DIW BERLIN

1750

Discussion Papers

Deutsches Institut für Wirtschaftsforschung 2018

Bootstrapping Impulse Responses of Structural Vector Autoregressive Models Identified through GARCH

Helmut Lütkepohl and Thore Schlaak

IMPRESSUM

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

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

ISSN electronic edition 1619-4535

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

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

Bootstrapping Impulse Responses of Structural Vector Autoregressive Models Identified through GARCH

Helmut Lütkepohl
DIW Berlin and Freie Universität Berlin, Mohrenstr. 58, 10117 Berlin, Germany
email: hluetkepohl@diw.de

and

Thore Schlaak
DIW Berlin, Mohrenstr. 58, 10117 Berlin, Germany
email: tschlaak@diw.de

Revised Version: January 29, 2019[1]

Abstract. Different bootstrap methods and estimation techniques for inference for structural vector autoregressive (SVAR) models identified by generalized autoregressive conditional heteroskedasticity (GARCH) are reviewed and compared in a Monte Carlo study. The bootstrap methods considered are a wild bootstrap, a moving blocks bootstrap and a GARCH residual based bootstrap. Estimation is done by Gaussian maximum likelihood, a simplified procedure based on univariate GARCH estimations and a method that does not re-estimate the GARCH parameters in each bootstrap replication. The latter method is computationally more efficient than the other methods and it is competitive with the other methods and often leads to the smallest confidence sets without sacrificing coverage precision. An empirical model for assessing monetary policy in the U.S. is considered as an example. It is found that the different inference methods for impulse responses lead to qualitatively very similar results.

Key Words: Structural vector autoregression, conditional heteroskedasticity, GARCH, identification via heteroskedasticity
JEL classification: C32

[1]The authors thank the ZEDAT High-Performance Computing Service of the Freie Universität Berlin for providing the computing facilities. The paper was presented at seminars at Monash University, Melbourne, University of Melbourne, Sydney University, Macquarie University, Sydney, and CREATES, Aarhus. We are grateful for comments to the seminar participants as well as Robert Taylor, Carsten Jentsch, Lutz Kilian and two anonymous referees.

1 Introduction

In the 1970s major advances in economic time series analysis initiated in part by seminal books such as Box and Jenkins (1976), Fuller (1976) and Granger and Newbold (1977) triggered the development of new tools for macroeconomic analysis. Important new insights by Granger and Newbold (1974) and others regarding problems related to inference in time series regression models involving trending variables led to a rethinking of the standard simultaneous equations model (SEM) methodology. In related research, there were major advances in time series forecasting at that time (e.g., Bates and Granger (1969), Newbold and Granger (1974), Box and Jenkins (1976)). The newly developed models and methods of that time showed that forecasts based on simple univariate time series models could outperform large scale SEMs (see, e.g., Nelson (1972), Cooper (1972), Granger and Newbold (1975) and the discussion in Granger and Newbold (1977, Section 8.4)). This insight and the difficulties related to identifying large-scale SEMs culminated in the proposal by Sims (1980) to use vector autoregressive (VAR) models instead of SEMs for economic analysis.

Since then a large part of time series econometrics has developed VAR based tools for macroeconometric analysis (see, e.g., Kilian and Lütkepohl (2017)). Impulse response analysis is one such tool. One branch of research in this area has focussed on tools for identifying structural shocks and related impulse responses that allow for meaningful economic analysis. In this context, using heteroskedasticity and conditional heteroskedasticity for identification has been proposed (see Rigobon (2003), Lanne and Lütkepohl (2008), Lütkepohl (2013)). The present study contributes to this literature by exploring inference methods related to impulse responses based on structural VAR (SVAR) models with generalized autoregressive conditionally heteroskedastic (GARCH) innovations, as proposed by Normandin and Phaneuf (2004) in this context.

Specifically, we consider bootstrap methods that capture GARCH type dynamics and compare their suitability for inference in SVAR models identified by conditional heteroskedasticity. Alternative bootstrap methods have been proposed in the literature that are capable of capturing conditional heteroskedasticity more generally. For example, a wild bootstrap was explored by Gonçalves and Kilian (2004, 2007) for conditionally heteroskedastic autoregressive models. For the VAR framework considered in the present study, Brüggemann, Jentsch and Trenkler (2016) point out, however, that the wild bootstrap is not valid asymptotically for structural impulse response analysis and propose an asymptotically valid residual-based moving blocks bootstrap instead for models with conditional heteroskedasticity of unknown form.

Assuming that the conditional heteroskedasticity is of GARCH type, yet another possibility is to use a GARCH residual-based bootstrap (see Hidalgo and Zaffaroni (2007), Jeong (2017) and Bruder (2018)). Some previous studies on bootstrap methods for conditionally heteroskedastic SVAR models focus on possible improvements for inference if identified structural shocks and impulse responses are considered. In contrast, in this study we investigate inference for the case where identification is obtained via conditional heteroskedasticity. Thus, in our setup, estimating the second moment structure well may be of particular importance because it is used for identifying the structural parameters. Therefore it is noteworthy that the GARCH residual-based bootstrap results in higher order precision gains for the GARCH parameters, if the GARCH model is correctly specified.

We explore and compare the small sample suitability of the alternative bootstrap methods for inference on structural impulse responses in the present study. In particular, we are interested in inference methods that are compatible with the identification of structural shocks through GARCH.

Given that a full maximum likelihood (ML) estimation of multivariate GARCH models is computationally demanding and, hence, is problematic in bootstrapping algorithms, we also explore other estimation methods that offer significant computational advantages to make the bootstrap methods operational. For example, we also consider an estimation method that has been proposed as a first step in a Gaussian ML procedure by Lanne and Saikkonen (2007).

It is found that the relative coverage frequencies for the impulse responses are quite heterogeneous when the structural parameters are identified purely via GARCH. They are partly below the nominal coverage rates and partly above, depending on the impulse response function considered. Bootstrap and estimation methods designed for more precise estimation of the GARCH structure have no advantages for the coverage precision of the confidence intervals and confidence bands. In fact, the methods that condition on the first round ML estimates of the GARCH parameters in the bootstrap tend to result in smaller intervals and bands with similar coverage properties which may be substantially smaller than the nominal level, however. Overall, the most accurate coverage is obtained if such a conditional approach is combined with a wild bootstrap.

We use the alternative bootstrap procedures and estimation methods to assess the effects of monetary policy shocks in the United States based on a benchmark study by Caldara and Herbst (2018). These authors use Bayesian estimation techniques and identify the structural shocks with an external instrument approach based on high frequency data. We use conditional heteroskedasticity for identification instead and consider a three-dimensional model consisting of the federal funds rate and variables measuring the excess bond premium and industrial production growth. We find plausible responses to a monetary policy and a financial shock.

The remainder of this study is organized as follows. In the next section the model setup is laid out. In Section 3 the alternative bootstrap and estimation methods are presented. In Section 4 the Monte Carlo study is described for comparing the methods in small samples and the simulation results are discussed. Section 5 considers the empirical example and Section 6 concludes. The Appendix contains computational details and supplementary results.

2 The Model

2.1 Model Setup

The reduced-form model is assumed to be a K-dimensional VAR(p) process,

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

where ν is a K-dimensional constant term, the A_j $(j = 1, \ldots, p)$ are $(K \times K)$ coefficient matrices and u_t is the conditionally heteroskedastic, serially uncorrelated error term with mean zero and unconditional covariance matrix Σ_u. Given that we are interested in the higher-order moment structure of the variables, we make the simplifying assumption that all variables are integrated of order zero and the VAR process is stable and stationary. In other words, the polynomial

$$\det(I_K - A_1 z - \cdots - A_p z^p)$$

has no roots inside and on the complex unit circle. Denoting the lag operator by L, stability of the VAR process ensures that the matrix operator $A(L) = I_K - A_1 L - \cdots - A_p L^p$ is invertible. This assumption simplifies the exposition but could be relaxed to allow for cointegration, for example.

The structural errors, ε_t, are assumed to be instantaneously uncorrelated with unconditional identity covariance matrix $\Sigma_\varepsilon = I_K$. They are determined as a linear transformation of u_t,

$$\varepsilon_t = B^{-1} u_t \quad \text{or} \quad u_t = B\varepsilon_t. \tag{2}$$

Of course, the structural errors are also white noise and, hence, serially uncorrelated but conditionally heteroskedastic. The transformation matrix B is such that $BB' = \Sigma_u$. It is the matrix of instantaneous effects or impact effects of the structural shocks on the observed variables y_t. This structural model will be called the B-model in the following.

Alternatively, we consider a structural model which is often referred to as A-model (see Lütkepohl (2005, Chapter 9)). It normalizes the diagonal elements of B^{-1} to one and leaves the unconditional variances of the structural shocks unrestricted. In other words, for the A-model, we have $\Sigma_u = B\Sigma_\varepsilon B'$, where the unconditional covariance matrix of the structural innovations $\Sigma_\varepsilon = \mathbb{E}(\varepsilon_t\varepsilon_t')$ is a diagonal matrix. This model is often preferred in the SVAR literature if the instantaneous relations between the variables represent economically meaningful relations (see, e.g., Christiano, Eichenbaum and Evans (1999) and Belongia and Ireland (2015)). In this model, the structural shocks have sizes that will typically differ from one (unconditional) standard deviation which is implicitly assumed by not restricting B but choosing it such that $BB' = \Sigma_u$. Technically, for our purposes the main difference between the A- and the B-model is that the matrix of impact effects is unrestricted in the latter model (apart from the constraint $BB' = \Sigma_u$).

The structural impulse responses are the elements of the coefficient matrices of the MA representation

$$y_t = \mu + \sum_{i=0}^{\infty} \Theta_i \varepsilon_{t-i} = \mu + \Theta(L)\varepsilon_t, \tag{3}$$

where $\Theta(L) = (I_K - A_1 L - \cdots - A_p L^p)^{-1} B$ and μ is the mean vector of y_t.

2.2 GARCH Structure

It is assumed that the GARCH structure of the error term is such that it identifies the structural parameters B. In the context of GARCH volatility, the GO-GARCH model originally proposed by van der Weide (2002) is convenient for this purpose. For identifying structural VAR models this setup was considered by Normandin and Phaneuf (2004), Bouakez and Normandin (2010), Lütkepohl and Milunovich (2016) and others. It assumes that

$$\mathbb{E}(u_t u_t' | \mathcal{F}_{t-1}) = \Sigma_{t|t-1} = B \Lambda_{t|t-1} B', \tag{4}$$

where $\mathcal{F}_t$ denotes the information available at time t and

$$\Lambda_{t|t-1} = \text{diag}(\sigma^2_{1,t|t-1}, \ldots, \sigma^2_{K,t|t-1})$$

is a diagonal matrix with univariate GARCH(1,1) diagonal elements,

$$\sigma^2_{k,t|t-1} = (1 - \gamma_k - g_k) + \gamma_k \varepsilon^2_{k,t-1} + g_k \sigma^2_{k,t-1|t-2}, \quad k = 1, \ldots, K. \tag{5}$$

Here $\varepsilon_{k,t}$ is the k^{th} component of ε_t which is assumed to have mean zero and identity covariance matrix in the B-model, $\varepsilon_t \sim (0, I_K)$. Thus, GARCH enters the model by each structural shock having univariate GARCH dynamics. This setup implies an unconditional residual covariance matrix

$$\mathbb{E}(u_t u_t') = \Sigma_u = BB'.$$

It is assumed that $g_k \geq 0$ and $\gamma_k > 0$ with $g_k + \gamma_k < 1$ for at least $K - 1$ of the univariate GARCH processes. Under this condition, the matrix B is locally identified (see Milunovich and Yang (2013) and Lütkepohl and Milunovich (2016)).

More precisely, the columns of B are identified up to sign and permutation. Since the k^{th} column of B contains the impact effects of the k^{th} structural shock on the components of y_t, this means that the signs and the order of the shocks are not uniquely determined by the GARCH structure. However, given the sign of a shock and its position in the vector ε_t, its effects on the variables on impact are uniquely specified in our B-model setup.

To obtain uniqueness of B, the shocks have to be ordered uniquely and their sign has to be fixed. In specific applications this should ideally be linked to economic considerations or features of economic variables. For example, in our application in Section 5 we choose the shock which explains the largest share of the variance of the policy interest rate as the monetary policy shock and fix the sign such that it increases the policy rate on impact. In our Monte Carlo study reported in Section 4, the shocks are ordered according to the size of the impact effects on specific variables and the sign is fixed by ensuring a positive impact response of specific variables (see Section 4 for the details).

As mentioned earlier, for the A-model it is assumed that B^{-1} has ones on the main diagonal and is such that $B^{-1}\Sigma_u B^{-1\prime} = \Sigma_\varepsilon$ is a diagonal matrix. In this case, $\varepsilon_t \sim (0, \Sigma_\varepsilon)$ and the GARCH(1,1) processes in (5) become

$$\sigma^2_{k,t|t-1} = \xi_k + \gamma_k \varepsilon^2_{k,t-1} + g_k \sigma^2_{k,t-1|t-2}, \quad k = 1, \ldots, K, \tag{6}$$

with unconditional variance $\sigma^2_{\varepsilon,k} = \xi_k/(1 - \gamma_k - g_k)$. Here $\sigma^2_{\varepsilon,k}$ denotes the k^{th} diagonal element of Σ_ε. In the A-model model, the signs of the rows and columns are identified by the requirement that the diagonal elements of B^{-1} are unity. In applied work, the row ordering is ideally also linked to economic or institutional considerations, as for the B-model. The specific orderings used in the Monte Carlo experiment and the application are discussed in Sections 4 and 5, respectively.

2.3 Estimation

Assuming a Gaussian conditional distribution, $u_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, \Sigma_{t|t-1})$, the log-likelihood of the model is

$$\log l = \sum_{t=1}^{T} \log f_{t|t-1}(y_t),$$

where the conditional densities have the form

$$f_{t|t-1}(y_t) = (2\pi)^{-K/2} \det(\Sigma_{t|t-1})^{-1/2} \exp\left(-\frac{1}{2} u_t' \Sigma^{-1}_{t|t-1} u_t\right). \tag{7}$$

Lanne and Saikkonen (2007) use a polar decomposition of B, $B = CR$, where C is a symmetric, positive definite $(K \times K)$ matrix and R is an orthogonal $(K \times K)$ matrix and observe that the expression in (7) can be rewritten as

$$f_{t|t-1}(y_t) = (2\pi)^{-K/2}\det(\Sigma_u)^{-1/2}\exp\left(-\frac{1}{2}u_t'\Sigma_u^{-1}u_t\right)\prod_{k=1}^{r}\sigma_{k,t|t-1}^{-1}$$
$$\times\exp\left(-\frac{1}{2}u_t'C^{-1}R(\Lambda_{t|t-1}^{-1}-I_K)R'C^{-1}u_t\right).$$

They note that $\Sigma_u = CC$ and propose to break down the estimation in two main steps. First, C is obtained as the unique square root of Σ_u, which is estimated as a sample covariance matrix. Second, the rows of R' and the GARCH equation parameters are estimated separately for $k = 1, \dots, K$, conditionally on the estimated C. In fact, equation $k+1$ is estimated conditionally on the previously estimated equation k. The exact procedure is described in Section 4 of Lanne and Saikkonen (2007). Thereby initial estimates of the parameters of the volatility model are obtained and in a second step a full, joint ML estimation of the parameters is performed starting from the initial estimates obtained in the first step.

This procedure is used for ML estimation in the simulation experiment reported later. The VAR slope coefficients $\nu, A_1, \dots, A_p$ are estimated by least squares (LS) and then the LS residuals are used in the likelihood function and the first step of the Lanne/Saikkonen procedure. Further details on the implementation of the likelihood optimization are given in Appendix B.

Lanne and Saikkonen (2007) show the consistency and asymptotic normality of the Gaussian ML estimator under assumptions that allow the true conditional distributions of the u_t to be non-Gaussian. In other words, standard asymptotic properties of the estimators are obtained even if the estimation procedure is only a Gaussian quasi-ML procedure. Lanne and Saikkonen (2007) also suggest that the estimator obtained in the first step of their procedure may be consistent under suitable conditions, although it may be inefficient relative to a full Gaussian quasi-ML estimator.

3 Bootstrapping Impulse Responses

In our setup, where the structural parameters are identified through conditional heteroskedasticity, a bootstrap procedure for inference on structural impulse responses has to mimic the GARCH structure in the innovations so that the structural parameters B are maintained and can be estimated from the bootstrap samples. Three bootstrap procedures that satisfy this condition are considered in the following: a recursive-design wild bootstrap, a recursive-design residual-based moving blocks bootstrap and a GARCH residual-based bootstrap. They can all be used to generate samples from which structural impulse responses can be estimated.

3.1 The Bootstrap procedures

Recursive-design wild bootstrap (WB)

Kreiss (1997) and Gonçalves and Kilian (2004, 2007) consider a wild bootstrap (WB) which preserves the second moment structure of autoregressive (AR) errors and, hence, maintains conditional heteroskedastic innovations of AR models. For our setup, bootstrap samples are constructed recursively as

$$y_t^* = \widehat{\nu} + \widehat{A}_1 y_{t-1}^* + \cdots + \widehat{A}_p y_{t-p}^* + u_t^* \tag{8}$$

for $t = 1, \ldots, T$, given a set of initial values $y_{-p+1}^*, \ldots, y_0^*$. Here $\widehat{\nu}$, $\widehat{A}_1, \ldots, \widehat{A}_p$ denote LS estimates and the errors $u_t^* = \eta_t \widehat{u}_t$, where the $\widehat{u}_t = y_t - \widehat{\nu} - \widehat{A}_1 y_{t-1} - \cdots - \widehat{A}_p y_{t-p}$ are LS residuals and the η_t are independent random variables with zero mean and unit variance which are independent of the VAR innovations. Hence, $\eta_t \widehat{u}_t$ has the same variance as $\widehat{u}_t$ and the bootstrap errors mimic the second moment structure of the original innovations.

For the WB, we always use the original initial values $y_{-p+1}^* = y_{-p+1}, \ldots, y_0^* = y_0$. Moreover, standard normal η_t are used, i.e., $\eta_t \sim$ i.i.d.$\mathcal{N}(0,1)$, as in Brüggemann et al. (2016). These authors provide a theoretical justification for this choice to be preferable to some other distributions for η_t. Such alternative distributions for η_t have been used in some related SVAR studies considering identification through heteroskedasticity without theoretical justification. For example, Herwartz and Lütkepohl (2014), Lütkepohl and Netšunajev (2017) and Netšunajev (2013) use a Rademacher distribution.

Gonçalves and Kilian (2004) are interested in inference for the slope parameters of AR processes. They use WB methods among others and prove the asymptotic validity in the presence of conditional heteroskedasticity. In a simulation study they find that their WB procedures improve inference in the presence of conditionally heteroskedastic innovations relative to an i.i.d. bootstrap which ignores heteroskedasticity or conditional heteroskedasticity. Clearly, such results do not ensure that the WB also improves inference for structural impulse responses. In fact, Brüggemann et al. (2016) show that the WB does not properly capture the higher-order moment properties of the original distribution, except possibly for larger propagation horizons. Since the structural matrix B is related to the residual variances, the WB is actually not asymptotically valid for inference on structural impulse responses. However, Brüggemann et al. (2016) show by simulation that it performs well in comparison to a valid bootstrap method. Moreover, the asymptotic problems of the WB are caused by the fact that confidence intervals for structural impulse responses of one standard deviation in size involve estimates of the variances of the second moments and hence higher-order moments. If the variances of the structural shocks are not standardized to one, as in the A-model setup, the impulse responses may not be affected by the problem of not capturing higher-order moments well. Therefore we include the

WB, although its theoretical basis is weak for some of the impulse responses considered in our Monte Carlo study.

In our setup we are actually facing an even more challenging problem than Brüggemann et al. (2016) because we want to identify the structural parameters through GARCH. We are nevertheless including the WB in our comparison because it has been used for inference in SVAR analysis. Initially, we have also included a fixed-design wild bootstrap in our comparison.[2] It turned out to be notably inferior to the presently considered WB procedure in terms of our performance criteria to be discussed later. The fixed-design WB was also found to be inferior to the recursive-design variant in a study by Gonçalves and Kilian (2004). Therefore we dropped it from the comparison.

VAR residual-based moving blocks bootstrap (MBB)

Brüggemann et al. (2016) show that a recursive-design residual-based moving blocks bootstrap (MBB) is asymptotically valid for structural impulse responses. They propose fitting a VAR(p) model by LS and consider sampling blocks of the LS residuals $\widehat{u}_t$, $t = 1, \ldots, T$. They choose a block length $l < T$ such that $s = [T/l]$ is the number of nonoverlapping blocks, where $ls \geq T$.[3] The blocks of length l of the LS residuals are arranged in the form of the matrix

$$\begin{bmatrix} \widehat{u}_1 & \widehat{u}_2 & \ldots & \widehat{u}_l \\ \widehat{u}_2 & \widehat{u}_3 & \ldots & \widehat{u}_{1+l} \\ \vdots & \vdots & & \vdots \\ \widehat{u}_{T-l+1} & \widehat{u}_{T-l+2} & \ldots & \widehat{u}_T \end{bmatrix}.$$

The bootstrap residuals are recentered by removing the columnwise mean to ensure that the bootstrap residuals have mean zero. More precisely, the recentering is done by constructing

$$\widetilde{u}_{jl+i} = \widehat{u}_{jl+i} - \frac{1}{T-l+1} \sum_{r=0}^{T-l} \widehat{u}_{i+r}$$

for $i = 1, 2, \ldots, l$ and $j = 0, 1, \ldots, s-1$. Bootstrap residuals are generated by drawing s times with replacement from the recentered rows of the matrix. These draws are combined in a time series of bootstrap residuals, $[u_1^*, \ldots, u_T^*]$, by joining them end-to-end and retaining the first T bootstrap residuals. Starting from a draw for the bootstrap presample observations $y_{-p+1}^*, \ldots, y_0^*$ obtained by randomly drawing p consecutive observations from

[2] A fixed-design wild bootstrap uses a fixed set of regressors in generating the bootstrap samples, that is, instead of using the recursive scheme in (8), the bootstrap samples are generated as $y_t^* = \widehat{\nu} + \widehat{A}_1 y_{t-1} + \cdots + \widehat{A}_p y_{t-p} + u_t^*$.

[3] $[x]$ denotes the smallest integer greater than or equal to the real number x.

the sample, the bootstrap sample, $y_1^*, \dots, y_T^*$, is generated recursively for $t = 1, \dots, T$ as

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

Brüggemann et al. (2016) use zero initial values. However, in their simulations they use very large samples compared to the smaller samples we will consider in our simulations. In smaller samples the initial values may be more important. Hence, we favour to vary them as well to get a better impression of the bootstrap uncertainty.

Brüggemann et al. (2016) show that this bootstrap method allows for asymptotically correct inference for statistics that depend on the slope parameters and the unconditional covariance matrix of stable VAR processes with conditionally heteroskedastic innovations because it replicates the fourth moment structure correctly if the impulse responses have proper nonsingular asymptotic distributions. Thus, this method provides valid inference for structural impulse responses. The asymptotic results are derived under the assumption that the block length goes to infinity with the sample size but much more slowly ($l^3/T \to 0$). In our benchmark simulations we make the block lengths dependent on the sample sizes considered. We also explore the impact of the block length by varying it for some Monte Carlo designs.

Brüggemann et al. (2016) also show by simulation that the MBB and the WB tend to underestimate the uncertainty in structural impulse responses in some situations. Thus, they tend to result in too narrow confidence intervals for the structural impulse responses with true coverage below the desired nominal coverage.

These results are obtained for SVAR models identified by conventional exclusion restrictions. In other words, in their bivariate setup, the B matrix is identified by setting one element to zero. Thus, their setup is very different from ours because they do not use the conditional heteroskedasticity for identifying B and, thus, their simulation results may differ from ours.

GARCH residual-based bootstrap (RBB)

A bootstrap method that is known to result in asymptotically valid inference for the GARCH parameters is based on the deep GARCH innovations. Bootstrap draws of the GARCH residual-based bootstrap (RBB) are obtained using the following steps:

(1) Get a quasi-ML estimate $\hat{B}^{-1}$ with rows $\hat{b}_k$ and GARCH estimates $\hat{\gamma}_k$, $\hat{g}_k$, $k = 1, \dots, K$.

(2) Generate sequences

$$\hat{\sigma}_{k,t|t-1}^2 = (1 - \hat{\gamma}_k - \hat{g}_k) + \hat{\gamma}_k (\hat{b}_k \hat{u}_{t-1})^2 + \hat{g}_k \hat{\sigma}_{k,t-1|t-2}^2 \tag{9}$$

recursively for $t = 1, \dots, T$ starting with $\hat{u}_0 = 0$ and $\hat{\sigma}^2_{k,0|-1} = 1$. Thereby we get estimates

$$\hat{\Lambda}_{t|t-1} = \text{diag}(\hat{\sigma}^2_{1,t|t-1}, \dots, \hat{\sigma}^2_{K,t|t-1})$$

from which we can compute deep innovations $\hat{\eta}_t = \hat{\Lambda}^{-1/2}_{t|t-1}\hat{B}^{-1}\hat{u}_t$ for $t = 1, \dots, T$.

(3) Draw bootstrap innovations $e^*_1, \dots, e^*_T$ from the demeaned innovations $\hat{\eta}_t$ and generate reduced-form errors as

$$u^*_t = \hat{B}\hat{\Lambda}^{1/2}_{t|t-1}e^*_t. \tag{10}$$

(4) Get bootstrap observations as

$$y^*_t = \widehat{\nu} + \widehat{A}_1 y^*_{t-1} + \cdots + \widehat{A}_p y^*_{t-p} + u^*_t.$$

Repeat Steps 3 and 4 a large number of times. It is worth pointing out that the same bootstrap time series are used for the B- and the A-model.

For univariate GARCH processes the asymptotic validity of the GARCH residual-based bootstrap was established by Jeong (2017) who shows that it can lead to higher order improvements and is asymptotically superior to a block bootstrap for inference regarding the GARCH parameters. Bruder (2018) considers a slightly different version of this bootstrap which generates $u^*_t = \hat{B}\Lambda^{*1/2}_{t|t-1}e^*_t$ in step (3), where $\Lambda^*_{t|t-1} = \text{diag}(\sigma^{*2}_{1,t|t-1}, \dots, \sigma^{*2}_{K,t|t-1})$ with

$$\sigma^{*2}_{k,t|t-1} = (1 - \hat{\gamma}_k - \hat{g}_k) + \hat{\gamma}_k(e^{*2}_{k,t-1} + \hat{g}_k)\sigma^{*2}_{k,t-1|t-2}.$$

Moreover, he standardizes the innovations $\hat{\eta}_t$ by mean- and variance-adjusting them to obtain the e^*_t errors. He shows the asymptotic validity of this bootstrap for the case of structural impulse responses identified by GARCH. We have also used this alternative RBB method for some simulations and we will comment on the results briefly in Section 4, where it is labelled as 'alt. RBB'. Bruder (2018) also studies a modified version of his RBB which turned out to have similar small sample properties and is therefore not considered in the following.

Other bootstrap procedures

There are also other bootstrap variants which maintain heteroskedasticity and conditional heteroskedasticity. For example, Gonçalves and Kilian (2004) also consider a pairwise bootstrap. They found, however, that even when the pairwise bootstrap is asymptotically valid, it tends to be less accurate in small samples than the recursive-design WB. Moreover, Brüggemann et al. (2016) show that, like the wild bootstrap, the pairwise bootstrap is not consistent for structural impulse responses. In their simulations it is inferior to their wild bootstraps in particular for persistent GARCH processes. We also did not consider the stationary bootstrap proposed by Politis and Romano (1994) which randomly chooses the block length of a block bootstrap. Although such a procedure might be adequate in the present context, it did not work as well as other related procedures in a simulation study reported by Berkowitz, Birgean and Kilian (2000) based on homoskedastic univariate time series. Thus, we do not include such alternative procedures in our comparison.

3.2 Estimation Methods

For each of these bootstraps, the structural VAR parameters and impulse responses have to be estimated in each replication. Since a full ML estimation of the model is computationally demanding, in particular for larger models, we consider four alternatives for estimating the impulse responses in the bootstrap replications.

Full ML (ML) A full ML estimation based on the Lanne/Saikkonen procedure mentioned in Section 2.3 is used in each bootstrap replication. Depending on the size of the model this option may in fact not always be feasible in practice. In our Monte Carlo study we always use Gaussian ML estimation, that is, we maximize the Gaussian likelihood function even for non-Gaussian processes. In that case, ML is more precisely Gaussian quasi-ML. There may be applications where other distributions are of interest.

First step Lanne/Saikkonen (L/S) The estimates in each bootstrap replication are based on the first step estimation of the Lanne/Saikkonen procedure for the GARCH parameters. As mentioned earlier, these authors expect the estimator to be consistent under suitable conditions but it is not necessarily efficient. Given that the computations of the procedure are much less demanding than a full ML estimation, it is included here to investigate the small sample efficiency losses when it is combined with a bootstrap procedure. In the following the procedure is abbreviated as L/S.

Conditioning on estimated GARCH parameters (CB) We also consider a computational short-cut where in each bootstrap replication we use the ML estimates of the GARCH parameters based on the original data. That is, we condition on the first round ML estimates of the GARCH parameters and estimate only the $\nu, A_1, \dots, A_p$ and B parameters in the bootstrap replications. Thereby substantial computational savings can be realized. The procedure is abbreviated by CB in the following.

Conditioning on true GARCH parameters (true GARCH) To see how much we loose by not knowing the true GARCH processes which drive the volatility changes, we also condition on the true GARCH parameters in the bootstraps and only estimate the VAR parameters and the structural matrix B in the bootstrap replications. Although this method is not feasible in practice, it may give an indication how much can be gained by refining the estimation methods for the second moment structure. The method is labelled as 'true GARCH' in the following.

The CB procedure requires only one full ML estimation. Such procedures were used as ad hoc methods in a related structural VAR context by Herwartz and Lütkepohl (2014), Lütkepohl and Netšunajev (2017) and others although the asymptotic validity has not been shown. It is a short-cut which reduces the computational complexity of the bootstraps considerably.

Instead of conditioning on the estimated GARCH parameters as in the CB method, one could alternatively condition on the estimated sequences of the conditional variances. Such a bootstrap was considered by Cavaliere, Pedersen and Rahbek (2018) in a different context for univariate processes. In our context of inference for structural impulse responses, we found in a limited simulation comparison that it produces very similar results to the CB method. Therefore we do not consider it in the following.

Bruder (2018) considers an alternative estimation method for the GARCH parameters and the structural parameters originally due to Boswijk and van der Weide (2011) which has computational advantages over our full ML procedure. We have not considered that estimator because refinements of the estimator of the GARCH parameters do not seem to be essential for improving inference for impulse responses, as we will see when we discuss our simulation results.

Generally, we use Gaussian ML type estimates of the reduced-form VAR slope coefficients although for inference on impulse responses, bias-adjusted estimates were found to be preferable in the related literature (e.g., Kilian (1998, 1999)). These results are related to homoskedastic VARs, however, and there is no evidence that they would be beneficial in the present context. In fact, it is not clear whether they may be detrimental to our objective of using heteroskedasticity for identification. Therefore we consider only the ML based estimates of the slope parameters without bias adjustment in the present study. An alternative way to reduce estimation bias may be the use of Hall's bootstrap confidence intervals (see Hall (1992) or Section 12.2.6 of Kilian and Lütkepohl (2017)). Given that the processes used in our simulations are such that estimation bias in the VAR slope coefficients may not be a severe problem and in a limited simulation experiment Hall's intervals were not uniformly better than percentile intervals, we prefer to use the more conventional percentile intervals in this study. They are presented in detail in Section 4.2.

4 Monte Carlo Comparison

A large-scale simulation comparison of the WB, the MBB and some other bootstrap methods for inference for structural impulse responses in the presence of conditional heteroskedasticity is also reported by Brüggemann et al. (2016). As mentioned earlier, these authors do not consider identification through heteroskedasticity, however. They just use heteroskedasticity to adjust variance estimates and confidence intervals and they identify the structural shocks by conventional exclusion restrictions. Thus, they face inference problems quite different from those of interest in the present study and therefore they also use different VAR and GARCH dynamics in their simulations. In the context of the present study, the second moment structure is used for parameter identification and, hence, it may be important to estimate it well.

4.1 Monte Carlo Setup

Our simulations are based on bivariate and trivariate DGPs, i.e., $K = 2, 3$, that evolve from

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad u_t \sim (0, BB'). \tag{11}$$

In order to generate GARCH innovations, we follow Lütkepohl and Schlaak (2018) and first generate random variates with zero mean and unit variance, $(e_{1t}, \ldots, e_{Kt})' \sim (0, I_K)$, and

$$\sigma^2_{k,t|t-1} = (1 - \gamma_k - g_k) + \gamma_k \epsilon^2_{k,t-1} + g_k \sigma^2_{k,t-1|t-2}, \quad k = 1, \ldots, K,$$

where $\epsilon_{k,t} = e_{k,t}\sigma_{k,t|t-1}$ for $t = 1, \ldots, T$. Finally, the $u_t = B\Lambda^{1/2}_{t|t-1} e_t$ are generated, where $\Lambda_{t|t-1} = \text{diag}(\sigma^2_{1,t|t-1}, \ldots, \sigma^2_{K,t|t-1})$. Thereby, the unconditional covariance matrix of u_t is $\Sigma_u = BB'$.

The number of replications of the simulation experiment for each design is 500. Although this is a rather modest number, the number of replications is limited by the substantial computation times for each replication. For further details see also Appendix B. The sample sizes used are $T = 200$ and 500, and the number of bootstrap replications is $N = 1000$.

4.1.1 Bivariate Benchmark DGPs

First, we use bivariate VAR(1) DGPs similar to those used in the related literature on constructing bootstrap confidence bands for impulse responses (e.g., Kilian (1998), Lütkepohl, Staszewska-Bystrova and Winker (2015a, 2015b), Lütkepohl and Schlaak (2018)). We choose

$$A_1 = \begin{bmatrix} \alpha & 0 \\ 0.5 & 0.5 \end{bmatrix}, \quad \nu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{12}$$

where $\alpha = 0.5$ and 0.9. Thus, the processes are stable and stationary because $|\alpha| < 1$. Although the constant terms in the DGPs are zero, they are always included in the models which are estimated. The matrix

$$B = \begin{bmatrix} 1 & 0 \\ -1 & 10 \end{bmatrix}.$$

This matrix is chosen such that the two columns are clearly distinct. Although the upper right-hand element is zero, a full, unrestricted B matrix is estimated, that is, the zero restriction is not imposed in the estimation. To ensure an identified B matrix in the B-model, where B is only identified up to sign and column permutation through GARCH, we normalize the diagonal elements to be positive which takes care of the sign and we order the columns such that the largest element appears in the lower right-hand corner of B. The latter choice ensures a unique column ordering. The elements of B have different estimation variance. Thereby we can get a better picture of the impact of estimation precision on the results. The structural matrix of the corresponding A-model is obtained by inverting our B matrix and standardizing its main diagonal,

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

We use GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$ and standard normal variates, $(e_{1t}, e_{2t})' \sim \mathcal{N}(0, I_2)$, for our benchmark setup. The choice of the GARCH parameters ensures rather persistent volatility patterns as they are often observed in practice (see, e.g., Bouakez and Normandin (2010, Table 1)). Both processes satisfy the necessary conditions for the existence of unconditional fourth moments (see He and Teräsvirta (1999)) which is relevant for the asymptotic estimation theory for GARCH processes. We will also briefly comment on results for other distributions for the innovations and alternative parameter combinations which correspond to less well identified models. The results are similar to our benchmark results and are therefore not discussed in detail.

4.1.2 Three-dimensional DGP

In contrast to the quite stylized simulation design of our bivariate DGPs we also simulate a more sophisticated DGP that mimics the properties of an observed dataset. Its parameters are based on estimates of a trivariate model that has been used in the literature to analyze the effects of monetary policy shocks (see, e.g., Caldara and Herbst (2018)). The model entails the federal funds rate, an excess bond premium variable, and the growth rate of industrial production. For a detailed description of the model and the data we refer to Section 5 where the same model is used in an empirical application. The data is at monthly frequency spanning from June 1993 to June 2007, hence, the sample size is $T = 169$. We choose a lag length of $p = 6$. The identification of the structural parameters is obtained by exploiting the time-varying volatility of the data using the GARCH model outlined in Section 2. The estimated coefficients are treated as parameters of our trivariate DGP.[4]

The estimates of the autoregressive coefficient matrices are presented in Appendix A. We ensure that our system is stable and stationary by checking that all eigenvalues of the companion matrix of the VAR have modulus less then 1. We find that the largest eigenvalue of the system is $\hat{\lambda} = 0.946$, thus, clearly less than 1. The impact effects matrix is

$$B = \begin{pmatrix} 0.208 & 0.020 & 0.020 \\ 0.003 & 0.119 & 0.000 \\ 0.000 & 0.000 & 0.004 \end{pmatrix}.$$

The coefficients on the main diagonal of B are normalized to have positive signs. Moreover, to achieve full identification, we impose a unique ordering by permuting the columns of B such that the column with the largest value in absolute terms in the first row of the matrix is ordered first. For the remaining columns of B, the largest value of the second row is ordered as second column which ensures identification.

Two pairs of estimated GARCH processes exhibit a similar persistence as our bivariate benchmark simulation designs ($(\gamma_1, g_1) = (0.063, 0.905)$ and $(\gamma_2, g_2) = (0.211, 0.757)$). In contrast, the persistence of the third GARCH process is much lower. The parameters for that process are $(\gamma_3, g_3) = (0.302, 0.136)$. The innovations are standard normal, $(e_{1t}, e_{2t}, e_{3t})' \sim \mathcal{N}(0, I_3)$.

4.2 Computing Bootstrapped Confidence Intervals and Confidence Bands

For all simulation designs we compute impulse response estimates as

$$\hat{\Theta}_i = \hat{\Phi}_i \hat{B}, \quad i = 0, 1, \ldots, H,$$

where the Φ_i are computed recursively as $\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j$ starting with $\Phi_0 = I_K$ and setting $A_j = 0$ for $j > p$. The propagation horizon is $H = 10$. We check how often the true impulse responses fall within the estimated confidence intervals and also how

[4] All estimated coefficients are truncated after the third digit.

often a full impulse response function, $\theta_{kl,i}$, $i = 0, 1, \ldots, H$, falls within a confidence band obtained by connecting the pointwise confidence intervals.

We consider structural impulse responses based on B-models as well as A-models. For the B-model the estimate $\hat{B}$ is such that it satisfies $\hat{B}\hat{B}' = \widehat{\Sigma}_u = T^{-1}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t'$. The estimate of B^{-1} for the A-model is obtained from this $\hat{B}$ estimate by inverting the matrix $\hat{B}$ and dividing each row of the inverse by the corresponding element of the main diagonal to obtain $\hat{A}$. Thus, all differences between the A- and the B-model are due to the standardization of diagonal elements in the matrix B^{-1} and not, for example, due to convergence to a different local optimum of the estimation algorithm.

The bootstrap confidence intervals for the individual impulse response coefficients are constructed as percentile intervals. For $0 < \gamma < 1$, a $1-\gamma$ interval is constructed as $[\theta^*_{\gamma/2}, \theta^*_{1-\gamma/2}]$, where θ^*_η is the η quantile of the bootstrap sample of the impulse response coefficient of interest. Percentile intervals are the most common bootstrap confidence intervals for structural impulse responses, although sometimes other types of intervals are used (see Kilian and Lütkepohl (2017, Chapter 12)). For our purposes the relative performance of the different bootstrap methods is of interest. Therefore the issue which confidence interval is used may be of limited importance because all methods will be affected in a similar way by the choice of confidence interval. Moreover, simulation evidence suggests that none of the other options uniformly dominates percentile intervals in terms of coverage and width (see Kilian and Lütkepohl (2017, p. 362)).

The confidence bands for an impulse response function, constructed by joining the pointwise confidence intervals, can be interpreted as Bonferroni confidence sets with a lower coverage probability than the individual intervals (see Lütkepohl, Staszewska-Bystrova and Winker (2015b) for further discussion). Note that, if a Bonferroni band is constructed for $H+1$ individual confidence intervals with nominal coverage probability $1-\gamma$, then, according to the Bonferroni inequality, the joint band has a coverage probability of at least $1-\gamma(H+1)$. In other words, in order to ensure a $1-\gamma$ joint confidence band, $1-\gamma/(H+1)$ individual confidence intervals have to be chosen.

In the simulations we use individual confidence levels of $1-\gamma = 0.9$ and 0.99 as well as impulse response functions up to propagation horizon $H = 10$. Thus, the Bonferroni inequality ensures that the nominal coverage rate of a joint band based on the individual 99% confidence intervals is bounded below by roughly 0.9. The joint bands reported later ideally should have coverage levels of 90% or more. Of course, the Bonferroni inequality provides only a lower bound for the coverage probability. The actual coverage level may be larger.

Although the actual asymptotic coverage probability of the joint bands obtained in this way is unknown, for our purposes the bands are still useful because an interval or band with prespecified coverage probability can be obtained from the given set by a multiplicative factor. Thus, if two intervals or bands have the same coverage, the smaller one is preferable. In turn, if two bands have the same widths, the one with the larger coverage is to be preferred. We measure the width of a joint confidence band by the sum of the widths of the individual confidence intervals involved. In other words, denoting the width of the confidence interval of the h^{th} impulse response coefficient associated with an impulse response function of interest by ℓ_h, the width of the band for $H+1$ impulse responses is measured as $\ell^{band} = \sum_{h=0}^{H} \ell_h$.

4.3 Monte Carlo Results

We first discuss the results for the bivariate DGPs and then report the results for the three-dimensional DGP. Some details are presented in tables and figures in the Appendix of Supplementary Results. A prefix A is attached to the numbers of these tables and figures.

4.3.1 Results for Bivariate DGPs

Since the estimates of the impulse responses crucially depend on the estimates of the structural parameters in the matrix $B = [b_{ij}]$ and because this matrix also represents the impact effects of the shocks, we first discuss relative coverage frequencies of bootstrap confidence intervals for the elements of the B matrix along with average interval widths for our benchmark DGPs. Then we discuss results for longer propagation horizons and finally we report on a robustness analysis based on alternative methods and bivariate processes.

Impact effects Relative coverage frequencies of confidence intervals for the elements of B and average interval widths for the bivariate DGP with $\alpha = 0.9$ are shown in Table 1. The table contains results for sample sizes $T = 200$ and 500. Results for nominal coverage levels of 90% are presented. We also determined intervals for confidence levels 95% and 68% which turned out to be qualitatively identical and therefore are not shown. A coverage level of 90% is quite common in SVAR studies in practice.

Looking first at the relative coverage frequencies for the B-model, it is seen in Table 1 that they vary substantially depending on the element under consideration. For example, in Table 1 for $T = 200$, combining ML estimation with the MBB gives relative coverage frequencies of 0.51, 0.98, 0.97 and 0.54 for b_{11}, b_{21}, b_{12} and b_{22}, respectively. Thus, the intervals are quite conservative for b_{21} and b_{12} while they are too small for b_{11} and b_{22}. This pattern is repeated for all combinations of estimation and bootstrap methods. It can also be seen for sample size $T = 500$, although for the larger sample size the actual coverage frequencies for b_{21} and b_{12} are closer to the nominal level of 90%, as one would expect. It should be noted that the same pattern is obtained when the true GARCH parameters are used instead of estimates.

Overall the WB generates the largest relative coverage frequencies which in some cases exceed the nominal coverage probabilities. The MBB and RBB result in lower coverage frequencies for $T = 200$ while the relative coverage frequencies of the MBB intervals are for most of the parameters closer to the nominal 90% than the RBB and WB intervals for $T = 500$. The focus of the RBB on estimating the GARCH parameters well does not lead to improved inference for the elements of the B matrix. In fact, knowing the true GARCH parameters does not help much to improve the coverage accuracy of the bootstrap confidence intervals.

Looking now at the interval widths for the B-model in Table 1, it is striking that the CB method, for a given bootstrap method and element of B, apart from very few exceptions, leads to the shortest intervals as compared to ML and L/S estimation. Even the intervals based on the true GARCH parameters are often wider and in the other cases very similar to the corresponding CB intervals. Thus, there is no scope for improving inference on the impact effects of the structural shocks by refining the estimation of the GARCH parameters.

The smallest intervals are typically obtained by combining CB with the WB or MBB. In contrast, RBB often leads to slightly larger intervals and is not competitive in terms of interval width. This is even true when we condition on the true GARCH parameters. The overall best procedure is a combination of the WB with the CB estimation method. The WB tends to have better coverage for elements with coverage below the nominal 90% level and MBB tends to provide shorter intervals. This often comes with lower coverage rates, however. Consequently, the computationally most efficient feasible method turns out to be the best. Even if the gains from using the WB/CB or MBB/CB methods in terms of coverage and interval length are often not substantial, there is little point in using any of the computationally more demanding estimation methods.

The situation for the A-model is similar in that the WB/CB method is typically at least as good as the other methods in terms of coverage and MBB/CB has a slight advantage in terms of interval width. For this model the WB/CB method tends to be conservative even for sample size $T = 200$. All methods result in conservative intervals with coverage levels close to 1.00 for b_{21} and b_{12} while, for example, the MBB method leads to lower than nominal coverage levels for b_{11} and b_{22} when $T = 200$. Another remarkable result for the A-model is perhaps that the intervals for b_{12}, which is actually zero, are quite small, meaning that this element is estimated very precisely by all methods. On the other hand, the confidence intervals for $b_{21} = -1$ are very wide, meaning that the element is not estimated very accurately.

Overall, based on the intervals for the impact effects and comparing feasible methods only, the WB/CB method is preferable in terms of coverage while MBB/CB tends to provide the smallest intervals. This result is remarkable because the WB method is asymptotically invalid at least for the B-model and the CB method does not re-estimate the GARCH parameters in each bootstrap replication. For the precision of the bootstrap confidence intervals for the impact effects of the shocks the estimation precision of the GARCH parameters appears to be of limited importance. Not even knowing the true GARCH parameters can help improving inference for the impact effects of the structural shocks.

Our results are to some extent in line with simulation results reported by Brüggemann et al. (2016) who found situations in which the WB and the MBB resulted in coverage rates below the nominal level. They also found that the WB often performs well in small samples despite its asymptotic invalidity. It is important to recall, however, that Brüggemann et al. (2016) are using a very different simulation setup. They identify the structural parameters with zero restrictions and the smallest sample size they consider is $T = 500$.

Impulse response functions Since for a VAR(1) process the impulse responses for propagation horizon i are estimated as $\hat{\Theta}_i = \hat{A}_1^i \hat{B}$, they all involve the estimated matrix of structural parameters $\hat{B}$. Thus, the estimation precision for this matrix will determine to some extent the confidence intervals and confidence bands for the impulse responses for larger propagation horizons. For the bivariate benchmark DGP with $\alpha = 0.9$, we present relative coverage frequencies of joint confidence bands for impulse response functions of propagation horizons $i = 0, 1, \dots, 10$ in Figure 1 along with relative band widths. Precise numbers are available in Table A.1 of the Appendix. Figure 1 contains results for both the B- and the A-models and sample sizes $T = 200$ and 500. The nominal coverage level of the individual impulse responses is now 99% so that we get a Bonferroni lower bound for the joint confidence level of about 90%.

In Figure 1, the widths of the confidence bands are divided by the width of the band obtained with the MBB in combination with the true GARCH parameters for each of the impulse response functions. Thus, the band widths are depicted relative to the corresponding MBB/true GARCH band and the width of the MBB/true GARCH band is 1. This normalization is useful to present very different interval widths in one figure and it also facilitates the comparison of intervals obtained with different methods. On the other hand, the normalization covers up any differences between the B- and the A-models. Therefore the actual bandwidths (not normalized) are shown in Table A.1 of the Appendix.

In Figure 1 it can be seen that there is some heterogeneity in the coverage levels for both the B- and the A-models across the different impulse response functions. However, in particular for the A-model the coverage levels are all remarkably close to 90% or even larger. This holds even for the smaller sample size of $T = 200$. The situation for the B-model is different in that, for example, for the impulse response function $\theta_{11,i}$, $i = 0, \ldots, 10$, the coverage is still only around 60% for $T = 200$ if MBB or RBB are used while WB achieves a better coverage of more than 70%. The differences in the coverage rates diminish for $T = 500$. The choice of estimation method has very little impact on the coverage rates. In fact, the plots corresponding to the alternative estimation methods are very similar and difficult to distinguish visually. This even holds when the true GARCH parameters are used in the bootstraps.

Turning now to the band widths, the relative widths in Figure 1 are all very close to one or greater than one, meaning that the MBB/true GARCH bands are always the smallest or very close to the smallest because all band widths are depicted relative to the corresponding MBB/true GARCH bands. It is remarkable, however, that also the MBB/CB band widths are very close to one. Thus, even in terms of band width, the CB method is as good as knowing the true GARCH parameters. In some cases the MBB and RBB bands are considerably larger than the WB bands (see, e.g., $\theta_{12,i}$). Comparing only the WB and the MBB, it turns out that the latter often leads to smaller band widths but also smaller coverage levels. In larger samples both methods are similar in terms of coverage and MBB tends to produce smaller band widths.

Comparing now the estimation methods in Figure 1, it is obvious that they provide very similar results for a given impulse response function, bootstrap and sample size. Again the CB method often leads to smaller bands than the other methods without sacrificing coverage. Since it is also the computationally most efficient method it is clearly preferable to the other methods. Thus, overall using the WB/CB method is recommended on the basis of our simulation results.

Since the MBB procedure depends on the block length used, we have also performed some of the simulations of the MBB/CB method with smaller block lengths. Some results are presented in Table 2. The table also contains the results for the previously used block lengths for comparison purposes. Table 2 shows coverage rates of nominal 90% confidence intervals for the impact effects and the relative coverage frequencies of joint bands based on pointwise nominal 99% confidence levels.

Comparing the results for different block lengths in Table 2, it can be seen that for $T = 200$ the coverage of the individual intervals and joint bands for $l = 10$ is typically very similar to the corresponding coverage for block length $l = 20$. On the other hand, the corresponding interval and band widths tend to be slightly smaller for $l = 20$ than for $l = 10$. Overall there is not much to choose between $l = 10$ and $l = 20$ for sample size $T = 200$.

Considering the intervals and bands for $T = 500$, it is found that a larger block size of $l = 50$ is better than $l = 10$ with respect to coverage and widths. The coverage level moves closer to the nominal 90% and the interval and band widths tend to decline for $l = 50$. The accuracy of the intervals and bands for the A-model is, in fact, quite remarkable. There is still some under-coverage for the B-model impulse responses, but even for that model the larger block size is helpful for improving coverage precision and often reduces width. In some cases the reduction in width is quite substantial (e.g., the θ_{21} bands for $l = 50$ maintain a coverage above 90% and reduce the band width from 28.29 for $l = 10$ to 22.96 for $l = 50$). Thus, for the larger sample size of $T = 500$, using a larger block length than $l = 10$ is beneficial and makes the MBB/CB method relatively more attractive. This finding motivated us to present the results for the larger block lengths in Table 1 and Figure 1.

Robustness analysis We have investigated the robustness of our results with respect to changes in the DGPs and the bootstrap methods used. A number of related results can be found in the Appendix of Supplementary Results. For example, we show results for $\alpha = 0.5$ in Tables A.2 and A.3 of that Appendix. They are qualitatively the same as those for $\alpha = 0.9$. Thus, the persistence of the process does not matter much, at least in the range considered in our simulations.

We have also considered alternative GARCH processes for the error terms. In particular, we have used bivariate VAR processes with $\alpha = 0.9$ and three alternative sets of GARCH parameters:

$$\begin{array}{lll} (\gamma_1, g_1) = (0, 0) & (\gamma_1, g_1) = (0, 0) & (\gamma_1, g_1) = (0.45, 0.1) \\ (\gamma_2, g_2) = (0.3, 0.5) & (\gamma_2, g_2) = (0.92, 0.05) & (\gamma_2, g_2) = (0.55, 0.05) \end{array}.$$

The first DGP has only one genuine GARCH component which, in addition, is not very persistent such that there is only weak conditional heteroskedasticity in the reduced-form errors. This may translate into weak identification of the structural parameters. Recall, however, that even with one GARCH component the structural shocks are still fully identified via GARCH. The second DGP also has just one nontrivial GARCH component which is more persistent, however. Finally, the third DGP has two nonpersistent GARCH components.

Relative coverage frequencies and interval and band widths for the first DGP are given in Tables A.5 and A.6 in the Appendix. It turns out that for this DGP the relative coverage frequencies are generally as good or even better than for the corresponding benchmark process (compare to Tables 1 and A.1) and also the interval and band widths are typically not larger. Thus, the rather weak GARCH in the errors is apparently sufficient to identify the structural parameters. Put differently, identification can be obtained even with little change in volatility. Results in the same range were also obtained with the other GARCH components and are therefore not reported in separate tables.

Overall, the main conclusions from the benchmark case are unaffected for these DGPs. The results in Tables A.5 and A.6 also show that the benchmark case is not a particularly easy one for our methods but there are processes where they perform better than in the benchmark case.

We have also considered a DGP with non-Gaussian structural errors. Since it is well-known that the asymptotic properties of Gaussian quasi-ML estimators even of univariate ARMA-GARCH processes with asymmetric error distributions are very different from symmetric error distributions (see Francq and Zakoïan (2004)), we have also done some simulations with skewed error distributions. Specifically we have used a standardized $\chi^2(4)$ distribution for the errors, that is, $e_{kt} \sim$ i.i.d. $-(\chi^2(4) - 4)/\sqrt{8}$, $k = 1, 2$. Such a negatively skewed distribution may mimic some left-skewed financial time series. The results are shown in Tables A.7 and A.8 of the Appendix and are quite similar to our benchmark case. They give rise to the same overall conclusions regarding the relative performance of the different bootstrap methods.

Our results are not fully comparable to those of Bruder (2018) who also considers bivariate DGPs in his simulations. However, he uses other parameter values and alternative non-Gaussian innovations. Moreover, he considers a different estimation method for the structural parameters. He includes bootstrap methods in his comparison which are similar to our MBB and RBB methods and finds that for some of his DGPs the coverage of the RBB is better than that of MBB which is to some extent in line with our results. As mentioned in Section 3, Bruder's version of the RBB method differs from ours. Therefore we have also used his RBB method (apart from the choice of initial values) on our benchmark setup. The corresponding relative coverage frequencies and interval/band widths are presented in Tables A.9 and A.10 of the Appendix. We find that Bruder's RBB method performs better than our RBB method. In fact, for larger sample sizes ($T = 500$) it performs as well as WB in terms of coverage accuracy and often leads to smaller interval and band widths (see, e.g., the results for b_{22} in Table A.9). For smaller sample sizes ($T = 200$) WB is still often clearly superior, however. Bruder (2018) does not consider the WB procedure. Therefore we do not know whether his RBB method is competitive to WB more generally. In any case, our results suggest that for computational efficiency it may be worth using it with the CB which is also not considered by Bruder (2018).[5]

4.3.2 Results for Three-dimensional DGP

Results for confidence intervals for the impact effects and confidence bands for impulse response functions for the 3-dimensional DGP are presented in Table 3 and Figure 2,

[5]We thank a referee for pointing out Bruder's alternative RBB method. Since we became aware of it only after completing large parts of our simulations, we maintain the original RBB results in the other tables and figures for comparability.

respectively. We only report results for the CB method and true GARCH parameters to reduce the computational burden and because ML and L/S estimation did not improve inference for the bivariate processes.

The results in Table 3 and Figure 2 paint in many respects a very similar picture as the results for the bivariate DGPs although the underlying models are much bigger. Since we have fitted VAR(6) processes the number of estimated parameters is substantially larger than for the bivariate DGPs and, hence, one would expect much larger estimation uncertainty because we are using the same sample sizes ($T = 200, 500$) as for the bivariate DGPs. In Table 3 it is seen that the heterogeneity in the coverage rates of the nominal 90% confidence intervals for the impact effects is again substantial. For example, for the B-model and RBB it ranges from 18% for b_{11} to 100% for some other impact effects. The very small coverage rates were not observed for the bivariate DGPs and are likely to be an implication of the much larger models we are dealing with now. The situation improves a little when the sample size increases from $T = 200$ to $T = 500$ but does not reach satisfactory levels. Even when the true GARCH parameters are used, the coverage rates are not improved for those structural parameters that have coverage rates much below the nominal 90%. As for the bivariate DGPs, the coverage rates are better in most cases for the A-model intervals. Again the pattern is the same for CB and when the true GARCH parameters are used.

Looking at the coverage rates and widths of joint bands for the impulse responses in Figure 2 the impression is also that there is not much difference between using estimated or true GARCH parameters. Again, using A-models leads to better coverage rates for some impulse response functions which are not covered well for the B-model.

Comparing the different bootstrap methods it becomes again clear that there is not much scope for improving them by using better estimation methods for the GARCH parameters because the results for CB are not much different from those for the true GARCH parameters. There is actually not much to choose between the different bootstraps now, although rather low relative coverage frequencies are achieved by RBB for b_{33} when the A-model is used.

The overall recommendation for inference for structural impulse responses identified via GARCH from our limited simulation results is to use estimation conditional on fixed GARCH parameters in the bootstrap replications. For smaller samples, WB leads to the best balance between coverage precision and width. If the size of the shocks to be traced through the system is not important, then considering the A-model is preferable to using the B-model.

5 Empirical Example

In this section, we compare the outcome of the different bootstrap procedures and estimation methods by means of an empirical example. We work with a trivariate model that was used by Caldara and Herbst (2018) to assess the effects of monetary policy shocks in the United States. These authors use Bayesian estimation techniques combined with an external instrument approach based on high frequency data to conduct a structural SVAR analysis. Their main focus is the identification of monetary policy shocks during the Great Moderation period. They show that by the inclusion of a financial market indicator the response of real activity to monetary policy shocks changes substantially compared to a bivariate benchmark model. In contrast to their identification technique, we use identification through GARCH in the following.

The dataset of the model consists of the effective nominal federal funds rate (FF) as monetary policy indicator, the excess bond premium (i.e., the spread between private corporates and government bond yields after having controlled for default risks) constructed by Gilchrist and Zakrajšek (2012) as financial market indicator (EBP) and the first differences of the logarithm of industrial production (Δip) as a measure for real activity growth.[6] The data is monthly and spans from June 1993 to June 2007 and, accordingly, the sample size is $T = 169$.

We follow Caldara and Herbst (2018) and fit a (homoskedastic) VAR model with six lags to the data. As suggested by Lütkepohl and Schlaak (2018), we check for the presence of time-varying volatility by testing the reduced form errors of the homoskedastic VAR(6) using Portmanteau-ARCH and ARCH-LM tests. These tests support the presence of conditional heteroskedasticity in the data. Moreover, the AIC information criterion clearly favors the GARCH-SVAR(6) model over a homoskedastic VAR(6) model. Based on this finding we fit the GARCH model from Section 2, i.e., we assume that each underlying structural shock is driven by a univariate GARCH(1,1) process and use the conditional heteroskedasticity for the identification of the structural shocks.

Since the simulation results suggest that better coverage rates of the confidence intervals are obtained for the A-model setup, we use that for inference. The model is estimated by the two-step Gaussian ML algorithm of Lanne and Saikkonen (2007) and we apply the three different bootstrap methods and estimation techniques from Section 3 to compute confidence intervals for the impulse responses of the structural shocks. We compare the intervals obtained with the different methods in the following.

Our comparison of bootstrap procedures is based on pointwise 90% confidence intervals for two reasons. First, in their analysis Caldara and Herbst (2018) use pointwise 90%

[6] We denote variables in levels by capital letters while variables in logarithms are signified by lower case letters.

Bayesian credible sets which are comparable to pointwise confidence intervals. Second, using 90% joint Bonferroni confidence bands would simply lead to wider confidence bands in our economic analysis. The relative performance of the different methods is not affected by changing the confidence level.

Of course, prior to a structural analysis we have to label the three shocks of our system identified by conditional heteroskedasticity. Our main interest lies in the monetary policy shock and the financial shock. After labeling these two shocks, the last shock of the system clearly resembles a real activity shock which we do not include in the analysis, however.

Because the federal funds rate is the main policy instrument of the central bank, the monetary policy shock is chosen to be the shock which explains the largest share of the variance of the federal funds rate on impact. A contractionary monetary policy shock is expected to tighten financial market conditions by leading to higher credit spreads, i.e., a rise of the excess bond premium. As a reaction to a tightening of monetary policy, real activity should eventually slow down as investments become more costly for the economic agents.

The financial shock is chosen to be the shock which, on impact, explains the largest share of the variance of the excess bond premium. It is expected to slacken output growth due to higher financing costs for corporations. Based on the findings of Rigobon and Sack (2003), we expect the central bank to ease the stance of monetary policy as a reaction to a financial shock.

5.1 Monetary Policy Shock

The responses of our variables to a contractionary monetary policy shock are depicted in Figure 3. After rising for about half a year, the federal funds rate gradually declines and the effect of the shock disappears after roughly two years. The excess bond premium increases initially in response to the monetary policy shock. It remains at this level before starting to return to zero after about three years. Judging on the basis of the confidence intervals, this response is not significant however. Also industrial production does not respond significantly to a monetary policy shock, although it shows a negative response some time after the monetary policy shock has hit the system, as one would expect.

We now turn to the comparison of the bootstrap confidence intervals. As Figure 3 shows, for the monetary policy shock all bootstrapping procedures yield qualitatively similar results which is in line with the findings of our simulation study that did not reveal severe differences between the different bootstrapping procedures. They all lead to qualitatively similar conclusions. Also the differences between the alternative estimation strategies remain small in general.

Economically, our findings somewhat differ from the results reported by Caldara and Herbst (2018). The monetary policy shock identified by conditional heteroskedasticity is comparable in the shape of the impulse responses (see Figure 2 (bottom row) in Caldara and Herbst (2018)). However, in contrast to these authors' conclusion that during the Great Moderation period monetary policy affected asset prices considerably and also played a prominent role in business cycle fluctuations, our analysis, based on three different bootstrap procedures for inference, reveals that the responses of credit spreads and the real activity to a contractionary monetary policy shock remain inconclusive since considerable uncertainty prevails about the reactions of the variables.

5.2 Financial Shock

We proceed with the financial shock which, on impact, explains the largest share of the excess bond premium in a variance decomposition. The corresponding impulse responses are presented in Figure 4. The shock leads to an initial increase of the excess bond premium, which tapers off thereafter and becomes negative after roughly two years before starting to return towards zero. The federal funds rate does not react on impact to a financial shock but then accommodative monetary policy leads to a decrease of the federal funds rate for about two years. After that, monetary policy remains expansionary while fading out. Real activity does not react contemporaneously to a financial shock but slackens in the following months.

Again, all bootstrapping and estimation approaches produce quite similar confidence intervals, as expected from our trivariate simulation study. Clearly, they all lead to qualitatively the same conclusions.

The point estimates of the impulse responses to a financial shock are remarkably similar to those of Caldara and Herbst (2018) (see Figure 5 bottom row in Caldara and Herbst (2018)) who use a very different identification approach. In their paper, Caldara and Herbst (2018) remain silent about inference on their financial shock. We fill this gap and find that a financial shock which increases the spreads on financial markets (EBP), reduces the federal funds rate and industrial production. In other words, monetary policy reacts expansionary with significant reductions of the policy rate at medium horizons. This result is robust across all bootstraps and estimation methods. Yet, the exact timing and strength of the estimated central bank reaction differs slightly depending on the bootstrap procedure. Similarly, our estimates suggest that output contracts significantly after about six to twelve months and remains depressed for eight to 24 months depending on the bootstrap and even more on the estimation method applied. The qualitative findings, however, are supported by all bootstrap procedures and estimation techniques.

6 Conclusions

In this study we have compared a range of bootstrap methods for inference in SVAR models identified by conditional heteroskedasticity. The model for conditional heteroskedasticity is a multivariate GARCH model which fully identifies the structural parameters of the SVAR model. Three types of bootstrap methods are included in the comparison: A recursive-design residual-based wild bootstrap (WB), a residual-based moving blocks bootstrap (MBB) and a bootstrap which draws samples from the GARCH innovations (RBB). They are all used for setting up confidence intervals and confidence bands for structural impulse responses. Based on previous theoretical and simulation studies all the methods are expected to do well for the case considered in the present study.

For all three bootstrap methods, full Gaussian ML and a computationally cheaper estimation method due to Lanne and Saikkonen (2007) as well as a method which conditions on the first round GARCH parameter estimates are used for estimating the structural parameters and impulse responses in the bootstrap replications. The last method, which conditions on one set of the GARCH parameter estimates in all bootstrap replications, is comparable to the first step of the Lanne/Saikkonen method in terms of computation time and it is about 10 times faster than using full ML in each bootstrap replication.

It is found that conditioning in the bootstrap replications on the GARCH parameters estimated from the original data and using the WB gives overall the best balance between accurate coverage probability and interval or band width. Thus, the simplest, computationally most efficient method provides the best confidence intervals and bands and, thus, there is no reason for re-estimating the GARCH parameters in each bootstrap replication. However, in particular for smaller samples as they are not uncommon in macroeconometric studies, the actual coverage can still be substantially below the nominal coverage especially for the B-model. In our simulations, the coverage rates were often better for the A-model.

Our results are roughly in line with earlier simulation evidence by Brüggemann et al. (2016) who found that the true coverage rates of confidence intervals for impulse responses tended to be below the nominal rates for some of the bootstrap methods considered in our comparison. The crucial difference to their study is, however, that we identify the structural parameters via GARCH while they use conventional identification by exclusion restrictions. In contrast, Bruder (2018) also studies inference in SVARs identified by heteroskedasticity. He uses a different estimation method for the structural parameters and also different bivariate DGPs in his simulation comparison of different methods. In his study, a method similar to our RBB is often preferable to other methods in terms of coverage frequency. He does not include the wild bootstrap in his comparison, however.

Our Monte Carlo study has, of course, limitations. Clearly, it may be of interest to explore the robustness of our results with respect to extensions of the simulation setup. For example, using alternative GARCH parameters, including other bootstrap procedures and considering other VAR slope coefficients would be of interest.

We also provide an empirical application to macroeconomic data from the United States. For that purpose we consider a three-dimensional model consisting of the federal funds rate, a measure for an excess bond premium and industrial production growth. We identify a monetary policy shock and a financial shock via GARCH which turn out to have plausible impulse responses. The confidence intervals for the impulse responses generated with our different estimation and bootstrap methods are all quite similar and lead to roughly the same qualitative conclusions.

Since our results hold for volatility changes generated by GARCH dynamics and since a number of other volatility models have been considered in the context of identification through heteroskedasticity in SVAR analysis, our results suggest that similar studies may be worthwhile for other volatility models. Such studies may be an interesting direction for future research. More generally, the reliability of identification through heteroskedasticity in SVAR models may be of interest in future research.

Appendix

A VAR parameters for three-dimensional DGP

$$\nu = \begin{pmatrix} -0.08 \\ 0.03 \\ 0 \end{pmatrix},$$

$$A_1 = \begin{pmatrix} 0.667 & 0.119 & 1.390 \\ -0.160 & 1.281 & 5.590 \\ 0 & 0.005 & -.149 \end{pmatrix}, A_2 = \begin{pmatrix} 0.262 & 0.038 & -3.015 \\ -0.024 & -0.197 & 0.026 \\ 0.002 & -0.011 & 0.107 \end{pmatrix},$$

$$A_3 = \begin{pmatrix} -0.022 & -0.274 & 1.202 \\ -0.063 & 0.036 & 4.550 \\ 0 & 0.006 & 0.172 \end{pmatrix}, A_4 = \begin{pmatrix} 0.180 & 0.078 & -1.894 \\ 0.026 & -0.124 & 3.013 \\ -0.002 & -0.005 & 0.004 \end{pmatrix},$$

$$A_5 = \begin{pmatrix} -0.199 & 0.031 & -0.071 \\ 0.053 & 0.065 & -0.132 \\ 0.001 & 0.005 & 0.118 \end{pmatrix}, A_6 = \begin{pmatrix} 0.049 & 0.024 & 1.891 \\ -0.020 & -0.077 & -0.489 \\ 0 & -0.001 & 0.147 \end{pmatrix}$$

B Notes on Computations

All estimations of this Monte Carlo study were conducted with the statistical computing software R3.2.3. For maximization of the log-likelihood function the R-package "nloptr" provides the optimization routine "slsqp", a sequential (least-squares) quadratic programming (SQP) algorithm for nonlinearly constrained, gradient-based optimization. The algorithm supports both equality and inequality constraints. The former are applied in the first step (L/S) and the full maximization (ML) procedure by Lanne and Saikkonen (2007). Inequality constraints are used in all three estimation procedures to ensure the (joint) restrictions $g_k \geq 0$ and $\gamma_k > 0$ with $g_k + \gamma_k < 1$ on the GARCH parameters γ_k and g_k for $k = 1, \dots, K$.

To generate starting values for the GARCH parameters γ_k and g_k for $k = 1, \dots, K$, we first draw γ_k from a uniform distribution on the interval (0,1). Second, conditional on the draw for γ_k the interval for drawing g_k is restricted to fulfill $\gamma_k + g_k < 1$. As starting values for the impact effects matrix B we use the square root of the unconditional reduced form covariance matrix Σ_u of the (bootstrapped) data.

After every optimization of the log-likelihood, convergence of the optimization algorithm is checked. In case no convergence was achieved the optimization is repeated with a fresh draw of starting values. In our setup, however, this check does not exclude the possibility of a local optimum of the highly nonlinear log-likelihood function.

The runtime of one simulation replication for the bivariate DGP with $\alpha = 0.5$ or 0.9 with parallelized bootstrap replications of sample size $T = 500$ (200) using 100 cores (Intel Xeon Westmere X5650 processors) on the high performance computing server at Freie Universität Berlin is approximately 140 (45) minutes. The runtime of one simulation replication of the three-dimensional DGP using the conditional estimation methods based on estimated and true GARCH parameters for $T = 500$ (200) is 90 (40) minutes. For each simulation design this time has to be multiplied by 500 because we are using 500 replications of the simulations.

References

Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts, *Operations Research Quarterly* **20**: 451–468.

Belongia, M. T. and Ireland, P. N. (2015). Interest Rates and Money in the Measurement of Monetary Policy., *Journal of Business & Economic Statistics* **33**(2): 255–269.

Berkowitz, J., Birgean, I. and Kilian, L. (2000). On the finite-sample accuracy of nonparametric resampling algorithms for economic time series, *Advances in Econometrics* **14**: 77–105.

Boswijk, H. P. and van der Weide, R. (2011). Method of moments estimation of GO-GARCH models, *Journal of Econometrics* **163**: 118–126.

Bouakez, H. and Normandin, M. (2010). Fluctuations in the foreign exchange market: How important are monetary policy shocks?, *Journal of International Economics* **81**: 139–153.

Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco.

Bruder, S. (2018). Inference for structural impulse responses in SVAR-GARCH models, *Technical Report Working Paper 281*, Department of Economics, University of Zurich.

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

Caldara, D. and Herbst, E. (2018). Monetary Policy, Real Activity, and Credit Spreads: Evidence from Bayesian Proxy SVARs, *American Economic Journal: Macroeconomics* **forthcoming**.

Cavaliere, G., Pedersen, R. S. and Rahbek, A. (2018). Fixed volatility bootstrap for a class of ARCH(q) models, *Journal of Time Series Analysis* **39**: 920–941.

Christiano, L. J., Eichenbaum, M. and Evans, C. (1999). Monetary policy shocks: What have we learned and to what end?, *in* J. B. Taylor and M. Woodford (eds), *Handbook of Macroeconomics*, Vol. 1A, Elsevier, Amsterdam, pp. 65–148.

Cooper, R. L. (1972). The predictive performance of quarterly econometric models of the United States, *in* B. G. Hickman (ed.), *Econometric Models of Cyclical Behavior*, NBER, pp. 813–947.

Francq, C. and Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, *Bernoulli* **10**: 605–637.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*, John Wiley & Sons, New York.

Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations, *American Economic Review* **102**(4): 1692–1720.

Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **123**: 89–120.

Gonçalves, S. and Kilian, L. (2007). Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity, *Econometric Reviews* **26**: 609–641.

Granger, C. W. J. and Newbold, P. (1974). Spurious regressions in eonometrics, *Journal of Econometrics* **2**: 111–120.

Granger, C. W. J. and Newbold, P. (1975). Economic forecasting: The atheist's viewpoint, *in* G. A. Renton (ed.), *Modelling the Economy*, Heinemann Educational Books, London, pp. 131–148.

Granger, C. W. J. and Newbold, P. (1977). *Forecasting Economic Time Series*, Academic Press, New York.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer, New York.

He, C. and Teräsvirta, T. (1999). Fourth moment structure of the GARCH(p, q) process, *Econometric Theory* **15**: 824–846.

Herwartz, H. and Lütkepohl, H. (2014). Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks, *Journal of Econometrics* **183**: 104–116.

Hidalgo, J. and Zaffaroni, P. (2007). A goodness-of-fit test for ARCH(∞) models, *Journal of Econometrics* **141**: 835–875.

Jeong, M. (2017). Residual-based GARCH bootstrap and second order asymptotic refinements, *Econometric Theory* **33**: 779–790.

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

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

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

Kreiss, J.-P. (1997). Asymptotical properties of residual bootstrap for autoregressions, *Technical report*, TU Braunschweig.

Lanne, M. and Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* **40**: 1131–1149.

Lanne, M. and Saikkonen, P. (2007). A multivariate generalized orthogonal factor GARCH model, *Journal of Business & Economic Statistics* **25**: 61–75.

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

Lütkepohl, H. (2013). Identifying structural vector autoregressions via changes in volatility, *Advances in Econometrics* **32**: 169–203.

Lütkepohl, H. and Milunovich, G. (2016). Testing for identification in SVAR-GARCH models, *Journal of Economic Dynamics and Control* **73**: 241–258.

Lütkepohl, H. and Netšunajev, A. (2017). Structural vector autoregressions with smooth transition in variances, *Journal of Economic Dynamics and Control* **84**: 43–57.

Lütkepohl, H. and Schlaak, T. (2018). Choosing between different time-varying volatility models for structural vector autoregressive analysis, *Oxford Bulletin of Economics and Statistics* **80**(4): 715–735.

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

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

Milunovich, G. and Yang, M. (2013). On identifying structural VAR models via ARCH effects, *Journal of Time Series Econometrics* **5**: 117–131.

Nelson, C. R. (1972). The prediction performance of the FRB-MIT-PENN model of the U.S. economy, *American Economic Review* **62**: 902–917.

Netšunajev, A. (2013). Reaction to technology shocks in Markov-switching structural VARs: Identification via heteroskedasticity, *Journal of Macroeconomics* **36**: 51–62.

Newbold, P. and Granger, C. W. J. (1974). Experience with forecasting univariate time series and combination of forecasts, *Journal of the Royal Statistical Society* **A137**: 131–146.

Normandin, M. and Phaneuf, L. (2004). Monetary policy shocks: Testing identification conditions under time-varying conditional volatility, *Journal of Monetary Economics* **51**: 1217–1243.

Politis, D. N. and Romano, J. P. (1994). The stationary bootstrap, *Journal of the American Statistical Association* **89**: 1303–1313.

Rigobon, R. (2003). Identification through heteroskedasticity, *Review of Economics and Statistics* **85**: 777–792.

Rigobon, R. and Sack, B. (2003). Measuring the reaction of monetary policy to the stock market, *The Quarterly Journal of Economics* **118**(2): 639–669.

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

van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal GARCH model, *Journal of Applied Econometrics* **17**: 549–564.

Table 1: Relative Coverage Frequencies of Impact Effects with Average Confidence Interval Widths in Parentheses for Nominal Level of 90% for Bivariate Benchmark DGP ($\alpha = 0.9$)

Estim. method		$T = 200$			$T = 500$		
		WB	MBB	RBB	WB	MBB	RBB
				B-Model			
ML	b_{11}	0.58 (0.39)	0.51 (0.34)	0.48 (0.33)	0.69 (0.27)	0.70 (0.26)	0.53 (0.21)
	b_{21}	1.00 (10.48)	0.98 (8.25)	1.00 (11.11)	1.00 (8.08)	0.92 (4.58)	1.00 (8.02)
	b_{12}	1.00 (1.03)	0.97 (0.80)	1.00 (1.08)	1.00 (0.79)	0.94 (0.44)	1.00 (0.78)
	b_{22}	0.72 (3.53)	0.54 (2.76)	0.59 (2.97)	0.74 (2.56)	0.68 (2.09)	0.62 (2.07)
L/S	b_{11}	0.58 (0.39)	0.52 (0.34)	0.49 (0.33)	0.70 (0.29)	0.71 (0.27)	0.55 (0.23)
	b_{21}	1.00 (10.68)	0.98 (8.74)	1.00 (11.25)	1.00 (9.03)	0.94 (5.33)	1.00 (8.76)
	b_{12}	1.00 (1.05)	0.98 (0.85)	1.00 (1.09)	1.00 (0.89)	0.95 (0.51)	1.00 (0.85)
	b_{22}	0.72 (3.55)	0.54 (2.82)	0.59 (2.99)	0.75 (2.71)	0.69 (2.18)	0.63 (2.23)
CB	b_{11}	0.57 (0.37)	0.50 (0.32)	0.49 (0.33)	0.68 (0.24)	0.70 (0.25)	0.55 (0.22)
	b_{21}	0.99 (9.47)	0.92 (7.21)	0.99 (11.01)	1.00 (6.77)	0.90 (4.01)	1.00 (8.31)
	b_{12}	0.98 (0.93)	0.91 (0.70)	0.99 (1.07)	1.00 (0.67)	0.91 (0.39)	1.00 (0.81)
	b_{22}	0.71 (3.40)	0.52 (2.63)	0.58 (2.94)	0.73 (2.33)	0.67 (2.00)	0.62 (2.10)
True GARCH	b_{11}	0.58 (0.37)	0.50 (0.32)	0.54 (0.32)	0.67 (0.24)	0.70 (0.25)	0.62 (0.20)
	b_{21}	1.00 (9.54)	0.95 (7.37)	1.00 (10.57)	1.00 (6.79)	0.92 (4.00)	1.00 (7.74)
	b_{12}	1.00 (0.93)	0.94 (0.71)	1.00 (1.03)	1.00 (0.66)	0.93 (0.38)	1.00 (0.76)
	b_{22}	0.71 (3.40)	0.53 (2.66)	0.59 (2.86)	0.74 (2.33)	0.67 (1.99)	0.66 (1.97)
				A-Model			
ML	b_{11}	0.93 (0.39)	0.77 (0.31)	0.93 (0.41)	0.98 (0.28)	0.90 (0.14)	0.98 (0.26)
	b_{21}	1.00 (9.32)	0.97 (7.66)	1.00 (9.52)	1.00 (7.30)	0.92 (4.39)	1.00 (7.16)
	b_{12}	1.00 (0.09)	0.97 (0.07)	1.00 (0.09)	1.00 (0.07)	0.94 (0.04)	1.00 (0.07)
	b_{22}	0.93 (0.39)	0.77 (0.31)	0.93 (0.41)	0.98 (0.28)	0.90 (0.14)	0.98 (0.26)
L/S	b_{11}	0.91 (0.40)	0.77 (0.33)	0.90 (0.41)	0.96 (0.32)	0.88 (0.17)	0.97 (0.30)
	b_{21}	1.00 (9.46)	0.98 (8.07)	1.00 (9.59)	1.00 (8.01)	0.94 (5.07)	1.00 (7.68)
	b_{12}	1.00 (0.09)	0.98 (0.07)	1.00 (0.09)	1.00 (0.08)	0.95 (0.05)	1.00 (0.07)
	b_{22}	0.91 (0.40)	0.77 (0.33)	0.90 (0.41)	0.96 (0.32)	0.88 (0.17)	0.97 (0.30)
CB	b_{11}	0.91 (0.35)	0.74 (0.26)	0.91 (0.40)	0.98 (0.21)	0.90 (0.11)	0.98 (0.27)
	b_{21}	0.98 (8.56)	0.92 (6.76)	0.99 (9.40)	1.00 (6.29)	0.90 (3.89)	1.00 (7.35)
	b_{12}	0.98 (0.08)	0.91 (0.06)	0.99 (0.09)	1.00 (0.06)	0.91 (0.04)	1.00 (0.07)
	b_{22}	0.91 (0.35)	0.74 (0.26)	0.91 (0.40)	0.98 (0.21)	0.90 (0.11)	0.98 (0.27)
True GARCH	b_{11}	0.94 (0.35)	0.81 (0.26)	0.97 (0.38)	0.99 (0.21)	0.92 (0.11)	0.99 (0.24)
	b_{21}	1.00 (8.65)	0.95 (6.93)	1.00 (9.19)	1.00 (6.31)	0.93 (3.90)	1.00 (6.96)
	b_{12}	1.00 (0.08)	0.94 (0.06)	1.00 (0.09)	1.00 (0.06)	0.93 (0.04)	1.00 (0.07)
	b_{22}	0.94 (0.35)	0.81 (0.26)	0.97 (0.38)	0.99 (0.21)	0.92 (0.11)	0.99 (0.24)

Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table 2: Comparison of Different Block Lengths for MBB/CB Method for Bivariate Benchmark DGP ($\alpha = 0.9$), Relative Coverage Frequencies with Average Interval/Band Widths in Parentheses

		$T = 200$		$T = 500$	
Block length		10	20	10	50
B-Model					
Confidence level 90%					
Impact effects	b_{11}	0.50 (0.31)	0.50 (0.32)	0.62 (0.21)	0.70 (0.25)
	b_{21}	0.96 (8.20)	0.92 (7.21)	0.98 (5.86)	0.90 (4.01)
	b_{12}	0.95 (0.80)	0.91 (0.70)	0.97 (0.57)	0.91 (0.39)
	b_{22}	0.51 (2.62)	0.52 (2.63)	0.61 (1.84)	0.67 (2.00)
Pointwise confidence level 99%					
Joint bands	θ_{11}	0.54 (5.61)	0.53 (5.46)	0.69 (4.17)	0.69 (4.12)
	θ_{21}	0.91 (36.83)	0.88 (34.32)	0.96 (28.29)	0.92 (22.96)
	θ_{12}	0.99 (8.67)	0.98 (7.89)	1.00 (7.07)	0.98 (5.32)
	θ_{22}	0.68 (22.53)	0.64 (21.20)	0.77 (15.57)	0.76 (13.99)
A-Model					
Confidence level 90%					
Impact effects	b_{11}	0.78 (0.30)	0.74 (0.26)	0.94 (0.11)	0.90 (0.11)
	b_{21}	0.96 (7.56)	0.92 (6.76)	0.98 (3.89)	0.90 (3.89)
	b_{12}	0.95 (0.07)	0.91 (0.06)	0.97 (0.04)	0.91 (0.04)
	b_{22}	0.78 (0.30)	0.74 (0.26)	0.94 (0.11)	0.90 (0.11)
Pointwise confidence level 99%					
Joint bands	θ_{11}	0.78 (5.59)	0.71 (5.30)	0.89 (4.20)	0.83 (3.59)
	θ_{21}	0.92 (34.54)	0.90 (32.92)	0.97 (25.77)	0.92 (22.14)
	θ_{12}	0.99 (0.74)	0.98 (0.69)	1.00 (0.61)	0.98 (0.49)
	θ_{22}	0.90 (2.28)	0.87 (2.11)	0.96 (1.60)	0.92 (1.30)

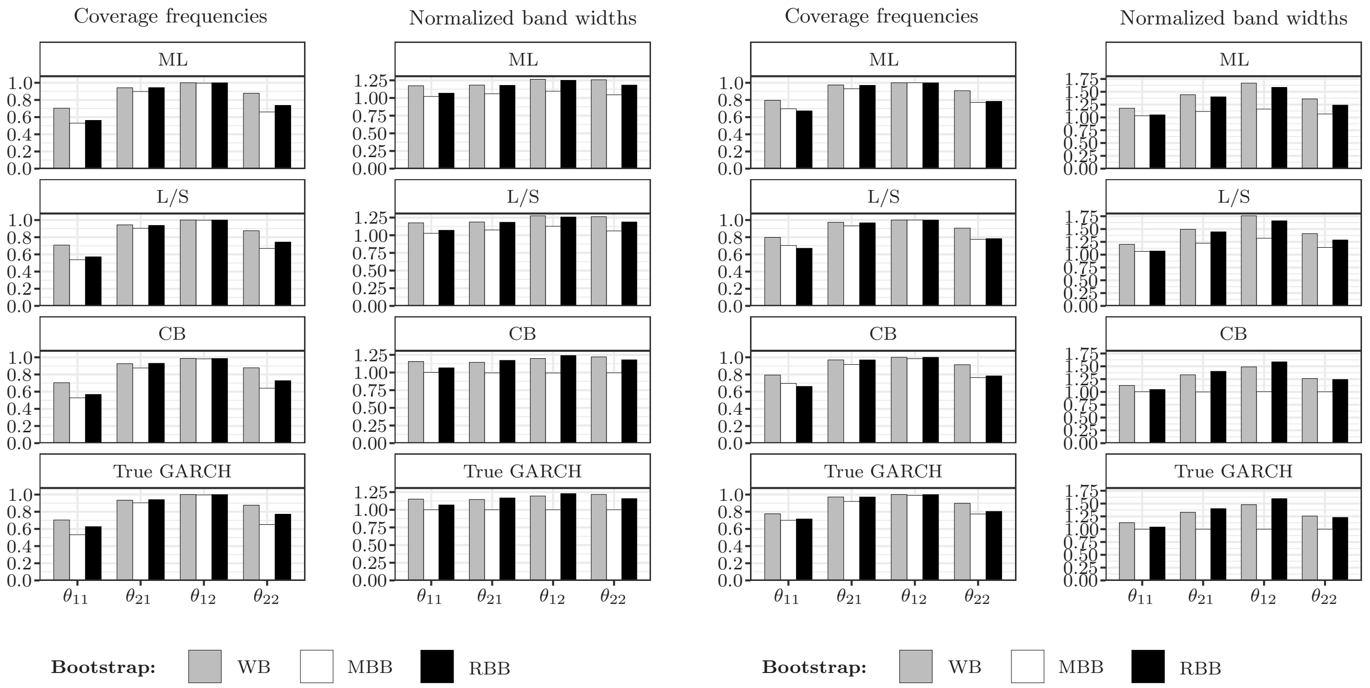
Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$.

Table 3: Relative Coverage Frequencies of Impact Effects with Average Confidence Interval Widths in Parentheses for Nominal Level of 90% for Three-dimensional DGP

Estimation method		$T = 200$ WB	$T = 200$ MBB	$T = 200$ RBB	$T = 500$ WB	$T = 500$ MBB	$T = 500$ RBB
				B-Model			
CB	b_{11}	0.21 (0.072)	0.27 (0.080)	0.19 (0.067)	0.27 (0.056)	0.44 (0.090)	0.23 (0.050)
	b_{21}	0.99 (0.100)	0.99 (0.091)	0.99 (0.117)	1.00 (0.062)	0.99 (0.045)	0.99 (0.075)
	b_{31}	0.99 (0.003)	0.99 (0.003)	0.99 (0.004)	1.00 (0.002)	0.94 (0.002)	0.99 (0.003)
	b_{12}	0.99 (0.148)	0.99 (0.132)	0.99 (0.169)	1.00 (0.094)	1.00 (0.066)	1.00 (0.107)
	b_{22}	0.30 (0.044)	0.28 (0.039)	0.37 (0.044)	0.51 (0.036)	0.53 (0.034)	0.58 (0.038)
	b_{32}	0.99 (0.004)	0.99 (0.004)	0.99 (0.005)	1.00 (0.004)	1.00 (0.004)	1.00 (0.005)
	b_{13}	0.99 (0.159)	0.99 (0.140)	0.99 (0.181)	1.00 (0.106)	0.97 (0.078)	1.00 (0.137)
	b_{23}	0.99 (0.126)	0.99 (0.121)	0.99 (0.135)	1.00 (0.122)	1.00 (0.106)	1.00 (0.141)
	b_{33}	0.26 (0.002)	0.20 (0.001)	0.19 (0.002)	0.53 (0.001)	0.45 (0.001)	0.46 (0.001)
True GARCH	b_{11}	0.20 (0.072)	0.26 (0.081)	0.18 (0.068)	0.27 (0.058)	0.45 (0.090)	0.22 (0.051)
	b_{21}	1.00 (0.101)	1.00 (0.091)	1.00 (0.120)	1.00 (0.066)	0.99 (0.046)	0.99 (0.080)
	b_{31}	1.00 (0.004)	1.00 (0.003)	1.00 (0.004)	1.00 (0.002)	0.94 (0.002)	0.99 (0.003)
	b_{12}	1.00 (0.148)	1.00 (0.130)	1.00 (0.171)	1.00 (0.099)	1.00 (0.066)	1.00 (0.113)
	b_{22}	0.35 (0.046)	0.30 (0.041)	0.39 (0.046)	0.51 (0.037)	0.54 (0.035)	0.58 (0.039)
	b_{32}	1.00 (0.004)	1.00 (0.004)	1.00 (0.005)	1.00 (0.004)	1.00 (0.004)	1.00 (0.005)
	b_{13}	1.00 (0.161)	1.00 (0.143)	1.00 (0.184)	1.00 (0.112)	0.97 (0.079)	1.00 (0.141)
	b_{23}	1.00 (0.126)	1.00 (0.121)	1.00 (0.136)	1.00 (0.125)	1.00 (0.107)	1.00 (0.143)
	b_{33}	0.28 (0.002)	0.23 (0.002)	0.20 (0.002)	0.53 (0.001)	0.45 (0.001)	0.44 (0.001)
				A-Model			
CB	b_{11}	0.66 (0.480)	0.61 (0.436)	0.39 (0.524)	0.87 (0.245)	0.81 (0.164)	0.67 (0.301)
	b_{21}	0.99 (0.523)	0.99 (0.496)	0.99 (0.562)	1.00 (0.324)	0.99 (0.256)	1.00 (0.364)
	b_{31}	0.99 (0.018)	0.99 (0.017)	0.99 (0.019)	1.00 (0.012)	0.94 (0.010)	0.99 (0.013)
	b_{12}	0.99 (1.164)	0.99 (1.057)	0.99 (1.218)	1.00 (0.742)	1.00 (0.545)	1.00 (0.751)
	b_{22}	0.27 (0.520)	0.28 (0.490)	0.12 (0.545)	0.46 (0.425)	0.52 (0.369)	0.33 (0.458)
	b_{32}	0.99 (0.032)	0.99 (0.032)	0.99 (0.032)	1.00 (0.030)	1.00 (0.027)	1.00 (0.030)
	b_{13}	0.99 (35.877)	0.99 (32.450)	0.99 (38.638)	1.00 (24.716)	0.97 (19.022)	0.99 (28.859)
	b_{23}	0.99 (28.358)	0.99 (27.641)	0.99 (28.887)	1.00 (26.456)	1.00 (24.010)	1.00 (28.471)
	b_{33}	0.19 (0.564)	0.21 (0.522)	0.05 (0.598)	0.31 (0.434)	0.40 (0.372)	0.10 (0.477)
True GARCH	b_{11}	0.70 (0.488)	0.67 (0.442)	0.38 (0.532)	0.86 (0.268)	0.82 (0.168)	0.63 (0.316)
	b_{21}	1.00 (0.527)	1.00 (0.495)	1.00 (0.570)	1.00 (0.341)	0.99 (0.259)	1.00 (0.382)
	b_{31}	1.00 (0.019)	1.00 (0.017)	1.00 (0.019)	1.00 (0.013)	0.94 (0.010)	0.99 (0.014)
	b_{12}	1.00 (1.167)	1.00 (1.047)	1.00 (1.235)	1.00 (0.778)	1.00 (0.545)	1.00 (0.791)
	b_{22}	0.32 (0.543)	0.34 (0.510)	0.10 (0.568)	0.44 (0.442)	0.52 (0.375)	0.28 (0.472)
	b_{32}	1.00 (0.032)	1.00 (0.031)	1.00 (0.032)	1.00 (0.030)	1.00 (0.028)	1.00 (0.030)
	b_{13}	1.00 (36.632)	1.00 (33.187)	1.00 (39.583)	1.00 (25.723)	0.97 (19.217)	0.99 (29.453)
	b_{23}	1.00 (28.109)	1.00 (27.431)	1.00 (28.700)	1.00 (26.669)	1.00 (24.142)	1.00 (28.525)
	b_{33}	0.18 (0.570)	0.23 (0.526)	0.04 (0.604)	0.27 (0.449)	0.38 (0.379)	0.09 (0.489)

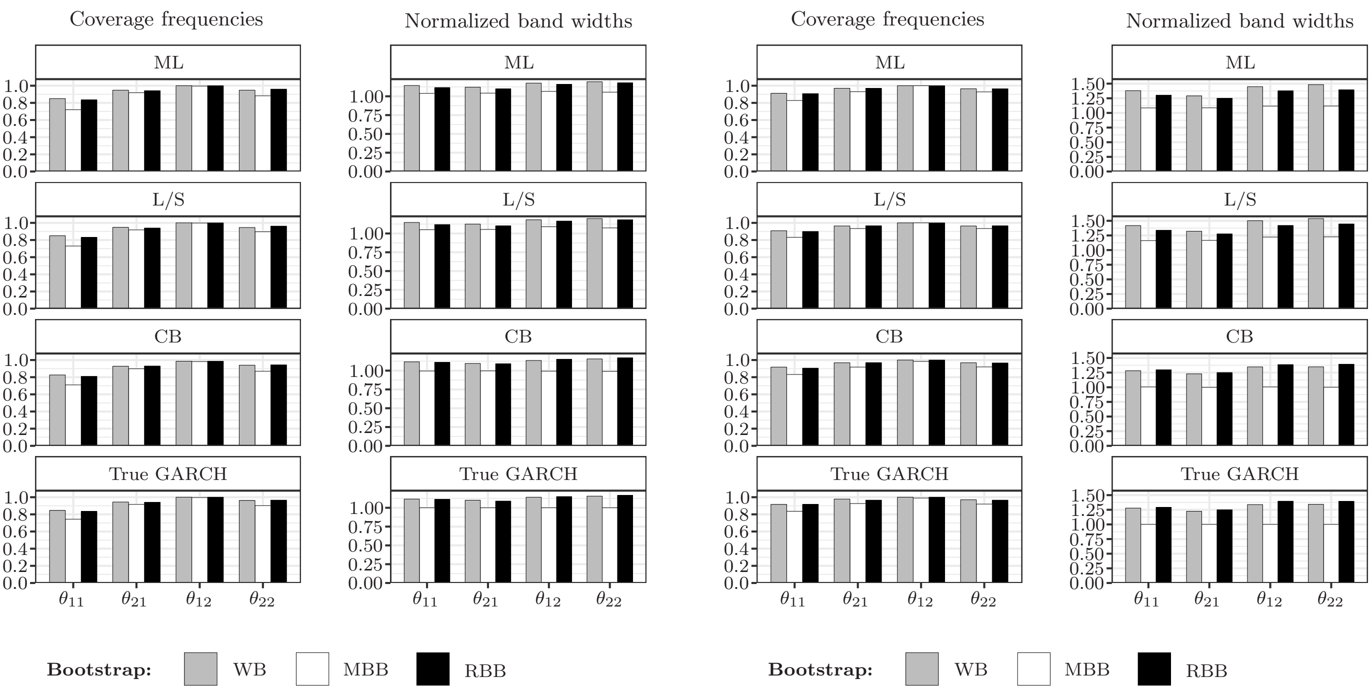
Note: Band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

B-Model



(a) $T = 200$ (b) $T = 500$

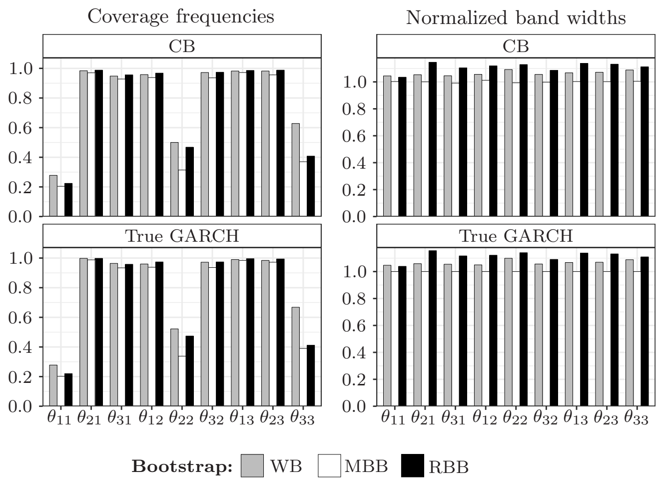
A-Model



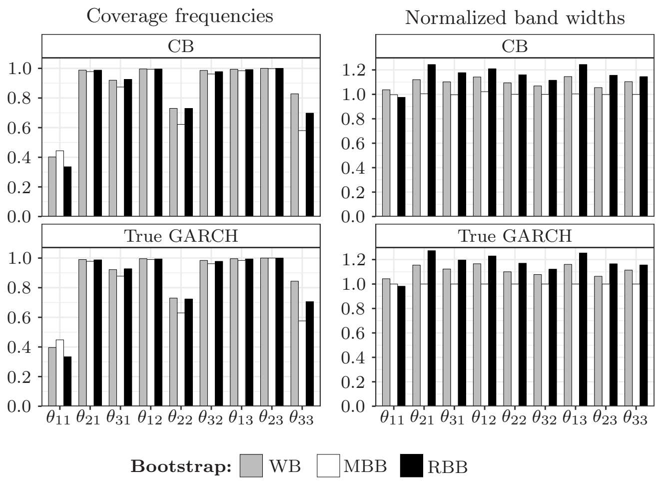
(c) $T = 200$ (d) $T = 500$

Figure 1: Relative coverage frequencies of joint confidence bands for impulse response functions with propagation horizon up to 10 and corresponding average normalized band widths for pointwise 99% confidence level for bivariate benchmark DGP ($\alpha = 0.9$) with GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$ (block lengths for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively).

B-Model

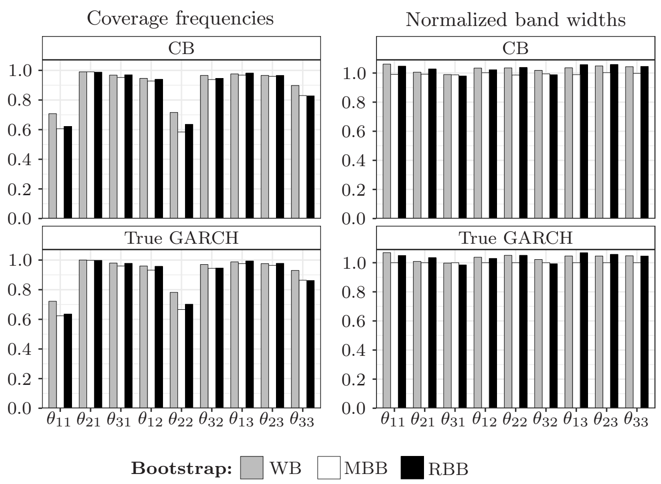


(a) $T = 200$

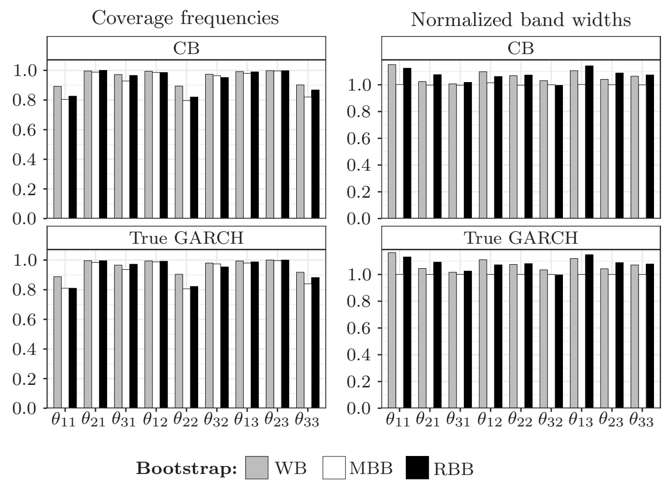


(b) $T = 500$

A-Model



(c) $T = 200$



(d) $T = 500$

Figure 2: Relative coverage frequencies of joint confidence bands for impulse response functions with propagation horizon up to 10 and corresponding average normalized band widths for pointwise 99% confidence level for three-dimensional DGP (block lengths for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively).

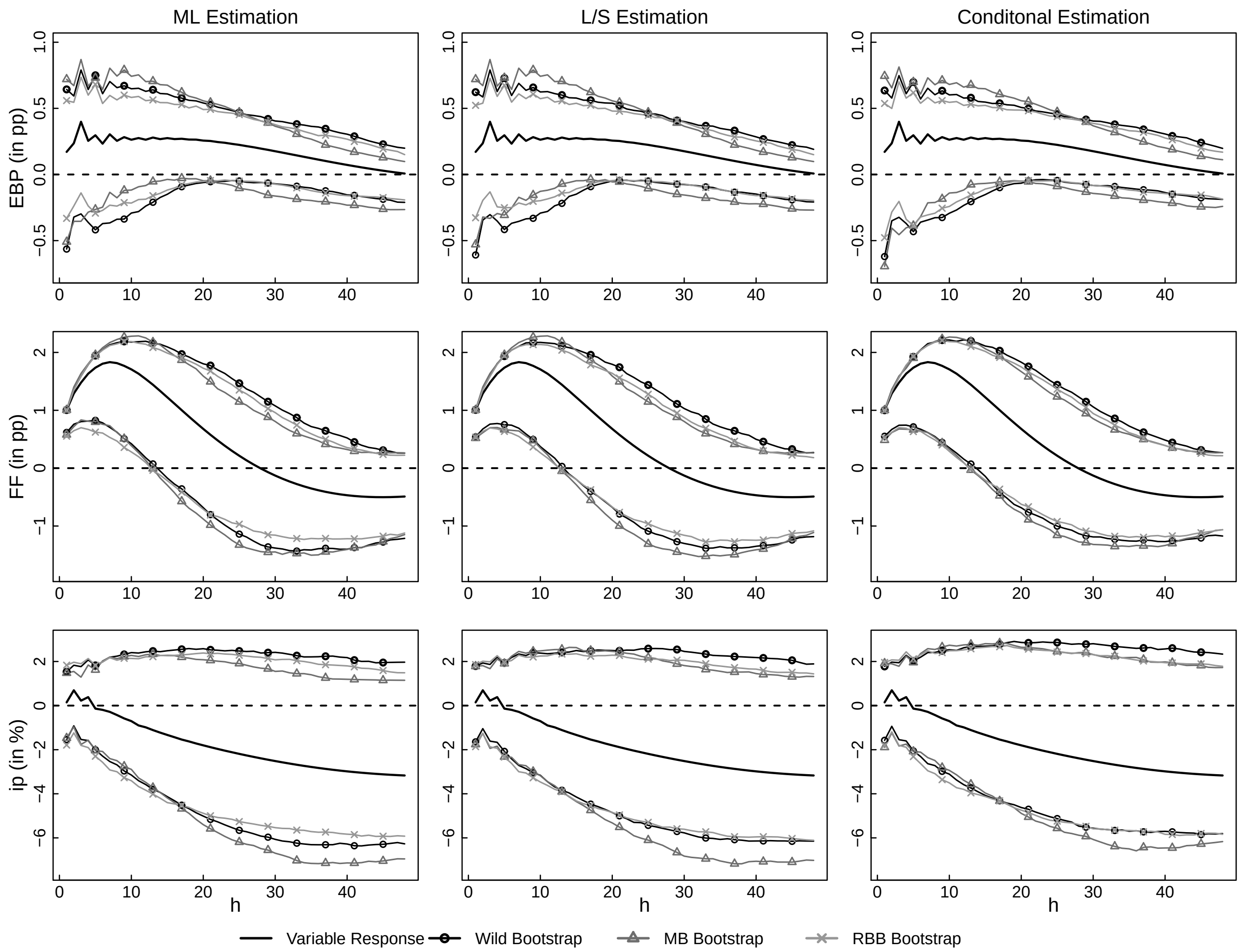


Figure 3: Comparison of 90% pointwise confidence intervals of different bootstrap procedures and estimation methods for a monetary policy shock.

Notes: The solid black line depicts the response of the variables to a shock for the A-model. The response of Δip has been cumulated and multiplied by 100. The symboled lines (black dot, dark grey triangle, light grey asterisk) represent the different bootstrapped confidence intervals (WB, MBB, RBB, respectively).

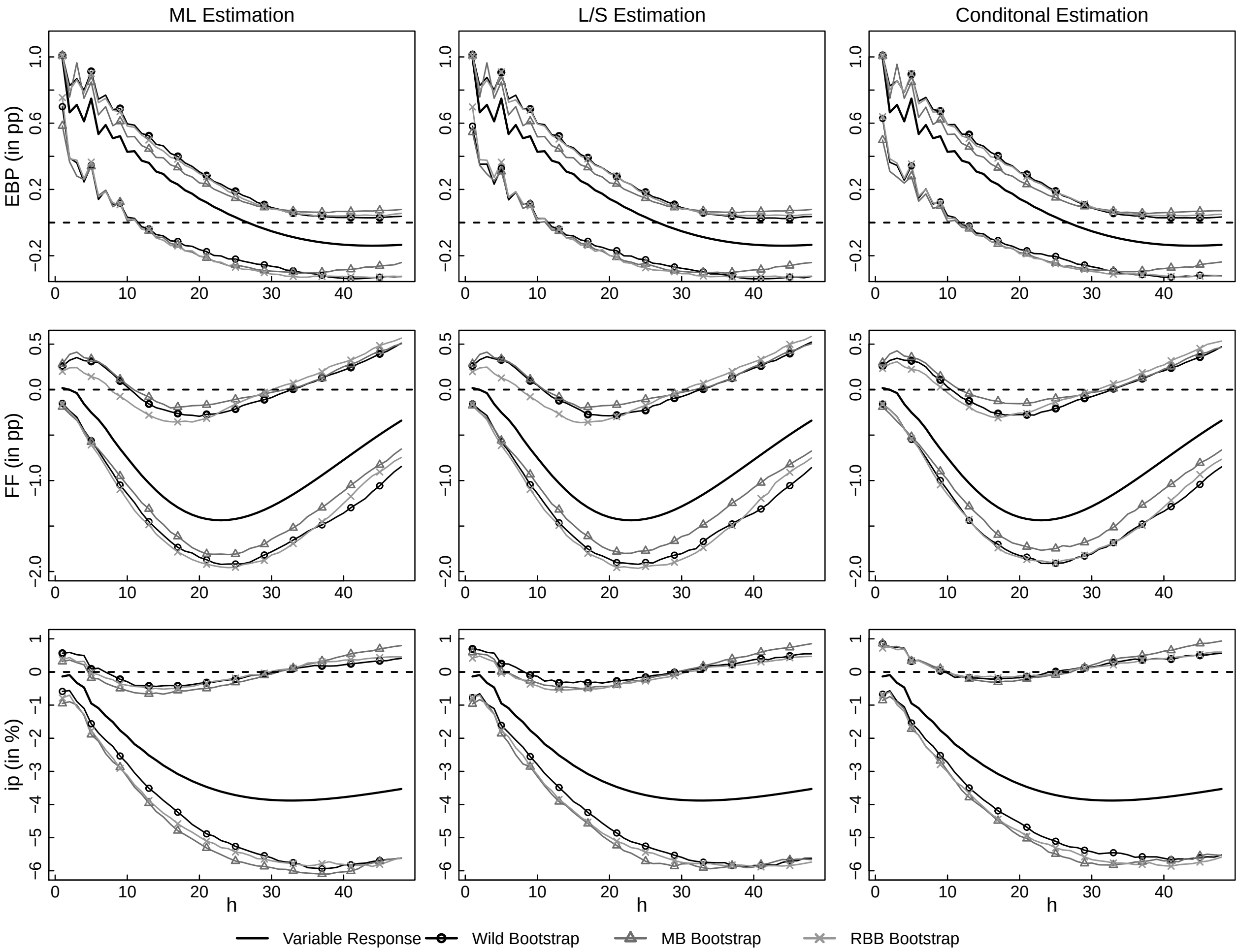


Figure 4: Comparison of 90 % Pointwise Confidence Intervals of Different Bootstrap Procedures and Estimation Methods for a Financial Shock.

Notes: The solid black line depicts the response of the variables to a shock for the A-model. The response of Δip has been cumulated and multiplied by 100. The symboled lines (black dot, dark grey triangle, light grey asterisk) represent the different bootstrapped confidence intervals (WB, MBB, RBB, respectively).

Appendix. Supplementary Results

Table A.1: Relative Coverage Frequencies of Joint Bands for Impulse Response Functions with Average Band Widths in Parentheses for Pointwise Nominal Level of 99% for Bivariate Benchmark DGP ($\alpha = 0.9$)

Estim. method		$T = 200$ WB	$T = 200$ MBB	$T = 200$ RBB	$T = 500$ WB	$T = 500$ MBB	$T = 500$ RBB
		B-Model					
ML	θ_{11}	0.70 (6.40)	0.53 (5.57)	0.56 (5.83)	0.80 (4.84)	0.70 (4.24)	0.67 (4.31)
	θ_{21}	0.94 (40.80)	0.90 (36.56)	0.94 (40.68)	0.97 (33.14)	0.93 (25.62)	0.97 (32.16)
	θ_{12}	1.00 (10.04)	1.00 (8.71)	1.00 (9.92)	1.00 (8.85)	1.00 (6.16)	1.00 (8.38)
	θ_{22}	0.88 (26.76)	0.66 (22.24)	0.74 (25.18)	0.91 (18.99)	0.77 (14.89)	0.78 (17.31)
L/S	θ_{11}	0.71 (6.41)	0.54 (5.61)	0.57 (5.84)	0.80 (4.94)	0.70 (4.37)	0.67 (4.40)
	θ_{21}	0.94 (40.97)	0.90 (37.09)	0.94 (40.86)	0.97 (34.31)	0.93 (28.09)	0.97 (33.22)
	θ_{12}	1.00 (10.14)	1.00 (8.95)	1.00 (10.00)	1.00 (9.32)	1.00 (7.00)	1.00 (8.80)
	θ_{22}	0.88 (26.89)	0.67 (22.59)	0.74 (25.30)	0.91 (19.70)	0.78 (15.93)	0.78 (17.96)
CB	θ_{11}	0.70 (6.30)	0.53 (5.46)	0.57 (5.82)	0.79 (4.63)	0.70 (4.12)	0.66 (4.30)
	θ_{21}	0.93 (39.47)	0.88 (34.32)	0.93 (40.44)	0.97 (30.64)	0.92 (22.96)	0.97 (32.20)
	θ_{12}	0.99 (9.52)	0.98 (7.89)	0.99 (9.85)	1.00 (7.89)	0.98 (5.32)	1.00 (8.41)
	θ_{22}	0.88 (26.01)	0.64 (21.20)	0.73 (25.12)	0.91 (17.61)	0.76 (14.00)	0.78 (17.36)
True GARCH	θ_{11}	0.70 (6.28)	0.53 (5.46)	0.63 (5.83)	0.78 (4.63)	0.70 (4.11)	0.72 (4.28)
	θ_{21}	0.93 (39.54)	0.90 (34.51)	0.94 (40.30)	0.97 (30.54)	0.92 (22.97)	0.97 (32.15)
	θ_{12}	1.00 (9.50)	1.00 (7.95)	1.00 (9.77)	1.00 (7.83)	0.99 (5.30)	1.00 (8.44)
	θ_{22}	0.88 (25.92)	0.65 (21.29)	0.77 (24.67)	0.90 (17.55)	0.77 (13.97)	0.80 (17.18)
		A-Model					
ML	θ_{11}	0.85 (6.09)	0.72 (5.53)	0.83 (5.94)	0.91 (4.93)	0.83 (3.88)	0.91 (4.66)
	θ_{21}	0.95 (37.05)	0.92 (34.44)	0.94 (36.31)	0.97 (28.62)	0.93 (24.09)	0.97 (27.72)
	θ_{12}	1.00 (0.81)	1.00 (0.74)	1.00 (0.80)	1.00 (0.71)	1.00 (0.54)	1.00 (0.67)
	θ_{22}	0.95 (2.54)	0.88 (2.25)	0.96 (2.51)	0.96 (1.92)	0.93 (1.45)	0.96 (1.81)
L/S	θ_{11}	0.85 (6.10)	0.73 (5.59)	0.83 (5.95)	0.91 (5.07)	0.83 (4.15)	0.90 (4.79)
	θ_{21}	0.95 (37.16)	0.92 (34.81)	0.94 (36.42)	0.96 (29.31)	0.93 (25.85)	0.97 (28.31)
	θ_{12}	1.00 (0.82)	1.00 (0.76)	1.00 (0.81)	1.00 (0.73)	1.00 (0.60)	1.00 (0.69)
	θ_{22}	0.95 (2.55)	0.90 (2.29)	0.96 (2.52)	0.96 (2.00)	0.93 (1.59)	0.97 (1.88)
CB	θ_{11}	0.83 (5.94)	0.71 (5.30)	0.81 (5.91)	0.92 (4.58)	0.83 (3.59)	0.91 (4.64)
	θ_{21}	0.93 (36.20)	0.90 (32.92)	0.93 (36.08)	0.97 (27.24)	0.92 (22.14)	0.97 (27.74)
	θ_{12}	0.99 (0.79)	0.98 (0.69)	0.99 (0.80)	1.00 (0.66)	0.98 (0.49)	1.00 (0.68)
	θ_{22}	0.94 (2.46)	0.87 (2.11)	0.94 (2.50)	0.97 (1.75)	0.92 (1.30)	0.97 (1.81)
True GARCH	θ_{11}	0.85 (5.94)	0.74 (5.32)	0.83 (5.92)	0.92 (4.57)	0.84 (3.58)	0.92 (4.62)
	θ_{21}	0.94 (36.37)	0.92 (33.04)	0.94 (36.00)	0.98 (27.14)	0.93 (22.17)	0.97 (27.65)
	θ_{12}	1.00 (0.79)	1.00 (0.69)	1.00 (0.80)	1.00 (0.65)	0.99 (0.49)	1.00 (0.68)
	θ_{22}	0.96 (2.46)	0.90 (2.13)	0.97 (2.49)	0.97 (1.74)	0.92 (1.30)	0.97 (1.81)

Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.2: Relative Coverage Frequencies of Impact Effects with Average Confidence Interval Widths in Parentheses for Nomial Level of 90% for Bivariate Benchmark DGP ($\alpha = 0.5$)

Estim. method		$T = 200$ WB	$T = 200$ MBB	$T = 200$ RBB	$T = 500$ WB	$T = 500$ MBB	$T = 500$ RBB
				B-Model			
ML	b_{11}	0.59 (0.39)	0.51 (0.34)	0.50 (0.33)	0.67 (0.27)	0.68 (0.26)	0.53 (0.21)
	b_{21}	1.00 (10.54)	0.97 (8.35)	1.00 (11.11)	1.00 (8.05)	0.92 (4.55)	1.00 (8.08)
	b_{12}	1.00 (1.04)	0.97 (0.81)	1.00 (1.08)	1.00 (0.78)	0.94 (0.43)	1.00 (0.78)
	b_{22}	0.72 (3.54)	0.54 (2.78)	0.56 (2.97)	0.76 (2.56)	0.70 (2.11)	0.63 (2.10)
L/S	b_{11}	0.58 (0.39)	0.52 (0.35)	0.50 (0.33)	0.68 (0.29)	0.69 (0.27)	0.56 (0.23)
	b_{21}	1.00 (10.77)	0.98 (8.92)	1.00 (11.29)	1.00 (9.07)	0.95 (5.33)	1.00 (8.84)
	b_{12}	1.00 (1.06)	0.98 (0.86)	1.00 (1.10)	1.00 (0.88)	0.95 (0.51)	1.00 (0.85)
	b_{22}	0.72 (3.56)	0.55 (2.84)	0.56 (2.99)	0.76 (2.72)	0.71 (2.21)	0.64 (2.25)
CB	b_{11}	0.58 (0.37)	0.51 (0.32)	0.50 (0.33)	0.67 (0.24)	0.68 (0.25)	0.54 (0.22)
	b_{21}	0.98 (9.41)	0.90 (7.20)	0.98 (10.93)	0.99 (6.68)	0.89 (3.93)	1.00 (8.39)
	b_{12}	0.98 (0.93)	0.90 (0.70)	0.98 (1.06)	0.99 (0.65)	0.90 (0.38)	1.00 (0.81)
	b_{22}	0.70 (3.39)	0.52 (2.62)	0.56 (2.93)	0.75 (2.33)	0.68 (2.01)	0.63 (2.13)
True GARCH	b_{11}	0.58 (0.37)	0.51 (0.32)	0.56 (0.32)	0.67 (0.24)	0.70 (0.25)	0.62 (0.20)
	b_{21}	1.00 (9.54)	0.94 (7.35)	1.00 (10.72)	1.00 (6.81)	0.93 (4.02)	1.00 (7.70)
	b_{12}	1.00 (0.93)	0.94 (0.71)	1.00 (1.05)	1.00 (0.67)	0.94 (0.38)	1.00 (0.76)
	b_{22}	0.70 (3.41)	0.54 (2.66)	0.61 (2.89)	0.74 (2.33)	0.68 (1.99)	0.66 (1.95)
				A-Model			
ML	b_{11}	0.93 (0.40)	0.76 (0.31)	0.92 (0.41)	0.98 (0.28)	0.89 (0.14)	0.98 (0.26)
	b_{21}	1.00 (9.34)	0.97 (7.72)	1.00 (9.52)	1.00 (7.30)	0.92 (4.39)	1.00 (7.22)
	b_{12}	1.00 (0.09)	0.97 (0.07)	1.00 (0.09)	1.00 (0.07)	0.94 (0.04)	1.00 (0.07)
	b_{22}	0.93 (0.40)	0.76 (0.31)	0.92 (0.41)	0.98 (0.28)	0.89 (0.14)	0.98 (0.26)
L/S	b_{11}	0.92 (0.41)	0.75 (0.33)	0.90 (0.41)	0.97 (0.32)	0.87 (0.17)	0.96 (0.30)
	b_{21}	1.00 (9.50)	0.98 (8.18)	1.00 (9.62)	1.00 (8.08)	0.95 (5.09)	1.00 (7.74)
	b_{12}	1.00 (0.09)	0.98 (0.08)	1.00 (0.09)	1.00 (0.08)	0.95 (0.05)	1.00 (0.07)
	b_{22}	0.92 (0.41)	0.75 (0.33)	0.90 (0.41)	0.97 (0.32)	0.87 (0.17)	0.96 (0.30)
CB	b_{11}	0.89 (0.35)	0.74 (0.26)	0.89 (0.40)	0.98 (0.21)	0.90 (0.11)	0.96 (0.27)
	b_{21}	0.98 (8.51)	0.90 (6.72)	0.98 (9.32)	0.99 (6.23)	0.89 (3.83)	1.00 (7.41)
	b_{12}	0.98 (0.08)	0.90 (0.06)	0.98 (0.09)	0.99 (0.06)	0.90 (0.04)	1.00 (0.07)
	b_{22}	0.89 (0.35)	0.74 (0.26)	0.89 (0.40)	0.98 (0.21)	0.90 (0.11)	0.96 (0.27)
True GARCH	b_{11}	0.95 (0.35)	0.80 (0.27)	0.97 (0.39)	0.98 (0.21)	0.92 (0.11)	1.00 (0.24)
	b_{21}	1.00 (8.64)	0.94 (6.90)	1.00 (9.25)	1.00 (6.32)	0.93 (3.92)	1.00 (6.93)
	b_{12}	1.00 (0.08)	0.94 (0.06)	1.00 (0.09)	1.00 (0.06)	0.94 (0.04)	1.00 (0.07)
	b_{22}	0.95 (0.35)	0.80 (0.27)	0.97 (0.39)	0.98 (0.21)	0.92 (0.11)	1.00 (0.24)

Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.3: Relative Coverage Frequencies of Joint Bands for Impulse Response Functions with Average Band Widths in Parentheses for Pointwise Nominal Level of 99% for Bivariate Benchmark DGP ($\alpha = 0.5$)

Estim. method		$T = 200$ WB	$T = 200$ MBB	$T = 200$ RBB	$T = 500$ WB	$T = 500$ MBB	$T = 500$ RBB
				B-Model			
ML	θ_{11}	0.82 (2.27)	0.62 (1.90)	0.64 (2.05)	0.84 (1.59)	0.75 (1.37)	0.69 (1.39)
	θ_{21}	0.99 (34.55)	0.96 (30.73)	0.99 (34.62)	0.99 (28.58)	0.95 (21.25)	0.99 (27.64)
	θ_{12}	1.00 (3.53)	0.97 (3.07)	1.00 (3.47)	1.00 (2.85)	0.98 (2.06)	1.00 (2.71)
	θ_{22}	0.88 (21.00)	0.66 (17.49)	0.75 (19.36)	0.92 (14.60)	0.79 (12.28)	0.77 (13.26)
L/S	θ_{11}	0.82 (2.27)	0.62 (1.92)	0.64 (2.06)	0.84 (1.63)	0.76 (1.42)	0.70 (1.43)
	θ_{21}	0.99 (34.79)	0.96 (31.33)	0.99 (34.80)	0.99 (29.85)	0.96 (23.84)	0.99 (28.91)
	θ_{12}	1.00 (3.56)	0.97 (3.14)	1.00 (3.49)	1.00 (2.99)	0.98 (2.31)	1.00 (2.84)
	θ_{22}	0.88 (21.05)	0.68 (17.65)	0.74 (19.41)	0.92 (14.93)	0.80 (13.65)	0.78 (13.59)
CB	θ_{11}	0.81 (2.22)	0.61 (1.84)	0.64 (2.04)	0.83 (1.50)	0.76 (1.32)	0.70 (1.39)
	θ_{21}	0.97 (32.93)	0.92 (27.97)	0.97 (34.16)	0.98 (25.78)	0.92 (18.41)	0.98 (27.74)
	θ_{12}	0.97 (3.36)	0.94 (2.79)	0.97 (3.43)	0.99 (2.55)	0.96 (1.79)	1.00 (2.72)
	θ_{22}	0.89 (20.65)	0.66 (16.99)	0.73 (19.28)	0.92 (13.90)	0.78 (11.82)	0.77 (13.28)
True GARCH	θ_{11}	0.81 (2.22)	0.62 (1.85)	0.71 (2.06)	0.84 (1.51)	0.76 (1.33)	0.77 (1.38)
	θ_{21}	0.99 (33.15)	0.94 (28.37)	0.99 (34.24)	0.98 (25.81)	0.94 (18.59)	0.98 (27.50)
	θ_{12}	1.00 (3.36)	0.97 (2.82)	0.99 (3.47)	1.00 (2.57)	0.96 (1.81)	1.00 (2.74)
	θ_{22}	0.88 (20.67)	0.66 (17.04)	0.78 (19.23)	0.91 (13.84)	0.78 (11.76)	0.80 (12.97)
				A-Model			
ML	θ_{11}	0.96 (2.15)	0.89 (1.87)	0.96 (2.07)	0.97 (1.60)	0.92 (1.23)	0.96 (1.48)
	θ_{21}	0.99 (30.39)	0.96 (28.13)	0.99 (29.95)	0.99 (24.00)	0.95 (19.70)	0.98 (23.14)
	θ_{12}	1.00 (0.29)	0.97 (0.27)	1.00 (0.29)	1.00 (0.23)	0.98 (0.18)	1.00 (0.22)
	θ_{22}	0.96 (2.09)	0.91 (1.86)	0.96 (2.05)	0.98 (1.57)	0.94 (1.22)	0.97 (1.48)
L/S	θ_{11}	0.96 (2.15)	0.90 (1.90)	0.95 (2.08)	0.97 (1.65)	0.93 (1.32)	0.96 (1.53)
	θ_{21}	0.99 (30.52)	0.96 (28.54)	0.99 (30.03)	0.99 (24.72)	0.96 (21.45)	0.99 (23.84)
	θ_{12}	1.00 (0.29)	0.97 (0.27)	1.00 (0.29)	1.00 (0.24)	0.98 (0.20)	1.00 (0.23)
	θ_{22}	0.96 (2.10)	0.91 (1.89)	0.96 (2.06)	0.97 (1.63)	0.95 (1.32)	0.97 (1.54)
CB	θ_{11}	0.94 (2.08)	0.85 (1.76)	0.92 (2.05)	0.97 (1.47)	0.92 (1.11)	0.95 (1.48)
	θ_{21}	0.97 (29.37)	0.92 (26.22)	0.96 (29.55)	0.98 (22.41)	0.92 (17.63)	0.98 (23.21)
	θ_{12}	0.97 (0.28)	0.94 (0.25)	0.97 (0.28)	0.99 (0.22)	0.96 (0.17)	1.00 (0.22)
	θ_{22}	0.93 (2.03)	0.86 (1.75)	0.94 (2.03)	0.97 (1.43)	0.93 (1.09)	0.97 (1.48)
True GARCH	θ_{11}	0.96 (2.09)	0.90 (1.78)	0.95 (2.07)	0.96 (1.47)	0.92 (1.12)	0.96 (1.48)
	θ_{21}	0.98 (29.66)	0.95 (26.57)	0.99 (29.52)	0.98 (22.35)	0.93 (17.72)	0.98 (22.93)
	θ_{12}	1.00 (0.29)	0.97 (0.25)	0.99 (0.29)	1.00 (0.22)	0.96 (0.17)	1.00 (0.22)
	θ_{22}	0.97 (2.04)	0.91 (1.77)	0.97 (2.05)	0.97 (1.43)	0.93 (1.10)	0.97 (1.47)

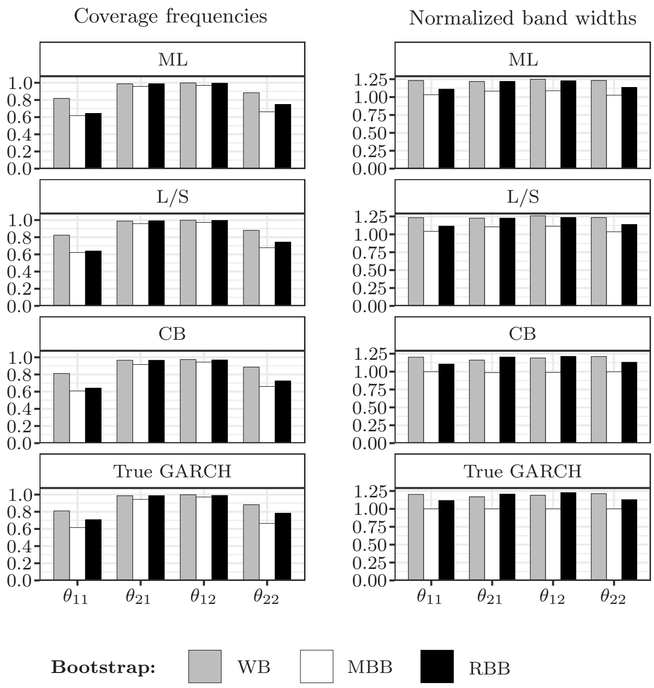
Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.4: Comparison of Different Block Lengths for MBB/CB Method for Bivariate Benchmark DGP ($\alpha = 0.5$), Relative Coverage Frequencies with Average Interval/Band Widths in Parentheses

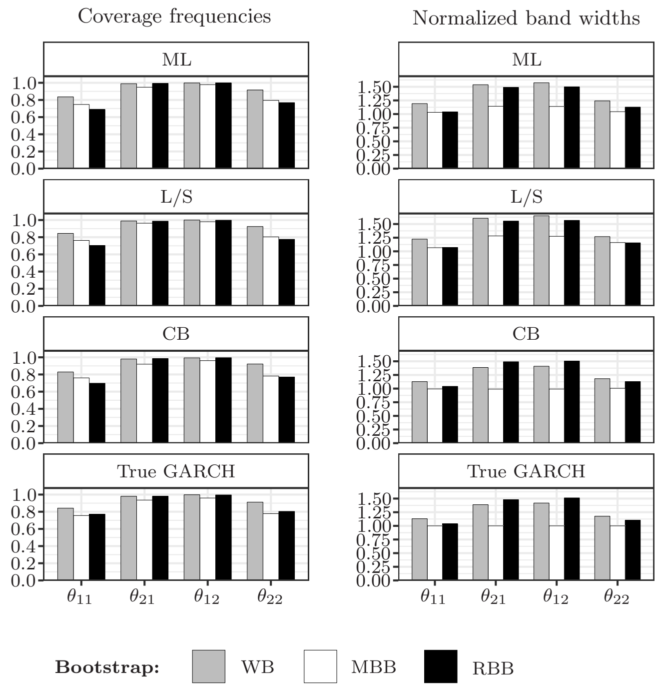
		$T = 200$		$T = 500$	
Block length		10	20	10	50
		B-Model			
		Confidence level 90%			
Impact effects	b_{11}	0.50 (0.31)	0.51 (0.32)	0.62 (0.21)	0.69 (0.25)
	b_{21}	0.95 (8.16)	0.90 (7.20)	0.97 (5.80)	0.89 (4.07)
	b_{12}	0.94 (0.80)	0.90 (0.70)	0.97 (0.57)	0.91 (0.39)
	b_{22}	0.50 (2.61)	0.52 (2.62)	0.61 (1.83)	0.66 (1.98)
		Pointwise confidence level 99%			
Joint bands	θ_{11}	0.63 (1.92)	0.61 (1.84)	0.75 (1.36)	0.77 (1.33)
	θ_{21}	0.94 (30.37)	0.92 (27.97)	0.98 (23.50)	0.93 (18.72)
	θ_{12}	0.97 (3.04)	0.94 (2.79)	0.99 (2.32)	0.96 (1.83)
	θ_{22}	0.68 (17.76)	0.66 (16.99)	0.77 (12.32)	0.77 (11.79)
		A-Model			
		Confidence level 90%			
Impact effects	b_{11}	0.78 (0.30)	0.74 (0.26)	0.94 (0.18)	0.90 (0.11)
	b_{21}	0.94 (7.48)	0.90 (6.72)	0.97 (5.42)	0.89 (3.94)
	b_{12}	0.94 (0.07)	0.90 (0.06)	0.97 (0.05)	0.91 (0.04)
	b_{22}	0.78 (0.30)	0.74 (0.26)	0.94 (0.18)	0.90 (0.11)
		Pointwise confidence level 99%			
Joint bands	θ_{11}	0.89 (1.90)	0.85 (1.76)	0.95 (1.34)	0.92 (1.13)
	θ_{21}	0.94 (27.75)	0.92 (26.22)	0.98 (20.92)	0.93 (17.80)
	θ_{12}	0.97 (0.27)	0.94 (0.25)	0.99 (0.20)	0.96 (0.17)
	θ_{22}	0.90 (1.87)	0.86 (1.75)	0.96 (1.32)	0.93 (1.11)

Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$

B-Model

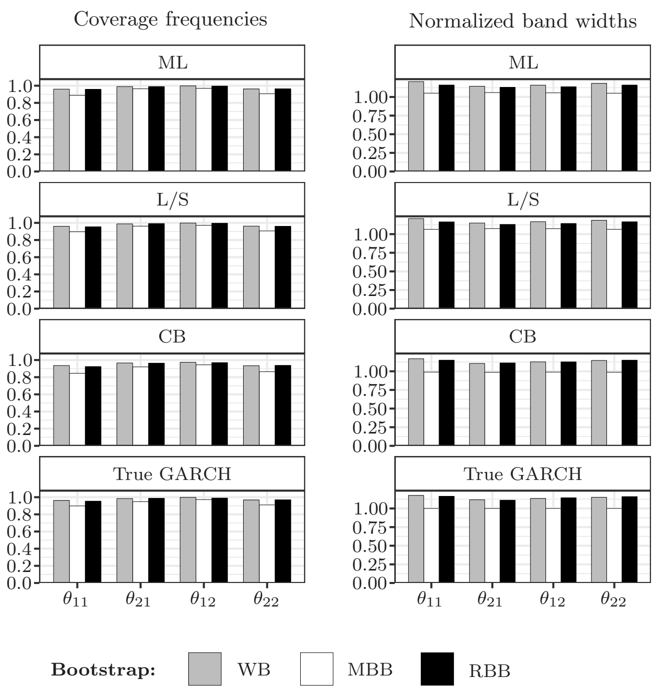


(a) $T = 200$

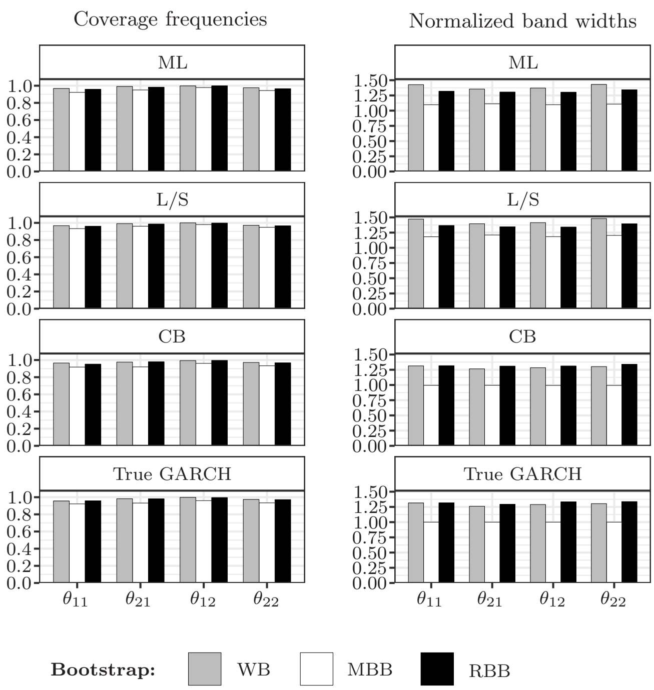


(b) $T = 500$

A-Model



(c) $T = 200$



(d) $T = 500$

Figure A.1: Relative coverage frequencies of joint confidence bands for impulse response functions with propagation horizon up to 10 and corresponding average normalized band widths for pointwise 99% confidence level for bivariate benchmark ($\alpha = 0.5$) DGP with GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$ (block lengths for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively).

Table A.5: Relative Coverage Frequencies of Impact Effects with Average Confidence Interval Widths in Parentheses for Nominal Level of 90% for Bivariate DGP ($\alpha = 0.9$) with Alternative GARCH Parameters

Estim. method		$T = 200$ WB	$T = 200$ MBB	$T = 200$ RBB	$T = 500$ WB	$T = 500$ MBB	$T = 500$ RBB
				B-Model			
ML	b_{11}	0.95 (0.36)	0.71 (0.21)	0.85 (0.41)	0.99 (0.22)	0.82 (0.11)	0.78 (0.29)
	b_{21}	1.00 (8.96)	0.94 (5.40)	1.00 (10.94)	1.00 (6.20)	0.88 (2.61)	1.00 (8.82)
	b_{12}	1.00 (0.95)	0.93 (0.60)	1.00 (1.20)	1.00 (0.66)	0.85 (0.30)	1.00 (0.98)
	b_{22}	0.69 (3.73)	0.64 (3.50)	0.52 (3.12)	0.72 (2.61)	0.72 (2.50)	0.55 (2.38)
L/S	b_{11}	0.95 (0.36)	0.70 (0.21)	0.87 (0.41)	0.98 (0.23)	0.83 (0.11)	0.82 (0.31)
	b_{21}	1.00 (9.09)	0.94 (5.72)	1.00 (11.21)	1.00 (6.41)	0.88 (2.76)	1.00 (9.71)
	b_{12}	1.00 (0.97)	0.94 (0.63)	1.00 (1.23)	1.00 (0.68)	0.85 (0.31)	1.00 (1.08)
	b_{22}	0.69 (3.75)	0.64 (3.52)	0.53 (3.14)	0.72 (2.64)	0.72 (2.52)	0.58 (2.54)
CB	b_{11}	0.95 (0.35)	0.70 (0.20)	0.88 (0.43)	0.99 (0.22)	0.82 (0.11)	0.88 (0.34)
	b_{21}	0.99 (8.64)	0.93 (5.01)	0.99 (11.71)	0.99 (5.78)	0.87 (2.47)	0.99 (10.55)
	b_{12}	0.99 (0.91)	0.90 (0.57)	0.99 (1.29)	0.99 (0.61)	0.83 (0.29)	0.99 (1.17)
	b_{22}	0.69 (3.70)	0.64 (3.46)	0.52 (3.21)	0.72 (2.54)	0.72 (2.49)	0.58 (2.68)
True GARCH	b_{11}	0.95 (0.36)	0.70 (0.21)	0.86 (0.43)	0.99 (0.22)	0.83 (0.11)	0.85 (0.33)
	b_{21}	1.00 (9.03)	0.95 (5.43)	1.00 (11.67)	1.00 (6.17)	0.88 (2.60)	1.00 (10.44)
	b_{12}	1.00 (0.96)	0.93 (0.61)	1.00 (1.29)	1.00 (0.65)	0.86 (0.30)	1.00 (1.16)
	b_{22}	0.68 (3.75)	0.64 (3.52)	0.51 (3.21)	0.72 (2.61)	0.73 (2.51)	0.56 (2.67)
				A-Model			
ML	b_{11}	0.96 (0.34)	0.87 (0.20)	0.93 (0.42)	0.99 (0.20)	0.93 (0.06)	0.98 (0.31)
	b_{21}	1.00 (7.90)	0.94 (4.97)	1.00 (8.40)	1.00 (5.73)	0.88 (2.52)	1.00 (6.99)
	b_{12}	1.00 (0.09)	0.93 (0.06)	1.00 (0.10)	1.00 (0.06)	0.85 (0.03)	1.00 (0.09)
	b_{22}	0.96 (0.34)	0.87 (0.20)	0.93 (0.42)	0.99 (0.20)	0.93 (0.06)	0.98 (0.31)
L/S	b_{11}	0.95 (0.35)	0.88 (0.21)	0.91 (0.43)	0.98 (0.22)	0.92 (0.07)	0.98 (0.35)
	b_{21}	1.00 (7.98)	0.94 (5.26)	1.00 (8.52)	1.00 (5.89)	0.88 (2.66)	1.00 (7.53)
	b_{12}	1.00 (0.09)	0.94 (0.06)	1.00 (0.10)	1.00 (0.06)	0.85 (0.03)	1.00 (0.09)
	b_{22}	0.95 (0.35)	0.88 (0.21)	0.91 (0.43)	0.98 (0.22)	0.92 (0.07)	0.98 (0.35)
CB	b_{11}	0.94 (0.33)	0.85 (0.19)	0.85 (0.45)	0.98 (0.18)	0.91 (0.06)	0.94 (0.38)
	b_{21}	0.99 (7.68)	0.92 (4.65)	0.99 (8.74)	0.99 (5.40)	0.86 (2.40)	0.99 (7.92)
	b_{12}	0.99 (0.08)	0.90 (0.06)	0.99 (0.10)	0.99 (0.06)	0.83 (0.03)	0.99 (0.10)
	b_{22}	0.94 (0.33)	0.85 (0.19)	0.85 (0.45)	0.98 (0.18)	0.91 (0.06)	0.94 (0.38)
True GARCH	b_{11}	0.96 (0.35)	0.87 (0.20)	0.84 (0.45)	0.99 (0.20)	0.91 (0.06)	0.92 (0.38)
	b_{21}	1.00 (7.95)	0.95 (5.00)	1.00 (8.75)	1.00 (5.70)	0.88 (2.51)	1.00 (7.85)
	b_{12}	1.00 (0.09)	0.93 (0.06)	1.00 (0.11)	1.00 (0.06)	0.86 (0.03)	1.00 (0.10)
	b_{22}	0.96 (0.35)	0.87 (0.20)	0.84 (0.45)	0.99 (0.20)	0.91 (0.06)	0.92 (0.38)

Note: GARCH parameters $(\gamma_1, g_1) = (0, 0)$ and $(\gamma_2, g_2) = (0.3, 0.5)$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.6: Relative Coverage Frequencies of Joint Bands for Impulse Response Functions with Average Band Widths in Parentheses for Pointwise Nominal Level of 99% for Bivariate DGP ($\alpha = 0.9$) with Alternative GARCH Parameters

Estim. method		$T = 200$			$T = 500$		
		WB	MBB	RBB	WB	MBB	RBB
		B-Model					
ML	θ_{11}	0.86 (6.05)	0.69 (4.95)	0.91 (6.92)	0.94 (4.27)	0.82 (3.19)	0.95 (5.29)
	θ_{21}	0.95 (39.43)	0.92 (31.54)	0.95 (40.97)	0.98 (30.12)	0.90 (18.65)	0.97 (33.84)
	θ_{12}	1.00 (9.73)	1.00 (7.49)	1.00 (11.13)	1.00 (8.02)	0.96 (4.36)	1.00 (10.18)
	θ_{22}	0.86 (29.47)	0.69 (23.52)	0.66 (26.34)	0.90 (20.74)	0.75 (15.61)	0.68 (19.19)
L/S	θ_{11}	0.86 (6.06)	0.69 (4.98)	0.91 (6.92)	0.93 (4.30)	0.82 (3.25)	0.96 (5.37)
	θ_{21}	0.95 (39.57)	0.92 (32.27)	0.94 (41.16)	0.98 (30.71)	0.90 (19.79)	0.97 (34.63)
	θ_{12}	1.00 (9.77)	1.00 (7.72)	1.00 (11.25)	1.00 (8.21)	0.96 (4.74)	1.00 (10.58)
	θ_{22}	0.86 (29.54)	0.69 (23.75)	0.67 (26.55)	0.90 (20.99)	0.75 (15.94)	0.68 (19.87)
CB	θ_{11}	0.86 (6.03)	0.68 (4.89)	0.90 (6.96)	0.94 (4.20)	0.81 (3.16)	0.97 (5.43)
	θ_{21}	0.94 (39.18)	0.90 (30.59)	0.93 (41.53)	0.97 (29.51)	0.89 (18.25)	0.96 (34.98)
	θ_{12}	0.99 (9.55)	0.98 (7.11)	0.99 (11.46)	0.99 (7.73)	0.95 (4.25)	0.99 (10.79)
	θ_{22}	0.87 (29.25)	0.69 (23.11)	0.65 (27.06)	0.89 (20.30)	0.74 (15.46)	0.68 (20.38)
True GARCH	θ_{11}	0.86 (6.07)	0.68 (4.93)	0.91 (6.96)	0.94 (4.24)	0.81 (3.17)	0.95 (5.42)
	θ_{21}	0.95 (39.52)	0.91 (31.26)	0.94 (41.62)	0.98 (30.02)	0.90 (18.86)	0.97 (34.96)
	θ_{12}	1.00 (9.73)	0.99 (7.38)	1.00 (11.48)	1.00 (7.92)	0.95 (4.46)	1.00 (10.75)
	θ_{22}	0.86 (29.47)	0.69 (23.45)	0.65 (27.05)	0.90 (20.56)	0.75 (15.65)	0.67 (20.29)
		A-Model					
ML	θ_{11}	0.84 (5.79)	0.78 (5.10)	0.86 (6.10)	0.91 (4.39)	0.85 (3.13)	0.93 (5.01)
	θ_{21}	0.95 (35.26)	0.92 (29.39)	0.92 (32.75)	0.98 (26.51)	0.90 (18.10)	0.95 (25.86)
	θ_{12}	1.00 (0.83)	1.00 (0.72)	1.00 (0.92)	1.00 (0.68)	0.96 (0.44)	1.00 (0.81)
	θ_{22}	0.97 (2.72)	0.90 (2.18)	0.90 (2.62)	0.97 (1.98)	0.89 (1.28)	0.89 (2.01)
L/S	θ_{11}	0.84 (5.80)	0.78 (5.15)	0.86 (6.12)	0.91 (4.44)	0.85 (3.25)	0.94 (5.11)
	θ_{21}	0.95 (35.37)	0.92 (29.91)	0.93 (32.82)	0.98 (26.86)	0.91 (18.95)	0.95 (26.22)
	θ_{12}	1.00 (0.83)	1.00 (0.73)	1.00 (0.92)	1.00 (0.69)	0.96 (0.47)	1.00 (0.83)
	θ_{22}	0.97 (2.72)	0.91 (2.22)	0.91 (2.63)	0.97 (2.01)	0.90 (1.34)	0.88 (2.08)
CB	θ_{11}	0.83 (5.75)	0.77 (4.98)	0.84 (6.11)	0.90 (4.28)	0.84 (3.07)	0.91 (5.13)
	θ_{21}	0.94 (35.14)	0.90 (28.80)	0.91 (32.89)	0.97 (26.23)	0.89 (17.81)	0.94 (26.36)
	θ_{12}	0.99 (0.82)	0.98 (0.69)	0.99 (0.93)	0.99 (0.67)	0.95 (0.44)	0.99 (0.85)
	θ_{22}	0.95 (2.69)	0.89 (2.11)	0.89 (2.66)	0.96 (1.92)	0.88 (1.25)	0.87 (2.11)
True GARCH	θ_{11}	0.83 (5.80)	0.77 (5.04)	0.85 (6.13)	0.91 (4.33)	0.84 (3.14)	0.92 (5.12)
	θ_{21}	0.95 (35.40)	0.92 (29.29)	0.93 (33.01)	0.98 (26.55)	0.91 (18.29)	0.95 (26.43)
	θ_{12}	1.00 (0.83)	0.99 (0.71)	1.00 (0.94)	1.00 (0.68)	0.95 (0.45)	1.00 (0.85)
	θ_{22}	0.97 (2.71)	0.91 (2.16)	0.89 (2.66)	0.97 (1.95)	0.89 (1.29)	0.88 (2.10)

Note: GARCH parameters $(\gamma_1, g_1) = (0, 0)$ and $(\gamma_2, g_2) = (0.3, 0.5)$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.7: Relative Coverage Frequencies of Impact Effects with Average Confidence Interval Widths in Parentheses for Nominal Level of 90% for Bivariate DGP ($\alpha = 0.9$) with χ^2 GARCH Innovations

Estim. method		$T = 200$ WB	$T = 200$ MBB	$T = 200$ RBB	$T = 500$ WB	$T = 500$ MBB	$T = 500$ RBB
				B-Model			
ML	b_{11}	0.59 (0.44)	0.53 (0.41)	0.48 (0.39)	0.59 (0.31)	0.62 (0.34)	0.54 (0.29)
	b_{21}	1.00 (9.92)	0.97 (8.48)	1.00 (11.05)	1.00 (7.46)	0.92 (5.14)	0.99 (8.65)
	b_{12}	1.00 (0.95)	0.97 (0.80)	1.00 (1.05)	1.00 (0.72)	0.93 (0.49)	0.99 (0.83)
	b_{22}	0.65 (4.00)	0.54 (3.32)	0.55 (3.42)	0.65 (2.96)	0.62 (2.84)	0.55 (2.68)
L/S	b_{11}	0.59 (0.44)	0.53 (0.42)	0.47 (0.40)	0.60 (0.32)	0.63 (0.35)	0.55 (0.30)
	b_{21}	1.00 (10.16)	0.97 (8.95)	1.00 (11.18)	1.00 (8.18)	0.95 (5.84)	0.99 (9.25)
	b_{12}	1.00 (0.98)	0.98 (0.84)	1.00 (1.07)	1.00 (0.79)	0.96 (0.55)	0.99 (0.88)
	b_{22}	0.64 (4.02)	0.55 (3.36)	0.55 (3.43)	0.66 (3.06)	0.62 (2.91)	0.57 (2.78)
CB	b_{11}	0.57 (0.43)	0.52 (0.40)	0.47 (0.40)	0.59 (0.30)	0.62 (0.34)	0.56 (0.32)
	b_{21}	0.95 (9.15)	0.91 (7.95)	0.95 (11.08)	0.99 (7.19)	0.90 (5.14)	0.99 (10.14)
	b_{12}	0.95 (0.89)	0.89 (0.76)	0.95 (1.07)	0.99 (0.70)	0.92 (0.49)	1.00 (0.97)
	b_{22}	0.63 (3.89)	0.53 (3.21)	0.52 (3.41)	0.66 (2.90)	0.62 (2.85)	0.57 (2.90)
True GARCH	b_{11}	0.55 (0.44)	0.51 (0.42)	0.46 (0.41)	0.59 (0.31)	0.61 (0.35)	0.56 (0.34)
	b_{21}	1.00 (10.14)	0.97 (9.44)	1.00 (12.09)	0.99 (7.74)	0.89 (5.71)	1.00 (10.64)
	b_{12}	1.00 (0.97)	0.97 (0.90)	1.00 (1.16)	1.00 (0.75)	0.92 (0.54)	1.00 (1.02)
	b_{22}	0.63 (3.88)	0.51 (3.34)	0.52 (3.46)	0.65 (2.94)	0.60 (2.92)	0.56 (2.95)
				A-Model			
ML	b_{11}	0.96 (0.38)	0.80 (0.32)	0.89 (0.41)	0.96 (0.26)	0.88 (0.17)	0.92 (0.30)
	b_{21}	1.00 (9.34)	0.97 (8.29)	1.00 (9.82)	1.00 (7.00)	0.91 (5.07)	0.99 (7.67)
	b_{12}	1.00 (0.08)	0.97 (0.07)	1.00 (0.09)	1.00 (0.07)	0.93 (0.05)	0.99 (0.07)
	b_{22}	0.96 (0.38)	0.80 (0.32)	0.89 (0.41)	0.96 (0.26)	0.88 (0.17)	0.92 (0.30)
L/S	b_{11}	0.94 (0.39)	0.78 (0.34)	0.86 (0.42)	0.94 (0.29)	0.88 (0.20)	0.90 (0.32)
	b_{21}	1.00 (9.53)	0.98 (8.68)	1.00 (9.91)	1.00 (7.58)	0.96 (5.74)	0.99 (8.10)
	b_{12}	1.00 (0.09)	0.98 (0.08)	1.00 (0.09)	1.00 (0.07)	0.96 (0.05)	0.99 (0.07)
	b_{22}	0.94 (0.39)	0.78 (0.34)	0.86 (0.42)	0.94 (0.29)	0.88 (0.20)	0.90 (0.32)
CB	b_{11}	0.88 (0.35)	0.71 (0.30)	0.78 (0.42)	0.95 (0.24)	0.86 (0.17)	0.86 (0.36)
	b_{21}	0.95 (8.63)	0.91 (7.72)	0.95 (9.66)	0.99 (6.76)	0.90 (5.05)	0.99 (8.63)
	b_{12}	0.95 (0.08)	0.89 (0.07)	0.95 (0.09)	0.99 (0.06)	0.92 (0.05)	1.00 (0.08)
	b_{22}	0.88 (0.35)	0.71 (0.30)	0.78 (0.42)	0.95 (0.24)	0.86 (0.17)	0.86 (0.36)
True GARCH	b_{11}	0.91 (0.38)	0.70 (0.36)	0.64 (0.46)	0.94 (0.27)	0.83 (0.20)	0.76 (0.38)
	b_{21}	1.00 (9.43)	0.97 (8.92)	1.00 (10.46)	0.99 (7.15)	0.90 (5.51)	1.00 (8.92)
	b_{12}	1.00 (0.08)	0.97 (0.08)	1.00 (0.09)	1.00 (0.07)	0.92 (0.05)	1.00 (0.08)
	b_{22}	0.91 (0.38)	0.70 (0.36)	0.64 (0.46)	0.94 (0.27)	0.83 (0.20)	0.76 (0.38)

Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$;
distribution of innovations: $e_{kt} \sim$ i.i.d. $-(\chi^2(4) - 4)/\sqrt{8}$, $k = 1, 2$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.8: Relative Coverage Frequencies of Joint Bands for Impulse Response Functions with Average Band Widths in Parentheses for Pointwise Nominal Level of 99% for Bivariate DGP ($\alpha = 0.9$) with χ^2 GARCH Errors

Estim. method		$T = 200$ WB	$T = 200$ MBB	$T = 200$ RBB	$T = 500$ WB	$T = 500$ MBB	$T = 500$ RBB
				B-Model			
ML	θ_{11}	0.68 (6.91)	0.50 (5.96)	0.57 (6.48)	0.74 (5.25)	0.62 (4.75)	0.67 (4.97)
	θ_{21}	0.95 (40.88)	0.89 (36.75)	0.95 (41.36)	0.97 (32.42)	0.91 (26.77)	0.97 (33.60)
	θ_{12}	1.00 (9.74)	1.00 (8.35)	1.00 (9.78)	1.00 (8.38)	0.99 (6.30)	1.00 (8.63)
	θ_{22}	0.82 (28.30)	0.60 (23.49)	0.69 (26.99)	0.82 (20.36)	0.69 (17.15)	0.73 (19.92)
L/S	θ_{11}	0.69 (6.92)	0.51 (6.01)	0.57 (6.48)	0.75 (5.33)	0.63 (4.84)	0.67 (5.03)
	θ_{21}	0.95 (41.13)	0.90 (37.47)	0.95 (41.48)	0.97 (33.60)	0.92 (28.98)	0.97 (34.44)
	θ_{12}	1.00 (9.86)	1.00 (8.61)	1.00 (9.86)	1.00 (8.81)	1.00 (7.09)	1.00 (8.94)
	θ_{22}	0.82 (28.45)	0.61 (23.81)	0.70 (27.07)	0.83 (20.95)	0.70 (18.85)	0.73 (20.32)
CB	θ_{11}	0.68 (6.83)	0.50 (5.90)	0.56 (6.48)	0.74 (5.17)	0.62 (4.75)	0.67 (5.09)
	θ_{21}	0.90 (39.38)	0.86 (35.23)	0.90 (40.91)	0.97 (31.64)	0.91 (26.34)	0.96 (35.08)
	θ_{12}	0.95 (9.34)	0.93 (8.01)	0.95 (9.80)	1.00 (8.13)	0.99 (6.16)	1.00 (9.25)
	θ_{22}	0.81 (27.84)	0.58 (23.21)	0.67 (27.21)	0.83 (19.97)	0.69 (17.05)	0.72 (20.97)
True GARCH	θ_{11}	0.66 (6.86)	0.49 (6.00)	0.55 (6.54)	0.73 (5.23)	0.62 (4.79)	0.66 (5.14)
	θ_{21}	0.94 (40.97)	0.89 (37.44)	0.95 (42.38)	0.97 (32.25)	0.91 (27.26)	0.96 (35.51)
	θ_{12}	1.00 (9.72)	1.00 (8.71)	1.00 (10.31)	1.00 (8.33)	0.99 (6.51)	1.00 (9.42)
	θ_{22}	0.79 (28.26)	0.58 (24.11)	0.67 (27.98)	0.82 (20.28)	0.69 (17.57)	0.72 (21.32)
				A-Model			
ML	θ_{11}	0.84 (6.10)	0.76 (5.60)	0.84 (6.07)	0.91 (4.92)	0.84 (4.09)	0.91 (4.89)
	θ_{21}	0.96 (39.20)	0.91 (36.77)	0.95 (38.34)	0.97 (29.31)	0.94 (25.91)	0.98 (29.16)
	θ_{12}	1.00 (0.83)	1.00 (0.74)	1.00 (0.82)	1.00 (0.70)	0.99 (0.57)	1.00 (0.69)
	θ_{22}	0.94 (2.54)	0.90 (2.29)	0.97 (2.57)	0.97 (1.92)	0.92 (1.56)	0.97 (1.94)
L/S	θ_{11}	0.84 (6.12)	0.75 (5.68)	0.83 (6.07)	0.90 (5.06)	0.83 (4.34)	0.90 (4.98)
	θ_{21}	0.96 (39.40)	0.92 (37.30)	0.96 (38.37)	0.98 (30.07)	0.94 (27.51)	0.97 (29.68)
	θ_{12}	1.00 (0.84)	1.00 (0.76)	1.00 (0.82)	1.00 (0.72)	1.00 (0.61)	1.00 (0.71)
	θ_{22}	0.94 (2.56)	0.90 (2.32)	0.96 (2.58)	0.97 (1.99)	0.93 (1.69)	0.97 (1.99)
CB	θ_{11}	0.80 (5.95)	0.71 (5.45)	0.79 (6.02)	0.91 (4.80)	0.82 (4.06)	0.89 (5.03)
	θ_{21}	0.91 (37.85)	0.87 (35.32)	0.91 (37.65)	0.98 (28.82)	0.92 (25.62)	0.97 (30.05)
	θ_{12}	0.95 (0.80)	0.93 (0.72)	0.95 (0.82)	1.00 (0.68)	0.99 (0.55)	1.00 (0.73)
	θ_{22}	0.90 (2.47)	0.84 (2.23)	0.91 (2.56)	0.96 (1.87)	0.89 (1.54)	0.97 (2.04)
True GARCH	θ_{11}	0.84 (6.07)	0.73 (5.66)	0.82 (6.15)	0.91 (4.87)	0.80 (4.16)	0.88 (5.09)
	θ_{21}	0.96 (39.18)	0.90 (36.99)	0.96 (38.73)	0.98 (29.18)	0.92 (26.21)	0.97 (30.22)
	θ_{12}	1.00 (0.83)	1.00 (0.76)	1.00 (0.85)	1.00 (0.69)	0.99 (0.58)	1.00 (0.74)
	θ_{22}	0.94 (2.54)	0.87 (2.34)	0.96 (2.64)	0.97 (1.90)	0.89 (1.61)	0.96 (2.07)

Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$;
distribution of innovations: $e_{kt} \sim$ i.i.d. $-(\chi^2(4) - 4)/\sqrt{8}$, $k = 1, 2$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.9: Relative Coverage Frequencies of Impact Effects with Average Confidence Interval Widths in Parentheses for Nominal Level of 90% for Bivariate Benchmark ($\alpha = 0.9$) with Alternative RBB Method

Estim. method		$T = 200$ WB	MBB	RBB	alt. RBB	$T = 500$ WB	MBB	RBB	alt. RBB
					B-Model				
ML	b_{11}	0.58 (0.39)	0.51 (0.34)	0.48 (0.33)	0.60 (0.44)	0.69 (0.27)	0.70 (0.26)	0.53 (0.21)	0.78 (0.32)
	b_{21}	1.00 (10.48)	0.98 (8.25)	1.00 (11.11)	0.99 (9.02)	1.00 (8.08)	0.92 (4.58)	1.00 (8.02)	0.95 (4.25)
	b_{12}	1.00 (1.03)	0.97 (0.80)	1.00 (1.08)	0.98 (0.85)	1.00 (0.79)	0.94 (0.44)	1.00 (0.78)	0.95 (0.40)
	b_{22}	0.72 (3.53)	0.54 (2.76)	0.59 (2.97)	0.60 (3.32)	0.74 (2.56)	0.68 (2.09)	0.62 (2.07)	0.75 (2.41)
L/S	b_{11}	0.58 (0.39)	0.52 (0.34)	0.49 (0.33)	0.61 (0.44)	0.70 (0.29)	0.71 (0.27)	0.55 (0.23)	0.78 (0.33)
	b_{21}	1.00 (10.68)	0.98 (8.74)	1.00 (11.25)	0.99 (9.38)	1.00 (9.03)	0.94 (5.33)	1.00 (8.76)	0.97 (4.87)
	b_{12}	1.00 (1.05)	0.98 (0.85)	1.00 (1.09)	0.99 (0.89)	1.00 (0.89)	0.95 (0.51)	1.00 (0.85)	0.97 (0.46)
	b_{22}	0.72 (3.55)	0.54 (2.82)	0.59 (2.99)	0.61 (3.34)	0.75 (2.71)	0.69 (2.18)	0.63 (2.23)	0.75 (2.47)
CB	b_{11}	0.57 (0.37)	0.50 (0.32)	0.49 (0.33)	0.60 (0.43)	0.68 (0.24)	0.70 (0.25)	0.55 (0.22)	0.78 (0.32)
	b_{21}	0.99 (9.47)	0.92 (7.21)	0.99 (11.01)	0.98 (8.25)	1.00 (6.77)	0.90 (4.01)	1.00 (8.31)	0.93 (3.97)
	b_{12}	0.98 (0.93)	0.91 (0.70)	0.99 (1.07)	0.97 (0.78)	1.00 (0.67)	0.91 (0.39)	1.00 (0.81)	0.94 (0.38)
	b_{22}	0.71 (3.40)	0.52 (2.63)	0.58 (2.94)	0.60 (3.23)	0.73 (2.33)	0.67 (2.00)	0.62 (2.10)	0.75 (2.36)
True GARCH	b_{11}	0.58 (0.37)	0.50 (0.32)	0.54 (0.32)	0.60 (0.44)	0.67 (0.24)	0.70 (0.25)	0.62 (0.20)	0.78 (0.32)
	b_{21}	1.00 (9.54)	0.95 (7.37)	1.00 (10.57)	0.99 (8.91)	1.00 (6.79)	0.92 (4.00)	1.00 (7.74)	0.95 (4.24)
	b_{12}	1.00 (0.93)	0.94 (0.71)	1.00 (1.03)	0.99 (0.85)	1.00 (0.66)	0.93 (0.38)	1.00 (0.76)	0.95 (0.40)
	b_{22}	0.71 (3.40)	0.53 (2.66)	0.59 (2.86)	0.61 (3.32)	0.74 (2.33)	0.67 (1.99)	0.66 (1.97)	0.75 (2.40)
					A-Model				
ML	b_{11}	0.93 (0.39)	0.77 (0.31)	0.93 (0.41)	0.83 (0.34)	0.98 (0.28)	0.90 (0.14)	0.98 (0.26)	0.92 (0.11)
	b_{21}	1.00 (9.32)	0.97 (7.66)	1.00 (9.52)	0.99 (8.57)	1.00 (7.30)	0.92 (4.39)	1.00 (7.16)	0.96 (4.24)
	b_{12}	1.00 (0.09)	0.97 (0.07)	1.00 (0.09)	0.98 (0.08)	1.00 (0.07)	0.94 (0.04)	1.00 (0.07)	0.95 (0.04)
	b_{22}	0.93 (0.39)	0.77 (0.31)	0.93 (0.41)	0.83 (0.34)	0.98 (0.28)	0.90 (0.14)	0.98 (0.26)	0.92 (0.11)
L/S	b_{11}	0.91 (0.40)	0.77 (0.33)	0.90 (0.41)	0.81 (0.35)	0.96 (0.32)	0.88 (0.17)	0.97 (0.30)	0.91 (0.14)
	b_{21}	1.00 (9.46)	0.98 (8.07)	1.00 (9.59)	0.99 (8.85)	1.00 (8.01)	0.94 (5.07)	1.00 (7.68)	0.97 (4.82)
	b_{12}	1.00 (0.09)	0.98 (0.07)	1.00 (0.09)	0.99 (0.08)	1.00 (0.08)	0.95 (0.05)	1.00 (0.07)	0.97 (0.04)
	b_{22}	0.91 (0.40)	0.77 (0.33)	0.90 (0.41)	0.81 (0.35)	0.96 (0.32)	0.88 (0.17)	0.97 (0.30)	0.91 (0.14)
CB	b_{11}	0.91 (0.35)	0.74 (0.26)	0.91 (0.40)	0.81 (0.30)	0.98 (0.21)	0.90 (0.11)	0.98 (0.27)	0.92 (0.10)
	b_{21}	0.98 (8.56)	0.92 (6.76)	0.99 (9.40)	0.98 (7.95)	1.00 (6.29)	0.90 (3.89)	1.00 (7.35)	0.94 (3.96)
	b_{12}	0.98 (0.08)	0.91 (0.06)	0.99 (0.09)	0.97 (0.07)	1.00 (0.06)	0.91 (0.04)	1.00 (0.07)	0.94 (0.04)
	b_{22}	0.91 (0.35)	0.74 (0.26)	0.91 (0.40)	0.81 (0.30)	0.98 (0.21)	0.90 (0.11)	0.98 (0.27)	0.92 (0.10)
True GARCH	b_{11}	0.94 (0.35)	0.81 (0.26)	0.97 (0.38)	0.85 (0.33)	0.99 (0.21)	0.92 (0.11)	0.99 (0.24)	0.92 (0.11)
	b_{21}	1.00 (8.65)	0.95 (6.93)	1.00 (9.19)	1.00 (8.46)	1.00 (6.31)	0.93 (3.90)	1.00 (6.96)	0.95 (4.22)
	b_{12}	1.00 (0.08)	0.94 (0.06)	1.00 (0.09)	0.99 (0.08)	1.00 (0.06)	0.93 (0.04)	1.00 (0.07)	0.95 (0.04)
	b_{22}	0.94 (0.35)	0.81 (0.26)	0.97 (0.38)	0.85 (0.33)	0.99 (0.21)	0.92 (0.11)	0.99 (0.24)	0.92 (0.11)

Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.10: Relative Coverage Frequencies of Joint Bands for Impulse Response Functions with Average Band Widths in Parentheses for Pointwise Nominal Level of 99% for Bivariate Benchmark DGP ($\alpha = 0.9$) with Alternative RBB Method

Estim. method		$T = 200$				$T = 500$			
		WB	MBB	RBB	alt. RBB	WB	MBB	RBB	alt. RBB
				B-Model					
ML	θ_{11}	0.70 (6.40)	0.53 (5.57)	0.56 (5.83)	0.70 (7.23)	0.80 (4.84)	0.70 (4.24)	0.67 (4.31)	0.87 (5.57)
	θ_{21}	0.94 (40.80)	0.90 (36.56)	0.94 (40.68)	0.94 (39.63)	0.97 (33.14)	0.93 (25.62)	0.97 (32.16)	0.97 (26.48)
	θ_{12}	1.00 (10.04)	1.00 (8.71)	1.00 (9.92)	1.00 (9.11)	1.00 (8.85)	1.00 (6.16)	1.00 (8.38)	1.00 (5.98)
	θ_{22}	0.88 (26.76)	0.66 (22.24)	0.74 (25.18)	0.79 (26.06)	0.91 (18.99)	0.77 (14.89)	0.78 (17.31)	0.84 (16.64)
L/S	θ_{11}	0.71 (6.41)	0.54 (5.61)	0.57 (5.84)	0.70 (7.25)	0.80 (4.94)	0.70 (4.37)	0.67 (4.40)	0.87 (5.68)
	θ_{21}	0.94 (40.97)	0.90 (37.09)	0.94 (40.86)	0.94 (40.02)	0.97 (34.31)	0.93 (28.09)	0.97 (33.22)	0.97 (29.14)
	θ_{12}	1.00 (10.14)	1.00 (8.95)	1.00 (10.00)	1.00 (9.25)	1.00 (9.32)	1.00 (7.00)	1.00 (8.80)	1.00 (6.87)
	θ_{22}	0.88 (26.89)	0.67 (22.59)	0.74 (25.30)	0.78 (26.20)	0.91 (19.70)	0.78 (15.93)	0.78 (17.96)	0.85 (17.52)
CB	θ_{11}	0.70 (6.30)	0.53 (5.46)	0.57 (5.82)	0.70 (7.17)	0.79 (4.63)	0.70 (4.12)	0.66 (4.30)	0.87 (5.49)
	θ_{21}	0.93 (39.47)	0.88 (34.32)	0.93 (40.44)	0.92 (38.64)	0.97 (30.64)	0.92 (22.96)	0.97 (32.20)	0.97 (24.88)
	θ_{12}	0.99 (9.52)	0.98 (7.89)	0.99 (9.85)	0.98 (8.70)	1.00 (7.89)	0.98 (5.32)	1.00 (8.41)	1.00 (5.47)
	θ_{22}	0.88 (26.01)	0.64 (21.20)	0.73 (25.12)	0.79 (25.49)	0.91 (17.61)	0.76 (14.00)	0.78 (17.36)	0.85 (16.10)
True GARCH	θ_{11}	0.70 (6.28)	0.53 (5.46)	0.63 (5.83)	0.71 (7.22)	0.78 (4.63)	0.70 (4.11)	0.72 (4.28)	0.87 (5.52)
	θ_{21}	0.93 (39.54)	0.90 (34.51)	0.94 (40.30)	0.94 (39.42)	0.97 (30.54)	0.92 (22.97)	0.97 (32.15)	0.97 (25.66)
	θ_{12}	1.00 (9.50)	1.00 (7.95)	1.00 (9.77)	1.00 (9.06)	1.00 (7.83)	0.99 (5.30)	1.00 (8.44)	1.00 (5.75)
	θ_{22}	0.88 (25.92)	0.65 (21.29)	0.77 (24.67)	0.79 (25.97)	0.90 (17.55)	0.77 (13.97)	0.80 (17.18)	0.85 (16.39)
				A-Model					
ML	θ_{11}	0.85 (6.09)	0.72 (5.53)	0.83 (5.94)	0.86 (5.96)	0.91 (4.93)	0.83 (3.88)	0.91 (4.66)	0.92 (4.12)
	θ_{21}	0.95 (37.05)	0.92 (34.44)	0.94 (36.31)	0.95 (37.99)	0.97 (28.62)	0.93 (24.09)	0.97 (27.72)	0.98 (25.48)
	θ_{12}	1.00 (0.81)	1.00 (0.74)	1.00 (0.80)	1.00 (0.78)	1.00 (0.71)	1.00 (0.54)	1.00 (0.67)	1.00 (0.54)
	θ_{22}	0.95 (2.54)	0.88 (2.25)	0.96 (2.51)	0.95 (2.47)	0.96 (1.92)	0.93 (1.45)	0.96 (1.81)	0.97 (1.47)
L/S	θ_{11}	0.85 (6.10)	0.73 (5.59)	0.83 (5.95)	0.85 (5.99)	0.91 (5.07)	0.83 (4.15)	0.90 (4.79)	0.91 (4.40)
	θ_{21}	0.95 (37.16)	0.92 (34.81)	0.94 (36.42)	0.95 (38.25)	0.96 (29.31)	0.93 (25.85)	0.97 (28.31)	0.99 (27.30)
	θ_{12}	1.00 (0.82)	1.00 (0.76)	1.00 (0.81)	1.00 (0.79)	1.00 (0.73)	1.00 (0.60)	1.00 (0.69)	1.00 (0.60)
	θ_{22}	0.95 (2.55)	0.90 (2.29)	0.96 (2.52)	0.95 (2.49)	0.96 (2.00)	0.93 (1.59)	0.97 (1.88)	0.97 (1.62)
CB	θ_{11}	0.83 (5.94)	0.71 (5.30)	0.81 (5.91)	0.83 (5.86)	0.92 (4.58)	0.83 (3.59)	0.91 (4.64)	0.92 (3.92)
	θ_{21}	0.93 (36.20)	0.90 (32.92)	0.93 (36.08)	0.93 (37.23)	0.97 (27.24)	0.92 (22.14)	0.97 (27.74)	0.99 (24.29)
	θ_{12}	0.99 (0.79)	0.98 (0.69)	0.99 (0.80)	0.98 (0.75)	1.00 (0.66)	0.98 (0.49)	1.00 (0.68)	1.00 (0.51)
	θ_{22}	0.94 (2.46)	0.87 (2.11)	0.94 (2.50)	0.94 (2.40)	0.97 (1.75)	0.92 (1.30)	0.97 (1.81)	0.97 (1.37)
True GARCH	θ_{11}	0.85 (5.94)	0.74 (5.32)	0.83 (5.92)	0.86 (5.95)	0.92 (4.57)	0.84 (3.58)	0.92 (4.62)	0.92 (4.01)
	θ_{21}	0.94 (36.37)	0.92 (33.04)	0.94 (36.00)	0.95 (37.80)	0.98 (27.14)	0.93 (22.17)	0.97 (27.65)	0.98 (24.85)
	θ_{12}	1.00 (0.79)	1.00 (0.69)	1.00 (0.80)	1.00 (0.78)	1.00 (0.65)	0.99 (0.49)	1.00 (0.68)	1.00 (0.53)
	θ_{22}	0.96 (2.46)	0.90 (2.13)	0.97 (2.49)	0.96 (2.46)	0.97 (1.74)	0.92 (1.30)	0.97 (1.81)	0.97 (1.42)

Note: GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)$ and $(\gamma_2, g_2) = (0.05, 0.92)$;
band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.

Table A.11: Relative Coverage Frequencies of Joint Bands for Impulse Response Functions with Band Widths in Parentheses for Pointwise Nominal Level of 99% of Three-dimensional DGP

		$T = 200$			$T = 500$		
Estimation	Element	WB	MBB	RBB	WB	MBB	RBB
				B-Model			
CB	θ_{11}	0.28 (1.49)	0.20 (1.43)	0.22 (1.48)	0.40 (1.25)	0.44 (1.20)	0.34 (1.18)
	θ_{21}	0.98 (2.96)	0.97 (2.81)	0.99 (3.22)	0.99 (2.27)	0.98 (2.04)	0.99 (2.52)
	θ_{31}	0.95 (0.02)	0.93 (0.02)	0.96 (0.02)	0.92 (0.01)	0.87 (0.01)	0.93 (0.01)
	θ_{12}	0.96 (1.78)	0.94 (1.71)	0.97 (1.89)	1.00 (1.42)	0.99 (1.27)	1.00 (1.50)
	θ_{22}	0.50 (2.23)	0.31 (2.03)	0.47 (2.30)	0.73 (1.87)	0.62 (1.71)	0.73 (1.98)
	θ_{32}	0.97 (0.02)	0.94 (0.02)	0.97 (0.02)	0.99 (0.01)	0.96 (0.01)	0.98 (0.02)
	θ_{13}	0.98 (1.85)	0.97 (1.74)	0.99 (1.97)	0.99 (1.55)	0.98 (1.36)	0.99 (1.69)
	θ_{23}	0.98 (3.07)	0.96 (2.87)	0.99 (3.24)	1.00 (2.94)	1.00 (2.78)	1.00 (3.22)
	θ_{33}	0.63 (0.02)	0.37 (0.02)	0.41 (0.02)	0.83 (0.01)	0.58 (0.01)	0.70 (0.01)
True GARCH	θ_{11}	0.28 (1.49)	0.20 (1.43)	0.22 (1.48)	0.40 (1.26)	0.45 (1.20)	0.33 (1.18)
	θ_{21}	1.00 (2.98)	0.99 (2.81)	1.00 (3.25)	0.99 (2.34)	0.98 (2.03)	0.99 (2.58)
	θ_{31}	0.96 (0.02)	0.93 (0.02)	0.96 (0.02)	0.92 (0.01)	0.88 (0.01)	0.93 (0.01)
	θ_{12}	0.96 (1.77)	0.94 (1.69)	0.97 (1.90)	1.00 (1.45)	0.99 (1.24)	0.99 (1.53)
	θ_{22}	0.52 (2.24)	0.34 (2.04)	0.47 (2.33)	0.73 (1.88)	0.63 (1.71)	0.72 (2.00)
	θ_{32}	0.97 (0.02)	0.94 (0.02)	0.97 (0.02)	0.98 (0.02)	0.96 (0.01)	0.98 (0.02)
	θ_{13}	0.99 (1.85)	0.98 (1.73)	1.00 (1.97)	1.00 (1.57)	0.98 (1.36)	0.99 (1.70)
	θ_{23}	0.98 (3.06)	0.97 (2.86)	0.99 (3.24)	1.00 (2.96)	1.00 (2.78)	1.00 (3.25)
	θ_{33}	0.67 (0.02)	0.39 (0.02)	0.41 (0.02)	0.84 (0.01)	0.58 (0.01)	0.71 (0.01)
				A-Model			
CB	θ_{11}	0.71 (7.98)	0.61 (7.45)	0.62 (7.88)	0.89 (6.04)	0.80 (5.26)	0.83 (5.90)
	θ_{21}	0.99 (14.84)	0.99 (14.63)	0.99 (15.17)	1.00 (10.69)	0.99 (10.40)	1.00 (11.22)
	θ_{31}	0.97 (0.10)	0.95 (0.10)	0.97 (0.10)	0.97 (0.06)	0.93 (0.06)	0.97 (0.06)
	θ_{12}	0.95 (14.05)	0.93 (13.62)	0.94 (13.90)	0.99 (10.75)	0.99 (9.94)	0.99 (10.40)
	θ_{22}	0.72 (19.31)	0.58 (18.40)	0.64 (19.38)	0.89 (16.50)	0.80 (15.39)	0.82 (16.56)
	θ_{32}	0.97 (0.15)	0.94 (0.15)	0.95 (0.15)	0.97 (0.11)	0.96 (0.11)	0.95 (0.11)
	θ_{13}	0.98 (425.36)	0.97 (406.05)	0.98 (434.37)	0.99 (338.67)	0.98 (307.00)	0.99 (349.64)
	θ_{23}	0.97 (699.22)	0.96 (668.60)	0.97 (706.02)	1.00 (621.40)	1.00 (597.26)	1.00 (649.72)
	θ_{33}	0.90 (4.36)	0.83 (4.17)	0.83 (4.37)	0.90 (3.04)	0.82 (2.85)	0.87 (3.07)
True GARCH	θ_{11}	0.72 (8.04)	0.62 (7.52)	0.64 (7.89)	0.89 (6.10)	0.81 (5.25)	0.81 (5.94)
	θ_{21}	1.00 (14.88)	1.00 (14.75)	1.00 (15.27)	1.00 (10.90)	0.98 (10.43)	1.00 (11.40)
	θ_{31}	0.98 (0.10)	0.96 (0.10)	0.98 (0.10)	0.97 (0.06)	0.94 (0.06)	0.97 (0.07)
	θ_{12}	0.96 (14.10)	0.93 (13.59)	0.96 (13.99)	0.99 (10.87)	0.99 (9.79)	0.99 (10.50)
	θ_{22}	0.78 (19.60)	0.67 (18.65)	0.70 (19.61)	0.90 (16.59)	0.81 (15.44)	0.82 (16.69)
	θ_{32}	0.97 (0.15)	0.94 (0.15)	0.95 (0.15)	0.98 (0.11)	0.97 (0.11)	0.95 (0.11)
	θ_{13}	0.99 (429.62)	0.98 (410.58)	0.99 (438.89)	0.99 (342.88)	0.98 (306.29)	0.99 (351.63)
	θ_{23}	0.98 (697.38)	0.96 (666.72)	0.98 (705.24)	1.00 (622.77)	1.00 (597.18)	1.00 (650.17)
	θ_{33}	0.93 (4.38)	0.86 (4.18)	0.86 (4.37)	0.92 (3.06)	0.84 (2.86)	0.88 (3.08)

Note: Band widths used for MBB: $l = 20$ and 50 for $T = 200$ and 500, respectively.