study.

Key Words: Impulse responses, Bayesian error bands, frequentist confidence bands, Wald statistic, vector autoregressive process

JEL classification: C32

[1]The paper was presented at the Macromodels International Conference in Warsaw, 21-23 October 2013. Helpful comments by the participants and in particular Karim Abadir are gratefully acknowledged. We thank Harald Uhlig for making the data available for the U.S. monetary policy example. This paper was written while the first author was a Bundesbank Professor at the Freie Universität Berlin. Financial support was provided by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk", the National Science Center, Poland (NCN) through 2013/08/A/HS4/00612, and by a MNiSW/DAAD PPP grant (56268818).

1 Introduction

The problem of constructing frequentist confidence bands or Bayesian credible sets around impulse responses of structural VAR processes is discussed extensively in the literature (see, e.g., Sims and Zha (1999), Staszewska (2007), Jordà (2009), and Lütkepohl, Staszewska-Bystrova and Winker (2013)). A recent Bayesian proposal is due to Inoue and Kilian (2013). They suggest constructing a credible set with a preassigned probability, say $1-\gamma$, based on the posterior of the VAR coefficients and map that into a credible set for the impulse responses by considering all impulse response functions corresponding to the parameter values within the credible set for the VAR parameters.

It is pointed out in a number of studies that simply connecting individual confidence intervals with the desired confidence level does not result in a band with a preassigned confidence level but will result in a band that is too narrow and contains the true impulse response function with probability less than the desired confidence level in a frequentist framework or likewise a Bayesian error band with posterior probability less than the posterior probability of the individual credible sets. Thus, the problem arises how to construct bands containing the true impulse response function with a preassigned probability.

In a frequentist framework one could construct an asymptotically valid confidence set for the estimated VAR parameters for a given confidence level $1-\gamma$ based on the Wald statistic. One could then consider the band that includes all impulse responses associated with VAR parameters within the Wald confidence set. Such a strategy leads in fact to conservative error bands for the impulse responses because the latter are constructed by considering the area between the minimum and the maximum of the impulse responses in the confidence set for each propagation horizon. To better understand this procedure let us consider just two impulse response coefficients jointly. Constructing the confidence band for the two impulse responses amounts to considering all impulse responses in a box that contains all points corresponding to VAR parameters in the Wald confidence set. Since impulse responses are nonlinear functions of the VAR parameters, the image of the VAR parameters will not be a box in the impulse response space but some other subset in the plane. Hence, the box might contain also other values than those in the confidence set corresponding to the Wald confidence set for the VAR parameters and, consequently, it is a conservative set. We show that such a confidence box may even have considerably more probability content than a set constructed according to the standard Bonferroni principle.

The same features arise in Bayesian estimation which is nowadays often used for VAR analysis. In a proper Bayesian analysis the joint posterior distribution of the impulse responses of interest can be constructed. This distribution can then be used to derive a credible set with preassigned probability $1-\gamma$. Then the question arises, however, how to map that set into an error band for the impulse responses that can be plotted in the usual way. The standard method appears to correspond to drawing a box around the credible set. Strictly speaking the Bayesian set may not even be dense in a higher dimensional Euclidean space (see Inoue and Kilian (2013)). However,

even then the interpretation of the impulse responses is often linked to a band that covers the full set of impulse response functions between the boundaries of the credible set. In that case, using arguments based on the Bonferroni inequality may result in smaller bands and, hence, more precise inference.

In this study we focus on a frequentist setting because it simplifies the discussion from our point of view and allows for a more rigorous evaluation of some of our methods. For Bayesians it should be apparent how to use analogous arguments and adjustments to their procedures to improve their inference.

In the following we first formally compare Wald and Bonferroni confidence bands for impulse responses. We demonstrate that in a standard setting a frequentist confidence band (or a Bayesian error band) around impulse responses constructed with Bonferroni's method may be smaller than the corresponding set obtained by drawing a box around the confidence region or credible set of parameter estimates obtained via the Wald statistics and projected into the impulse response space. While both methods are conservative, we show that Bonferroni typically results in smaller error bands than using the Wald statistic. We conclude that studies constructing error bands based on the Wald statistic should be reconsidered. We also point out adjustments to both Bonferroni and Wald bands that result in more precise and at the same time smaller bands than the unadjusted, conservative methods. Since we have to rely on asymptotic arguments in a frequentist setting, a small sample simulation experiment is carried out to study the small sample implications of our asymptotic results. We also illustrate the methods by a structural VAR analysis of the market for crude oil from Kilian (2009) and a monetary economic system from Uhlig (2005). The examples show that the method for constructing impulse responses is important in practice because alternative methods may lead to different conclusions.

The problem of constructing confidence bands also occurs in forecasting where often a sequence of different forecast horizons is of interest. Constructing confidence ellipsoids based on the Wald statistic has been discussed in the literature for joint forecast regions (see, e.g., Lütkepohl (2005, Sec. 2.2.3), Kim (1999), Kim (2004), Grigoletto (2005)). In this literature the ellipsoids are obtained for the forecasts, that is, the objects of direct interest. Even then, if a plot of a forecast band is needed the same problem occurs and the analysis of the present study becomes relevant in that context as well. We leave a detailed analysis of the specific issues related to forecasting to future work.

The structure of this study is as follows. In the next section the Wald and Bonferroni bands are first reviewed and compared in a general but idealized setting based on a normal distribution assumption for the parameter estimators. The advantage of using such an idealized setting is that precise results can be obtained. Then the bands are placed in a more realistic setting where only asymptotic normality is obtained in a frequentist framework. Impulse response analysis for structural vector autoregressive (VAR) models is considered as a specific area where the results are relevant. In Section 3 a

Monte Carlo investigation of error bands for impulse responses is presented and examples are considered in Section 4. Section 5 concludes.

2 Confidence Bands for Functions of Parameters

For simplicity we present the procedure in an idealized setting first and then present the more realistic version that is relevant in practice.

2.1 The Idealized General Procedures

Consider a $(G \times 1)$ parameter vector θ and a normally distributed estimator $\hat{\theta}$ such that

$$\hat{\theta} \sim \mathcal{N}(\theta, \Sigma_\theta), \tag{2.1}$$

where Σ_θ is a nonsingular $(G \times G)$ covariance matrix that we assume to be known for the moment. Suppose we are interested in a $(M \times 1)$ vector $\phi(\theta) = (\phi_1(\theta), \ldots, \phi_M(\theta))'$ that is a function of θ and we would like to construct a $1-\gamma$ confidence or error band for $\phi(\theta)$. Then, the following procedures can be used.

For the first approach, let $\chi^2(G)_{1-\gamma}$ be the $1-\gamma$ quantile of a χ^2 distribution with G degrees of freedom. Then, the set

$$\mathcal{W}^\theta_{1-\gamma} = \{\theta | W = (\hat{\theta} - \theta)' \Sigma_\theta^{-1} (\hat{\theta} - \theta) \leq \chi^2(G)_{1-\gamma}\}. \tag{2.2}$$

is a $1-\gamma$ confidence set for θ based on the Wald statistic W. It contains the true parameter vector with probability $1-\gamma$ in a repeated sampling sense. Hence, the set

$$\mathcal{W}^{\phi(\theta)}_{1-\gamma} = \{\phi(\theta) | \theta \in \mathcal{W}^\theta_{1-\gamma}\} \tag{2.3}$$

is an exact $1-\gamma$ confidence set for $\phi(\theta)$ if the function is one-to-one or it has at least $1-\gamma$ confidence level if the function $\phi(\cdot)$ is not one-to-one.

If $\phi(\theta)$ is more than 3-dimensional, plotting the confidence set $\mathcal{W}^{\phi(\theta)}_{1-\gamma}$ is not practical and usually one plots a band around the elements of $\phi(\theta)$. In other words, instead of $\mathcal{W}^{\phi(\theta)}_{1-\gamma}$ we choose

$$l_m = \min\{\phi_m(\theta) | \theta \in \mathcal{W}^\theta_{1-\gamma}\}$$

and

$$u_m = \max\{\phi_m(\theta) | \theta \in \mathcal{W}^\theta_{1-\gamma}\}$$

for $m = 1, \ldots, M$, and use

$$\mathcal{W}^{band}_{1-\gamma} = [l_1, u_1] \times \cdots \times [l_M, u_M] \tag{2.4}$$

as an error band. Clearly,

$$\Pr[\phi(\theta) \in \mathcal{W}_{1-\gamma}^{band}] \geq \Pr[\phi(\theta) \in \mathcal{W}_{1-\gamma}^{\phi(\theta)}] \geq 1-\gamma.$$

Hence, the Wald error band $\mathcal{W}_{1-\gamma}^{band}$ is generally conservative.

Alternatively, we may construct an error band using Bonferroni's principle. In that case the band is constructed for the individual elements of $\phi(\theta)$ by ignoring any information on a possible dependence between the elements. The approach is to construct a $1-\gamma/M$ confidence interval for each element. Thus, the Bonferroni band is

$$\mathcal{B}_{1-\gamma} = [L_1, U_1] \times \cdots \times [L_M, U_M], \tag{2.5}$$

where, for $m = 1, \ldots, M$, L_m and U_m are the $\gamma/2M$ and $1-\gamma/2M$ quantiles of the distribution of $\phi_m(\hat{\theta})$, respectively. It is well-known that

$$\Pr[\phi(\theta) \in \mathcal{B}_{1-\gamma}] \geq 1-\gamma.$$

Hence, the band is also conservative. It may be worth emphasizing that equality does not even hold and, hence, the band is conservative, if the elements of $\phi(\hat{\theta})$ are stochastically independent.

Because the Bonferroni error band is based on the marginal distributions of the components of $\phi(\hat{\theta})$ only and ignores any stochastic dependence, the band can in fact be rather conservative. This problem was pointed out by a number of authors in the past (e.g., Lütkepohl et al. (2013)). Since the Wald statistic fully utilizes the joint dependence, it appears useful to compare the Wald and the Bonferroni bands. Before we do so systematically in more generality, it is instructive to focus on a very special case.

Consider a bivariate vector θ and the identity transformation, that is, $\phi(\theta) = \theta$. Moreover, suppose that the variances are unity so that

$$\Sigma_\theta = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

In that case the Wald statistic in (2.2) becomes

$$W = [(\hat{\theta}_1 - \theta_1)^2 + 2\rho(\hat{\theta}_1 - \theta_1)(\hat{\theta}_2 - \theta_2) + (\hat{\theta}_2 - \theta_2)^2]/(1-\rho^2)$$

and the corresponding confidence set $\mathcal{W}_{1-\gamma}^{\phi(\theta)}$ is an ellipse that depends on the correlation ρ and the confidence level $1-\gamma$. Of course, a circle is obtained for $\rho = 0$. The corresponding Wald confidence bands $\mathcal{W}_{1-\gamma}^{band}$ are obtained as squares or boxes that fully cover the ellipses.

In contrast, the Bonferroni confidence bands are

$$\mathcal{B}_{1-\gamma} = [\hat{\theta}_1 - c_{1-\gamma/4}, \hat{\theta}_1 + c_{1-\gamma/4}] \times [\hat{\theta}_2 - c_{1-\gamma/4}, \hat{\theta}_2 + c_{1-\gamma/4}],$$

where c_η is the $\eta \times 100\%$ quantile of a standard normal distribution. In this case $\mathcal{W}_{1-\gamma}^{band}$ is larger than $\mathcal{B}_{1-\gamma}$. Thus, the Bonferroni confidence band can well be smaller than the band based on the Wald statistic.

Table 1: Relative Widths of Bonferroni and Wald Confidence Boxes for Independent iid $\mathcal{N}(0,1)$ Variables

	Confidence level			
	$1-\gamma = 0.68$		$1-\gamma = 0.90$	
M	$\mathcal{B}_{1-\gamma}$	$\mathcal{W}^{band}_{1-\gamma}$	$\mathcal{B}_{1-\gamma}$	$\mathcal{W}^{band}_{1-\gamma}$
2	1.0369	1.1140	1.0057	1.1012
3	1.0395	1.2064	1.0066	1.1827
4	1.0387	1.2857	1.0068	1.2528
5	1.0375	1.3561	1.0068	1.3153
6	1.0363	1.4200	1.0067	1.3720
7	1.0352	1.4787	1.0066	1.4243
8	1.0343	1.5333	1.0065	1.4731
9	1.0334	1.5845	1.0064	1.5188
10	1.0326	1.6329	1.0064	1.5621
15	1.0298	1.8433	1.0060	1.7513
20	1.0279	2.0188	1.0057	1.9098
30	1.0255	2.3091	1.0054	2.1732
40	1.0239	2.5501	1.0051	2.3929
50	1.0228	2.7598	1.0050	2.5846

To assess the magnitude of the differences between the Wald and Bonferroni boxes, let us consider the case of independent components of the M-dimensional vector $\hat{\theta}$ with covariance matrix I_M. For this case we can determine a precise confidence box with confidence level exactly $1-\gamma$. Such a precise box is obtained if individual confidence intervals with confidence level $1-\gamma_{ind}$ are chosen with $\gamma_{ind} = 1-(1-\gamma)^{1/M}$. For this choice we get $(1-\gamma_{ind})^M = 1-\gamma$. A precise confidence box has width $2 \times c_{1-\gamma_{ind}/2}$, while the width of the Bonferroni box is $2 \times c_{1-\gamma/2M}$ and the width of the $\mathcal{W}^{band}_{1-\gamma}$ box is $2 \times \sqrt{\chi^2(M)_{1-\gamma}}$.

Examples for different M and confidence levels are presented in Table 1. The numbers in the table are the widths of $\mathcal{B}_{1-\gamma}$ and $\mathcal{W}^{band}_{1-\gamma}$ relative to the width of the perfect box that has exact confidence level $1-\gamma$. For this example the widths of the Bonferroni boxes are considerably smaller than those of the Wald boxes. For both confidence levels, $1-\gamma = 0.68$ or 0.90 the Bonferroni boxes are about as wide as the perfect boxes. Note that 68% error bands are not uncommon in the VAR literature which is why we have included that level in the table. In fact, since both γ_{ind} and γ/M converge to zero for $M \to \infty$, the Bonferroni and the perfect boxes become indistinguishable for large M. In contrast, the Wald boxes remain rather large and grow in width relative to the perfect box when $M \to \infty$.

Of course, this example deals with a very special case that is unrealistic in various respects. It illustrates the principle, however, and it clearly shows that the Bonferroni boxes can be very close to the perfect box with exact coverage. In fact, when the widths of the boxes are very similar, also the coverage levels are very close together. In other words, the Bonferroni boxes

may have almost exact coverage although they do not use any information on the dependence of the underlying estimators while the perfect box is constructed under the assumption that independence of the components is known. On the other hand, the example also shows that the confidence boxes constructed from the Wald statistics can be much wider than the Bonferroni boxes.

In practice the situation is more complex in different dimensions. In particular, the dependence structure of the impulse responses is not known precisely. We look into that problem in the context of impulse response analysis in the following.

2.2 In Practice

In frequentist statistics, the distribution of the estimator $\hat{\theta}$ in (2.1) is typically not known. Instead an asymptotically normally distributed estimator $\hat{\theta}$ based on a sample of size T, that is,

$$\sqrt{T}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \Sigma_\theta), \tag{2.6}$$

may be used, where Σ_θ is an unknown nonsingular $(G \times G)$ covariance matrix. The Wald statistic, $T(\hat{\theta} - \theta)' \Sigma_\theta^{-1} (\hat{\theta} - \theta)$, based on such an asymptotic result often has a distribution that is not well approximated by a $\chi^2(G)$ in small samples. To obtain a better small sample approximation to the distribution of the Wald statistic, the following bootstrap procedure can be used.

Bootstrap the estimator $\hat{\theta}$ so as to get bootstrap estimates $\hat{\theta}_n$ and $\hat{\Sigma}_\theta(n)$, $n = 1, \ldots, N$, and define

$$w_n = T(\hat{\theta}_n - \hat{\theta})' \hat{\Sigma}_\theta(n)^{-1} (\hat{\theta}_n - \hat{\theta}). \tag{2.7}$$

Order the bootstrap estimates such that $w_1 \leq \cdots \leq w_N$ and choose the Wald confidence set such that

$$\mathcal{W}_{1-\gamma}^\theta = \{\theta | W = T(\hat{\theta} - \theta)' \hat{\Sigma}_\theta^{-1} (\hat{\theta} - \theta) \leq w_{(1-\gamma)N}\}. \tag{2.8}$$

Clearly, as T and N go to infinity,

$$\Pr[T(\hat{\theta} - \theta)' \hat{\Sigma}_\theta^{-1} (\hat{\theta} - \theta) \leq w_{(1-\gamma)N}] \to 1 - \gamma,$$

if the usual conditions underlying the bootstrap are satisfied. Hence, asymptotically

$$\Pr[\theta \in \mathcal{W}_{1-\gamma}^\theta] \to 1 - \gamma,$$

so that $\mathcal{W}_{1-\gamma}^\theta$ is an asymptotically valid confidence set. The corresponding confidence band can be obtained by determining

$$l_m = \min\{\phi_m(\hat{\theta}_n) | n = 1, \ldots, (1-\gamma)N\}$$

and

$$u_m = \max\{\phi_m(\hat{\theta}_n) | n = 1, \ldots, (1-\gamma)N\},$$

for $m = 1, \ldots, M$, and setting

$$\mathcal{W}_{1-\gamma}^{band} = [l_1, u_1] \times \cdots \times [l_M, u_M].$$

Finally, the corresponding Bonferroni band is obtained as in (2.5), where, however, for $m = 1, \ldots, M$, L_m and U_m are now the $\gamma/2M$ and $1 - \gamma/2M$ bootstrap quantiles, respectively, of the empirical bootstrap distribution of $\phi(\hat{\theta}_n)$.

In the following subsection these general considerations are applied in constructing confidence bands for impulse responses of structural VAR models.

2.3 Confidence Bands for VAR Impulse Responses

Consider the reduced form VAR(p) model for $y_t = (y_{1t}, \ldots, y_{Kt})'$ in lag operator notation

$$A(L)y_t = \nu + u_t \tag{2.9}$$

with $A(L) = I_K - A_1 L - \cdots - A_p L^p$. The A_i $(i = 1, \ldots, p)$ are $(K \times K)$ parameter matrices and $u_t \sim (0, \Sigma_u)$ is a K-dimensional zero mean white noise process with covariance matrix $E(u_t u_t') = \Sigma_u$. Other deterministic terms than the K-dimensional intercept vector ν are possible but not essential for the following discussion. Stability and stationarity of the process is ensured if

$$\det A(z) = \det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \text{ for } z \in \mathbb{C}, |z| \leq 1. \tag{2.10}$$

If this condition is satisfied, the process has the moving average (MA) representation

$$y_t = A(1)^{-1}\nu + A(L)^{-1}u_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \tag{2.11}$$

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

Let $\varepsilon_t = \mathbf{B}^{-1}u_t$ be a set of structural shocks, obtained by a linear transformation of u_t. For example, $\mathbf{B}$ may be a lower-triangular matrix obtained by a Cholesky decomposition of Σ_u. Replacing u_t in (2.11) by $\mathbf{B}\varepsilon_t$ shows that the responses to structural shocks are obtained as $\Theta_i = \Phi_i \mathbf{B}$, $i = 0, 1, \ldots$. Thus, the impulse responses of the kth variable to the ith shock are functions of the VAR parameters and the framework of the previous subsection can be used to construct confidence bands for the impulse responses.

In the following we investigate the relative performance of Wald and Bonferroni confidence bands in small samples. The main issue of interest in this context is to check whether the asymptotic properties are also seen in small sample situations. In other words, we are interested in exploring which of the methods is more conservative in specific situations that are relevant in applied work.

Of course, we are not only interested in the coverage of the two types of bands but also in their widths. Our precise measure for the width of a confidence band is specified in the next section. At this point it may be worth emphasizing that a confidence band being conservative is not a bad property per se. Only if that also leads to excessively wide bands this may be problematic because it can cover up interesting features in the data generation mechanism. For example, the response of a variable to a shock may appear insignificantly different from zero although in practice the variable reacts quite strongly to the shock. In the literature modifications to the Bonferroni band have been proposed to alleviate this problem (see Lütkepohl et al. (2013)). Similar modifications are also possible for the Wald bands and are considered in the Monte Carlo section.

For the evaluation of the bands it is of central importance to be clear about the functions of interest, $\phi(\theta)$. If the responses of all K variables to all K shocks for propagation horizon H are of interest, that is, all elements of $[\Theta_0, \Theta_1, \ldots, \Theta_H]$ are of interest, then, $\phi(\theta)$ has dimension $M = K^2H + K(K+1)/2$ if a triangular initial effects matrix Θ_0 is considered. Thus, even for a VAR of moderate dimension the number of individual confidence intervals considered in the Bonferroni approach becomes large. For example, if there are only $K = 3$ variables and the propagation horizon is $H = 12$ (e.g., the responses are traced for one year for monthly data), the total number of impulse responses of interest is $M = 114$. Clearly that is a problem for the Bonferroni approach. For example, if we choose a confidence level of 68%, that is, $\gamma = 0.32$ one would have to choose individual confidence intervals with level $1 - \gamma/M = 0.9972$ for the aforementioned example of a three-dimensional VAR process. Clearly, even if only a rather small overall confidence level is desired, the individual confidence level of the Bonferroni procedure is very close to one in this example which is not unrealistic for empirical analysis.

Of course, it is also possible that only the responses to specific shocks are of interest. For instance, in monetary analysis the responses of some of the variables to monetary shocks may be of interest only. That reduces the dimension of $\phi(\theta)$ substantially and such reductions should be taken into account in the Bonferroni procedure.

Another feature worth taking into account in impulse response analysis is that there is a one-to-one mapping between the VAR parameters (including the white noise covariance matrix) and the impulse response matrices up to propagation horizon $H = p$. More precisely, it follows easily from the derivations in Lütkepohl (1988) that the mapping

$$[A_1, \ldots, A_p, \Sigma_u] \to [\Theta_0, \Theta_1, \ldots, \Theta_p]$$

is one-to-one onto if triangularity of Θ_0 is taken into account. That result implies that all Θ_i with $i > p$ are completely determined by $[\Theta_0, \Theta_1, \ldots, \Theta_p]$. Hence, for constructing Bonferroni confidence bands, we can restrict attention to propagation horizons of at most up to the VAR lag order. In other words, the maximum M to be considered is $M = K^2p + K(K+1)/2$ if all

responses to all shocks are of interest. Of course, for moderate dimensional VAR systems even that number can be quite substantial, in particular, since choosing relatively large VAR orders is not uncommon.

3 Monte Carlo Comparison with other Methods

3.1 Monte Carlo Design

We follow Lütkepohl et al. (2013) and use a bivariate DGP from Kilian (1998) for our Monte Carlo comparison of different methods for constructing confidence bands for impulse responses. The DGP is a VAR(1),

$$y_t = \begin{bmatrix} \alpha & 0 \\ 0.5 & 0.5 \end{bmatrix} y_{t-1} + u_t, \quad u_t \sim \text{iid } \mathcal{N}\left(0, \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}\right), \tag{3.1}$$

with $\alpha \in \{-0.95, -0.9, -0.5, 0, 0.5, 0.9, 0.95, 1\}$. Thus, most of the processes are stationary ($|\alpha| < 1$) but some are quite persistent with $|\alpha| = 0.95$ and one process even has a unit root ($\alpha = 1$).

Sample sizes $T = 50, 100$ or 200 are used and the number of bootstrap replications is $N = 2000$, $\gamma = 0.1$ and 2000 Monte Carlo replications are used. Responses to orthogonal innovations, where $\mathbf{B}$ is a lower-triangular matrix obtained from the Cholesky decomposition of the estimated residual covariance matrix, are investigated. We have run experiments where the true lag order is used and we have also used the popular model selection criterion AIC for lag order selection. The maximum lag length used with lag order selection depends on the sample size and is given by 10, 12, and 14 for samples of 50, 100, and 200 observations, respectively.

3.2 Methods for Constructing Confidence Bands

In addition to the basic Wald and Bonferroni bands we also include modified versions in the comparison that are inspired by adjustments considered by Lütkepohl et al. (2013). They are meant to account for the conservativeness of the Bonferroni and Wald bands. In these methods the residual based bootstrap procedure is implemented as in the paper by Lütkepohl et al. (2013). Overall we consider the following alternative bands.

Bonferroni (B) The abbreviation B is used for the basic Bonferroni band constructed for an individual impulse response function. In that case the band is constructed from individual $(1 - \gamma/H) \times 100\%$ or $(1 - \gamma/(H + 1)) \times 100\%$ bootstrap confidence intervals.

Joint Bonferroni (JB) JB stands for the basic Bonferroni bands constructed for all impulse response functions jointly. These are obtained from $(1 - \gamma/(K^2H + K(K + 1)/2)) \times 100\%$ bootstrap confidence intervals.

In other words, the full set of impulse response functions of all K variables to the K shocks is considered jointly and the confidence level is chosen such that the bands include all the true response functions with at least the desired probability $1-\gamma$.

Adjusted Bonferroni (B^a) As mentioned earlier, the Bonferroni bands are in theory conservative bands and in practice usually contain more than $(1-\gamma)N$ bootstrap response functions. Thus, it is possible to further adjust these bands. If a single response function is investigated, an adjusted Bonferroni band is proposed by Lütkepohl et al. (2013). The band is abbreviated as B^a and constructed as follows. The bootstrap response functions fully within the Bonferroni band are counted and their number is denoted by N_B. If $N_B > (1-\gamma)N$, a sequential procedure aimed at removing $N_B - (1-\gamma)N$ bootstrap impulse responses is applied. In each step, the bootstrap impulse responses are identified which provide at least one point on the current bounds. There are at most $2(H+1)$ such functions. The function which contributes the most to the size of the current band (measured as sum of widths) is rejected. The procedure terminates after $N_B-(1-\gamma)N$ functions have been eliminated. The band is obtained as an envelope of the remaining $(1-\gamma)N$ bootstrap impulse responses.

Adjusted Joint Bonferroni (JB^a) A similar adjustment procedure can also be applied in the case all impulse response functions are of interest. The resulting bands are denoted by JB^a. The bands are calculated in the following way. The joint Bonferroni bands are constructed. The number of bootstrap samples providing bootstrap response functions inside all the bands is denoted by N_B. If $N_B > (1-\gamma)N$, a sequential procedure aimed at removing superfluous bootstrap samples is applied. In each step, the bootstrap impulse responses are identified that provide at least one point on the bounds of at least one band. The function which contributes the most to the sum of widths of all the bands is identified and the bootstrap impulse responses corresponding to the given bootstrap sample are rejected from all the bands. The procedure terminates when the joint coverage with respect to the bootstrap samples is equal to $(1-\gamma)N$. The bands are obtained as envelopes of the remaining bootstrap impulse responses.

Reduced Joint Bonferroni (JB_*) As already mentioned in Section 2.3, for a VAR(p) process, the joint confidence bands are completely determined by the first $p+1$ impulse responses. More precisely, the Θ_i for $i > p$ are fully determined by $\Theta_0, \ldots, \Theta_p$. Hence, when joint sets of impulse responses are considered we may focus on the latter set of impulse responses only and construct Bonferroni bands for initial $p+1$ periods accordingly from individual $(1-\gamma/M)\times 100\%$ confidence intervals where $M = K^2p + K(K+1)/2$. The bounds for the remaining periods are given by the envelope of the bootstrap impulse responses

covered up to $h = p$ by these intervals. These bands are referred to as reduced Bonferroni bands in the following and abbreviated as JB_*.

Adjusted Reduced Joint Bonferroni (JB_*^a) An adjustment to the reduced joint Bonferroni bands can be computed exactly as for the joint Bonferroni bands. The resulting bands are denoted by JB_*^a.

Wald (W) The basic Wald band is abbreviated as W in the following. The band is joint by construction and so it does not differ when individual or all impulse responses are considered. W bands usually cover more than $(1-\gamma)N$ bootstrap response functions (both when an individual impulse response function is investigated and also when all impulse responses are considered jointly). Thus, adjustments of these bands can be considered. Alternative adjustments are plausible and, hence, are included in the Monte Carlo comparison.

Adjusted Wald (W^a) The first adjustment consists in rejecting bootstrap functions corresponding to $w_{(1-\gamma)N-1}$, $w_{(1-\gamma)N-2}$, $w_{(1-\gamma)N-3}$, ... until the bootstrap coverage is as close to $(1-\gamma)\times 100\%$ as possible (and at least $(1-\gamma)\times 100\%$). In the case a single impulse response function is of interest, the bands constructed using this adjustment are denoted by W^a.

Joint Adjusted Wald (JW^a) If all impulse responses are considered jointly, a joint adjusted band may be obtained by investigating the coverage of each band and rejecting the bootstrap samples until the bootstrap coverage of one of the bands becomes as close to the desired coverage level as possible. This band is abbreviated as JW^a.

Bonferroni-adjusted Wald (WB^a) Alternatively the Wald band can be adjusted just as the Bonferroni band by eliminating impulse response functions until only $(1-\gamma)\times 100\%$ are left over. The resulting band is abbreviated as WB^a.

Joint Bonferroni-adjusted Wald (JWB^a) Focussing on all K^2 impulse response functions jointly and applying the same reduction method as used for WB^a, results in the joint Bonferroni-adjusted Wald band which we abbreviate as JWB^a. It is considered in the following simulation comparison for completeness.

Lütkepohl et al. (2013) also look at some other frequentist methods for constructing confidence bands for impulse responses. They find that most of these are inferior to the Bonferroni or adjusted Bonferroni bands, while some exhibit a similar performance. Therefore we do not consider them here. There are also other adjustments for the Bonferroni inequality that amount to using different significance levels for different propagation horizons (see, e.g., Holm (1979), Hommel (1988), Hochberg (1988), Benjamini and Hochberg (1995)). Such refinements may be worth exploring in future research.

3.3 Evaluation Criteria

In comparing the Wald and Bonferroni confidence bands for the impulse responses, we use the actual coverage rate and the widths of the bands measured as the sum of the widths of the intervals around the individual impulse response coefficients of interest. As regards the coverage rates it is important to note that if responses to all shocks up to horizon H are of interest and the matrix of initial effects is lower-triangular, the function of interest, $\phi(\theta)$, has dimension $M = 4H + 3$. If only the response of one variable to a specific shock is of interest, $M = H$ or $M = H + 1$, depending on whether the initial effect is estimated or restricted to zero. We consider both types of situations because the theoretical comparison in Section 2 suggests that M is crucial for the relative widths of the bands.

3.4 Monte Carlo Results

We discuss results separately for individual impulse responses (i.e., for the case when the response of a specific variable to a specific shock is of interest, see Section 3.4.1) and for all impulse responses considered jointly (i.e., for the case when responses of all variables to every shock are of interest, see Section 3.4.2).

When individual responses are considered, the coverage probability is estimated as the fraction of Monte Carlo (MC) replications in which the band covers the true response function. The width is calculated as the sum of the widths of the band for $h = 0, 1, \ldots, H$. When all impulse responses are considered jointly, the coverage probability is estimated as the proportion of MC replications in which all the bands cover the true response functions. The width is calculated as the sum of widths of all the bands.

3.4.1 Individual impulse response functions

Selected results for individual impulse response functions are reported in Tables 2 and 3. Both tables report results for sample size $T = 100$ and propagation horizon $H = 10$. They differ in that the true VAR order $p = 1$ is used in Table 2 while the VAR order is chosen with the AIC criterion in Table 3. The coverage rates are reported and the average widths are given in parentheses.

The following main results emerge from Table 2.

1. As expected, both Bonferroni and Wald bands are conservative. For all MC designs reported in the table, except when $\alpha = 1$, the actual relative coverage frequency is larger than 90%. The W bands typically have larger coverage rates than the B bands. Correspondingly, the widths of the B bands are smaller than those of the W bands. For the integrated process ($\alpha = 1$) the coverage rates of the B bands for the responses of the variables to the first shock are slightly below 90%. Apart from that the situation is the same as for the stationary processes.

2. All adjustments lead to smaller coverage rates that are still close to 90% in most cases. The exception is again the case of an integrated, nonstationary process ($\alpha = 1$). In most cases the adjusted bands are considerably smaller than B and W. Thus, the adjustments serve the purpose of bringing the actual coverage closer to the nominal coverage rates and thereby reduce the costs in terms of band width.

3. The B^a and WB^a bands in most cases have very similar coverage rates and band widths. This result is not surprising because both the B and the W bands are in a sense unnecessarily wide and are reduced by the same procedure. So they end up with similar bands. The B^a and WB^a bands have the smallest width and are for that reason attractive from an applied point of view.

In Table 3 it can be seen that the results for known order do not change much if the order is estimated by AIC. The relative ordering of the methods in terms of coverage rates and widths of the bands remains the same. The coverage rates are reduced a little and the widths are slightly larger than in the known order case. Again these results are not surprising given the additional uncertainty due to not knowing the true VAR order. One implication is, however, that the adjusted bands now tend to have coverage rates a bit smaller than 90%. Although they are still typically above 85%, it is of course an issue how much undercoverage one wants to tolerate. For the case of an integrated process the actual coverage rate for the impulse responses to the first shock are in fact quite low and do not even reach 75%. Since in practice processes with unit root properties are not uncommon, the issue may be important, at least when only moderately long time series are available. Note that we are discussing results for samples of size $T = 100$ now which may not even be available for macroeconomic time series.

As mentioned in Section 3.1, we have also considered a wide range of other MC designs. The results are not reported because they are qualitatively very similar to those in the tables for $T = 100$. We just summarize the main deviations here.

1. Coverage rates go down substantially when the sample size is reduced to $T = 50$. Thus, in that case the unadjusted bands B and W become more attractive if a precise coverage rate is a major concern. They still are considerably wider than the adjusted bands, however. If the sample size is increased to $T = 200$ the situation is again very similar to that in Tables 2 and 3. That is, B and W are very conservative and, hence, these bands are unnecessarily wide.

2. If the propagation horizon increases to $H = 20$, the situation is again very similar to that for $H = 10$. By construction, B becomes even more conservative while W is not much affected. The latter band still tends to be more conservative than the Bonferroni band.

Thus, overall the original B and W methods can be recommended when one wants to be on the safe side regarding coverage with B being slightly less

conservative than W. The price to pay in terms of width of the bands around the impulse responses can be high, however.

3.4.2 All impulse responses considered jointly

So far we have considered results for the case that the response function of a single variable to a single shock is of interest. That situation appears to be relevant in many applied studies. It has been argued, however, notably by Inoue and Kilian (2013) that in principle, if all impulse responses are of interest, the error bands should be such that all impulse responses are covered jointly with a preassigned probability. Hence, in this section we discuss results taking this requirement into account.

As discussed in Section 3.2, it is in principle easy to adjust our methods so as to account for the fact that impulse responses of all variables to all shocks are of interest jointly. In particular, the Bonferroni bands just have to be widened in a suitable way to ensure that a larger set of impulse responses falls within the confidence regions constructed in this way. Some coverage and width results for the bands constructed for the overall set of impulse responses are shown in Tables 4 and 5. The former table presents results for known VAR order while the order is estimated by AIC in the latter table. Since the coverage evaluation criterion now refers to all impulse responses simultaneously, there is no need to distinguish between different variables and impulse responses. Hence, only one coverage rate is reported for each MC design. Results for three different sample sizes are presented ($T = 50, 100, 200$) but the propagation horizon is chosen to be $H = 10$.

The following results emerge from Table 4.

1. The actual coverage rates of the JB and W regions are remarkably close to the nominal 90% for all stationary processes ($\alpha \neq 1$) even for sample size $T = 50$. The coverage is excellent for $T = 100$ and 200 even for the unit root process. Now the joint Bonferroni and the Wald bands do not differ much in coverage but often JB is slightly wider than W. Clearly this result is due to the fact that a large number of estimated impulse responses is considered jointly and the Bonferroni band has to be expanded accordingly. Thus, comparing only JB and W, the latter has a slight advantage in both coverage precision and band width.

2. We have argued in Section 2.3 that adjusting the Bonferroni region by simply counting all estimated impulse response coefficients ignores the fact that they depend on a smaller number of estimated VAR parameters. The JB_* bands take this fact into account. For the presently considered simulation setup with propagation horizon $H = 10$ it can be seen in Table 4, however, that JB_* bands do not have an advantage over JB. The gains in band width are very moderate if they exist at all and the coverage rates of JB_* are usually considerably lower than those of JB.

3. The adjustments of the JB and W methods work in the expected way.

In other words, they reduce the coverage but also the widths of the bands. In fact, their coverage precision is quite remarkable for processes that are not very persistent if the sample size is at least $T = 100$. While the coverage is only reduced slightly, the width is often reduced substantially. For example, for $\alpha = 0.9$ and $T = 200$, JB has a coverage of 94.8 and a band width of 22.55 while JB^a has a coverage of 86.9 and a width of 16.9. Thus, the width is reduced by roughly one fourth whereas the coverage of both methods is close to the nominal 90%.

4. JB^a and JWB^a are usually very similar in terms of coverage and width. Thus, this result carries over from the case when only single impulse response functions are considered.

Comparing these results to those in Table 5 where the VAR order is estimated by AIC, it becomes clear that the additional uncertainty in VAR order estimation tends to reduce the coverage levels. This is especially true for small sample sizes of $T = 50$ observations only. Even for $T = 100$ the coverage rates are often substantially lower than in the known order case. However, for a sample size of $T = 200$ the bands are again remarkably precise in terms of coverage. Gains in terms of width can be obtained by considering the adjustments. However, if the coverage level is viewed as the dominant criterion, using the JB or W bands is the recommended strategy. They both lead to very similar coverage levels for $T = 100$ and 200, while W has a slight advantage for $T = 50$.

We have also considered longer propagation horizons of the shocks and do not report the detailed results because they are similar to the results reported in the tables. A main difference is that the JB bands are even more conservative for $H = 20$ than for $H = 10$. This increases the precision of the coverage level when it is below 90% in the tables but also increases the widths of the bands. For longer propagation horizons it can pay to utilize the relation between the VAR parameters and the impulse responses and use JB$_*$ instead of JB. For $H = 20$ the widths of the JB$_*$ bands are typically slightly smaller than those of the JB bands and for large sample sizes the coverage rates are also satisfactory. Overall the gains from using JB$_*$ are limited, however. Therefore we do not show the corresponding results in detail.

Thus, as an overall summary of our MC experiment we can conclude that the Bonferroni and Wald methods both result in coverage levels that are close to or above the nominal level if the sample sizes are not very small and the processes are not very persistent. Whether individual impulse response functions or the whole set of all impulse responses is considered jointly, the methods are in fact remarkably precise as far as coverage is concerned. The price to pay is of course a rather large band width. The adjustments proposed address this problem by reducing the width sometimes substantially. In more difficult circumstances (small sample size, large propagation horizon, estimated VAR order) they may however lead to undesirably low coverage rates. Hence, in practice the original B and W methods may be the preferred choice to be on the safe side.

4 Examples

In this section two examples are presented. The first one concerns a model for the market for crude oil from Kilian (2009) and the second example investigates the effects of monetary policy shocks in the U.S. (see, e.g., Uhlig (2005)). Both examples are also considered by Inoue and Kilian (2013) to illustrate their methods for constructing Bayesian error bands for impulse responses. The prime objective of our analysis is to compare how different approaches for constructing frequentist confidence bands for the impulse responses affect the conclusions drawn from the analysis.

4.1 Oil Market

We begin with a model for the market of crude oil from Kilian (2009). It is a three-dimensional system of the percent change in global crude oil production, $\Delta prod$, an index of real economic activity, rea, and the real price of oil, rpo. We use monthly observations from Kilian (2009) for the period 1973M1-2007M12. Hence, the sample size is $T = 420$. The system was also used by Lütkepohl et al. (2013) to illustrate different types of frequentist confidence bands around impulse responses.

Kilian uses a VAR(24) with intercept and so do we. Following Kilian we use a response horizon of $H = 18$ and a recursive identification scheme with lower-triangular $\mathbf{B}$ matrix obtained from a Cholesky decomposition of the reduced form residual covariance matrix. Kilian (2009) constructs pointwise confidence intervals around his impulse responses with a bootstrap method that differs from ours. Moreover, he reports one- and two-standard error bounds around his impulse responses which roughly correspond to 68 and 95% confidence bounds in a normal distribution setting. For better comparability with our confidence bands we have computed 90% confidence intervals for the individual impulse response coefficients with our residual based bootstrap method and show the resulting impulse responses and bands obtained by simply connecting the individual intervals in Figure 1. The zero line, corresponding to no response of a particular variable to a specific shock is obviously not fully covered by some of the bands. Such a result is commonly interpreted as evidence for a significant response of the variable. Specifically we draw attention to the response of the real price of oil to an aggregate demand shock and the reaction of real activity to an oil-market specific demand shock. Both responses would be classified as significant based on the bands in Figure 1.

For comparison the 90% confidence bands calculated for each impulse response function separately by the B, B^a, W, and W^a methods are shown in Figure 2. Furthermore, joint 90% bands based on JB, JB^a, JB_*, and JW^a are depicted in Figure 3. In both figures the Wald bands are overall the widest which is well in line with our simulation results. Obviously, in some cases the other bands are considerably smaller than the W bands, see for example the response of rea to an oil-market specific demand shock. Of course, our simulation results raise the question which of the bands is more reliable in

the present case. The sample size is quite large but also the VAR order is substantial. Thus, to be on the safe side, one may want to rely on the B or W bands if the focus is on individual impulse response functions or on JB and W if the whole set of impulse responses is considered jointly.

Now looking at the response of the real price of oil to an aggregate demand shock it is seen in Figures 2 and 3 that the zero line is fully inside the Wald bands and just barely outside the adjusted Bonferroni and Wald bands. Thus, relying on a proper confidence band, the evidence for a positive response of the real price of oil is much less convincing than in Kilian's study. Similarly, the significant reaction of real activity to an oil-market specific demand shock seen in Figure 1 has disappeared in Figures 2 and 3. Now the zero line is even covered by the adjusted bands. Thus, a reaction of real activity to an oil-market specific demand shock has much less support when the full uncertainty in the response function is taken into account. Clearly, constructing full bands as in our study rather than considering individual confidence intervals makes a difference for the interpretation of the results.

4.2 U.S. Monetary Policy Analysis

For analysing the effects of monetary policy, we use the monthly U.S. benchmark dataset from Uhlig (2005) for a system consisting of the six variables real GDP, GDP deflator, commodity prices, federal funds rate, nonborrowed reserves and total reserves. The sample period is 1965M1 - 2003M12. All series apart from the federal funds rate are in logs. Similar systems for shorter sample periods have also been analyzed by Bernanke and Mihov (1998b, 1998a). As in Uhlig (2005) we use a VAR order of $p = 12$. While Uhlig does not use any deterministic terms, we include an intercept. Also we use a recursive identification as in Bernanke and Mihov (1998b) that contrasts with the sign identification approach used by Uhlig. Note that Uhlig also considers a recursive identification scheme for comparison purposes and we will compare our results to that benchmark system.

Interest centers primarily on the responses of the variables to a monetary policy shock which is the fourth shock in our vector of structural shocks. One could even argue that we are mainly interested in the GDP and price level responses to monetary policy shocks. The total number of responses considered together makes a difference for the confidence bands. Therefore, we present two types of bands for response functions: first, bands constructed for individual response functions to a monetary policy shock, and second, bands constructed for the whole set of 36 response functions considered jointly.

The estimated impulse-responses (for the monetary policy shock) and the 68% confidence bands calculated for each function separately (B, B^a, W, W^a) are shown in Figure 4. Joint 68% bands (JB, JB^a, JB_*, W, JW^a) are depicted in Figure 5. A confidence level of 68% is used now because that level is quite common in the related literature and it is also used by Uhlig in his benchmark analysis. Again the Wald band is an outer envelope for most of the impulse response functions if individual bands are used. In some cases the other bands are much more narrow. If joint bands are used, the

unadjusted bands JB, JB_* and W are usually quite similar and form an outer envelope for the other bands (see Figure 5).

In the upper left-hand panels of Figures 4 and 5 the response of real GDP is seen to be barely significant. In particular, an initial significantly positive reaction to a contractionary monetary policy shock as observed in Uhlig's benchmark study (his Figure 5) is not seen. Based on our simulation results and taking into account that a very large model is estimated and a large number of impulse responses is considered jointly even if individual impulse response functions are considered, the most reliable choices might be the B and W bands. Based on them the evidence for a negative response of real GDP to a contractionary monetary policy shock is not very strong. Moreover, no significant response of the GDP deflator to a monetary policy response is found. In particular, a price puzzle, that is, a positive response of the price level after a contractionary monetary policy shock, is not observed. This result is in sharp contrast with Uhlig's benchmark analysis where a significant price puzzle is obtained using 68% Bayesian error bands. Our results are well in line with Inoue and Kilian (2013, Figure 7) who use different Bayesian 68% error bands and find that the initial positive reaction of real GDP may not be significant and that there is no significant price puzzle.

5 Conclusions

In this study we have compared different methods for constructing error bands around multivariate estimates with the example of impulse responses in VAR analysis in mind. The main objective is to ensure a prespecified coverage probability. In other words, the analyst is assumed to desire a band in which the true impulse response function is contained with a prespecified probability in a frequentist setting. The same problem is equally relevant in a Bayesian setting where a posterior probability band may be of interest.

We have identified two main approaches to construct such bands. The first one is based on the Bonferroni principle and ensures the desired coverage probability by increasing confidence intervals for individual impulse response coefficients sufficiently. The method depends to some extent on the number of quantities included in the analysis. For impulse response analysis this means that Bonferroni bands differ depending on whether individual impulse response functions, that is, the impulse responses of one variable to one specific shock, or all impulse response functions jointly are of interest.

The second approach uses the Wald statistic to specify a confidence set of sufficient coverage and maps that into a confidence band around the impulse response functions. This method is also used by Bayesians who construct error bands on the basis of the posterior distribution.

Both methods are theoretically conservative and a number of adjustments are proposed to move the actual confidence level closer to the desired level with the prime objective to reduce the widths of the bands. The width is measured as the sum of the widths of the intervals around the individual impulse response coefficients.

We have compared the Bonferroni and the Wald bands theoretically and empirically. Theoretical considerations suggest that Bonferroni bands may well be considerably smaller than Wald bands in an idealized setting where normally distributed estimators are available. This result is remarkable because Bonferroni does not use information on the dependence structure among the quantities of interest considered jointly while the Wald method uses such information.

The framework for our theoretical analysis does not account for important aspects of applied impulse response analysis, however. In particular, in a frequentist VAR analysis only asymptotic distributions of estimators are known and the implications of not knowing the precise underlying distributions for constructing the alternative error bands are of interest. We have therefore conducted a MC simulation experiment based on bivariate VAR processes. The main results are as follows. Both Bonferroni and Wald methods result in coverage levels for impulse response functions that are close to or above the nominal level if the sample sizes are moderate or large and the processes are not very persistent. The methods tend to be slightly conservative but are overall remarkably precise in terms of coverage level regardless of whether individual impulse response functions or the whole set of all impulse responses are considered jointly. The Wald bands tend to be wider than the Bonferroni bands.

Unfortunately, both types of error bands are rather wide. Therefore the adjustments proposed in this study are worth contemplating. If the sample size is small, the propagation horizon large or the VAR order is estimated, our adjustments may however lead to coverage rates markedly below the nominal level. Hence, in practice the original Bonferroni and Wald methods may be preferrable to be on the safe side. The price in terms of band width can be high, however. Taking together all the evidence, the Bonferroni methods seem to have a slight advantage in terms of coverage precision and width. This result suggests that also Bayesian inference can be improved by using the Bonferroni results when doing inference regarding multivariate parameters.

We have also illustrated the methods for constructing error bands for impulse responses with two examples from the literature. The first one is a model for the market of crude oil and the second one is a macro system for investigating U.S. monetary policy. For both systems it turns out that our bands are wider than those previously considered in the literature based on individual confidence intervals. Such bands are known to have an actual confidence level far below the desired one. Using our more realistic bands some of the previously drawn conclusions are not supported any more or at least have much less support from the data.

A natural extensions of our present work is to use the methods for constructing joint forecast bands for forecasts of different horizons. Such bands have been discussed extensively in the literature (e.g., Jordà and Marcellino (2010), Staszewska-Bystrova (2011), Wolf and Wunderli (2012), Staszewska-Bystrova and Winker (2013)). Forecasts based on VAR processes are very similar to impulse responses and the methods considered here can be used

straightforwardly in that context. They may also be useful, however, in considering sequences of forecasts obtained with alternative models such as factor or dynamic regression models.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society B* **57**: 289–300.

Bernanke, B. S. and Mihov, I. (1998a). The liquidity effect and long-run neutrality, *Carnegie-Rochester Conference Series on Public Policy* **49**: 149–194.

Bernanke, B. S. and Mihov, I. (1998b). Measuring monetary policy, *Quarterly Journal of Economics* **113**: 869–902.

Grigoletto, M. (2005). Bootstrap prediction regions for multivariate autoregressive processes, *Statistical Methods & Applications* **14**: 179–207.

Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance, *Biometrika* **75**: 800–802.

Holm, S. (1979). A simple sequentially rejective multiple test procedure, *Scandinavian Journal of Statistics* **6**: 65–70.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test, *Biometrika* **75**: 383–386.

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

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

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

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

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

Kim, J. H. (1999). Asymptotic and bootstrap prediction regions for vector autoregression, *International Journal of Forecasting* **15**: 393–403.

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

Lütkepohl, H. (1988). Asymptotic distribution of the moving average coefficients of an estimated vector autoregressive process, *Econometric Theory* **4**: 77–85.

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

Lütkepohl, H., Staszewska-Bystrova, A. and Winker, P. (2013). Comparison of methods for constructing joint confidence bands for impulse response functions, *International Journal of Forecasting* **forthcoming**.

Sims, C. A. and Zha, T. (1999). Error bands for impulse responses, *Econometrica* **67**: 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. and Winker, P. (2013). Constructing narrowest pathwise bootstrap prediction bands using threshold accepting, *International Journal of Forecasting* **29**: 221–233.

Uhlig, H. (2005). What are the effects of monetary policy on output? Results from an agnostic identification procedure, *Journal of Monetary Economics* **52**: 381–419.

Wolf, M. and Wunderli, D. (2012). Bootstrap joint prediction regions, *Working Paper 748*, National Center of Competence in Research Financial Valuation and Risk Management.

Table 2: Estimated Coverage Probabilities and Total Width (in brackets) for Nominal 90% Confidence Bands for Individual Impulse Responses for bivariate VAR(1) Processes (based on sample size $T = 100$, 2000 MC replications and 2000 bootstrap replications, $H = 10$, true lag order)

	first shock				second shock			
band	y_1		y_2		y_1		y_2	
	$\alpha = -0.95$							
B	94.50	(8.09)	93.05	(3.69)	99.15	(2.36)	92.70	(2.58)
B^a	87.05	(5.75)	88.75	(3.09)	93.95	(1.54)	86.75	(2.00)
W	96.35	(9.00)	95.95	(4.22)	99.30	(2.46)	95.45	(2.75)
W^a	85.85	(6.23)	88.90	(3.39)	93.35	(1.61)	88.35	(2.25)
WB^a	86.55	(5.73)	88.20	(3.10)	93.95	(1.54)	87.45	(2.00)
	$\alpha = -0.9$							
B	94.45	(7.68)	93.20	(3.64)	99.20	(2.01)	93.20	(2.58)
B^a	84.10	(5.73)	85.55	(3.11)	94.40	(1.32)	87.55	(1.97)
W	96.75	(8.44)	95.65	(4.13)	99.25	(2.11)	96.40	(2.76)
W^a	85.50	(6.13)	86.65	(3.40)	94.45	(1.39)	87.70	(2.23)
WB^a	83.95	(5.71)	84.80	(3.13)	94.35	(1.32)	88.00	(1.97)
	$\alpha = -0.5$							
B	95.35	(2.58)	95.30	(2.68)	99.00	(0.78)	94.25	(2.42)
B^a	88.70	(1.95)	89.70	(2.22)	96.75	(0.56)	87.45	(1.76)
W	97.90	(2.76)	97.45	(3.03)	99.10	(0.80)	97.15	(2.56)
W^a	91.70	(2.17)	89.35	(2.53)	97.20	(0.59)	91.85	(2.00)
WB^a	88.55	(1.95)	89.80	(2.23)	96.95	(0.56)	87.30	(1.76)
	$\alpha = 0$							
B	95.30	(1.63)	94.85	(3.15)	99.15	(0.94)	93.90	(2.64)
B^a	91.45	(1.35)	89.25	(2.54)	96.00	(0.65)	88.05	(1.93)
W	97.55	(1.79)	97.15	(3.51)	99.20	(0.97)	97.40	(2.80)
W^a	92.50	(1.48)	89.85	(2.86)	96.00	(0.68)	91.40	(2.13)
WB^a	92.00	(1.36)	89.35	(2.55)	95.95	(0.65)	88.15	(1.92)
	$\alpha = 0.5$							
B	94.30	(3.30)	95.10	(5.01)	99.25	(1.86)	93.70	(3.41)
B^a	88.55	(2.61)	88.65	(4.04)	95.40	(1.28)	87.80	(2.57)
W	96.75	(3.54)	97.50	(5.44)	99.40	(1.93)	96.55	(3.62)
W^a	89.90	(2.82)	89.95	(4.40)	95.35	(1.36)	90.60	(2.78)
WB^a	89.15	(2.61)	88.45	(4.04)	95.30	(1.28)	88.40	(2.57)
	$\alpha = 0.9$							
B	93.15	(8.69)	93.40	(8.81)	99.00	(4.94)	93.85	(5.61)
B^a	86.90	(6.65)	87.75	(6.93)	93.85	(3.33)	86.90	(4.22)
W	95.50	(9.36)	95.85	(9.62)	99.10	(5.28)	96.80	(6.06)
W^a	89.35	(7.10)	89.80	(7.42)	92.90	(3.54)	88.55	(4.49)
WB^a	86.90	(6.63)	87.45	(6.91)	93.85	(3.33)	88.05	(4.20)
	$\alpha = 0.95$							
B	92.85	(9.43)	93.20	(9.09)	99.15	(5.63)	93.60	(6.00)
B^a	85.90	(7.01)	86.40	(6.98)	93.35	(3.77)	85.75	(4.50)
W	95.80	(10.21)	95.75	(9.99)	99.20	(6.06)	97.05	(6.52)
W^a	90.20	(7.64)	89.60	(7.59)	91.75	(4.01)	88.00	(4.81)
WB^a	86.05	(6.99)	86.85	(6.95)	93.35	(3.77)	86.50	(4.48)
	$\alpha = 1$							
B	87.20	(9.35)	86.85	(8.70)	98.95	(6.05)	93.05	(6.17)
B^a	75.10	(6.64)	74.55	(6.46)	94.35	(4.00)	85.85	(4.60)
W	92.55	(10.29)	92.45	(9.71)	99.15	(6.55)	96.30	(6.72)
W^a	83.05	(7.38)	82.10	(7.12)	92.00	(4.28)	86.85	(4.94)
WB^a	75.05	(6.61)	74.25	(6.43)	94.30	(4.00)	86.40	(4.58)

Table 3: Estimated Coverage Probabilities and Total Width (in brackets) for Nominal 90% Confidence Bands for Individual Impulse Responses for bivariate VAR(1) Processes (based on sample size $T = 100$, 2000 MC replications and 2000 bootstrap replications, $H = 10$, lag order estimated using AIC)

band	first shock				second shock			
	y_1		y_2		y_1		y_2	
	$\alpha = -0.95$							
B	92.05	(8.32)	91.45	(3.91)	97.55	(2.93)	89.20	(2.88)
B^a	85.20	(5.98)	86.20	(3.32)	90.75	(1.99)	83.40	(2.31)
W	95.30	(9.38)	95.60	(4.53)	98.70	(3.22)	93.60	(3.16)
W^a	84.90	(6.50)	86.00	(3.62)	90.20	(2.12)	83.95	(2.56)
WB^a	85.20	(5.96)	86.45	(3.33)	90.60	(1.99)	84.30	(2.31)
	$\alpha = -0.9$							
B	92.90	(7.84)	91.80	(3.85)	97.35	(2.51)	90.55	(2.86)
B^a	81.80	(5.92)	83.05	(3.33)	91.25	(1.73)	84.45	(2.28)
W	95.90	(8.73)	95.35	(4.43)	98.80	(2.78)	95.15	(3.15)
W^a	84.10	(6.36)	85.05	(3.61)	91.65	(1.84)	85.65	(2.53)
WB^a	81.95	(5.91)	83.80	(3.34)	91.15	(1.73)	85.70	(2.28)
	$\alpha = -0.5$							
B	93.00	(2.80)	93.90	(2.93)	96.50	(1.05)	91.85	(2.69)
B^a	86.85	(2.19)	88.05	(2.47)	93.65	(0.82)	84.60	(2.05)
W	97.05	(3.07)	96.95	(3.37)	98.25	(1.16)	96.45	(2.94)
W^a	89.20	(2.42)	86.75	(2.79)	93.70	(0.86)	88.55	(2.29)
WB^a	86.95	(2.19)	88.20	(2.49)	93.95	(0.82)	84.65	(2.05)
	$\alpha = 0$							
B	93.25	(1.83)	93.25	(3.37)	97.80	(1.11)	91.90	(2.87)
B^a	89.60	(1.55)	87.80	(2.77)	94.25	(0.83)	85.75	(2.18)
W	97.00	(2.06)	96.50	(3.82)	98.95	(1.21)	96.00	(3.12)
W^a	89.70	(1.68)	89.00	(3.08)	93.70	(0.87)	89.20	(2.39)
WB^a	90.25	(1.56)	87.60	(2.78)	94.25	(0.84)	86.20	(2.17)
	$\alpha = 0.5$							
B	92.70	(3.47)	94.00	(5.18)	97.10	(2.05)	91.20	(3.64)
B^a	86.75	(2.79)	86.65	(4.23)	92.60	(1.47)	85.60	(2.81)
W	95.70	(3.80)	96.70	(5.73)	98.55	(2.20)	95.75	(3.96)
W^a	88.15	(3.00)	88.85	(4.58)	92.05	(1.57)	88.70	(3.03)
WB^a	87.40	(2.80)	86.80	(4.23)	92.95	(1.48)	86.10	(2.81)
	$\alpha = 0.9$							
B	90.85	(8.82)	92.05	(8.95)	97.20	(5.22)	91.10	(5.91)
B^a	84.35	(6.81)	85.40	(7.11)	90.55	(3.62)	83.65	(4.53)
W	94.40	(9.67)	95.35	(9.95)	98.35	(5.75)	94.85	(6.55)
W^a	87.30	(7.30)	87.75	(7.62)	89.80	(3.88)	84.90	(4.84)
WB^a	84.85	(6.79)	85.55	(7.08)	90.65	(3.62)	84.75	(4.51)
	$\alpha = 0.95$							
B	90.80	(9.61)	91.45	(9.27)	98.00	(5.90)	90.95	(6.28)
B^a	84.35	(7.21)	84.95	(7.18)	90.65	(4.03)	82.95	(4.77)
W	94.45	(10.57)	94.80	(10.33)	98.90	(6.52)	95.30	(6.96)
W^a	88.50	(7.87)	88.00	(7.82)	89.25	(4.34)	84.15	(5.12)
WB^a	84.95	(7.18)	85.15	(7.15)	90.55	(4.03)	84.00	(4.75)
	$\alpha = 1$							
B	83.60	(9.67)	85.50	(9.03)	97.10	(6.42)	90.70	(6.55)
B^a	73.20	(6.96)	73.30	(6.78)	91.20	(4.35)	82.40	(4.95)
W	90.70	(10.82)	91.45	(10.22)	98.55	(7.14)	95.20	(7.29)
W^a	80.20	(7.75)	81.30	(7.49)	89.80	(4.72)	84.25	(5.34)
WB^a	73.25	(6.93)	73.20	(6.75)	91.30	(4.35)	83.10	(4.92)

Table 4: Estimated Coverage Probabilities and Total Width (in brackets) for Nominal 90% Confidence Bands for All Impulse Responses considered jointly for bivariate VAR(1) Processes (based on 2000 MC replications and 2000 bootstrap replications, $H = 10$, true lag order)

	$T = 50$		$T = 100$		$T = 200$	
	$\alpha = -0.95$					
JB	87.75	(28.06)	91.45	(19.50)	93.70	(13.69)
JB^a	79.70	(21.05)	84.10	(14.62)	85.20	(10.23)
JB$_*$	77.50	(27.40)	83.95	(18.92)	86.45	(13.39)
JB$^a_*$	75.15	(22.78)	81.55	(15.78)	84.00	(11.05)
W	88.15	(26.86)	90.65	(18.44)	92.05	(12.88)
JW^a	81.95	(23.98)	85.45	(16.52)	86.40	(11.53)
JWB^a	80.25	(21.12)	84.10	(14.67)	85.35	(10.26)
	$\alpha = -0.9$					
JB	88.00	(27.16)	92.05	(18.46)	93.75	(12.72)
JB^a	79.95	(20.80)	82.10	(14.25)	85.20	(9.75)
JB$_*$	77.40	(26.87)	84.70	(18.08)	88.00	(12.53)
JB$^a_*$	75.95	(22.43)	82.25	(15.36)	85.20	(10.53)
W	88.40	(26.00)	91.85	(17.44)	92.65	(11.82)
JW^a	83.40	(23.26)	86.05	(15.74)	87.50	(10.75)
JWB^a	80.70	(20.84)	83.25	(14.28)	85.65	(9.78)
	$\alpha = -0.5$					
JB	90.05	(17.02)	93.80	(10.08)	95.85	(6.55)
JB^a	83.40	(12.20)	86.35	(7.63)	85.70	(5.08)
JB$_*$	78.10	(17.42)	84.60	(9.92)	87.30	(6.30)
JB$^a_*$	77.35	(13.68)	83.30	(8.24)	85.50	(5.38)
W	89.65	(15.35)	93.30	(9.15)	93.85	(6.01)
JW^a	86.25	(13.75)	89.55	(8.39)	90.00	(5.55)
JWB^a	83.65	(12.28)	86.45	(7.67)	86.45	(5.11)
	$\alpha = 0$					
JB	89.05	(16.94)	93.60	(9.99)	95.45	(6.45)
JB^a	82.45	(11.87)	85.85	(7.44)	85.65	(4.92)
JB$_*$	78.25	(17.45)	83.80	(9.90)	87.60	(6.24)
JB$^a_*$	77.50	(13.52)	82.40	(8.10)	85.80	(5.23)
W	89.65	(15.36)	93.35	(9.06)	93.25	(5.88)
JW^a	85.70	(13.44)	88.85	(8.15)	88.40	(5.34)
JWB^a	83.10	(11.94)	86.15	(7.47)	86.25	(4.94)
	$\alpha = 0.5$					
JB	89.50	(26.82)	92.65	(16.17)	95.30	(10.46)
JB^a	82.15	(19.29)	84.75	(12.13)	87.85	(7.99)
JB$_*$	79.45	(28.06)	83.80	(16.55)	87.55	(10.51)
JB$^a_*$	78.20	(22.20)	82.40	(13.46)	86.25	(8.70)
W	89.90	(24.71)	92.40	(14.54)	94.20	(9.38)
JW^a	85.45	(21.64)	87.80	(13.04)	90.25	(8.48)
JWB^a	83.00	(19.42)	85.85	(12.17)	88.30	(8.00)
	$\alpha = 0.9$					
JB	85.45	(45.29)	92.35	(32.54)	94.80	(22.55)
JB^a	76.75	(34.27)	83.25	(24.64)	86.90	(16.90)
JB$_*$	76.00	(46.54)	84.30	(33.01)	87.45	(22.90)
JB$^a_*$	73.75	(37.85)	81.90	(27.09)	85.15	(18.60)
W	87.30	(44.11)	91.30	(30.32)	93.65	(20.50)
JW^a	81.95	(38.92)	85.90	(26.54)	88.90	(17.94)
JWB^a	78.10	(34.35)	83.70	(24.62)	87.10	(16.87)
	$\alpha = 0.95$					
JB	84.10	(47.37)	91.65	(35.03)	93.85	(24.80)
JB^a	74.15	(35.49)	82.15	(25.98)	85.80	(18.12)
JB$_*$	74.95	(48.66)	82.40	(35.33)	86.50	(25.02)
JB$^a_*$	71.70	(39.16)	80.20	(28.61)	84.35	(19.98)
W	85.60	(46.39)	91.75	(32.79)	92.45	(22.66)
JW^a	79.55	(40.69)	85.30	(28.51)	87.10	(19.50)
JWB^a	75.40	(35.54)	82.50	(25.96)	85.95	(18.10)
	$\alpha = 1$					
JB	78.15	(47.29)	87.15	(35.46)	89.95	(25.01)
JB^a	62.60	(34.90)	73.75	(25.40)	76.15	(17.40)
JB$_*$	70.80	(48.47)	79.90	(35.71)	83.60	(25.17)
JB$^a_*$	63.80	(38.65)	74.25	(28.14)	77.35	(19.28)
W	80.60	(46.58)	87.65	(33.27)	90.00	(22.87)
JW^a	74.70	(40.50)	80.60	(28.46)	81.05	(19.32)
JWB^a	63.65	(34.93)	74.35	(25.38)	76.15	(17.38)

Table 5: Estimated Coverage Probabilities and Total Width (in brackets) for Nominal 90% Confidence Bands for All Impulse Responses considered jointly for bivariate VAR(1) Processes (based on 2000 MC replications and 2000 bootstrap replications, $H = 10$, lag order estimated using AIC)

	$T = 50$		$T = 100$		$T = 200$	
	$\alpha = -0.95$					
JB	76.35	(33.02)	87.65	(21.07)	91.55	(14.54)
JB^a	69.15	(25.85)	80.10	(16.08)	81.15	(11.02)
JB$_*$	66.75	(32.77)	79.75	(20.50)	84.55	(14.24)
JB$_*^a$	64.35	(27.49)	76.95	(17.26)	81.55	(11.84)
W	78.80	(33.13)	87.75	(20.29)	90.10	(13.87)
JW^a	71.35	(29.41)	81.10	(18.11)	82.70	(12.37)
JWB^a	69.70	(25.86)	80.25	(16.12)	81.45	(11.04)
	$\alpha = -0.9$					
JB	74.90	(32.25)	89.05	(19.84)	91.80	(13.45)
JB^a	67.85	(25.73)	78.60	(15.59)	82.15	(10.44)
JB$_*$	65.40	(32.06)	81.00	(19.48)	84.30	(13.28)
JB$_*^a$	64.00	(27.26)	78.70	(16.69)	81.65	(11.22)
W	77.25	(32.44)	89.25	(19.09)	90.20	(12.68)
JW^a	70.75	(28.86)	82.45	(17.14)	83.50	(11.49)
JWB^a	68.20	(25.73)	78.90	(15.61)	82.00	(10.47)
	$\alpha = -0.5$					
JB	79.40	(21.04)	91.25	(11.27)	93.60	(7.04)
JB^a	73.55	(16.40)	82.70	(8.84)	84.05	(5.56)
JB$_*$	68.70	(21.46)	80.45	(11.14)	82.55	(6.93)
JB$_*^a$	67.50	(17.78)	79.30	(9.45)	80.65	(6.00)
W	80.70	(20.36)	90.80	(10.54)	92.35	(6.57)
JW^a	75.65	(18.21)	84.55	(9.64)	87.60	(6.04)
JWB^a	74.05	(16.46)	82.70	(8.88)	84.75	(5.59)
	$\alpha = 0$					
JB	76.30	(20.97)	90.85	(10.97)	92.35	(7.00)
JB^a	69.85	(16.18)	82.70	(8.44)	82.75	(5.47)
JB$_*$	66.15	(21.47)	80.85	(10.91)	85.20	(6.79)
JB$_*^a$	65.40	(17.69)	79.60	(9.10)	83.45	(5.77)
W	78.45	(20.39)	91.05	(10.21)	91.35	(6.52)
JW^a	73.75	(18.02)	85.65	(9.19)	84.50	(5.91)
JWB^a	69.95	(16.23)	82.95	(8.46)	82.85	(5.48)
	$\alpha = 0.5$					
JB	78.15	(29.82)	90.35	(17.06)	92.50	(10.95)
JB^a	71.75	(22.79)	81.70	(13.07)	83.90	(8.51)
JB$_*$	69.20	(30.91)	80.75	(17.45)	85.60	(11.07)
JB$_*^a$	67.45	(25.28)	78.90	(14.36)	83.25	(9.27)
W	79.55	(29.08)	89.90	(15.69)	92.35	(10.03)
JW^a	73.90	(25.50)	84.45	(14.03)	87.20	(9.03)
JWB^a	72.25	(22.88)	82.50	(13.09)	84.75	(8.52)
	$\alpha = 0.9$					
JB	72.90	(48.49)	88.75	(33.61)	93.15	(22.99)
JB^a	64.40	(37.52)	78.40	(25.70)	84.30	(17.35)
JB$_*$	64.85	(49.64)	80.80	(34.09)	84.25	(23.42)
JB$_*^a$	62.20	(40.56)	77.30	(28.02)	81.10	(19.08)
W	75.75	(49.19)	88.10	(31.92)	92.35	(21.22)
JW^a	69.80	(43.15)	81.65	(27.83)	85.95	(18.49)
JWB^a	65.90	(37.55)	79.15	(25.67)	84.80	(17.32)
	$\alpha = 0.95$					
JB	70.45	(51.58)	88.35	(36.13)	92.35	(25.45)
JB^a	60.90	(39.44)	77.55	(27.04)	83.50	(18.75)
JB$_*$	61.90	(52.71)	79.50	(36.44)	86.15	(25.52)
JB$_*^a$	58.90	(42.55)	76.10	(29.58)	82.65	(20.42)
W	73.35	(52.66)	88.30	(34.37)	91.60	(23.62)
JW^a	67.05	(45.94)	81.05	(29.76)	85.40	(20.26)
JWB^a	62.40	(39.43)	78.25	(27.00)	83.85	(18.72)
	$\alpha = 1$					
JB	67.05	(52.57)	83.80	(37.13)	89.50	(25.87)
JB^a	53.20	(39.58)	70.70	(26.94)	74.10	(18.15)
JB$_*$	65.00	(49.56)	76.45	(37.35)	81.80	(26.16)
JB$_*^a$	65.00	(41.20)	70.30	(29.56)	76.00	(20.11)
W	71.00	(53.83)	85.00	(35.46)	89.35	(23.98)
JW^a	63.65	(46.64)	76.45	(30.22)	79.10	(20.21)
JWB^a	54.70	(39.55)	70.75	(26.91)	74.00	(18.12)

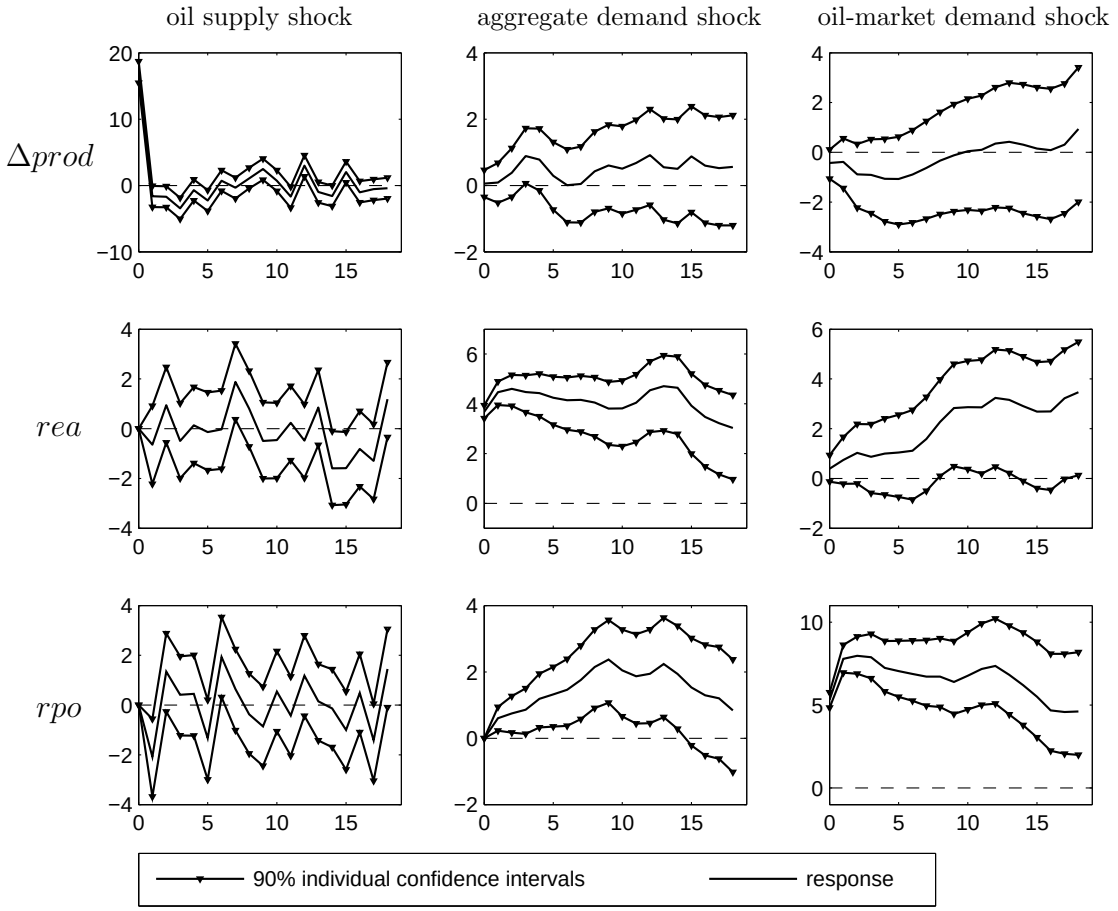

Figure 1: Impulse response analysis for the empirical VAR(24) model based on data from Kilian (2009) and $H = 18$. Responses of $\Delta prod$, rea and rpo to the oil supply shock, the aggregate demand shock and the oil-market specific demand shock are given, respectively, in the first, second and third rows of the graph. The point estimates are supplemented with the 90% individual confidence intervals around the impulse-responses.

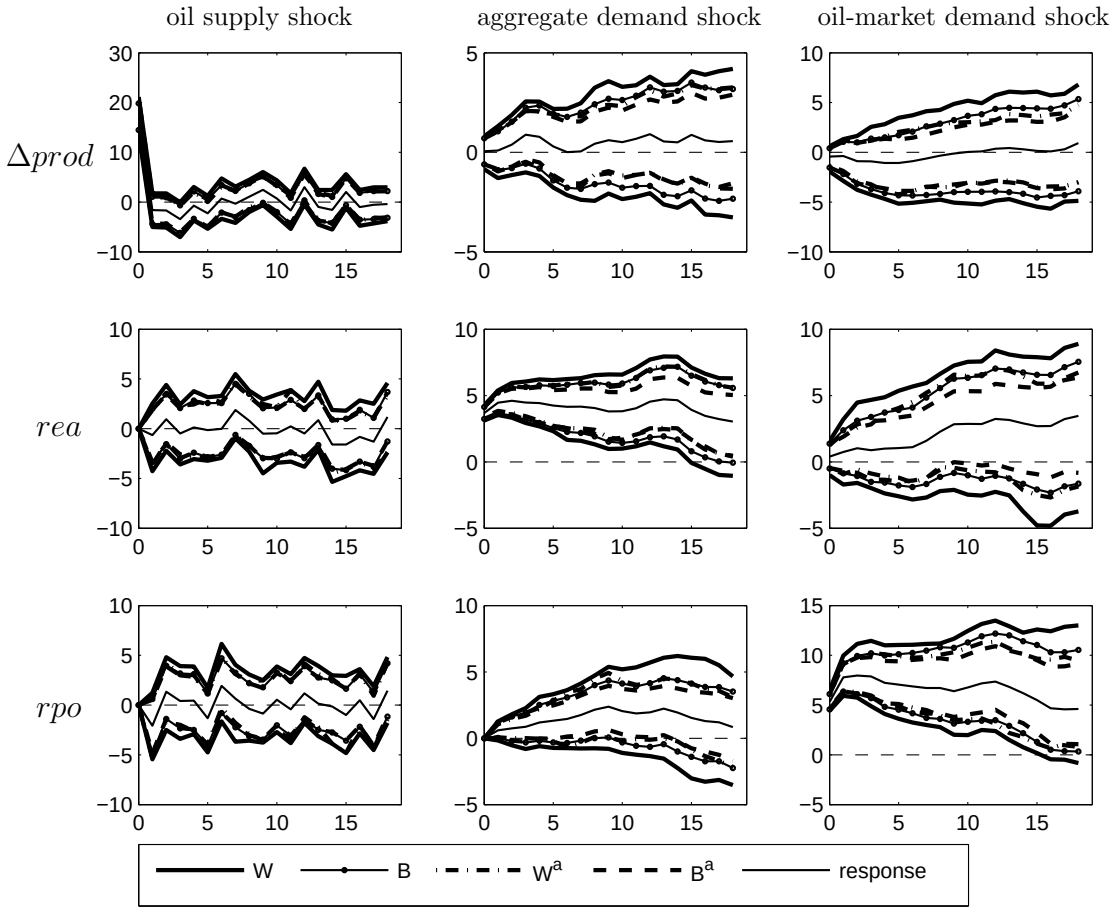

Figure 2: Impulse response analysis for the empirical VAR(24) model based on data from Kilian (2009) and $H = 18$. Responses of $\Delta prod$, rea and rpo to the oil supply shock, the aggregate demand shock and the oil-market specific demand shock are given, respectively, in the first, second and third rows of the graph. The point estimates are supplemented with the 90% confidence bands obtained individually for each impulse-response function.

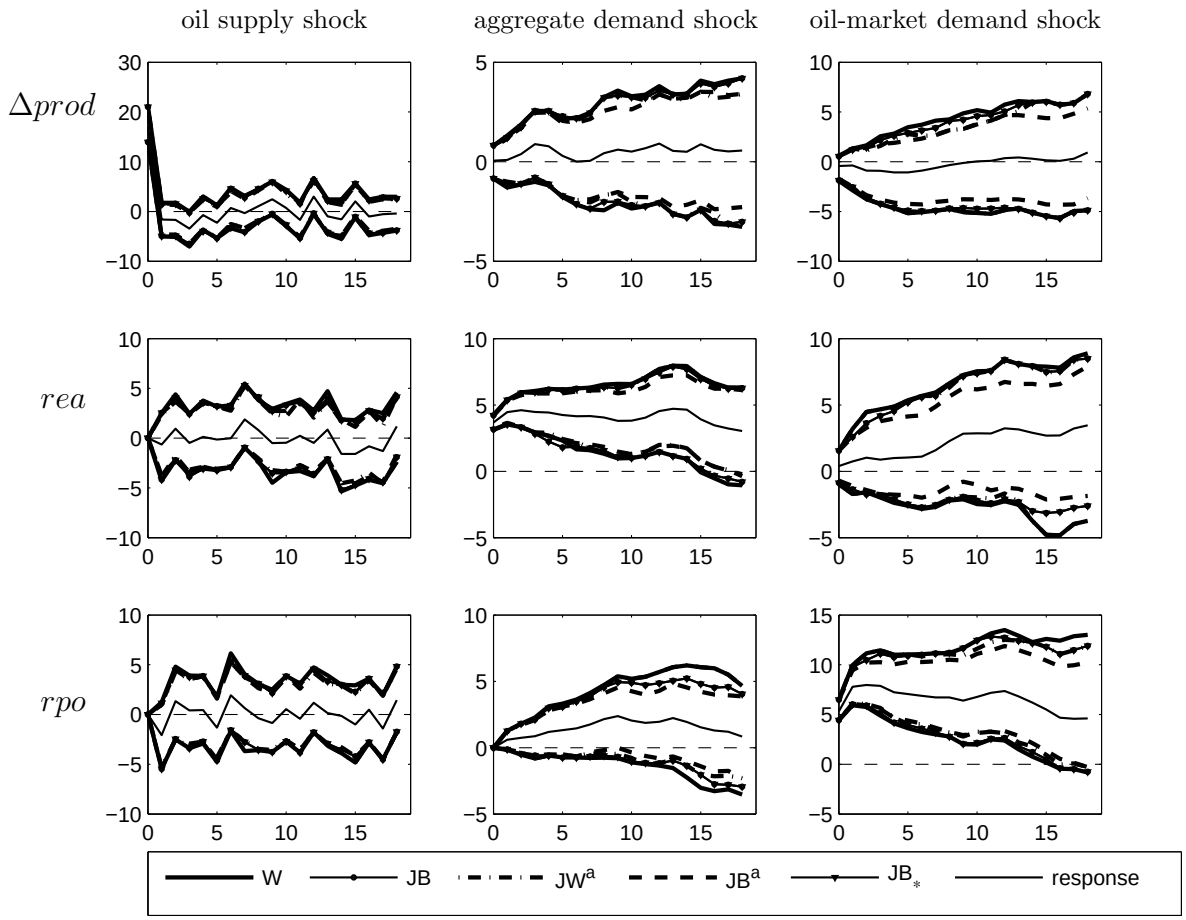

Figure 3: Impulse response analysis for the empirical VAR(24) model based on data from Kilian (2009) and $H = 18$. Responses of $\Delta prod$, rea and rpo to the oil supply shock, the aggregate demand shock and the oil-market specific demand shock are given, respectively, in the first, second and third rows of the graph. The point estimates are supplemented with the 90% joint confidence bands.

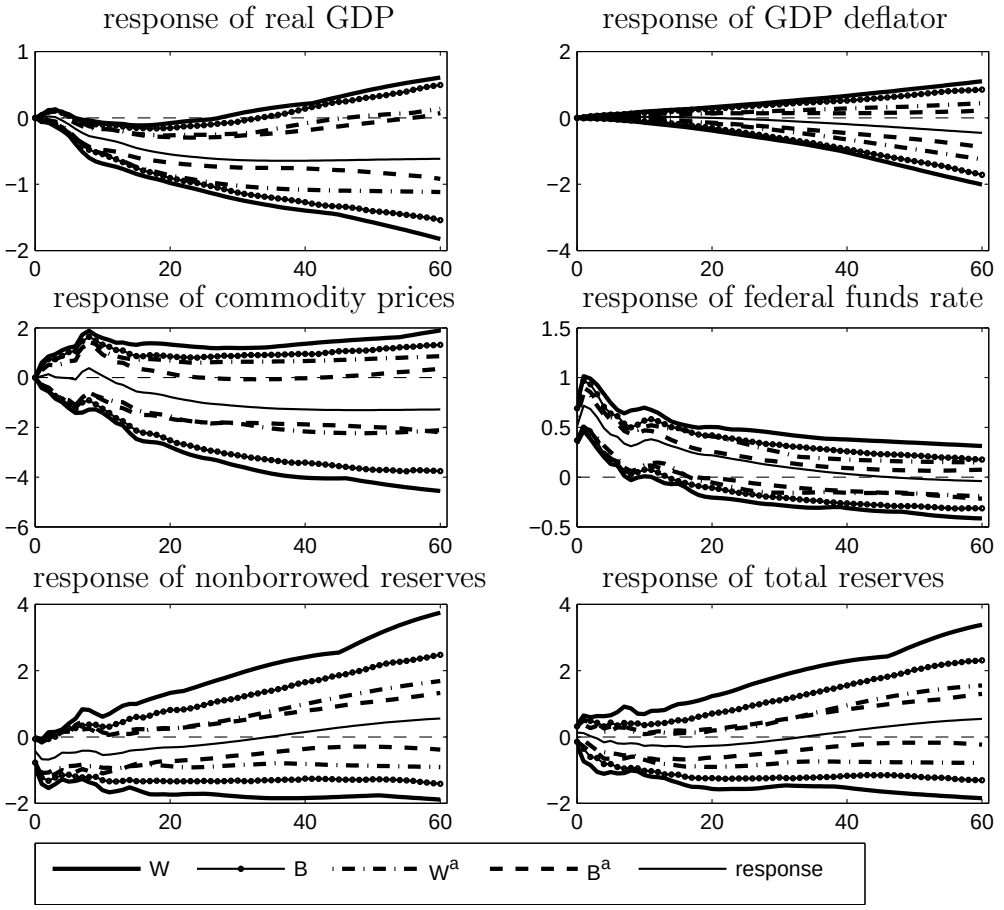

Figure 4: Impulse response analysis for the empirical VAR(12) model based on data from Uhlig (2005) and $H = 60$. The graph shows responses of all the modelled variables to the monetary policy shock together with the 68% confidence bands obtained individually for each impulse-response function.

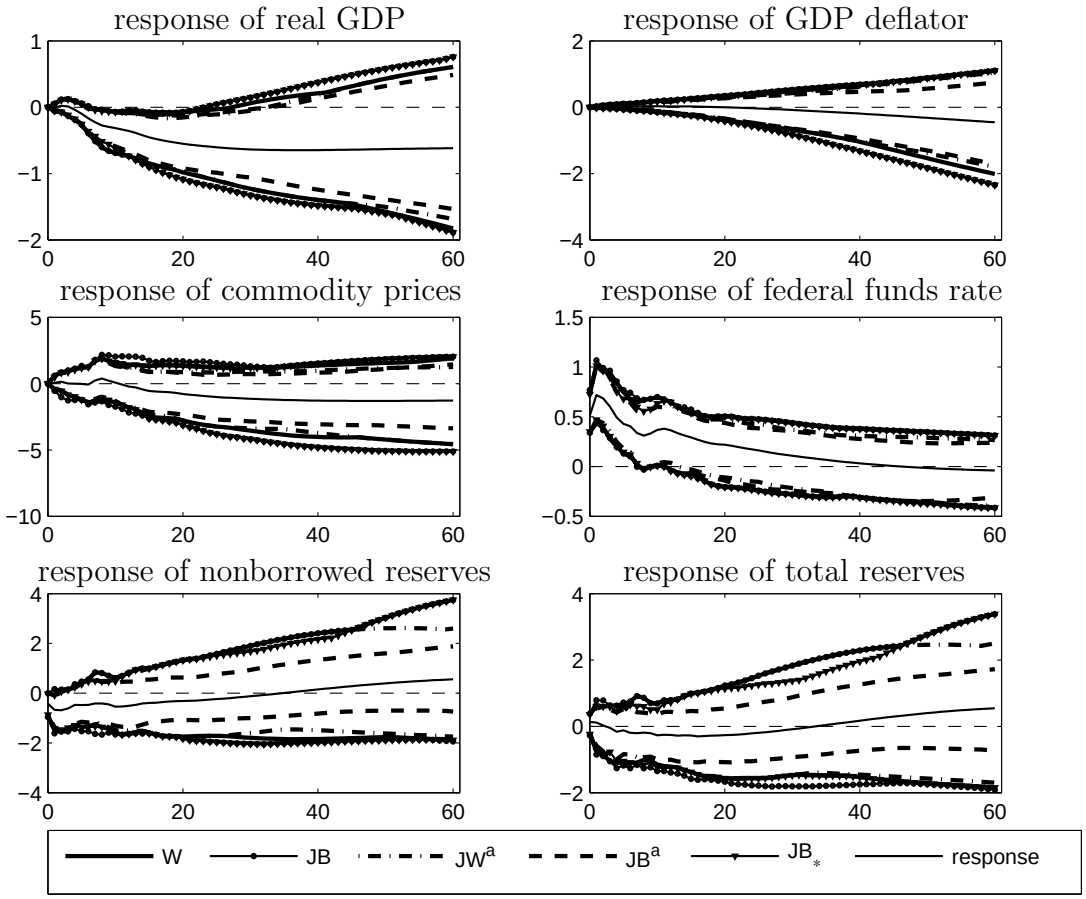

Figure 5: Impulse response analysis for the empirical VAR(12) model based on data from Uhlig (2005) and $H = 60$. The graph shows responses of all the modelled variables to the monetary policy shock together with the joint 68% confidence bands.