Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions

Helmut Lütkepohl*
Anna Staszewska-Bystrova**
Peter Winker***

* Freie Universität Berlin and DIW, Germany
** University of Lodz, Poland
*** Universität Gießen, Germany

BERLIN

ECONOMIC RISK

SFB 649

Comparison of Methods for Constructing Joint Confidence Bands for Impulse Response Functions[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

April 12, 2013

Abstract. In vector autoregressive analysis confidence intervals for individual impulse responses are typically reported to indicate the sampling uncertainty in the estimation results. A range of methods are reviewed and a new proposal is made for constructing joint confidence bands, given a prespecified coverage level, for the impulse responses at all horizons considered simultaneously. The methods are compared in a simulation experiment and recommendations for empirical work are provided.

Key Words: Vector autoregressive process, impulse responses, bootstrap, confidence band

JEL classification: C32

1 Introduction

In vector autoregressive (VAR) analyses, impulse responses are commonly used for investigating the effects of shocks to the system. Because the impulse responses are functions of the VAR parameters, they have to be estimated in practice. Estimation uncertainty is usually indicated by showing confidence intervals around the individual impulse response coefficients. Asymptotic, bootstrap or Bayesian methods are typically used for setting up such intervals (see, e.g., Lütkepohl (2005)). If individual confidence intervals for a given confidence level are constructed around the impulse response coefficients for each response horizon separately, there is no guarantee that the overall coverage level for all impulse responses of one variable will correspond to the prespecified confidence level. In other words, the probability of the band containing the true impulse response function of a specific variable will, in general, not be $1 - \gamma$ if the confidence band is constructed as the union of individual $(1 - \gamma) \times 100\%$ confidence intervals. Hence, it is desirable to construct confidence bands with an overall prespecified coverage probability. A range of suitable methods are reviewed in this study and a new proposal is considered. A simulation experiment is used to compare the methods and recommendations for applied work are given. Our criteria for assessing the bands are the coverage level and the size of the confidence band. While different measures of size are conceivable, in this study we calculate it as the sum of the widths of all individual intervals.

This is not the first study to consider the problem of constructing confidence bands for impulse responses. For example, Sims and Zha (1999) propose a method based on Bayesian principles. In this study we will remain within a classical framework where one could use, for example, the Bonferroni inequality for constructing confidence bands with a joint coverage level at least as large as the desired one. The drawback of this method is that it may deliver very conservative bands that provide a much larger coverage than desired and, consequently, be unnecessarily wide. Therefore we propose a strategy for reducing the bands by adjusting the Bonferroni bands. Another proposal is due to Jordà (2009). He constructs the bands on the basis of so-called Scheffé bounds. Unfortunately, the underlying inequalities are only approximate and may fail to deliver correct coverage levels even under ideal conditions, as convincingly argued by Wolf and Wunderli (2012) in the context of constructing joint forecast bands. Yet another approach is proposed by Staszewska (2007), who uses numerical search methods to find smallest confidence bands for a given coverage level. The disadvantage is that it requires a rather substantial computational effort. Moreover, no general results are available showing that the desired coverage level is actually obtained at least asymptotically. All these proposals will be compared in a simulation experiment.

Bands with a given coverage level are also of interest in computing forecast paths over a number of horizons. Constructing bands around path forecasts has been considered, for instance, by Jordà and Marcellino (2010), Wolf and Wunderli (2012), Staszewska-Bystrova (2011) and Staszewska-Bystrova and

Winker (2013). Since impulse responses are conditional forecasts, there is an obvious relation to the forecast literature and we will draw on it by adopting the method proposed by Wolf and Wunderli (2012) to our framework of constructing confidence bands around impulse responses. The difference to the literature on path forecasts is that there are two components of uncertainty attached to forecasts of specific variables even if the data generation process (DGP) is known apart from its parameters: the intrinsic uncertainty from the DGP and the estimation uncertainty obtained from using estimated instead of true parameters. In contrast, since impulse responses are conditional forecasts that consider only the marginal effect of a specific shock for a given process, only estimation uncertainty is relevant in the context of evaluating impulse responses if the correct model is used. Of course, in practice there is the usual uncertainty about the DGP in both types of analysis. In any case, our results are also of interest for constructing bands around multiple-horizon forecasts for specific variables although we focus on the impulse response context.

The remainder of the study is organized as follows. In Sec. 2 the model setup is presented. Sec. 3 reviews the methods for constructing joint confidence bands for impulse responses and in Sec. 4 a simulation comparison is discussed. An illustrative example is given in Sec. 5 and Sec. 6 concludes. A number of technical details can be found in the Appendix.

2   Model Setup

A standard reduced form VAR setup is used with variables $y_t = (y_{1t}, \ldots, y_{Kt})'$ being generated by a K-dimensional VAR(p) process,

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

The A_i $(i = 1, \ldots, p)$ are $(K \times K)$ parameter matrices and the error process $u_t = (u_{1t}, \ldots, u_{Kt})'$ is a K-dimensional zero mean white noise process with covariance matrix $\mathbb{E}(u_t u_t') = \Sigma_u$, that is, $u_t \sim (0, \Sigma_u)$. The K-dimensional intercept vector ν is the only deterministic term because such terms are of limited relevance for the following arguments. Adding other terms such as linear trends or seasonal dummy variables would not change the substance of the argument. They need to be included in practice as required for a good description of the data, of course.

In lag operator notation the process (2.1) can be written as

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

with $A(L) = I_K - A_1 L - \cdots - A_p L^p$. The process is *stable* if

$$\det A(z) = \det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \text{ for } z \in \mathbb{C}, |z| \leq 1. \tag{2.3}$$

Structural shocks ε_t are obtained from the reduced form errors by a linear transformation, $\varepsilon_t = B^{-1} u_t$, such that the structural shocks are instantaneously uncorrelated and have variance one. In other words, $\varepsilon_t \sim (0, I_K)$.

The $(K \times K)$ matrix B is the matrix of impact effects of the shocks. It is assumed to be identified by a set of suitable restrictions. These can be exclusion restrictions on the impact effects or constraints for the long-run effects of the shocks. For stable processes the impulse responses are just the coefficients of the moving average (MA) representation of y_t,

$$y_t = A(1)^{-1}\nu + A(L)^{-1}B\varepsilon_t = \mu + \sum_{i=0}^{\infty} \Phi_i \varepsilon_{t-i}, \qquad (2.4)$$

where $\mu = A(1)^{-1}\nu$, $\Phi_0 = B$ and $\sum_{i=0}^{\infty} \Phi_i L^i = A(L)^{-1}B$. Hence, the impulse response coefficients $\Phi_i = \Phi_i(A_1, \ldots, A_p, B)$ are nonlinear functions of the reduced form parameters and B (see, e.g., Lütkepohl (2005, Chapter 9) for details). For integrated and cointegrated processes where the VAR operator $A(L)$ has unit roots, that is, $\det A(1) = 0$, the impulse responses are computed using the same functions of the VAR coefficients as for the stable case although y_t cannot be represented in terms of the infinite order MA representation (2.4).

3 Methods for Constructing Confidence Bands for Impulse Responses

In empirical studies VAR parameters are usually estimated using standard methods with estimates for the impulse responses obtained as (nonlinear) functions of these estimates. The estimated impulse responses are typically plotted with confidence bands. In most cases these confidence bands are obtained by simply connecting confidence intervals for individual impulse responses. In other words, the joint distribution is ignored in setting up the confidence bands. Such confidence bands are called *naïve* bands in the following. It is well-known that, in general, individual confidence intervals with level $1 - \gamma$ ($0 < \gamma < 1$) do not ensure a joint coverage probability of $1 - \gamma$. Therefore a number of proposals have been made for constructing confidence bands with a more precise coverage level. We will review some of them in the following and also propose a new variant.

Most of the methods are based on standard residual based bootstraps. Also, where not explicitly stated otherwise, instead of the usual least-squares estimates for the VAR coefficients we use bias-corrected estimates because they were found to work well in this context (see Kilian (1998, 2001)). The precise details of these methods are presented in the Appendix. The following procedures for constructing confidence bands around impulse responses are compared in a Monte Carlo study reported in Sec. 4.

3.1 The naïve band (naïve)

As mentioned earlier, the naïve confidence band is given by a collection of confidence intervals for individual impulse response coefficients. Let $s^*_{\gamma/2}$ and $s^*_{1-\gamma/2}$ denote the $\gamma/2$ and $1 - \gamma/2$ quantiles of the bootstrap distribution

of the estimated impulse response coefficient at horizon h ($\phi_{ij,h}$), where $h = 0, \ldots, H$. Then the $(1 - \gamma) \times 100\%$ confidence interval for $\phi_{ij,h}$ is just given by

$$[s^*_{\gamma/2}, s^*_{1-\gamma/2}]$$

for all horizons $h = 0, \ldots, H$.

Although the method does not ensure a joint coverage probability of $1 - \gamma$, it is often used in applied work, possibly on the grounds that the confidence bands are just meant to give an indication of the sampling uncertainty related to the estimated impulse responses.

3.2 Traditional Bonferroni band (B)

In contrast to the naïve bands, the Bonferroni bounds account for the stochastic dependence in the estimated impulse response coefficients by increasing the size of the bands so as to ensure a prespecified lower bound of the confidence level at least asymptotically, that is, they are generally conservative. Formally the Bonferroni band is constructed from confidence intervals for individual impulse response coefficients. The nominal confidence level of each interval is given by $(1 - \gamma/L) \times 100\%$, where $L = H + 1$ if the impact effect is not restricted to zero and $L = H$ if the impact effect is restricted to zero. For each horizon h ($h = 0, \ldots, H$), the $(1 - \gamma/L) \times 100\%$ confidence interval for $\phi_{ij,h}$, is given by

$$\left[s^*_{\gamma/2L}, s^*_{1-\gamma/2L}\right],$$

where $s^*_{\gamma/2L}$ and $s^*_{1-\gamma/2L}$ are the $\gamma/2L$ and $1 - \gamma/2L$ quantiles of the bootstrap distribution of the estimator of $\phi_{ij,h}$ under consideration.

3.3 An adjusted Bonferroni band (B-adj)

Since the traditional Bonferroni bands are conservative by construction, we propose an adjustment that reduces the bands but still observes the overall coverage level with respect to the bootstrapped impulse responses. In what follows, the number of bootstrap replicates is denoted by b. First a Bonferroni band is constructed as in Sec. 3.2. This band covers both complete and incomplete bootstrap response functions. Bootstrap impulse responses fully within the band are identified and maintained and a new band is obtained as an envelope of these functions. The resulting band is narrower, or at least not wider, than the traditional Bonferroni band. In a final stage, the number of bootstrap impulse responses covered by the band is calculated and denoted by nb. If $nb > (1 - \gamma)b$, a sequential procedure aimed at removing $nb - (1 - \gamma)b$ bootstrap impulse responses is applied. In each step, the bootstrap impulse responses are identified that provide at least one point on the current bounds. There are at most $2(H+1)$ such functions. The function that contributes the most to the size of the current band (measured as sum of widths of the individual intervals) is rejected. The procedure terminates

after $nb - (1 - \gamma)b$ functions are eliminated. The band is obtained as an envelope of the remaining $(1 - \gamma)b$ bootstrap impulse responses.

This procedure is called adjusted Bonferroni method and abbreviated as B-adj in the following. It has asymptotically correct coverage whenever the bootstrap provides a confidence set with asymptotically precise coverage level because our method maintains $(1 - \gamma)b$ bootstrap samples and uses the outer envelope as confidence set (see Benkwitz, Lütkepohl and Neumann (2000) for a discussion of the conditions for the bootstrap to work for impulse responses). Another advantage of the method is that by construction the full bootstrap impulse response functions all lie within the confidence bands which meets the Fry and Pagan (2011) critique of standard Bayesian confidence bands in the context of sign restrictions.

3.4 The Scheffé band (Scheffé)

The so-called Scheffé band is proposed by Jordà (2009). It is based on the asymptotic distribution of the impulse responses (see Appendix A.2 for details). In our comparison it is calculated using the step-down procedure, proposed by Jordà and Marcellino (2010) in the context of constructing joint prediction bands. To describe the method more formally, we use the following notation: $\boldsymbol{\phi}_{ij} = (\phi_{ij,0}, \ldots, \phi_{ij,H})$ is the $(H + 1)$-dimensional vector of responses of variable i to the jth shock, $\boldsymbol{\phi}_{ij,1,\ldots,H} = (\phi_{ij,1}, \ldots, \phi_{ij,H})$ denotes the corresponding vector when the impact effect is restricted to zero and, hence, not estimated. The estimated asymptotic covariance matrix of $\widehat{\boldsymbol{\phi}}_{ij}$, $\widehat{\Omega}_{ij}/T$ can be decomposed into $\widehat{\Omega}_{ij}/T = PP'$, where P denotes the lower triangular Cholesky factor (see Appendix A.2). The quantity $c_\gamma^2(h + 1)$ is the $(1 - \gamma)$-quantile of the χ^2 distribution with $h + 1$ degrees of freedom and $\left[\sqrt{c_\gamma^2(h+1)/(h+1)}\right]$ is an $((H + 1) \times 1)$ vector whose $(h + 1)$-th entry is $\sqrt{c_\gamma^2(h+1)/(h+1)}$ (for $h = 0, 1, \ldots, H$). Using this notation the band is given by

$$\widehat{\boldsymbol{\phi}}_{ij} \pm P \left[\sqrt{\frac{c_\gamma^2(h+1)}{h+1}}\right], \tag{3.1}$$

if the impact effect of the shock is unrestricted and by

$$\widehat{\boldsymbol{\phi}}_{ij,1,\ldots,H} \pm P^* \left[\sqrt{\frac{c_\gamma^2(h)}{h}}\right], \tag{3.2}$$

if the impact effect of the shock is 0 by construction. Here P^* is a lower triangular matrix resulting from the Cholesky decomposition of a matrix obtained from $\widehat{\Omega}_{ij}/T$ by removing the first column and the first row and $\left[\sqrt{c_\gamma^2(h)/h}\right]$ is an $(H \times 1)$ vector with h-th entry $\sqrt{c_\gamma^2(h)/h}$ (for $h = 1, 2, \ldots, H$).

Notice that the Scheffé bands are based on least-squares estimates of the VAR coefficients, as originally proposed, and not on bias-corrected estimates.

While calculating the Scheffé bands for large H (e.g., for $H = 10$ in our experiments), it is common to face some numerical problems while trying to obtain P (as $\widehat{\Omega}_{ij}/T$ may be rendered as not positive definite). In the experiments, to overcome this problem and to enable the decomposition, all eigenvalues of the estimated covariance matrix that are smaller or equal to 0 are replaced with the value 10^{-6}.

We also obtained some results for the version of the original Scheffé band described by Jordà (2009), which does not use the step-down procedure. We do not report them, as the method turns out to be inferior to the method used here under our performance criteria. As mentioned earlier, the inequalities underlying the Scheffé bands are only approximate and may fail to deliver correct coverage levels even under ideal conditions. This point is made by Wolf and Wunderli (2012) in the context of constructing joint forecast bands and it is equally valid for the present situation.

3.5 The neighbouring paths band (NP)

The neigbouring paths method (NP) proposed by Staszewska (2007) is a procedure based on b bootstrap response paths $\widehat{\boldsymbol{\phi}}_{ij}^{c*}$. In a sequential algorithm, γb bootstrap paths are eliminated. The envelope of the remaining $(1 - \gamma)b$ paths provides the confidence band. While eliminating individual paths the method aims at achieving a narrow confidence band (still containing $(1-\gamma)b$ bootstrap paths). Therefore, candidate paths for elimination in each step are those containing at least one point outside the envelope of the other bootstrap paths. Out of the group of these candidate paths, the algorithm selects the one most distant (in terms of Euclidean distance) from the impulse response obtained from the original data. After eliminating this extreme path, the elimination step is repeated until γb paths have been excluded.

3.6 The Wolf & Wunderli band (WW)

Wolf and Wunderli (2012) originally proposed their bands for multiple horizon forecasts but they can be adapted for the present purposes. The band is based on b bootstrap response paths $\widehat{\boldsymbol{\phi}}_{ij}^{c*}$ and calculated as:

$$\widehat{\boldsymbol{\phi}}_{ij}^{c} \pm \left[d_{|\cdot|,1-\gamma}^{\max,c} \sqrt{\widehat{\omega}_{h+1,h+1}^{c}} \right],$$

where $\left[d_{|\cdot|,1-\gamma}^{\max,c} \sqrt{\widehat{\omega}_{h+1,h+1}^{c}} \right]$ is a vector with dimensions $((H + 1) \times 1)$, whose $(h + 1)$-th entry (for $h = 0, 1, \ldots, H$) is given by $d_{|\cdot|,1-\gamma}^{\max,c} \sqrt{\widehat{\omega}_{h+1,h+1}^{c}}$. Here $\widehat{\omega}_{h+1,h+1}^{c}$ is the $(h + 1)$-th element of the main diagonal of $\widehat{\Omega}_{ij}^{c}$ (see Appendix A.2 for details on this estimate) and $d_{|\cdot|,1-\gamma}^{\max,c}$ is obtained as follows:

- In a single iteration of the bootstrap method,

$$\widehat{s}_{ij}^{c*}(h) = \frac{\widehat{\phi}_{ij,h}^{c*} - \widehat{\phi}_{ij,h}^{c}}{\sqrt{\widehat{\omega}_{h+1,h+1}^{c*}}}, \quad h = 0, \ldots, H,$$

6

are computed. The quantity $\max_{|\cdot|}^{c*} \equiv \max(|\widehat{S}^{c*}(H)|)$ is calculated, where $\widehat{S}^{c*}(H) = (\widehat{s}_{ij}^{c*}(0), \ldots, \widehat{s}_{ij}^{c*}(H))'$ and $\max(X)$ returns the largest value from the vector X.

- After performing b iterations of the bootstrap method and obtaining $\max_{|\cdot|,1}^{c*}$, ..., $\max_{|\cdot|,b}^{c*}$, $d_{|\cdot|,1-\gamma}^{\max,c}$ is computed, which stands for the $1 - \gamma$ empirical quantile of the statistics $\max_{|\cdot|,1}^{c*}$, ..., $\max_{|\cdot|,b}^{c*}$.

We emphasize that Wolf and Wunderli (2012) did not propose these bands for impulse responses but for multiple horizon forecasts. In fact, they may not be ideal for impulse responses because the estimated standard deviation $\sqrt{\widehat{\omega}_{h+1,h+1}^{c*}}$ appears in the denominator of $\widehat{s}_{ij}^{c*}(h)$. This standard deviation may become very small for long response horizons and hence the $\widehat{s}_{ij}^{c*}(h)$ may become quite unstable by dividing by a quantity close to zero. This problem is not likely to have an effect for forecasts because forecast error variances reflect not only estimation uncertainty but also systematic uncertainty that will not decrease with the forecast horizon. Thus, what works well for forecasts may not work well for impulse responses. The method is included in our comparison for completeness. The following simulations are meant to shed light on the small sample performance of the various proposals.

4 Monte Carlo Comparison with other Methods

For our small sample comparison we use the DGPs from Kilian (1998),

$$y_t = \begin{bmatrix} \alpha_{11} & 0 \\ 0.5 & 0.5 \end{bmatrix} y_{t-1} + u_t, \quad u_t \sim \text{i.i.d.}\, \mathcal{N}\left(0, \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}\right), \qquad (4.1)$$

with $\alpha_{11} \in \{-0.95, -0.9, -0.5, 0, 0.5, 0.9, 0.95, 1\}$. This type of DGP is referred to as DGP1.

A second DGP (denoted by DGP2) is based on the empirical model described in Section 5. The DGP is a three-dimensional VAR(3) and the errors are assumed to follow a multivariate normal distribution with covariance matrix as estimated on the basis of the data.

The experiments are conducted for different sample sizes $T = 50, 100$ or 200 for samples generated from DGP1 and $T = 100, 200$ or 400 in the case of DGP2. The response horizons considered are $H = 10$ and 20 for DGP1 and $H = 18$ and 36 for DGP2. The number of bootstrap replications is $b = 5000$ because a large number of replications is recommended for confidence intervals in the related literature (e.g., Efron and Tibshirani (1993)). We choose $\gamma = 0.1$ because that leaves room for over- and under-estimation of the nominal coverage level. The number of Monte Carlo replications is 1000. Responses to orthogonal innovations, where B is obtained using a lower-triangular 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 model selection criteria for lag order selection. Clearly the latter approach is more common in practice where the VAR order is typically unknown. In particular, we have used the rather popular AIC criterion for lag order selection. It is more generous than some of the other criteria and is hence perhaps more suitable when the objective is to uncover the dynamics of a process rather than optimize forecast performance. Such considerations may also explain the popularity of AIC in applied structural VAR analysis. When lag order selection criteria are used, the maximum lag length depends on the sample size and is given by 10, 12, 14 and 16 for samples of 50, 100, 200 and 400 observations, respectively. We now discuss the results of our simulation study.

4.1   Results for DGP1

Selected results of the experiments based on DGP1 are given in Tables 1 - 3 and Figures 1 - 4. In the tables values without parentheses are the estimated coverage probabilities of the bands, while the values in parentheses correspond to the widths of the bands, calculated as the sum of the bands' spreads at horizons $0, 1, \ldots, H$. Notice that the coverage rates now refer to the fraction of full impulse response functions for $h = 0, 1, \ldots, H$ being inside the band, that is, the bands are joint confidence regions. Thus, this criterion differs from the usual consideration of fractions of individual coefficients lying within their respective confidence intervals. Of course, one could think of other measures of the size of the bands. The one used here is a natural extension of looking at the widths of confidence intervals for individual impulse response coefficients as considered in Kilian (1998).

Table 1 presents results for DGP1 with selected values of α_{11} obtained for a maximum response horizon of $H = 10$, for samples of size $T = 100$. The VAR order is chosen by AIC. A propagation horizon of $H = 10$ is relatively short in empirical studies. We still consider it here because a large H will be more problematic for some of the methods. For example, in constructing Bonferroni bands the width depends directly on H and it is clear that, at least indirectly, the other bands are also affected by the choice of propagation horizon. Thus, this quantity is critical. We therefore present corresponding results for a longer response horizon ($H = 20$) in Table 2. The nominal coverage rate is 90% throughout. Although there is a lot of variety in the empirical literature with nominal coverage rates ranging from 68% to 95% in most studies, a level of 90% is suitable here because it leaves room for actual rates to be smaller or larger.

Looking first at the results in Table 1, there are some striking differences in the coverage rates obtained by the different methods. First, it is obvious that the naïve and Scheffé bands have coverage rates far below the nominal rate of 90% for some parameter values, for instance if α_{11} is negative. This result is not unexpected for the naïve bands, given the Bonferroni inequality that suggests that the individual confidence levels have to be increased in order to obtain a joint confidence level of 90%. Also, for the Scheffé bands,

the results correspond to what is reported in other related literature (see in particular Wolf and Wunderli (2012) and Staszewska-Bystrova (2011)). For all other bands (B, B-adj, NP, WW) the actual coverage levels come close to the nominal level of 90% for all parameter values considered. In fact, in all cases they are above 80%. As expected, the traditional Bonferroni bands are conservative and tend to produce coverage levels of more than 90%. Also the WW bands are in some cases rather conservative. For example, the response path of the first variable to the second shock is inside the WW band in more than 94% of the cases, except when $\alpha_1 = -0.95$. Generally, there is some variability of the coverage rates when α_{11} changes.

Now when considering the widths of the bands, it turns out that they vary substantially and there is no clear winner. This is, perhaps, more easily seen in Figure 1, where the coverage levels and widths of the confidence bands are depicted for four methods that produce reasonable coverage levels: B, B-adj, NP, and WW. In Figure 1 it becomes apparent that all four methods result in coverage levels close to the nominal one. The B and WW bands tend to be the largest, however. Note that in the figure results for a unit root process ($\alpha_{11} = 1$) are also presented. For that case, the coverage levels of all bands for the responses to the first shock tend to deteriorate compared to stable, stationary processes.

The width of the WW band becomes worse relative to the other bands when the sample size is reduced to $T = 50$. For this case the coverage levels and sizes of the B, B-adj, NP, and WW bands are depicted in Figure 2. It is, in fact, striking that in most cases the Bonferroni bands are considerably smaller than the WW bands. Note, however, that now the actual coverage rates of all bands are less than 90% in most cases and the WW bands sometimes, but not always, come closer to the nominal level than the other bands. To further investigate the impact of the sample size, results for $T = 200$ are shown in Figure 3. These results look rather similar to those for $T = 100$ (see Figure 1). In other words, the coverage levels are close to the nominal one, at least for stationary processes, with the B-adj and NP methods having advantages in terms of the sizes of the bands.

To study the impact of the response horizon H, we present some results for $H = 20$ in Table 2. The sample size is again $T = 100$, the VAR order is chosen by AIC and the nominal level is 90%, as before. Again all methods but the naïve and Scheffé produce reasonable coverage levels of more than 80% for almost all values of α_{11} reported in the table; with the naïve and Scheffé band again having partly unacceptably low coverage. For example, the response of the first variable to the first shock is never fully contained in the Scheffé band when $\alpha_{11} = -0.9$.

Comparing the sizes of the bands of the four most promising methods there is again no clear winner. However, the WW bands are now excessively wide in some cases. This is seen even more clearly in Figure 4, where the coverage levels and sizes of the four best methods are depicted. For example, for $\alpha_{11} = -0.5$ or 0 the WW bands are extremely wide compared to other methods. Clearly the WW method produces overall much larger bands even

than the Bonferroni method, which has a handicap for large H. The situation does not improve much for $T = 200$ (results not shown).[2] Hence, based on these results and taking into account that in empirical work large response horizons are not uncommon, the WW method may not be the best choice for constructing confidence bands for impulse responses.

We have also considered a number of other MC designs, as mentioned earlier. In particular, we have also performed comparisons of the different methods using the true VAR order. Summary results for different designs based on DGP1 and the four best methods are presented in Table 3. In that table the mean absolute deviations of the empirical coverages from the nominal coverage across all values of α_{11} and all impulse responses are given together with the average band sizes over all experiments. Clearly, if the true lag length is used in the simulations, the B-adj and NP methods have overall the smallest deviations from the nominal coverage level and also the smallest band size on average. The B-adj method is a little better in terms of band width while NP has a more precise coverage level. The traditional Bonferroni method tends to lead to wider bands. The same is true for the WW method. It shows considerable variation and results in rather wide bands for long horizons and small sample sizes. By and large the situation remains unchanged when the lag order is estimated. There is more variability in the results in that case, however. Thus, overall we can summarize the results for DGP1 as follows:

1. The naïve and Scheffé methods are, in some situations, very poor in terms of coverage. In particular, the Scheffé method produces extremely low coverage rates for some Monte Carlo designs. Therefore these two methods cannot be recommended if a reasonable coverage rate is desired. A modification of this method, based on the absolute value of the Cholesky factor used in the construction of the band, that improves the performance of the Scheffé method is proposed by Staszewska-Bystrova (2012). We also used this modification in our simulations (results not reported) and found that it does better but is still unsatisfactory.

2. Among the four remaining methods the adjusted Bonferroni and NP bands tend to be the best in terms of deviations from the nominal coverage and width of confidence band.

3. The WW method may result in much wider bands than B, B-adj, and NP, especially when the sample size is small.

4. The traditional Bonferroni method tends to be conservative and, hence, produces coverage levels in excess of the nominal level for many Monte

---

[2]Notice that for $\alpha_{11} = 0$ the response of the first variable to the first shock is zero for propagation horizon $h > 0$. If the true response is estimated fairly precisely, the standard deviation $\sqrt{\widehat{\omega}_{h+1,h+1}^{c*}}$ is small in the denominator of $\widehat{s}_{ij}^{c*}(h)$. Hence, this quantity may become unstable in the WW method, as mentioned in Sec. 3.6. This may explain the excessively large confidence bands produced by the WW method when α_{11} is close to zero.

Carlo designs. Despite this fact, the width of the bands is often competitive with the best methods. Unlike the WW method it does not result in excessively wide bands.

We now turn to the results for the second DGP.

4.2 Results for DGP2

The second DGP is meant to mimic a real-life data generation process. As a three-dimensional VAR(3) process it has a larger dimension and a higher lag order than DGP1. Apart from that the two DGPs are difficult to compare because the parameters of DGP2 are set to the values estimated from a real data set (as described in Section 5). We do not reproduce the actual parameter values here but note that the largest autoregressive root is close to one (0.9681). Hence, the process has some persistence and is, in that respect, comparable to DGP1 with α_{11} close to the unit circle. Coverage rates and band sizes for different sample sizes and propagation horizons $H = 18$ and 36 are presented in Figures 5 and 6. Given the poor performance of the naïve and Scheffé methods for DGP1, we only present results for the B, B-adj, NP, and WW methods.

In Figure 5 the results for response horizon $H = 18$ are depicted. It is obvious that the coverage of the bands leaves something to be desired for the smallest sample size, $T = 100$. In particular, the methods seem to have difficulties with responses of the ith variable to the ith shock, $i = 1, 2, 3$. Apart from that the coverage levels of all methods are impressively close to the nominal level for all sample sizes. In other words one may want to choose between the methods on the basis of the band widths. There is no clear winner in terms of this criterion either, although the B-adj method delivers slightly smaller bands than the other methods in many cases.

If the propagation horizon is longer, there are again situations where the WW method produces much wider bands than the other methods, especially for the smallest sample size ($T = 100$) (see Figure 6). They are not as extreme, however, as for DGP1. Apart from that, it is again difficult to determine a clear winner on the basis of the coverage levels and band widths. A summary with averages across all impulse responses is given in Table 4. In that table it is seen that the B-adj and NP methods tend to be best in terms of coverage, that is, they tend to deliver confidence bands with actual coverage closest to the nominal coverage probability. They also produce the smallest bands on average.

In summary, the results of DGP2 conform with those for DGP1. Note that our simulations for DGP2 are based on a single DGP and vary only with respect to sample size, response horizon, and lag length determination. Therefore they are more limited than those for DGP1 and are perhaps best compared to results with $|\alpha_{11}|$ close to 1. For this more limited setup we find that in most situations all four methods do reasonably well in terms of coverage level. Overall the smallest confidence bands are obtained by B-adj and NP.

4.3 Conclusions from MC Simulations

Thus, from our MC experiment we can draw the following general conclusions:

1. The naïve and Scheffé methods may have very poor coverage rates much smaller than the nominal rate. This is true in particular for the Scheffé method. These two methods cannot be recommended if a good coverage rate is desired. Certainly the Scheffé method should be dismissed for constructing confidence bands for impulse responses. This does not mean, of course, that the method may not be useful, for instance, for testing hypotheses about contrasts in the analysis of variance, as originally proposed by Scheffé (1953).

2. The WW method may give extremely wide confidence bands compared to Bonferroni type and NP methods, especially when the sample size is small. Thus, this method should be used with caution, in particular for relatively small samples.

3. The traditional Bonferroni method tends to produce slightly wider bands than the adjusted Bonferroni and NP methods, especially if the propagation horizon is large.

4. The B-adj and NP methods perform well in terms of coverage level and band width for most scenarios. An exception are unit root processes when the sample size is small.

Thus, overall the MC experiment confirms that the B-adj method is quite competitive and in some respects even best. While it has good coverage it also leads to reasonably small bands. The NP method is the closest competitor, however, so far it lacks a theoretical basis. An empirical illustration of how the different methods compare in practice is given next.

5 Illustrative Example

In this section the different methods for constructing confidence bands are applied in the context of a structural VAR model previously analyzed by Kilian (2009). It was used to investigate the world crude oil market. The dataset is the same as that used by Kilian (2009) and consists of monthly observations for the period 1973:1-2007:12. In other words, we have a sample size of $T = 420$. The variables are 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 consider a VAR(3) model with intercepts suggested by AIC while Kilian uses a VAR(24) in his study. The estimated VAR(3) model is the same as DGP2. Recall that the model has some persistence, the modulus of the largest autoregressive root being 0.9681. Following Kilian we use a response horizon of $H = 18$ but also consider a longer horizon of $H = 36$.

As in Kilian, a recursive identification scheme with lower-triangular B matrix obtained from a Cholesky decomposition of the reduced form residual covariance matrix is used. The estimated structural impulse response functions for $H = 18$ and $H = 36$ are presented in Figures 7 and 8, respectively. Confidence bands obtained with the B, B-adj, NP, and WW methods are shown in the figures.

Figure 7 reflects what was seen in the MC results for DGP2. The different methods give bands of different size for different impulse responses. Of course, it is not clear from this example which bands are most reliable. However, the bands for responses to the first shock (the left-hand column of Figure 7) from all methods are very similar. The same is not true for the responses to the second and third shocks. For these the B-adj and NP bands tend to be the smallest, in line with the simulation results. In fact, if one judges significance of the response by the bands covering zero or not, the B band suggests that there is no significant response of the price of oil to the oil market specific demand shock (the third shock). On the other hand, using the same criterion, a significant response is diagnosed when the B-adj or NP bands are considered. Thus, it can make a difference for the interpretation of the results which method is used for constructing confidence bands around the impulse responses.

The situation is similar in Figure 8 where the impulse responses and estimated 90% confidence bands for propagation horizon $H = 36$ are presented. In this case increasing the propagation horizon does not increase the width of the bands very much. The confidence bands are similar to those in Figure 7 for the first 18 propagation periods. The B-adj and NP bands are among the smallest for all impulse responses.

The example is meant to illustrate the different methods for constructing confidence bands for impulse responses. It shows that the differences can be remarkable. Hence, in applied work the choice of confidence bands may matter, as can be seen in our particular example.

6   Conclusions

Impulse response functions are popular tools in structural vector autoregressive analysis. They are used to investigate the reactions of the variables of a VAR process to specific shocks. Typically the propagation of a shock is traced over a number of periods and it is desirable to construct confidence bands for these propagation paths for a prespecified confidence level. A number of proposals for constructing such bands in a classical setting are reviewed and it is argued that they either may not obtain the desired coverage level, are conservative or lack a theoretical justification based on asymptotic theory. We have also proposed an adjustment of the Bonferroni method that at least partly accounts for these deficiencies. Our adjusted Bonferroni band is shown to have some practical advantages.

In a large scale MC study the methods are compared and the adjusted Bonferroni method is found to perform very well in terms of coverage level

and width of the band. Hence, it can be recommended for applied work. In our simulations we also find that the so-called Scheffé bands proposed by Jordà (2009) and Jordà and Marcellino (2010) may lead to strongly distorted bands and should not be used for constructing confidence bands around impulse responses. Also the standard bands that just connect confidence intervals for individual impulse response coefficients may have a very distorted coverage substantially below the nominal level. Moreover, bands adapted from ideas of Wolf and Wunderli (2012) in the context of multiple horizon forecasting have satisfactory coverage levels but may be even wider than the very conservative traditional Bonferroni bands in some situations. Finally, the NP bands proposed by Staszewska (2007) have comparable coverage and size characteristics to the adjusted Bonferroni bands in our simulations.

Since impulse responses are conditional forecasts, the adjusted Bonferroni method can also be used for constructing confidence bands for multiple horizon forecasts. It is not obvious that it will perform better than other methods in that context because the uncertainty in such forecasts is partly due to factors that are not present in impulse responses. It is left to future research to investigate the suitability of the method for constructing forecast bands.

A   Appendix: Methods Used

In this appendix we present some technical matters following Staszewska-Bystrova and Winker (2013) related to the techniques presented and discussed in the main body of the text. In particular, we present details on the bias-corrected estimates for the VAR parameters, asymptotic and bootstrap estimation methods.

A.1   Bias-corrected VAR estimates

The OLS estimators of the VAR parameters are denoted by $\widehat{\nu}, \widehat{A}_1, \ldots, \widehat{A}_p$ and the corresponding impulse responses by $\widehat{\Phi}_0, \ldots, \widehat{\Phi}_H$. In contrast, $\widehat{\nu}^c, \widehat{A}_1^c, \ldots, \widehat{A}_p^c$ stand for the estimators with the bias-correction based on the asymptotic bias formula given by Nicholls and Pope (1988) and Pope (1990) and $\widehat{\Phi}_0^c, \ldots, \widehat{\Phi}_H^c$ denote the resulting estimators of the response coefficients. When calculating the bias-corrected estimates, the stationarity correction, described by Kilian (1998), is used.

The following notation and presentation follows closely Kim (2004). The bias-correction is based on the VAR(1) representation of (2.1) for the mean-corrected variables $x_t = y_t - \mathbb{E}(y_t)$. The representation is given by:

$$W_t = \Pi W_{t-1} + U_t, \tag{A.1}$$

where $W_t = (x_t', \ldots, x_{t-p+1}')'$ and $U_t = (u_t', 0, \ldots, 0)'$ have dimensions $(Kp \times 1)$, and Π is a $(Kp \times Kp)$ matrix given by

$$\Pi = \begin{bmatrix} A_* & A_p \\ I & \mathbf{0} \end{bmatrix},$$

where I is a $K(p-1)$-dimensional identity matrix, $\mathbf{0}$ is a $(K(p-1) \times K)$ null matrix and $A_* = [A_1, \ldots, A_{p-1}]$. Then, the small sample bias of the least squares estimator $\widehat{\Pi}$ for Π is given by:

$$B_\Pi = -b_\Pi/T + O(T^{-3/2}),\tag{A.2}$$

where $b_\Pi = G[(I - \Pi')^{-1} + \Pi'(I - \Pi'^2)^{-1} + \sum_{\lambda \in Spec(\Pi)} \lambda(I - \lambda\Pi')^{-1}]\Gamma(0)^{-1}$, $\Gamma(0) = \mathrm{E}(W_t W_t')$, $G = \mathrm{E}(U_t U_t')$ and $Spec(\Pi)$ stands for the set of eigenvalues λ of the Π matrix. The bias-corrected parameter estimators are obtained using $\widehat{\Pi}^c = \widehat{\Pi} - \widehat{B}$, where $\widehat{B} = -\widehat{b}_{\widehat{\Pi}}/T$.

The $\widehat{\Pi}^c$ matrix contains as sub-matrices the bias-corrected least squares estimators $\widehat{A}_1^c, \ldots, \widehat{A}_p^c$, while the bias-corrected estimator $\widehat{\nu}^c$ can be obtained from $\overline{y} - \widehat{A}_1^c \overline{y} \ldots - \widehat{A}_p^c \overline{y}$, where $\overline{y}$ stands for the vector of sample means.

A.2 Asymptotic Confidence Bands

Let $\boldsymbol{\phi}_{ij} = (\phi_{ij,0}, \ldots, \phi_{ij,H})$ be the $(H+1)$-dimensional vector of responses of variable i to the jth shock. If the impact effect of the shock is restricted to zero it is sometimes useful to consider only the vector of the last H elements. It will be denoted as $\boldsymbol{\phi}_{ij,1,\ldots,H} = (\phi_{ij,1}, \ldots, \phi_{ij,H})$.

The Scheffé bands are based on the asymptotic distributions of the impulse responses. Hence, they are based on the multivariate normal approximation

$$\sqrt{T}(\widehat{\boldsymbol{\phi}}_{ij} - \boldsymbol{\phi}_{ij}) \xrightarrow{d} \mathcal{N}(0, \Omega_{ij}).$$

The corresponding covariance estimates $\widehat{\Omega}_{ij}$ and $\widehat{\Omega}_{ij}^c$ are obtained using least-squares and bias-corrected VAR estimates, respectively, and the results from Lütkepohl (2005, Sec. 3.7.1). The estimators $\widehat{\Sigma}_u$ and $\widehat{\Sigma}_u^c$ for the residual covariance matrix are based on the formula (3.2.18) in Lütkepohl (2005).

A.3 Bootstrap Confidence Bands

The bootstrap methods are based on b bootstrap replicates of $\widehat{\boldsymbol{\phi}}_{ij}^c$, denoted by $\widehat{\boldsymbol{\phi}}_{ij}^{c*}$. These are obtained as follows:

- The vectors of residuals corresponding to $\widehat{\nu}^c, \widehat{A}_1^c, \ldots, \widehat{A}_p^c$ are calculated and rescaled by a factor of $[T/(T - Kp - 1)]^{1/2}$ (see, e.g., Davidson and MacKinnon (2004)). The rescaled residuals are denoted by $\{\widehat{u}_t^c\}$.

- Bootstrap samples are generated from

$$y_t^* = \widehat{\nu}^c + \widehat{A}_1^c y_{t-1}^* + \ldots + \widehat{A}_p^c y_{t-p}^* + u_t^*.\tag{A.3}$$

For the required p initial observations of y_t^* it is assumed that they correspond to the first p observed values from the real data. As usual, the u_t^* are randomly drawn with replacement from the residuals $\widehat{u}_t^c$. Once a bootstrap sample is generated, a VAR(p) model is estimated. By $\widehat{\nu}^*, \widehat{A}_1^*, \ldots, \widehat{A}_p^*$ we denote the resulting least squares estimators. The

bias correction is applied to these estimates resulting in $\widehat{\nu}^{c*}, \widehat{A}_1^{c*}, \ldots, \widehat{A}_p^{c*}$. If necessary, the stationarity correction of Kilian (1998) is applied. Then the covariance matrix estimator $\widehat{\Sigma}_{u^*}^c$ is calculated.

- Structural impulse responses, $\widehat{\phi}_{ij}^{c*}$, are derived based on $\widehat{\nu}^{c*}, \widehat{A}_1^{c*}, \ldots, \widehat{A}_p^{c*}$ and $\widehat{\Sigma}_{u^*}^c$.

References

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

Davidson, R. and MacKinnon, J. G. (2004). *Econometric Theory and Methods*, Oxford University Press, New York.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Fry, R. and Pagan, A. (2011). Sign restrictions in structural vector autoregressions: A critical review, *Journal of Economic Literature* **49**: 938–960.

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

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

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

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

Nicholls, D. F. and 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.

Scheffé, H. (1953). A method for judging all contrasts in the analysis of variance, *Biometrika* **40**: 87–104.

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* **80**: 721–735.

Staszewska-Bystrova, A. (2012). Modified Scheffé's prediction bands, *Unpublished manuscript*, University of Lodz, Lodz, Poland.

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

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 1: Estimated Coverage Probabilities and Total Widths (in brackets) for Nominal 90% Confidence Bands for Impulse Responses of DGP1 for Response Horizon $H = 10$ (based on sample size $T = 100$, 1000 MC replications and 5000 bootstrap replications, lag length estimated using AIC)

band	first shock y_1		first shock y_2		second shock y_1		second shock y_2	
				$\alpha_{11} = -0.95$				
naïve	73.30	(5.29)	61.40	(2.47)	85.30	(1.77)	58.00	(1.76)
Scheffé	0.00	(0.48)	1.20	(1.72)	0.00	(0.28)	9.50	(1.72)
B	91.60	(8.31)	91.50	(3.92)	97.50	(2.92)	89.70	(2.88)
B-adj	84.90	(5.99)	87.80	(3.36)	90.00	(2.00)	85.30	(2.34)
NP	82.00	(6.35)	87.70	(3.46)	89.20	(2.08)	87.30	(2.49)
WW	86.10	(7.36)	87.50	(3.72)	92.70	(2.08)	86.40	(2.64)
				$\alpha_{11} = -0.9$				
naïve	71.90	(5.20)	58.80	(2.46)	84.90	(1.51)	55.60	(1.73)
Scheffé	0.00	(0.44)	1.00	(1.73)	0.00	(0.30)	18.40	(1.70)
B	94.10	(7.87)	92.60	(3.84)	97.60	(2.52)	88.80	(2.83)
B-adj	84.80	(5.96)	86.50	(3.35)	89.60	(1.76)	84.70	(2.27)
NP	85.10	(6.29)	88.00	(3.44)	88.50	(1.82)	86.90	(2.41)
WW	88.20	(7.36)	88.10	(3.72)	94.90	(1.82)	86.10	(2.61)
				$\alpha_{11} = -0.5$				
naïve	64.90	(1.77)	57.50	(1.84)	81.90	(0.65)	66.20	(1.71)
Scheffé	0.70	(0.56)	50.70	(2.07)	35.70	(0.43)	58.90	(1.83)
B	93.50	(2.86)	93.40	(3.01)	96.70	(1.11)	91.80	(2.78)
B-adj	86.50	(2.28)	88.10	(2.58)	94.10	(0.89)	85.90	(2.15)
NP	88.10	(2.39)	89.20	(2.68)	87.00	(0.96)	88.00	(2.23)
WW	86.40	(4.32)	92.30	(3.64)	96.80	(1.06)	85.30	(4.68)
				$\alpha_{11} = 0.5$				
naïve	66.50	(2.09)	68.20	(3.21)	83.50	(1.22)	65.40	(2.21)
Scheffé	54.50	(2.17)	61.60	(3.78)	55.70	(1.16)	59.50	(2.44)
B	94.40	(3.42)	95.90	(5.15)	97.20	(2.05)	92.00	(3.66)
B-adj	90.10	(2.79)	89.60	(4.25)	93.70	(1.50)	86.90	(2.86)
NP	90.70	(3.05)	90.60	(4.55)	90.10	(1.61)	88.30	(3.12)
WW	91.40	(3.35)	88.90	(5.19)	97.20	(1.93)	89.90	(3.62)
				$\alpha_{11} = 0.9$				
naïve	70.80	(5.80)	65.50	(5.81)	84.60	(3.17)	63.10	(3.57)
Scheffé	65.40	(6.95)	65.80	(7.22)	86.90	(3.35)	72.20	(4.05)
B	89.80	(8.76)	91.90	(8.86)	96.80	(5.17)	90.20	(5.84)
B-adj	82.40	(6.81)	85.00	(7.08)	90.60	(3.60)	84.70	(4.50)
NP	82.20	(7.33)	84.70	(7.65)	88.70	(3.78)	86.90	(4.76)
WW	85.00	(8.67)	86.70	(8.66)	94.90	(3.92)	89.10	(5.02)
				$\alpha_{11} = 0.95$				
naïve	70.50	(6.24)	65.60	(5.94)	87.40	(3.61)	65.10	(3.83)
Scheffé	63.60	(7.80)	65.70	(7.60)	87.50	(3.89)	74.00	(4.40)
B	90.00	(9.53)	90.70	(9.20)	98.40	(5.86)	90.70	(6.24)
B-adj	82.30	(7.19)	82.60	(7.18)	93.80	(4.01)	86.30	(4.77)
NP	84.30	(7.88)	83.60	(7.88)	91.00	(4.20)	89.40	(5.04)
WW	85.10	(9.69)	86.30	(9.14)	96.50	(4.48)	88.30	(5.46)

Table 2: Estimated Coverage Probabilities and Total Widths (in brackets) for Nominal 90% Confidence Bands for Impulse Responses of DGP1 for Response Horizon $H = 20$ (based on sample size $T = 100$, 1000 MC replications and 5000 bootstrap replications, lag length estimated using AIC)

band	first shock y_1		y_2		second shock y_1		y_2	
				$\alpha_{11} = -0.95$				
naïve	72.60	(12.84)	61.00	(4.90)	85.30	(3.18)	57.90	(2.28)
Scheffé	0.00	(0.50)	0.00	(1.76)	0.00	(0.31)	0.50	(1.78)
B	93.20	(20.09)	94.20	(7.91)	98.70	(6.01)	92.90	(4.19)
B-adj	85.10	(14.10)	86.20	(6.21)	92.90	(3.66)	87.00	(3.14)
NP	76.90	(14.96)	79.50	(6.66)	92.20	(3.85)	88.30	(3.48)
WW	87.00	(22.00)	88.50	(8.80)	96.50	(4.16)	89.60	(3.69)
				$\alpha_{11} = -0.9$				
naïve	71.40	(11.07)	58.10	(4.36)	84.80	(2.43)	55.00	(2.10)
Scheffé	0.00	(0.45)	0.10	(1.79)	0.00	(0.34)	2.60	(1.75)
B	95.00	(17.21)	94.40	(7.04)	98.40	(4.76)	91.50	(3.89)
B-adj	84.30	(12.23)	86.70	(5.60)	92.80	(2.91)	86.90	(2.88)
NP	85.90	(12.94)	87.90	(5.96)	93.00	(3.09)	87.50	(3.21)
WW	91.10	(19.75)	91.10	(7.98)	98.10	(3.45)	90.30	(3.61)
				$\alpha_{11} = -0.5$				
naïve	63.00	(1.93)	54.90	(2.00)	81.80	(0.72)	65.60	(1.87)
Scheffé	0.30	(0.58)	27.10	(2.14)	15.10	(0.45)	30.30	(1.90)
B	95.30	(3.51)	95.60	(3.69)	98.00	(1.42)	94.60	(3.46)
B-adj	88.50	(2.53)	90.50	(2.86)	95.40	(1.03)	87.20	(2.40)
NP	89.10	(2.67)	89.60	(3.01)	87.40	(1.14)	88.80	(2.50)
WW	90.70	(61.40)	96.50	(84.94)	98.00	(13.62)	89.70	(95.68)
naïve	65.70	(2.38)	67.20	(3.79)	81.80	(1.38)	64.80	(2.54)
Scheffé	31.90	(2.28)	42.60	(4.03)	26.20	(1.21)	34.40	(2.57)
B	95.70	(4.50)	97.00	(7.04)	98.10	(2.73)	94.20	(4.93)
B-adj	92.00	(3.22)	91.40	(5.07)	95.40	(1.74)	89.20	(3.32)
NP	91.60	(3.58)	91.60	(5.51)	91.10	(1.92)	89.20	(3.69)
WW	94.40	(11.19)	89.70	(16.62)	98.80	(7.52)	92.80	(12.67)
				$\alpha_{11} = 0.9$				
naïve	70.10	(12.15)	64.50	(12.68)	84.40	(5.29)	62.70	(6.02)
Scheffé	52.70	(12.35)	54.50	(13.41)	82.70	(5.08)	69.80	(6.06)
B	92.30	(19.01)	93.90	(19.92)	98.50	(10.22)	93.20	(11.50)
B-adj	83.60	(13.88)	85.00	(14.89)	93.90	(6.28)	87.90	(7.63)
NP	81.30	(15.14)	82.50	(16.22)	92.10	(6.68)	90.50	(8.08)
WW	87.60	(30.39)	88.20	(28.81)	98.30	(9.07)	95.40	(10.56)
				$\alpha_{11} = 0.95$				
naïve	69.90	(14.49)	65.10	(14.36)	87.30	(6.63)	64.90	(7.07)
Scheffé	53.90	(16.47)	54.00	(16.65)	86.50	(6.81)	73.30	(7.47)
B	92.60	(22.10)	93.10	(22.18)	99.30	(12.55)	93.50	(13.30)
B-adj	82.50	(16.27)	82.50	(16.61)	95.30	(7.63)	87.90	(8.74)
NP	83.90	(18.37)	83.80	(18.67)	93.30	(8.07)	91.50	(9.19)
WW	88.80	(36.92)	88.20	(33.61)	99.50	(11.13)	93.60	(12.34)

Table 3: Mean Absolute Deviations of Empirical Coverage from the Nominal Coverage and Mean Widths of the Bands Calculated Across all Values of α_{11} and all Combinations of Shocks and Response Variables for DGP1

		mean absolute deviations				mean band width			
		B	B-adj	NP	WW	B	B-adj	NP	WW
		true lag length							
$T = 50$	$H = 10$	4.35	5.70	3.86	4.81	6.78	5.09	5.60	9.73
$T = 100$	$H = 10$	5.39	3.68	2.44	3.75	4.72	3.55	3.81	4.85
$T = 200$	$H = 10$	5.81	2.63	1.24	2.50	3.26	2.44	2.59	2.88
$T = 50$	$H = 20$	5.48	5.67	4.62	6.44	13.97	9.19	10.50	189.76
$T = 100$	$H = 20$	6.56	4.43	4.08	5.17	9.48	6.55	7.17	42.94
$T = 200$	$H = 20$	7.18	2.96	1.97	4.23	6.63	4.53	4.85	10.75
		estimated lag length							
$T = 50$	$H = 10$	7.14	10.46	8.14	8.41	7.64	6.02	6.50	9.94
$T = 100$	$H = 10$	3.68	4.44	3.20	4.01	4.98	3.82	4.08	5.01
$T = 200$	$H = 10$	4.51	3.18	2.19	2.67	3.38	2.57	2.72	2.99
$T = 50$	$H = 20$	5.78	9.80	7.71	6.63	16.55	11.44	12.74	148.75
$T = 100$	$H = 20$	5.32	4.72	4.06	4.43	10.03	7.03	7.68	38.47
$T = 200$	$H = 20$	6.11	3.01	2.35	3.74	6.85	4.73	5.06	10.51

Table 4: Mean Absolute Deviations of Empirical Coverage from the Nominal Coverage and Mean Widths of the Bands Calculated Across all Combinations of Shocks and Response Variables for DGP2

|		mean absolute deviations				mean band width			
		B	B-adj	NP	WW	B	B-adj	NP	WW
		true lag length							
$T = 100$	$H = 18$	6.53	7.24	5.86	5.10	133.79	98.24	106.01	126.94
$T = 200$	$H = 18$	5.83	3.27	2.16	2.31	90.59	66.33	71.10	75.52
$T = 400$	$H = 18$	6.74	1.82	0.99	1.31	61.09	44.64	47.61	48.28
$T = 100$	$H = 36$	6.40	7.48	6.46	4.66	321.18	202.99	225.69	416.85
$T = 200$	$H = 36$	6.77	3.50	2.82	3.02	212.61	141.71	153.97	192.97
$T = 400$	$H = 36$	7.81	2.36	1.34	1.36	142.42	96.06	102.73	112.20
		estimated lag length							
$T = 100$	$H = 18$	5.66	9.81	8.83	10.01	131.99	95.96	103.84	125.37
$T = 200$	$H = 18$	5.47	5.84	5.99	6.27	90.72	66.09	70.87	75.11
$T = 400$	$H = 18$	6.19	1.62	1.31	1.49	61.33	44.87	47.84	48.50
$T = 100$	$H = 36$	5.07	8.79	8.60	7.91	317.16	199.78	222.23	419.07
$T = 200$	$H = 36$	5.80	5.54	5.88	6.57	212.84	141.47	153.60	191.21
$T = 400$	$H = 36$	7.23	2.00	1.68	1.46	142.84	96.39	103.05	112.53
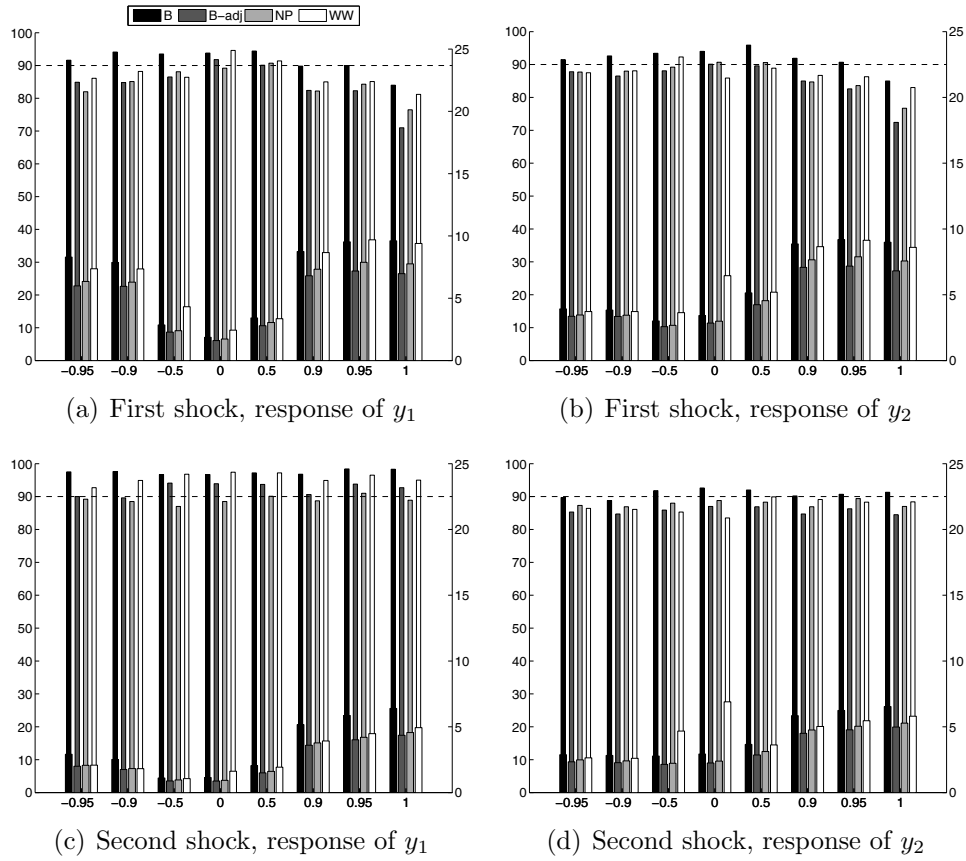
Figure 1: Coverage (narrower bars, left axis) and width (wider bars, right axis) of the 90% Bonferroni (B), adjusted Bonferroni (B-adj), neighbouring paths (NP), and Wolf & Wunderli (WW) bands for DGP1 (estimated lag length, $T = 100$ and $H = 10$).
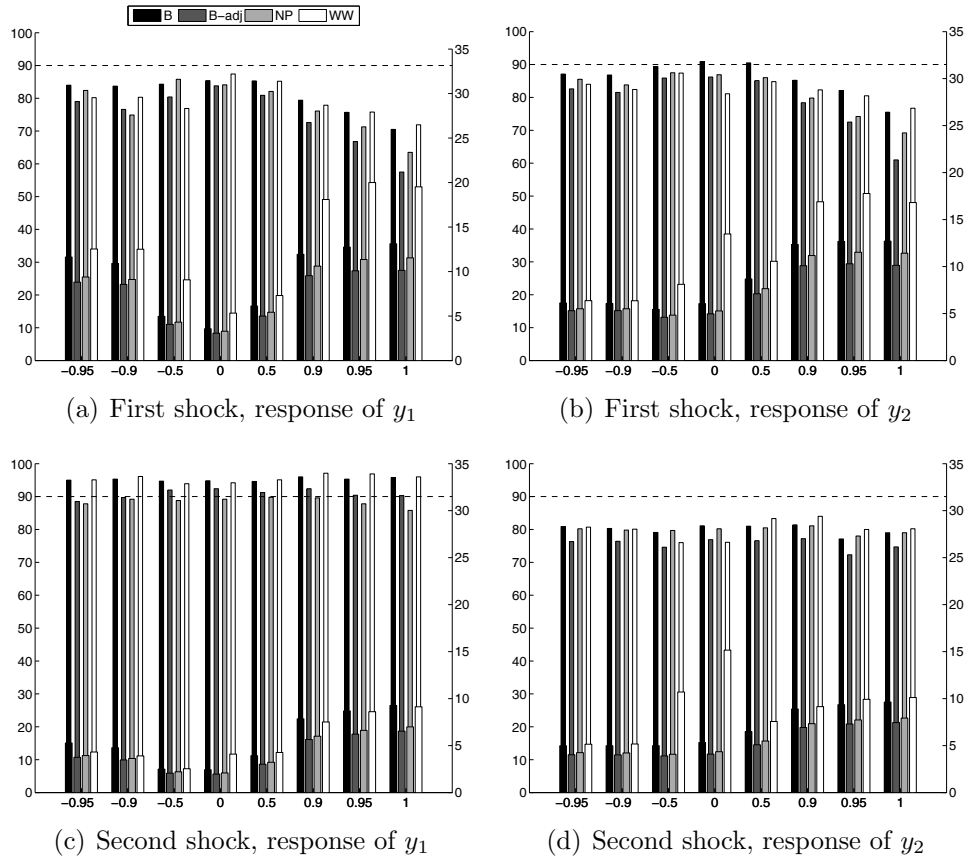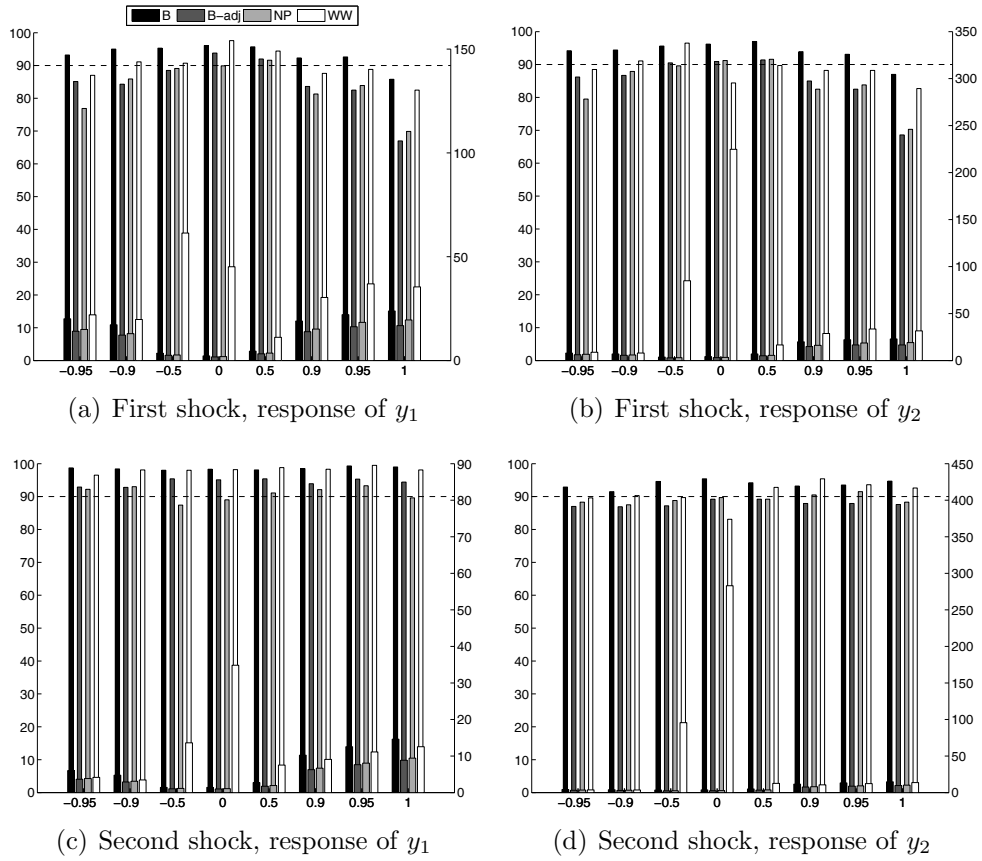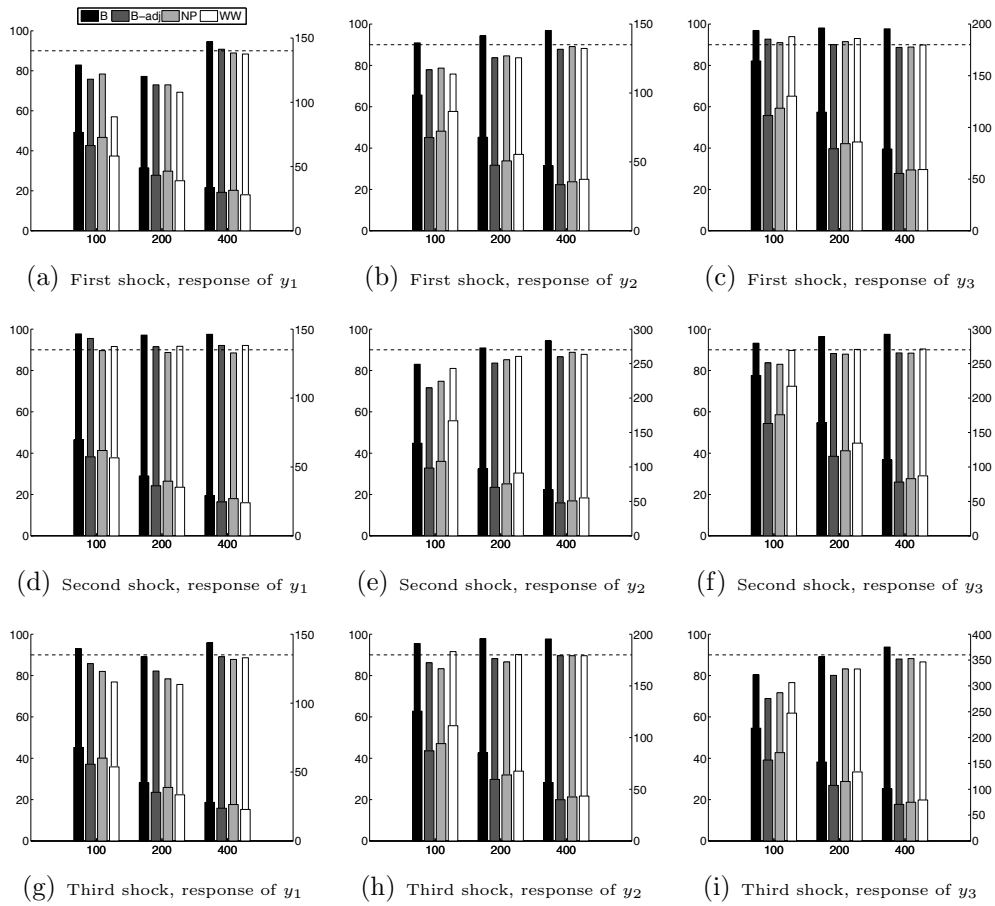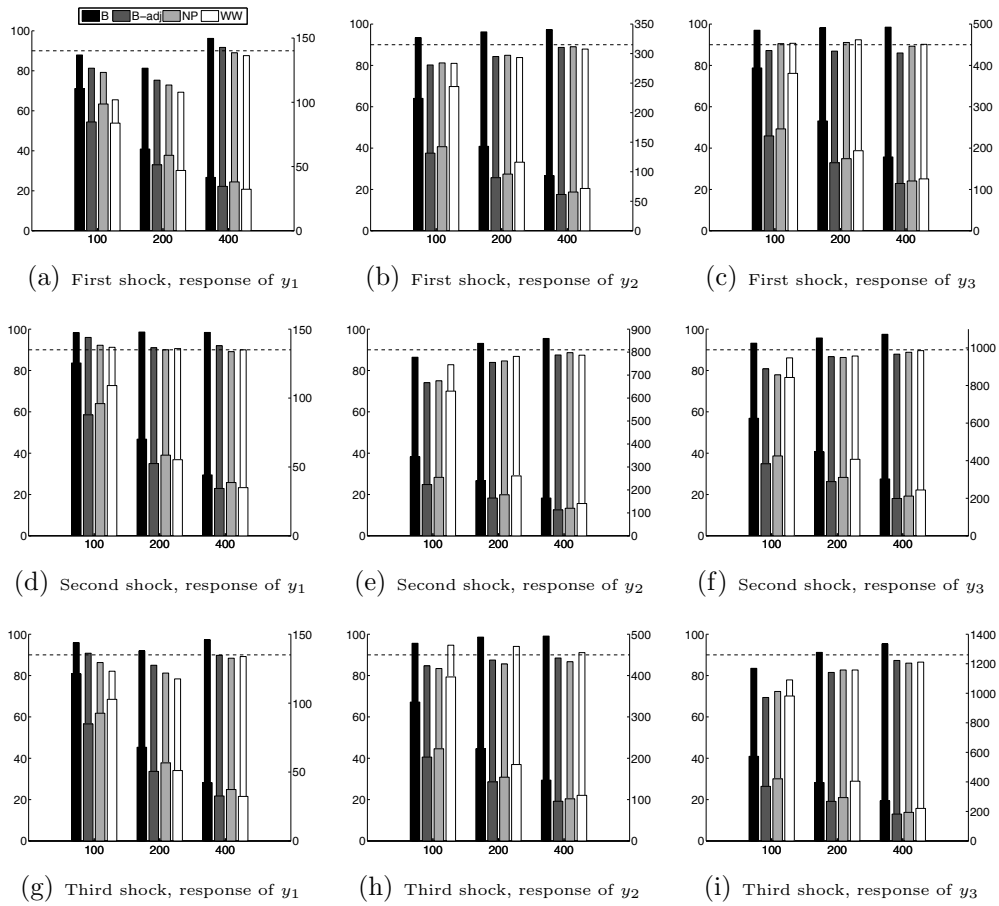
Figure 2: Coverage (narrower bars, left axis) and width (wider bars, right axis) of the 90% Bonferroni (B), adjusted Bonferroni (B-adj), neighbouring paths (NP), and Wolf & Wunderli (WW) bands for DGP1 (estimated lag length, $T = 50$ and $H = 10$).
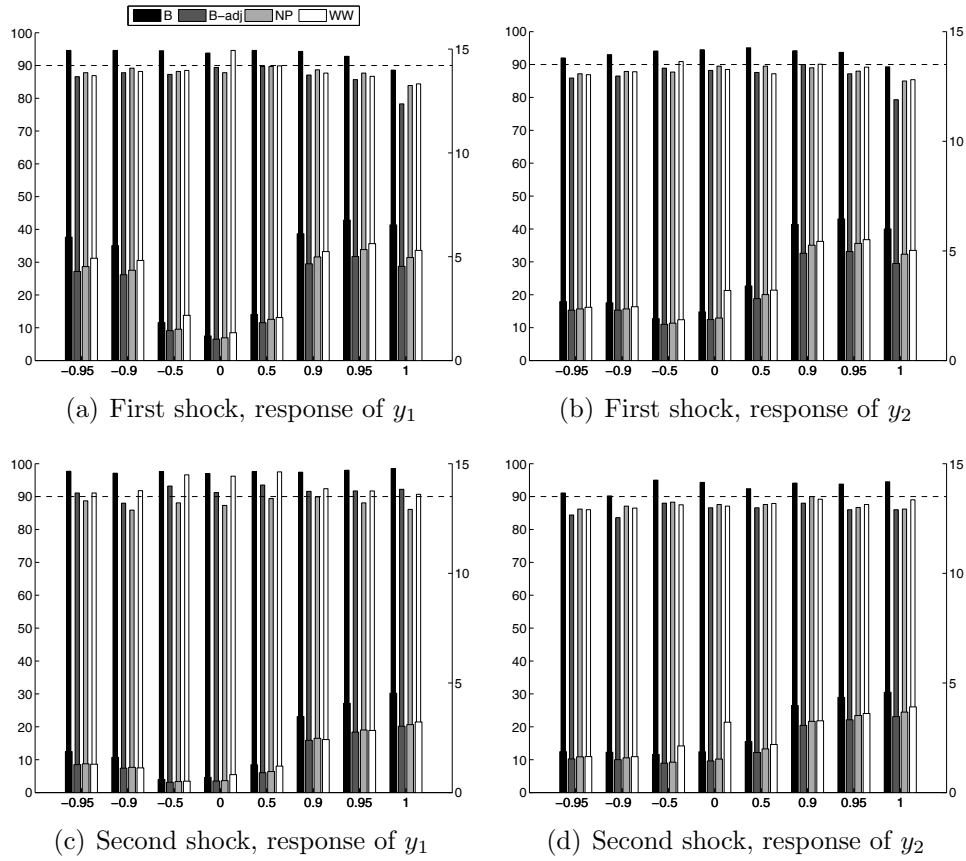
Figure 3: Coverage (narrower bars, left axis) and width (wider bars, right axis) of the 90% Bonferroni (B), adjusted Bonferroni (B-adj), neighbouring paths (NP), and Wolf & Wunderli (WW) bands for DGP1 (estimated lag length, $T = 200$ and $H = 10$).

Figure 4: Coverage (narrower bars, left axis) and width (wider bars, right axis) of the 90% Bonferroni (B), adjusted Bonferroni (B-adj), neighbouring paths (NP), and Wolf & Wunderli (WW) bands for DGP1 (estimated lag length, $T = 100$ and $H = 20$).

Figure 5: Coverage (narrower bars, left axis) and width (wider bars, right axis) of the 90% Bonferroni (B), adjusted Bonferroni (B-adj), neighbouring paths (NP), and Wolf & Wunderli (WW) bands for DGP2 (estimated lag length and $H = 18$).

Figure 6: Coverage (narrower bars, left axis) and width (wider bars, right axis) of the 90% Bonferroni (B), adjusted Bonferroni (B-adj), neighbouring paths (NP), and Wolf & Wunderli (WW) bands for DGP2 (estimated lag length and $H = 36$).
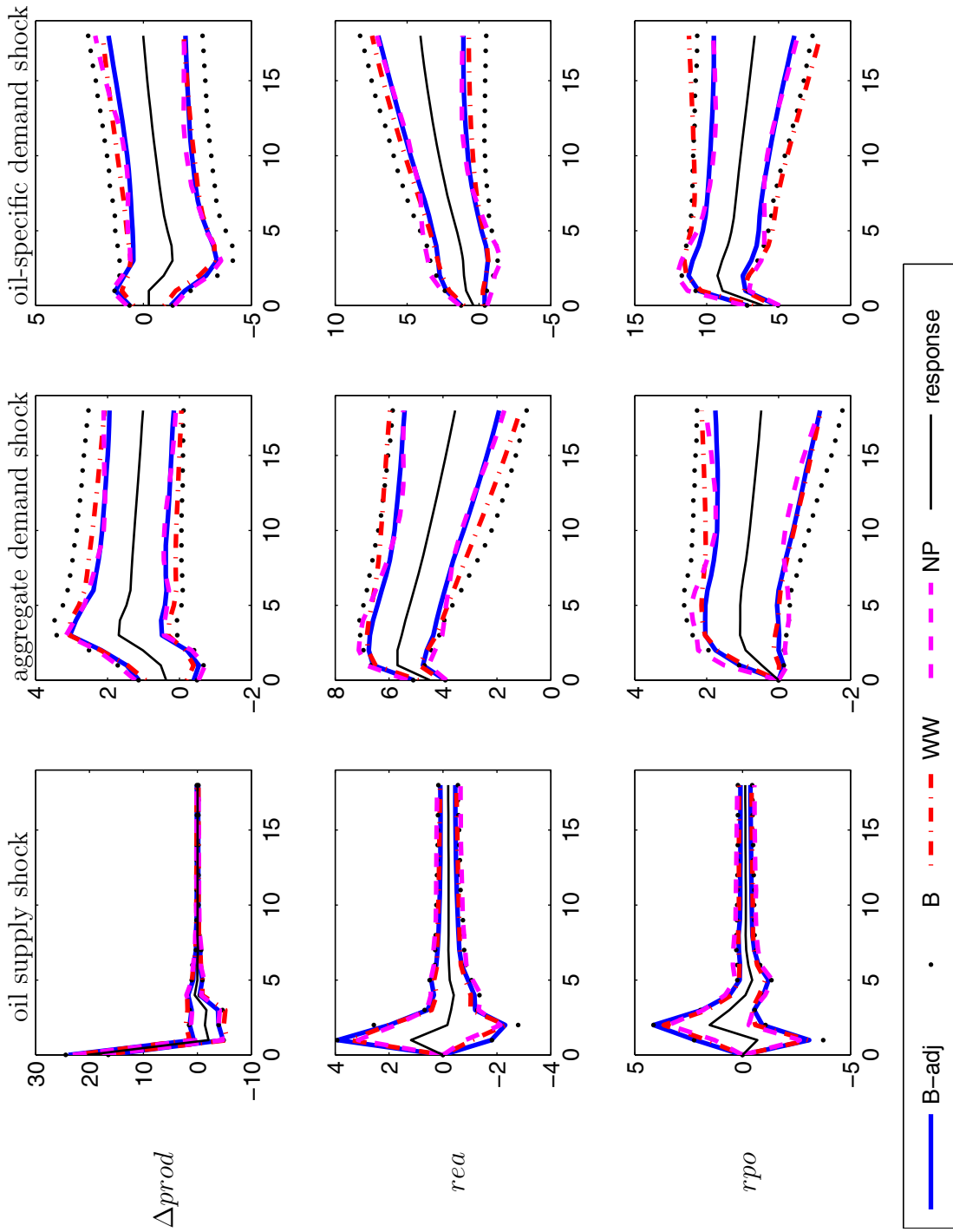
Figure 7: Orthogonalized impulse response analysis for the empirical VAR(3) model and $H = 18$. Responses of $\Delta prod$, rea and rpo to the oil supply shock, the aggragate demand shock and the oil-specific demand shock are given, respectively, in the first, second and third rows of the graph.

Figure 8: Orthogonalized impulse response analysis for the empirical VAR(3) model and $H = 36$. Responses of $\Delta prod$, rea and rpo to the oil supply shock, the aggregate demand shock and the oil-specific demand shock are given, respectively, in the first, second and third rows of the graph.

SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "Functional Data Analysis of Generalized Quantile Regressions" by Mengmeng Guo, Lhan Zhou, Jianhua Z. Huang and Wolfgang Karl Härdle, January 2013.

002 "Statistical properties and stability of ratings in a subset of US firms" by Alexander B. Matthies, January 2013.

003 "Empirical Research on Corporate Credit-Ratings: A Literature Review" by Alexander B. Matthies, January 2013.

004 "Preference for Randomization: Empirical and Experimental Evidence" by Nadja Dwenger, Dorothea Kübler and Georg Weizsäcker, January 2013.

005 "Pricing Rainfall Derivatives at the CME" by Brenda López Cabrera, Martin Odening and Matthias Ritter, January 2013.

006 "Inference for Multi-Dimensional High-Frequency Data: Equivalence of Methods, Central Limit Theorems, and an Application to Conditional Independence Testing" by Markus Bibinger and Per A. Mykland, January 2013.

007 "Crossing Network versus Dealer Market: Unique Equilibrium in the Allocation of Order Flow" by Jutta Dönges, Frank Heinemann and Tijmen R. Daniëls, January 2013.

008 "Forecasting systemic impact in financial networks" by Nikolaus Hautsch, Julia Schaumburg and Melanie Schienle, January 2013.

009 "'I'll do it by myself as I knew it all along': On the failure of hindsight-biased principals to delegate optimally" by David Danz, Frank Hüber, Dorothea Kübler, Lydia Mechtenberg and Julia Schmid, January 2013.

010 "Composite Quantile Regression for the Single-Index Model" by Yan Fan, Wolfgang Karl Härdle, Weining Wang and Lixing Zhu, February 2013.

011 "The Real Consequences of Financial Stress" by Stefan Mittnik and Willi Semmler, February 2013.

012 "Are There Bubbles in the Sterling-dollar Exchange Rate? New Evidence from Sequential ADF Tests" by Timo Bettendorf and Wenjuan Chen, February 2013.

013 "A Transfer Mechanism for a Monetary Union" by Philipp Engler and Simon Voigts, March 2013.

014 "Do High-Frequency Data Improve High-Dimensional Portfolio Allocations?" by Nikolaus Hautsch, Lada M. Kyj and Peter Malec, March 2013.

015 "Cyclical Variation in Labor Hours and Productivity Using the ATUS" by Michael C. Burda, Daniel S. Hamermesh and Jay Stewart, March 2013.

016 "Quantitative forward guidance and the predictability of monetary policy – A wavelet based jump detection approach –" by Lars Winkelmann, April 2013.

017 "Estimating the Quadratic Covariation Matrix from Noisy Observations: Local Method of Moments and Efficiency" by Markus Bibinger, Nikolaus Hautsch, Peter Malec and Markus Reiss, April 2013.

018 "Fair re-valuation of wine as an investment" by Fabian Y.R.P. Bocart and Christian M. Hafner, April 2013.

019 "The European Debt Crisis: How did we get into this mess? How can we get out of it?" by Michael C. Burda, April 2013.

SFB 649 Discussion Paper Series 2013

For a complete list of Discussion Papers published by the SFB 649,
please visit http://sfb649.wiwi.hu-berlin.de.