University of Konstanz
Department of Economics

Inference in VARs with Conditional Heteroskedasticity of Unknown Form

Ralf Brüggemann, Carsten Jentsch, and Carsten Trenkler

Working Paper Series
2014-13

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-0-257706

http://www.wiwi.uni-konstanz.de/econdoc/working-paper-series/

Inference in VARs with Conditional Heteroskedasticity of Unknown Form*

Ralf Brüggemann[a] — *University of Konstanz*

Carsten Jentsch[b] — *University of Mannheim*

Carsten Trenkler[c] — *University of Mannheim*, *IAB Nuremberg*

August 4, 2014

Abstract

We derive a framework for asymptotically valid inference in stable vector autoregressive (VAR) models with conditional heteroskedasticity of unknown form. We prove a joint central limit theorem for the VAR slope parameter and innovation covariance parameter estimators and address bootstrap inference as well. Our results are important for correct inference on VAR statistics that depend both on the VAR slope and the variance parameters as e.g. in structural impulse response functions (IRFs). We also show that wild and pairwise bootstrap schemes fail in the presence of conditional heteroskedasticity if inference on (functions) of the unconditional variance parameters is of interest because they do not correctly replicate the relevant fourth moments' structure of the error terms. In contrast, the residual-based moving block bootstrap results in asymptotically valid inference. We illustrate the practical implications of our theoretical results by providing simulation evidence on the finite sample properties of different inference methods for IRFs. Our results point out that estimation uncertainty may increase dramatically in the presence of conditional heteroskedasticity. Moreover, most inference methods are likely to understate the true estimation uncertainty substantially in finite samples.

JEL classification: C30, C32

Keywords: VAR, Conditional heteroskedasticity, Residual-based moving block bootstrap, Pairwise bootstrap, Wild bootstrap

---

*We thank Silvia Gonçalves, Jens-Peter Kreiss, Ulrich K. Müller and the participants of the conference "Recent Developments for Bootstrap Methods in Time Series" in Copenhagen, 8-10 September 2013, for very helpful comments.

[a]University of Konstanz, Chair for Statistics and Econometrics, Box 129, 78457 Konstanz, Germany; ralf.brueggemann@uni-konstanz.de

[b]University of Mannheim, Department of Economics, Chair of Statistics, L7, 3-5, 68131 Mannheim, Germany; cjentsch@mail.uni-mannheim.de

[c]Corresponding author, University of Mannheim, Department of Economics, Chair of Empirical Economics, L7, 3-5, 68131 Mannheim, Germany; trenkler@uni-mannheim.de, phone: +49-621-1811852, fax: +49-621-1811931; and Institute for Employment Research (IAB) Nuremberg

1 Introduction

Many financial and macroeconomic time series exhibit evidence of heteroskedasticity. Examples include e.g. daily financial time series of asset returns but also macroeconomic time series as the monthly growth rates in industrial production, money, exchange rates, interest or inflation rates. Conditional heteroskedasticity patterns have been documented in many empirical examples in the literature, see for instance Gonçalves & Kilian (2004). Moreover, these time series are often analyzed within vector autoregressive (VAR) models. VAR models are a popular econometric tool to summarize the dynamic interaction between the variables included in the VAR system. Many applications in applied macroeconomics and finance (see e.g. Sims (1992), Bernanke & Blinder (1992), Christiano, Eichenbaum & Evans (1999), Kim & Roubini (2000), Brüggemann, Härdle, Mungo & Trenkler (2008), Alter & Schüler (2012)) use VARs and conclusions are based on statistics obtained from the estimated VAR model. These statistics include e.g. Wald tests for Granger-causality, impulse response functions (IRFs) and forecast error variance decompositions (FEVDs). Inference on these statistics is typically based either on first order asymptotic approximations or on different bootstrap methods. The presence of heteroskedasticity invalidates a number of standard inference procedures for the quantities of interest, such that the application of these methods may lead to conclusions that are not in line with the true underlying dynamics. Therefore, in many VAR applications there is a need for inference methods that are valid even in the presence of heteroskedasticity.

In the time series context the existing literature makes some suggestions for valid inference under conditional heteroskedasticity. For instance, Gonçalves & Kilian (2004, 2007) consider inference on autoregressive (AR) parameters in univariate autoregressions with conditional heteroskedasticity. They show that wild and pairwise bootstrap approaches are asymptotically valid (under suitable assumptions) and may be used to set up t-tests and confidence intervals for individual parameters. In addition, they also document that in finite samples the bootstrap methods are typically more accurate than the usual first-order asymptotic approximations based on robust standard errors. Hafner & Herwartz (2009) focus on Wald tests for Granger-causality within VAR models. They use both heteroskedasticity-consistent asymptotic inference as well as wild bootstrap methods, and find that especially the bootstrap methods provide more reliable inference.

Although the presence of heteroskedasticity in time series data has been exploited in the VAR context for structural identification of shocks, see e.g. Rigobon (2003), Normandin & Phaneuf (2004) and Herwartz & Lütkepohl (2014), the implications for inference e.g. on structural impulse responses have not been analyzed in detail yet. To be more precise, the theoretical results for models with conditional heteroskedasticity available in the literature so far do not cover inference on a number of VAR statistics that are also functions of the residual covariance matrix. Examples include popular statistics like responses to orthogonalized shocks, forecast error variance decompositions and tests for instantaneous causality, see e.g. Lütkepohl (2005, Chapter 2). Inference on these statistics is more complicated as it requires to consider the joint asymptotic behavior of estimators for both VAR slope parameters and the parameters of the VAR innovation covariance matrix. While the joint distribution is well explored in the case of i.i.d. innovations, see e.g. Lütkepohl (2005, Chapter 3), there is a gap in the econometric literature for the case of conditional heteroskedastic VAR innovations.

To fill this gap in the literature, we analyze how the introduction of conditional heteroskedasticity into stable VAR models affects the limiting properties of estimators of both the VAR slope parameters and the unconditional innovation covariance matrix. In the following we refer to the vector autogressive slope parameter matrices simply as the 'VAR parameters', while the unconditional innovation covariance matrix is referred to as 'variance parameters'. We provide results for conventional least squares (LS) as well as bootstrap estimators. Thereby, our analysis provides a framework for asymptotically valid inference in stable VAR models with conditional heteroskedasticity of unknown form. In fact, our asymptotic results suggest important differences compared to a set-up with i.i.d. errors as well as to situations with conditional heteroskedasticity in which only inference on the VAR parameters is conducted.

We derive the joint limiting distribution of the LS estimators of the VAR and variance parameters in case the innovation vector forms a martingale difference sequence (mds) and satisfy certain mixing and moment conditions. Thereby, we complement Hafner & Herwartz (2009) by providing a complete proof for the asymptotic results in the VAR case. In contrast to an i.i.d. error term set-up which leads to a block-diagonal asymptotic covariance matrix, see Lütkepohl (2005, Chapter 3), it turns out that the estimators of the mean and variance parameters are asymptotically correlated in general. A result corresponding to ours has been found by Ling & McAleer (2003) and Francq & Zakoïan (2004) for (vector) autoregressive moving average ((V)ARMA) models with generalized autoregressive conditional heteroscedastic (GARCH) innovations in terms of the estimators of the (V)ARMA and GARCH parameters.

We also analyze the theoretical properties of different bootstrap approaches commonly used in the VAR context. We find that the recursive- and fixed-design wild bootstrap as well as the pairwise bootstrap that have been considered by Gonçalves & Kilian (2004, 2007) and Hafner & Herwartz (2009) turn out to lead to asymptotically invalid inference on (functions of) the innovation covariance matrix in the presence of conditional heteroskedasticity. The same holds true for the blockwise wild bootstrap that was recently proposed by Shao (2011). In detail, these bootstrap approaches fail in replicating the asymptotic variance of the innovation covariance estimator, which is a function of the fourth moments' structure of the innovations. Moreover, the wild bootstrap turns out to be inappropriate even in case of i.i.d. errors.

As an alternative to the asymptotically invalid bootstrap methods mentioned above, we suggest to use a residual-based moving block bootstrap. The idea of the block bootstrap has been proposed by Künsch (1989) and Liu & Singh (1992) to extend the seminal bootstrap idea of Efron (1979) to dependent data. This and related approaches that resample blocks of time series data have been studied extensively in the literature, see e.g. Lahiri (2003) for an overview. In this paper, we prove that the residual-based moving block bootstrap (MBB) results in asymptotically valid joint inference on the VAR and variance parameters if suitable mixing and moment assumptions are imposed. Since the block length in the MBB is assumed to grow to infinity with the sample size (at an appropriate rate), the MBB is capable of capturing the higher moment structure of the innovation process asymptotically. Therefore, the MBB is indeed able to correctly replicate the limiting covariance matrix of the innovation covariance estimator.

We illustrate the importance and implications of the theoretical results by studying inference on IRFs that are functions of both the VAR parameters and the innovation covariance parameters. This type of IRFs are of major importance in typical applied VAR studies. We provide simulation evidence on the finite-sample properties of corresponding first-order asymptotic approximations and of various bootstrap approaches. We draw two main lessons from our simulation study. First, applied researchers have to be aware that estimation uncertainty may dramatically increase if conditional heteroskedasticity is present. Second, in many situations the true sampling variation of the IRF estimators is understated by most of the inference procedures. This, in turn, leads to (bootstrap) confidence intervals for impulse response coefficients being too narrow. Accordingly, applied researchers should interpret their results with caution.

The remainder of the paper is structured as follows. Section 2 provides the modeling framework while the asymptotic results for the LS estimators of the VAR and unconditional variance parameters are discussed in Section 3. We show the invalidity of the wild and pairwise bootstrap schemes in Section 4 and present the residual-based MBB scheme and its asymptotic properties in Section 5. Section 6 contains a discussion on structural impulse response analysis and presents the simulation results. Finally, Section 7 concludes. The proofs and calculations related to the data generating process (DGP) used in Section 6 are deferred to Appendices A and B, respectively.

2 Modeling Framework

2.1 Notation and preliminaries

Let $(u_t, t \in \mathbb{Z})$ be a K-dimensional sequence of martingale differences defined on a probability space $(\Omega, \mathcal{F}, P)$ such that each $u_t = (u_{1t}, \dots, u_{Kt})'$ is assumed to be measurable with respect to $\mathcal{F}_t$, where $(\mathcal{F}_t)$ is a sequence of increasing σ-fields of $\mathcal{F}$. We observe a data sample $(y_{-p+1}, \dots, y_0, y_1, \dots, y_T)$ of sample size T plus p pre-sample values from the following DGP for the K-dimensional time series $y_t = (y_{1t}, \dots, y_{Kt})'$,

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad t \in \mathbb{Z}, \tag{2.1}$$

or $A(L)y_t = \nu + u_t$ in compact representation. Here, $A(L) = I_K - A_1 L - A_2 L^2 - \cdots - A_p L^p$, $A_p \neq 0$, L is the lag operator such that $Ly_t = y_{t-1}$, the lag order p is finite and known, and $\det(A(z))$ is assumed to have all roots outside the unit circle. Hence, we are dealing with a stable (invertible and causal) VAR model of order p.

In order to simplify the exposition we assume a zero intercept vector $\nu = 0$ throughout this paper and focus on estimators for the VAR parameters $A_1, \dots, A_p$ and the unconditional innovation covariance matrix $\Sigma_u = E(u_t u_t')$. Our results can be generalized to a set-up with a non-zero intercept vector. We will make some remarks in this respect later on. We introduce the following notation, where the dimensions of the defined quantities are also given in parentheses:

$$\begin{aligned}
\mathbf{y} &= \text{vec}(y_1,\dots,y_T) \quad (KT\times 1)\\
Z_t &= \text{vec}(y_t,\dots,y_{t-p+1}) \quad (Kp\times 1)\\
Z &= (Z_0,\dots,Z_{T-1}) \quad (Kp\times T) \qquad (2.2)\\
\boldsymbol{\beta} &= \text{vec}(A_1,\dots,A_p) \quad (K^2p\times 1)\\
\mathbf{u} &= \text{vec}(u_1,\dots,u_T) \quad (KT\times 1),
\end{aligned}$$

where the vec-operator stacks the columns of a matrix below each other. The parameter of interest is $\boldsymbol{\beta}$ which is estimated by $\widehat{\boldsymbol{\beta}} = \text{vec}(\widehat{A}_1,\dots,\widehat{A}_p)$ via multivariate LS using observations $y_1,\dots,y_T$. Hence, we have, see e.g. Lütkepohl (2005, p. 71),

$$\widehat{\boldsymbol{\beta}} = ((ZZ')^{-1}Z\otimes I_K)\mathbf{y}, \qquad (2.3)$$

and $\mathbf{y} = (Z'\otimes I_K)\boldsymbol{\beta}+\mathbf{u}$ leads to

$$\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta} = ((ZZ')^{-1}Z\otimes I_K)\mathbf{u}. \qquad (2.4)$$

Here, $A\otimes B = (a_{ij}B)_{ij}$ denotes the Kronecker product of matrices $A=(a_{ij})$ and $B=(b_{ij})$ and I_K is the K-dimensional identity matrix. Since the process $(y_t, t\in\mathbb{Z})$ is stable, y_t has a vector moving-average (VMA) representation

$$y_t = \sum_{j=0}^{\infty}\Phi_j u_{t-j}, \quad t\in\mathbb{Z}, \qquad (2.5)$$

where $\Phi_j, j\in\mathbb{N}_0$, is a sequence of (exponentially fast decaying) $(K\times K)$ coefficient matrices with $\Phi_0 = I_K$ and $\Phi_i = \sum_{j=1}^{i}\Phi_{i-j}A_j$, $i=1,2,\dots$. Further, we define $(Kp\times K)$ matrices $C_j = (\Phi'_{j-1},\dots,\Phi'_{j-p})'$ and the $(Kp\times Kp)$ matrix $\Gamma = \sum_{j=1}^{\infty}C_j\Sigma_u C_j'$. The standard estimator of Σ_u is

$$\widehat{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t\widehat{u}_t', \qquad (2.6)$$

where $\widehat{u}_t = y_t - \widehat{A}_1 y_{t-1} - \cdots - \widehat{A}_p y_{t-p}$ are the residuals obtained from the estimated VAR(p) model. We set $\boldsymbol{\sigma} = \text{vech}(\Sigma_u)$ and $\widehat{\boldsymbol{\sigma}} = \text{vech}(\widehat{\Sigma}_u)$. The vech-operator is defined to stack columnwise the elements on and below the main diagonal of a square matrix below each other.

2.2 Assumptions

For the theory established in this paper we need the following assumptions on the process $(y_t, t\in\mathbb{Z})$ in addition to the stability condition for the DGP (2.1).

Assumption 2.1 (mds innovations)**.**

(i) It holds $E(u_t|\mathcal{F}_{t-1}) = 0$ *almost surely, where* $\mathcal{F}_{t-1} = \sigma(u_{t-1}, u_{t-2},\dots)$ *is the* σ*-field gen-*

erated by $(u_{t-1}, u_{t-2}, \ldots)$.

(ii) *The* $(K \times K)$ *innovation covariance matrix* $\Sigma_u = E(u_t u_t')$ *exists and is positive definite.*

(iii) *It holds* $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(u_t u_t' | \mathcal{F}_{t-1}) = \Sigma_u$ *in probability.*

(iv) *For* $a, b, c \in \mathbb{Z}$ *define* $(K^2 \times K^2)$ *matrices*

$$\tau_{0,a,b,c} = E\left(\operatorname{vec}(u_t u_{t-a}')\operatorname{vec}(u_{t-b}u_{t-c}')'\right) \tag{2.7}$$

and assume that (the entries of) $\tau_{0,r,0,s}$ *are uniformly bounded for all* $r, s \geq 1$ *as well as positive definiteness of* $L_K \tau_{0,r,0,r} L_K'$ *for all* $r \geq 1$. *Here* L_K *is the* $\left(K(K+1)/2 \times K^2\right)$ *elimination matrix which is defined such that* $\operatorname{vech}(A) = L_K \operatorname{vec}(A)$ *holds for a* $(K \times K)$ *matrix* A, *see e.g. Lütkepohl (2005, Sect. A.12.2).*

(v) *It holds* $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(\operatorname{vec}(u_t u_{t-r}')\operatorname{vec}(u_t u_{t-s}')' | \mathcal{F}_{t-1}) = \tau_{0,r,0,s}$ *in probability for all* $r, s \geq 1$.

(vi) *For some* $r > 1$, *we have that* $E|u_t|_{4r}^{4r}$ *is uniformly bounded, where* $|A|_p = (\sum_{i,j} |a_{ij}|^p)^{1/p}$ *for some matrix* $A = (a_{ij})$.

Here, the common i.i.d. assumption for the innovation process $(u_t, t \in \mathbb{Z})$ is replaced by the less restrictive mds condition in Assumption 2.1. In particular, 2.1(i) and 2.1(ii) cover a large class of dependent, but uncorrelated second-order stationary innovation processes and allow for conditional heteroskedasticity. For $\tau_{0,a,b,c}$ being well-defined in 2.1(iv) it is assumed that $E\left(\operatorname{vec}(u_t u_{t-a}')\operatorname{vec}(u_{t-b}u_{t-c}')'\right)$ is independent of t This is implied by fourth-order stationarity of $(u_t, t \in \mathbb{Z})$, but is somewhat weaker. As the inverse of $\Gamma = \sum_{j=1}^{\infty} C_j \Sigma_u C_j'$ occurs in the following, we assume Σ_u to be positive definite, which together with the stability condition of the DGP leads also to invertibility of Γ. Non-singularity of $L_K \tau_{0,r,0,r} L_K'$ and the moment condition in 2.1(iv) is required for the central limit theorem (CLT) for mds that is used to prove asymptotic normality of $\widehat{\boldsymbol{\beta}}$ in Theorem 3.1(i) below.

Assumption 2.1 is a vector-valued analogue to Gonçalves & Kilian (2004, Assumption A). In comparison to Hafner & Herwartz (2009) we do not require u_t to be mixing in order to derive the limiting distribution of $\widehat{\boldsymbol{\beta}}$. Rather, we impose the following mixing condition for obtaining the joint limiting results for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\sigma}}$.

Assumption 2.2 (mixing innovations)**.**

(i) *The innovations process* $(u_t, t \in \mathbb{Z})$ *is strictly stationary.*

(ii) *The process* $(u_t, t \in \mathbb{Z})$ *is* α*-mixing and satisfies*

$$\sum_{m=1}^{\infty} (\alpha_u(m))^{\delta/(2+\delta)} < \infty,$$

where $\mathcal{F}_{-\infty}^{t} = \sigma(\ldots, u_{t-2}, u_{t-1}, u_t)$, $\mathcal{F}_{t+m}^{\infty} = \sigma(u_{t+m}, u_{t+m+1}, \ldots)$ *and*

$$\alpha_u(m) = \sup_{t\in\mathbb{Z}} \sup_{A\in\mathcal{F}_{-\infty}^{t}, B\in\mathcal{F}_{t+m}^{\infty}} |P(A\cap B) - P(A)P(B)|.$$

(iii) It holds that $\sum_{h=-\infty}^{\infty} L_K\{\tau_{0,0,h,h} - \text{vec}(\Sigma_u)\,\text{vec}(\Sigma_u)'\}L_K'$ *exists and is positive definite.*

The strict stationarity and mixing conditions imposed on $(u_t, t \in \mathbb{Z})$ in Assumption 2.2 are required to prove a joint CLT for $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\sigma}}$ in Theorem 3.1(ii). Note that a CLT for mds is not applicable here. This is due to the fact that $\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}$ includes $\text{vech}(u_t u_t')$ whereas $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ contains only terms of the form $\text{vech}(u_t u_{t-j}')$ with $j \geq 1$ and, therefore, $\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}$ is not an mds. Further, the summability condition in Assumption 2.2(ii) together with the moment condition in Assumption 2.1(vi) is sufficient for $\sum_{h=-\infty}^{\infty}\{\tau_{0,0,h,h} - \text{vec}(\Sigma_u)\,\text{vec}(\Sigma_u)'\}$ to exist which can be shown with the help of Corollary 14.3 in Davidson (1994).

3 Asymptotic Inference

In this section, we give two unconditional CLTs in Theorem 3.1. The first CLT is for the VAR parameter estimator $\widehat{\boldsymbol{\beta}}$ as defined in (2.3) under the mds-type Assumption 2.1. Under the additional mixing condition in Assumption 2.2, the second CLT is concerned with joint asymptotic normality of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\sigma}}$.

Theorem 3.1 (Unconditional CLTs)**.**

(i) Under Assumption 2.1, we have

$$\sqrt{T}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \overset{\mathcal{D}}{\to} \mathcal{N}(0, V^{(1,1)}),$$

where $\overset{\mathcal{D}}{\to}$ *denotes convergence in distribution.*

$$V^{(1,1)} = (\Gamma^{-1} \otimes I_K)\left(\sum_{i,j=1}^{\infty}(C_i \otimes I_K)\tau_{0,i,0,j}(C_j \otimes I_K)'\right)(\Gamma^{-1} \otimes I_K)'. \tag{3.1}$$

(ii) Under Assumptions 2.1 and 2.2, we have

$$\sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \end{pmatrix} \overset{\mathcal{D}}{\to} \mathcal{N}(0, V),$$

where

$$V = \begin{pmatrix} V^{(1,1)} & V^{(1,2)} \\ V^{(2,1)} & V^{(2,2)} \end{pmatrix}$$

with $V^{(2,2)} = \sum_{h=-\infty}^{\infty} L_K\left\{\tau_{0,0,h,h} - \text{vec}(\Sigma_u)\,\text{vec}(\Sigma_u)'\right\}L_K'$, $V^{(2,1)} = V^{(1,2)\prime}$ *and*

$$V^{(2,1)} = \sum_{j=1}^{\infty}\sum_{h=0}^{\infty} L_K \tau_{0,0,h,h+j}(C_j \otimes I_K)'(\Gamma^{-1} \otimes I_K)'. \tag{3.2}$$

The proof of Theorem 3.1 is provided in Appendix A.

Remark 3.1. The result on V in part (ii) of Theorem 3.1 is a generalization of the case where $u_t \sim i.i.d.(0, \Sigma_u)$ that is discussed e.g. in Lütkepohl (2005, Chapter 3).[1] Note in particular that block-diagonality of V is generally lost if u_t is conditionally heteroskedastic. A corresponding finding has been obtained by Francq & Zakoïan (2004) and Ling & McAleer (2003) in relation to (vector) ARMA-GARCH processes.

In Francq & Zakoïan (2004) a univariate framework is considered in which the error term is expressed as $u_t = \sigma_t \varepsilon_t$, where $\varepsilon_t \sim i.i.d.(0, 1)$ and σ_t^2 has a strictly stationary GARCH(m,n) representation. In this set-up the estimators of the ARMA and of the GARCH parameters are asymptotically correlated in general. However, if ε_t has a symmetric distribution, then the joint asymptotic covariance matrix of the estimators is block-diagonal. This finding extends to the vector ARMA-GARCH case if a corresponding vector version of ε_t follows a spherically symmetric distribution, see Ling & McAleer (2003) and Hafner (2004, Lemma 1). The spherical symmetry assumption assures that all mixed N-th order moments $E\left[\Pi_{j=1}^{N} \varepsilon_j^{s_j}\right]$ are zero if at least one s_j is odd.[2]

Hence, if u_t in (2.1) follows e.g. a stable vector GARCH(m,n) process with ε_t having a spherically symmetric distribution, then $V^{(2,1)} = 0$ since $\tau_{0,0,h,h+j} = 0$ for all $h \geq 0$ and $j \geq 1$, compare Francq & Zakoïan (2004, Lemma 4.1). Two comments are in order. First, a spherical symmetry assumption on the distribution of ε_t is stronger than necessary to obtain a block-diagonal covariance matrix structure. In fact, symmetry assures that all mixed 'odd-moments' of u_t behave as those of an independent sequence, compare Deo (2000, Condition A.(vii)) and its interpretation therein. Second, the set-up of a block-diagonal covariance matrix still differs from a situation with $u_t \sim i.i.d.(0, \Sigma_u)$ since $V^{(1,1)}$ and $V^{(2,2)}$ are also affected by the presence of conditional heteroskedasticity.

Remark 3.2. Defining $\boldsymbol{u}_t^2 = \text{vech}(u_t u_t')$ one can also write

$$V^{(2,2)} = \text{Var}(\boldsymbol{u}_t^2) + \sum_{\substack{h=-\infty \\ h \neq 0}}^{\infty} \text{Cov}\left(\boldsymbol{u}_t^2, \boldsymbol{u}_{t-h}^2\right).$$

Hence, $V^{(2,2)}$ has a long-run variance representation in terms of $\boldsymbol{u}_t^2$ that captures the (linear) dependence structure in the sequence $(\boldsymbol{u}_t^2)$. If the error terms are i.i.d., we obviously have $V^{(2,2)} = \text{Var}(\boldsymbol{u}_t^2) = L_K \tau_{0,0,0,0} L_K' - \boldsymbol{\sigma}\boldsymbol{\sigma}'$.

Remark 3.3. Implementing asymptotic inference based on Theorem 3.1 requires estimation of V. The blocks $V^{(1,1)}$ and $V^{(2,2)}$ may be estimated consistently by a White-type estimator as in Hafner & Herwartz (2009) and a VARHAC approach of Den Haan & Levin (1996) for $\boldsymbol{u}_t^2$, respectively. Estimation of $V^{(1,2)}$ is less straightforward and needs to be investigated in future research that is beyond the scope of the current paper. Against this background, a bootstrap approach as discussed below may be useful as it avoids estimating $V^{(1,2)}$ directly.

[1] The joint limiting result in Lütkepohl (2005, Proposition 3.4) is based on additionally assuming that u_t is normally distributed. The normality assumption only affects the asymptotic variance of $\widehat{\boldsymbol{\sigma}}$ since $\tau_{0,0,0,0} = 3\,\text{vec}(\Sigma_u)\,\text{vec}(\Sigma_u)'$ in this case. Also compare Remark 3.2 below in this respect.

[2] The standard multivariate normal and t-distributions belong e.g. to the class of spherical distributions. The result on the mixed N-th order moments is also obtained for elliptically symmetric distributions with a mean equal to zero, compare Berkane & Bentler (1986).

Remark 3.4. For the theory provided in Theorem 3.1, we assume that the intercept term in (2.1) is known and equals zero, i.e. $\nu = 0$ such that $\boldsymbol{\mu} = E(y_t) = 0$ holds. This is in order to simplify the exposition. However, we remark that it is straightforward to allow for arbitrary intercepts and to include the sample mean $\overline{\boldsymbol{y}} = \frac{1}{T}\sum_{t=1}^{T} y_t$ into the analysis. Joint normality for $\sqrt{T}(\overline{\boldsymbol{y}} - \boldsymbol{\mu}, \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})$ can be derived by similar arguments. In view of the generally non-vanishing covariance structure of $V^{(2,1)}$, it can be observed that this property remains true in the limit also for $T\,\mathrm{Cov}(\overline{\boldsymbol{y}} - \boldsymbol{\mu}, \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $T\,\mathrm{Cov}(\overline{\boldsymbol{y}} - \boldsymbol{\mu}, \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})$. According to the findings summarized in Remark 3.1, a spherical symmetry assumption implies the asymptotic covariance matrix of $\sqrt{T}(\overline{\boldsymbol{y}} - \boldsymbol{\mu}, \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})$ to be block diagonal.

4 Asymptotic Invalidity of the Wild and Pairwise Bootstraps

Since the finite sample properties of asymptotic-based VAR inference approaches can be rather poor, the use of bootstrap methods is often advocated, see e.g. Kilian (1998*b*,*a*, 1999) in relation to impulse response analysis. The results of Gonçalves & Kilian (2004) for univariate autoregressions and of Hafner & Herwartz (2009) for Wald-tests in VARs indicate that bootstrap methods can also be very beneficial in case of conditional heteroskedasticity. In our set-up, the use of bootstrap methods additionally avoids the cumbersome estimation of the asymptotic covariance matrix V, compare Remark 3.3.

In order to obtain valid bootstrap approximations for statistics that are only functions of the VAR parameters in $\boldsymbol{\beta}$, it suffices for a certain bootstrap procedure to mimic the CLT in Theorem 3.1(i). This would apply e.g. to forecast error impulse responses (FEIRs) or restriction tests on the VAR parameters as considered e.g. in Hafner & Herwartz (2009). However, to get valid bootstrap approximations for statistics that depend on parameters both in $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$, as e.g. in the case of structural impulse responses, we need a bootstrap scheme capable of mimicking the CLT in Theorem 3.1(ii). In view of the papers by Gonçalves & Kilian (2004, 2007) on the univariate case, it is clear that an i.i.d. resampling of the residuals does not work in general. Due to their results, it seems obvious to check the following schemes applied to the residuals obtained from fitting a VAR(p) model to the data:

(a) recursive-design wild bootstrap

(b) fixed-design wild bootstrap

(c) pairwise bootstrap

(d) blockwise wild bootstrap

In Gonçalves & Kilian (2004, 2007) the bootstrap schemes (a)-(c) are particularly used under an mds assumptions for the innovations. Bootstrap (d) has been recently proposed by Shao (2011) to white noise testing and applied in the context of unit root testing by Smeekes & Urbain (2014). We will show in this section that all procedures (a), (b), (c), and (d) actually fail to mimic the proper distribution in Theorem 3.1(ii). To show this, it suffices to consider the notationally simpler univariate case $K = 1$ and $\sqrt{T}(\widehat{\sigma}_u^2 - \sigma_u^2)$, where $\sigma_u^2 = E(u_t^2)$ and $\widehat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t^2$.

Furthermore, these bootstrap schemes will not be able to replicate the covariance block $V^{(1,2)}$ either.

As shown in Hafner & Herwartz (2009), the fixed-design wild bootstrap correctly mimics the CLT in Theorem 3.1(i) by making appropriate non-i.i.d. error term assumptions. Corresponding results are obtained for the bootstrap schemes (a) and (c) by extending the results of Gonçalves & Kilian (2004) to the multivariate case. Hence, they are asymptotically valid for inference that only refers to the VAR parameters, including the case of FEIRs.

4.1 Recursive-design and fixed-design wild bootstrap

As the recursive- and fixed-design wild bootstrap schemes rely on the same set of residuals

$$\widehat{u}_t = y_t - \widehat{A}_1 y_{t-1} - \cdots - \widehat{A}_p y_{t-p}, \quad t = 1, \ldots, T,$$

and as the estimator $\widehat{\sigma}_u^2$ is computed from those residuals exclusively, both approaches coincide here and yield the same bootstrap estimator $\widehat{\sigma}_{WB}^{2*} = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t^{*2}$ to be discussed further. For the wild bootstrap, we set $\widehat{u}_t^* = \widehat{u}_t \eta_t$, where $(\eta_t, t \in \mathbb{Z})$ are i.i.d. random variables with $E^*(\eta_t) = 0$, $E^*(\eta_t^2) = 1$ and $E^*(\eta_t^4) < \infty$. From $E^*(\eta_t^2) = 1$, we get $E^*(\sqrt{T}(\widehat{\sigma}_{WB}^{2*} - \widehat{\sigma}_u^2)) = 0$ and

$$\begin{aligned} E^*\left(\sqrt{T}\left(\widehat{\sigma}_{WB}^{2*} - \widehat{\sigma}_u^2\right)\right)^2 &= \frac{1}{T}\sum_{t=1}^{T} E^*\left(\widehat{u}_t^{*4}\right) + \frac{1}{T}\sum_{\substack{t_1,t_2=1 \\ t_1 \neq t_2}}^{T} E^*\left(\widehat{u}_{t_1}^{*2}\right) E^*\left(\widehat{u}_{t_2}^{*2}\right) - \frac{1}{T}\sum_{t_1,t_2=1}^{T} \widehat{u}_{t_1}^2 \widehat{u}_{t_2}^2 \\ &= \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_t^4 \left(E^*\left(\eta_t^4\right) - 1\right). \end{aligned}$$

Replacing $\widehat{u}_t$ by u_t above does not affect the asymptotics such that the last right-hand side converges in probability to

$$V_{WB}^{(2,2)} := E(u_t^4)\left\{E^*\left(\eta_t^4\right) - 1\right\} = \tau_{0,0,0,0}\left\{E^*\left(\eta_t^4\right) - 1\right\} \neq \sum_{h=-\infty}^{\infty}\left\{\tau_{0,0,h,h} - \sigma_u^4\right\} = V^{(2,2)},$$

which indicates the invalidity of the wild bootstrap for the estimator of the innovation variance. Note that even if $u_t \sim i.i.d.(0, \sigma_u^2)$, the wild bootstrap would be invalid since $\tau_{0,0,0,0}\left\{E^*\left(\eta_t^4\right) - 1\right\} \neq \tau_{0,0,0,0} - \sigma_u^4$, compare Remark 3.2. The latter has been already observed in Kreiss (1997) for linear processes. Similarly, one can show that in general $V_{WB}^{(2,1)} = 0 \neq V^{(2,1)}$ holds, that is, the wild bootstrap estimates the potentially non-zero limiting covariances always as being zero. Further, it is worth noting that the more natural approach of using re-calculated residuals $\widehat{\widehat{u}}_t^* := y_t^* - \widehat{A}_1^* y_{t-1}^* - \cdots - \widehat{A}_p^* y_{t-p}^*$ for the bootstrap estimator does not alter the asymptotics and this leads to the same invalidity results as shown above.

4.2 Pairwise bootstrap

Let $\left\{(y_t^*, Y_{t-1}^{*\prime}) := (y_t^*, \ldots, y_{t-p}^*), t = 1, \ldots, T\right\}$ be a bootstrap sample drawn independently from $\left\{(y_t, Y_{t-1}') := (y_t, \ldots, y_{t-p}), t = 1, \ldots, T\right\}$. Based on these bootstrap tupels, we define bootstrap residuals

$$\widehat{u}_t^{**} = y_t^* - \left(\widehat{A}_1^*, \ldots, \widehat{A}_p^*\right) Y_{t-1}^* =: (1, -\widehat{B}^*) \begin{pmatrix} y_t^* \\ Y_{t-1}^* \end{pmatrix}, \quad t = 1, \ldots, T.$$

By standard arguments, it is valid to replace $\widehat{B}^*$ by $\widehat{B} = (\widehat{A}_1, \ldots, \widehat{A}_p)$ and to consider corresponding residuals $\widehat{u}_1^*, \ldots, \widehat{u}_T^*$ and the bootstrap estimator $\widehat{\sigma}_{PB}^{2*} = \frac{1}{T}\sum_{t=1}^T \widehat{u}_t^{*2}$ in the following. Due to i.i.d. resampling we get $E^*(\sqrt{T}(\widehat{\sigma}_{PB}^{2*} - \widehat{\sigma}_u^2)) = 0$ and

$$\begin{aligned}
E^*\left(\sqrt{T}\left(\widehat{\sigma}_{PB}^{2*} - \widehat{\sigma}_u^2\right)\right)^2 &= \frac{1}{T}\sum_{t=1}^T E^*(\widehat{u}_t^{*4}) + \frac{1}{T}\sum_{\substack{t_1,t_2=1 \\ t_1\neq t_2}}^T E^*(\widehat{u}_{t_1}^{*2}\widehat{u}_{t_2}^{*2}) - \frac{1}{T}\sum_{t_1,t_2=1}^T \widehat{u}_{t_1}^2\widehat{u}_{t_2}^2 \\
&= \frac{1}{T}\sum_{t=1}^T \frac{1}{T}\sum_{s=1}^T \widehat{u}_s^4 + \frac{1}{T}\sum_{\substack{t_1,t_2=1 \\ t_1\neq t_2}}^T \left(\frac{1}{T}\sum_{s=1}^T \widehat{u}_s^2\right)^2 - T\left(\frac{1}{T}\sum_{t=1}^T \widehat{u}_t^2\right)^2 \\
&= \frac{1}{T}\sum_{s=1}^T \widehat{u}_s^4 - \left(\frac{1}{T}\sum_{s=1}^T \widehat{u}_s^2\right)^2.
\end{aligned}$$

Again replacing $\widehat{u}_t$ by u_t above does not affect the asymptotics and the last right-hand side converges in probability to

$$V_{PB}^{(2,2)} := E(u_t^4) - \sigma_u^4 = \tau_{0,0,0,0} - \sigma^4 \neq \sum_{h=-\infty}^{\infty} \left\{\tau_{0,0,h,h} - \sigma_u^4\right\} = V^{(2,2)},$$

which also proves also the general inconsistency of the pairwise bootstrap. Observe here that the pairwise bootstrap is equivalent to an i.i.d. bootstrap applied to the residuals. Similarly, one can show that

$$V_{PB}^{(2,1)} = \sum_{j=1}^{\infty} \tau_{0,0,0,j}(C_j \otimes I_K)'(\Gamma^{-1} \otimes I_K)' \neq \sum_{j=1}^{\infty}\sum_{h=0}^{\infty} \tau_{0,0,h,h+j}(C_j \otimes I_K)'(\Gamma^{-1} \otimes I_K)' = V^{(2,1)}$$

holds. That is, in comparison to the wild bootstrap, the pairwise bootstrap does not estimate the limiting covariances as being zero. Yet, the limiting covariances are not correctly estimated in general if the innovations are not i.i.d.. However, the pairwise bootstrap will asymptotically be valid if $u_t \sim i.i.d.(0, \sigma_u)$ in contrast to the wild bootstrap approaches.

4.3 Blockwise wild bootstrap

For notational convenience, suppose that $T = N\ell$, where $\ell \in \mathbb{N}$ denotes the block length and N the number of blocks. For the blockwise wild bootstrap, let $\eta_1, \ldots, \eta_N$ be i.i.d. random variables with $E^*(\eta_t) = 0$, $E^*(\eta_t^2) = 1$ and $E^*(\eta_t^4) < \infty$ and define $\widehat{u}_t^* = \widehat{u}_t \eta_{\lceil t/\ell \rceil}$. In other words, we cut $\widehat{u}_1, \ldots, \widehat{u}_T$ in N blocks of length ℓ and multiply the jth block with η_j to get the bootstrap sample $\widehat{u}_1, \ldots, \widehat{u}_T$ and the corresponding estimator $\widehat{\sigma}_{BWB}^{2*} = \frac{1}{T}\sum_{t=1}^T \widehat{u}_t^{*2}$. From $E^*(\eta_t^2) = 1$, we have $E^*(\sqrt{T}(\widehat{\sigma}_{BWB}^{2*} - \widehat{\sigma}_u^2)) = 0$ and

$$
\begin{aligned}
E^*\left(\sqrt{T}\left(\widehat{\sigma}^{2*}_{BWB}-\widehat{\sigma}^2_u\right)\right)^2 &= \frac{1}{T}\sum_{r_1,r_2=1}^{N}\sum_{s_1,s_2=1}^{\ell} E^*(\widehat{u}^{*2}_{s_1+(r_1-1)\ell}\widehat{u}^{*2}_{s_2+(r_2-1)\ell}) - \frac{1}{T}\sum_{t_1,t_2=1}^{T}\widehat{u}^2_{t_1}\widehat{u}^2_{t_2} \\
&= \frac{1}{T}\sum_{r=1}^{N}\sum_{s_1,s_2=1}^{\ell}\widehat{u}^2_{s_1+(r-1)\ell}\widehat{u}^2_{s_2+(r-1)\ell}E^*(\eta_r^4) - \frac{1}{T}\sum_{t_1,t_2=1}^{T}\widehat{u}^2_{t_1}\widehat{u}^2_{t_2} \\
&\quad +\frac{1}{T}\sum_{\substack{r_1,r_2=1\\ r_1\neq r_2}}^{N}\sum_{s_1,s_2=1}^{\ell}\widehat{u}^2_{s_1+(r_1-1)\ell}\widehat{u}^2_{s_2+(r_2-1)\ell}E^*(\eta^2_{r_1})E^*(\eta^2_{r_2}) \\
&= \left(\frac{1}{T}\sum_{r=1}^{N}\sum_{s_1,s_2=1}^{\ell}\widehat{u}^2_{s_1+(r-1)\ell}\widehat{u}^2_{s_2+(r-1)\ell}\right)\left(E^*(\eta_r^4)-1\right).
\end{aligned}
$$

Interestingly, it turns out that the first factor on the last right-hand side above is of order $O_P(\ell)$ and diverges for $\ell\to\infty$. This can be seen by the following calculation. We get

$$
\begin{aligned}
&\frac{1}{T}\sum_{r=1}^{N}\sum_{s_1,s_2=1}^{\ell}\widehat{u}^2_{s_1+(r-1)\ell}\widehat{u}^2_{s_2+(r-1)\ell} \\
&= \frac{1}{T}\sum_{r=1}^{N}\sum_{s_1,s_2=1}^{\ell}(\widehat{u}^2_{s_1+(r-1)\ell}-E(\widehat{u}^2_{s_1+(r-1)\ell}))(\widehat{u}^2_{s_2+(r-1)\ell}-E(\widehat{u}^2_{s_2+(r-1)\ell})) \\
&\quad +\frac{1}{T}\sum_{r=1}^{N}\sum_{s_1,s_2=1}^{\ell}E(\widehat{u}^2_{s_1+(r-1)\ell})E(\widehat{u}^2_{s_2+(r-1)\ell}) \\
&= O_P(1)+\ell E^2(\widehat{u}^2_1) = O_P(\ell).
\end{aligned}
$$

The latter result indicates that the blockwise wild bootstrap is not only unable to mimic the proper limiting variance but also that the conditional variance of $\widehat{\sigma}^{2*}_{BWB}$ is not even finite in the limit if $\ell\to\infty$ as $T\to\infty$. Consequently, the blockwise wild bootstrap fails drastically here. Therefore, we do not consider this bootstrap scheme any further in the paper.

4.4 Numerical evaluation of asymptotic bias

We have numerically evaluated the bias when replacing the asymptotic covariance matrix $V^{(2,2)}$ by the variance expressions obtained from the wild or pairwise bootstrap. To this end, we again focus on the univariate case and consider a simple GARCH(1,1) model for u_t:

$$
u_t=\sigma_t\varepsilon_t,\quad \sigma_t^2=a_0+a_1u^2_{t-1}+b_1\sigma^2_{t-1},\ \text{with } a_0=1-a_1-b_1 \text{ and } \varepsilon_t\sim i.i.d.\ N(0,1). \tag{4.1}
$$

In line with Remark 3.2 we write for the univariate case $V^{(2,2)}=\mathrm{Var}(u_t^2)+2\sum_{h=1}^{\infty}\gamma_{u^2}(h)$, where $\gamma_{u^2}(h)=\mathrm{Cov}(u_t^2,u^2_{t-h})$. From Francq & Zakoïan (2010, Chapter 2) and using some algebra we get

$$
V^{(2,2)}=\mathrm{Var}(u_t^2)+2\,\mathrm{Var}(u_t^2)\rho_{u^2}(1)\frac{1}{1-a_1-b_1},
$$

Table 1: Moments for GARCH(1,1) model (4.1)

Case	a_1	b_1	$\mathrm{Var}(u_t^2)$	$2\sum_{h=1}^{\infty}\gamma_{u^2}(h)$	$V^{(2,2)}$	$V_{WB}^{(2,2)}$	$V_{PB}^{(2,2)}$
G0	0.00	0.00	2	0	2	6	2
G1	0.05	0.94	3.007	93.154	96.161	8.013	3.007
G2	0.05	0.90	2.162	6.270	8.432	6.324	2.162
G3	0.50	0.00	8.000	16.000	24.000	18.000	8.000
G4	0.30	0.60	56.000	552.00	608.00	114.000	56.000
G5	0.20	0.75	15.714	262.86	278.57	33.429	15.714

Note: The results for $V_{WB}^{(2,2)}$ are based on $\eta_t \sim i.i.d.N(0,1)$.

where the first-order autocorrelation of u_t^2 is given by

$$\rho_{u^2}(1) = \frac{a_1\{1 - b_1(a_1 + b_1)\}}{1 - 2a_1 b_1 - b_1^2}.$$

Moreover, $\mathrm{Var}(u_t^2) = E(u_t^4) - \sigma_u^4$. Since $\sigma_u^4 = 1$ in our case we obtain from Francq & Zakoïan (2010, Chapter 2)

$$E(u_t^4) = \frac{1 - (a_1 + b_1)^2}{1 - (a_1 + b_1)^2 - a_1^2(\kappa_\varepsilon - 1)}\kappa_\varepsilon,$$

where $\kappa_\varepsilon = E(\varepsilon_t^4) = 3$.

From the previous subsections we have for the pairwise bootstrap $V_{PB}^{(2,2)} = Var(u_t^2)$. For the wild bootstraps we get $V_{WB}^{(2,2)} = E(u_t^4)(E^*(\eta_t^4) - 1)$. Typical choices for the distribution of η_t are the standard normal or the Rademacher distribution. In case of $\eta_t \sim i.i.d.N(0,1)$ one has $E^*(\eta_t^4) = 3$ such that $V_{WB}^{(2,2)} = 2E(u_t^4)$. In contrast, the Rademacher distribution implies $E^*(\eta_t^4) = 1$ such that $V_{WB}^{(2,2)} = 0$ independent of the conditional variance model for u_t. Therefore, we do not consider the Rademacher distribution any further in the paper.

Table 1 summarizes the results for different values of the GARCH parameters a_1 and b_1. The choices are mainly motivated by the parameters considered in Gonçalves & Kilian (2004). Obviously, the asymptotic bias with respect to $V^{(2,2)}$ can be tremendous. Nevertheless, the absolute but also the relative bias depend quite importantly on a_1 and b_1. E.g. in Cases G2 and G3 $V_{WB}^{(2,2)}$ are relatively close to $V^{(2,2)}$. Both bootstrap variants always underestimate the correct asymptotic variance since the sum of the covariances dominate. However, in a multivariate set-up the asymptotic variance of the estimator of interest are linear combinations of the matrix version of $V^{(2,2)}$ such that the relevant variance may be even overestimated. Finally, note the potential dramatic increase in the asymptotic variance $V^{(2,2)}$ when switching from the i.i.d. case G0 to a GARCH set-up.

5 Residual-Based Moving Block Bootstrap

Block bootstrap methods have been used for several purposes in time series econometrics. In the literature, the block bootstrap has been applied to suitably defined residuals that are obtained after fitting a certain model or differencing the data. For instance, Paparoditis & Politis (2001) and Paparoditis & Politis (2003) apply the MBB to unit root testing and prove bootstrap consistency, where Jentsch, Paparoditis & Politis (2014) provide theory for residual-based block bootstraps in multivariate integrated and co-integrated models. In this section, we propose to use the moving block bootstrap techniques for the residuals obtained from a fitted VAR(p) model to approximate the proper distribution of $\sqrt{T}((\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})',(\widehat{\boldsymbol{\sigma}}-\boldsymbol{\sigma})')'$ derived in Theorem 3.1(ii), which leads to bootstrap consistency in Theorem 5.1 below.

Bootstrap Scheme I

Step 1. Fit a VAR(p) model to the data to get $\widehat{A}_1,\ldots,\widehat{A}_p$ and compute the residuals $\widehat{u}_t = y_t - \widehat{A}_1 y_{t-1} - \cdots - \widehat{A}_p y_{t-p}$, $t = 1,\ldots,T$.

Step 2. Choose a block length $\ell < T$ and let $N = \lceil T/\ell \rceil$ be the number of blocks needed such that $\ell N \geq T$. Define $(K\times\ell)$-dimensional blocks $B_{i,\ell} = (\widehat{u}_{i+1},\ldots,\widehat{u}_{i+\ell})$, $i = 0,\ldots,T-\ell$ and let $i_0,\ldots,i_{N-1}$ be i.i.d. random variables uniformly distributed on the set $\{0,1,2,\ldots,T-\ell\}$. Lay blocks $B_{i_0,\ell},\ldots,B_{i_{N-1},\ell}$ end-to-end together and discard the last $N\ell - T$ values to get bootstrap residuals $\widehat{u}_1^*,\ldots,\widehat{u}_T^*$.

Step 3. Center $\widehat{u}_1^*,\ldots,\widehat{u}_T^*$ according to the rule

$$u_{j\ell+s}^* = \widehat{u}_{j\ell+s}^* - E^*(\widehat{u}_{j\ell+s}^*) = \widehat{u}_{j\ell+s}^* - \frac{1}{T-\ell+1}\sum_{r=0}^{T-\ell}\widehat{u}_{s+r} \tag{5.1}$$

for $s = 1,2,\ldots,\ell$ and $j = 0,1,2,\ldots,N-1$ to get $E^*(u_t^*) = 0$ for all $t = 1,\ldots,T$.

Step 4. Set bootstrap pre-sample values $y_{-p+1}^*,\ldots,y_0^*$ equal to zero and generate the bootstrap sample $y_1^*,\ldots,y_T^*$ according to

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

Step 5. Compute the bootstrap estimator

$$\widehat{\boldsymbol{\beta}}^* = \mathrm{vec}(\widehat{A}_1^*,\ldots,\widehat{A}_p^*) = ((Z^*Z^{*\prime})^{-1}Z^* \otimes I_K)\mathbf{y}^*, \tag{5.2}$$

where Z^* and $\mathbf{y}^*$ are defined analogously to Z and $\mathbf{y}$ in (2.2), respectively, but based on $y_{-p+1}^*,\ldots,y_0^*,y_1^*,\ldots,y_T^*$. Further, we define the bootstrap analogue of $\widehat{\Sigma}_u$ as

$$\widehat{\Sigma}_u^* = \frac{1}{T}\sum_{t=1}^{T}\widehat{u}_t^*\widehat{u}_t^{*\prime}, \tag{5.3}$$

where $\widehat{u}_t^* = y_t^* - \widehat{A}_1^* y_{t-1}^* - \cdots - \widehat{A}_p^* y_{t-p}^*$ are the bootstrap residuals obtained from the VAR(p) fit. We set $\widehat{\boldsymbol{\sigma}}^* = \mathrm{vech}(\widehat{\Sigma}_u^*)$.

Remark 5.1. Contrary to a bootstrap scheme that uses i.i.d. resampling of the residuals, the standard centering $\widehat{u}_t = \widetilde{u}_t - \frac{1}{T}\sum_{s=1}^{T}\widetilde{u}_s$, $t = 1,\ldots,T$, does in general lead to $E^*(u_t^*) \neq 0$ when a MBB is applied to resample the residuals. To get properly centered residuals, the centering as described in Step 3. has to be executed. Note that (5.1) is tailor-made for the MBB and adjusted centering has to be applied for other approaches as e.g. non-overlapping block bootstrap, cyclical block bootstrap or stationary bootstrap. However, the effect of not properly centered residuals vanishes asymptotically and we expect only a slight loss in performance in practice.

Remark 5.2. In Bootstrap Scheme I we rely on pre-whitening the data which should be much more efficient than drawing from blocks of y_t. As for the wild bootstrap approach one may also consider a fixed-design MBB rather than relying on the recursive structure in Step 4. As discussed in Gonçalves & Kilian (2004), the fixed-design wild bootstrap requires weaker assumptions on the error terms than a recursive version. To prove asymptotic validity of the MBB, however, we require stronger assumptions than needed for an appropriate wild bootstrap framework such that the use of a fixed-design MBB would not simplify the setting here. Therefore, we do not consider this bootstrap scheme in the following.

We make the following assumption.

Assumption 5.1 (cumulants)**.** *The K-dimensional innovation process $(u_t, t \in \mathbb{Z})$ has absolutely summable cumulants up to order eight. More precisely, we have for all $j = 2, \ldots, 8$ and $a_1, \ldots, a_j \in \{1, \ldots, K\}$, $\mathbf{a} = (a_1, \ldots, a_j)$ that*

$$\sum_{h_2,\ldots,h_j=-\infty}^{\infty} |\mathrm{cum}_{\mathbf{a}}(0, h_2, \ldots, h_j)| < \infty \tag{5.4}$$

holds, where $\mathrm{cum}_{\mathbf{a}}(0, h_2, \ldots, h_j)$ denotes the jth joint cumulant of $u_{0,a_1}, u_{h_2,a_2}, \ldots, u_{h_j,a_j}$, see e.g. Brillinger (1981). In particular, this condition includes the existence of eight moments of $(u_t, t \in \mathbb{Z})$.

Such a condition has been imposed e.g. by Gonçalves & Kilian (2007) to prove consistency of wild and pairwise bootstrap methods applied to univariate AR(∞) processes. In terms of α-mixing conditions, Assumption 5.1 is implied by

$$\sum_{m=1}^{\infty} m^{n-2} (\alpha_u(m))^{\delta/(2n-2+\delta)} < \infty$$

for $n = 8$ if all moments up to order eight of $(u_t, t \in \mathbb{Z})$ exist, see Künsch (1989). For example, GARCH processes are known to be geometrically strong mixing under mild assumptions on the conditional distribution. This result goes back to Boussama (1998), compare also the discussion in Lindner (2009). Hence, one can focus on verifying whether the 8-th moment of a GARCH process exists for given GARCH parameters and the conditional distribution, compare Ling & McAleer (2002), Lindner (2009).

Now, we can state

Theorem 5.1 (Residual-based MBB consistency)**.**
Under Assumptions 2.1, 2.2 and 5.1 and if $\ell \to \infty$ such that $\ell^3/T \to 0$ as $T \to \infty$, we have

$$\sup_{x \in \mathbb{R}^{\tilde{K}}} \left| P^* \left(\sqrt{T} \left((\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})', (\widehat{\boldsymbol{\sigma}}^* - \widehat{\boldsymbol{\sigma}})' \right)' \leq x \right) - P \left(\sqrt{T} \left((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})', (\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})' \right)' \leq x \right) \right| \to 0$$

in probability, where P^* denotes the probability measure induced by the residual-based MBB and $\bar{K} = K^2p + (K^2 + K)/2$. The short-hand $x \leq y$ for some $x, y \in \mathbb{R}^d$ is used to denote $x_i \leq y_i$ for all $i = 1, \ldots, d$.

Remark 5.3. The proof of bootstrap consistency in Theorem 5.1 is provided in Appendix A. It is not restricted to innovation processes $(u_t, t \in \mathbb{Z})$ being martingale difference sequences and can be achieved under suitable mixing and moment conditions alone. However, by dropping the mds condition the covariance matrix V will differ in general to that derived in Theorem 3.1 where the mds structure is heavily exploited.

6 Inference on Impulse Response Functions

The theoretical results derived above are important if inference is done on quantities that depend on both the VAR parameters and the innovation covariance matrix. This is e.g. the case if inference on structural impulse responses is of interest. Since these impulse responses are very often used in empirical VAR studies, we illustrate the implication of our results in the context of structural impulse responses. Inference based on asymptotic theory for an i.i.d. error term set-up is discussed e.g. in Lütkepohl (1990) while bootstrap methods for impulse response inference are considered e.g. by Runkle (1987), Fachin & Bravetti (1996), Kilian (1998*b*), Benkwitz, Lütkepohl & Wolters (2001), and Benkwitz, Lütkepohl & Neumann (2000). The properties of bootstrap confidence intervals for this type of IRFs in the case of non-i.i.d. innovations have also been investigated by Monte Carlo simulations in Kilian (1998*a*, 1999).

In this section, we first obtain the asymptotic distribution of the impulse response estimators under conditional heteroskedasticity by relying on the Delta method. Following this, we adapt the MBB bootstrap scheme in order to obtain confidence intervals for the impulse response coefficients. Third, we present a simulation study on the finite sample properties of various bootstrap and asymptotic confidence intervals.

6.1 Asymptotic distribution of impulse response functions

In what follows, we use structural impulse responses obtained from recursive VAR systems that imply a Wold causal ordering. These recursive VARs are popular in empirical work in macroeconomics and finance, see e.g. Sims (1992), Bernanke & Blinder (1992), Christiano et al. (1999), Breitung, Brüggemann & Lütkepohl (2004), Kilian (2009). In recursive VARs the structural shocks w_t are identified by using the Choleski decomposition $\Sigma_u = PP'$, where P is lower-triangular with positive diagonal elements. The shocks are $w_t = P^{-1}u_t$, $t = 1, 2, \ldots$, with $w_t \sim (0, I_K)$. In this framework the structural IRFs are given by $\Theta_i = \Phi_i P$, $i = 0, 1, 2, \ldots$, see e.g. Lütkepohl (2005, Section 2.3). In the following we refer to the parameters Θ_i simply as IRFs. Clearly, the impulse responses in Θ_i are continuously differentiable functions of the parameters in $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$. The estimators of the VMA coefficient matrices, $\widehat{\Phi}_i$, $i = 0, 1, 2, \ldots$, are obtained from the LS estimators of the VAR parameters in $\boldsymbol{\beta}$. Applying the Choleski decomposition to $\widehat{\Sigma}_u$ provides us with the estimator $\widehat{P}$ such that the IRFs estimators are $\widehat{\Theta}_i = \widehat{\Phi}_i\widehat{P}$,

$i = 0, 1, 2, \ldots$. Consequently, their limiting distribution is easily obtained via the Delta method. Following Lütkepohl (2005, Proposition 3.6) on the i.i.d. set-up, one can deduce the following corollary from Theorem 3.1

Corollary 6.1 (CLT for Structural IRFs).
Under Assumptions 2.1 and 2.2 we have

$$\sqrt{T}\,\mathrm{vec}\left(\widehat{\Theta}_i - \Theta_i\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma_{\widehat{\Theta}_i}\right), \quad i = 0, 1, 2, \ldots,$$

where

$$\Sigma_{\widehat{\Theta}_i} = C_{i,\boldsymbol{\beta}} V^{(1,1)} C'_{i,\boldsymbol{\beta}} + C_{i,\boldsymbol{\sigma}} V^{(2,2)} C'_{i,\boldsymbol{\sigma}} + C_{i,\boldsymbol{\beta}} V^{(1,2)} C'_{i,\boldsymbol{\sigma}} + C_{i,\boldsymbol{\sigma}} V^{(1,2)'} C'_{i,\boldsymbol{\beta}} \tag{6.1}$$

with $C_{0,\boldsymbol{\beta}} = 0$, $C_{i,\boldsymbol{\beta}} = \dfrac{\partial\,\mathrm{vec}(\Theta_i)}{\partial \boldsymbol{\beta}'} = (P' \otimes I_K) G_i$, $i = 1, 2, \ldots$, $C_{i,\boldsymbol{\sigma}} = \dfrac{\partial\,\mathrm{vec}(\Theta_i)}{\partial \boldsymbol{\sigma}'} = (I_K \otimes \Phi_i) H$, $i = 0, 1, \ldots$, $G_i = \dfrac{\partial\,\mathrm{vec}(\Phi_i)}{\partial \boldsymbol{\beta}'} = \sum_{m=0}^{i-1} J(\boldsymbol{A}')^{i-1-m} \otimes \Phi_m$, $i = 0, 1, \ldots$, *where* $J = (I_K, 0, \ldots, 0)$ *is a* $(K \times Kp)$ *matrix,* $\boldsymbol{A}$ *is the companion matrix of the VAR process defined in Appendix B, and* $H = \dfrac{\partial\,\mathrm{vec}(P)}{\partial \boldsymbol{\sigma}'}$.

Compared to an i.i.d. error term set-up, different limiting covariance matrices $V^{(1,1)}$ and $V^{(2,2)}$ as well as two additional terms occur in $\Sigma_{\widehat{\Theta}_i}$. These are the last two terms in (6.1) that are present whenever the off-diagonal blocks in V are non-zero, compare Remark 3.1.

6.2 Bootstrap inference on impulse response functions

The implementation of the asymptotic approximation in Corollary 6.1 for inference on the impulse response coefficients can be rather cumbersome since it requires estimation of $V^{(2,2)}$ and $V^{(1,2)}$. As a valid alternative we consider the residual-based MBB for inference.

Let $\theta_{jk,i}$ be the response of the j-th variable to the k-th structural shock that occurred i periods ago, $j, k = 1, \ldots, K$, $i = 0, 1, \ldots$ with $j \leq k$ if $i = 0$. To simplify notation we suppress the subscripts in the following and simply use θ and $\widehat{\theta}$ to represent a specific structural impulse response coefficient and its estimator, respectively. Bootstrap confidence intervals for θ can be obtained by the following scheme that relies on Hall's percentile intervals, compare e.g. Hall (1992) and Lütkepohl (2005, Appendix D).

Bootstrap Scheme II

Step 1. Fit a VAR(p) model to the data in order to obtain the estimator $\widehat{\theta}$ as a function of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\sigma}}$.

Step 2. Apply the Bootstrap Scheme I as described in Section 5 B times, where B is large, in order to obtain B bootstrap versions of $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\boldsymbol{\sigma}}^*$.

Step 3. Compute $\widehat{\theta}^*$ using $\widehat{\boldsymbol{\beta}}^*$ and $\widehat{\boldsymbol{\sigma}}^*$ for each of the B bootstrap versions corresponding to $\widehat{\theta}$. Obtain the $\gamma/2$- and $(1-\gamma/2)$-quantiles of $[\widehat{\theta}^* - \widehat{\theta}]$, $\gamma \in (0, 1)$, labelled as $c^*_{\gamma/2}$ and $c^*_{(1-\gamma/2)}$, respectively.

Step 4. Determine Hall's percentile interval by

$$\left[\widehat{\theta} - c^*_{(1-\gamma/2)}; \widehat{\theta} - c^*_{\gamma/2}\right].$$

Since Θ_i, $i = 0, 1, 2, \ldots$, are continuously differentiable functions of $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}$, the asymptotic validity of the Bootstrap Scheme II follows from Theorem 5.1 corresponding to arguments in Kilian (1998*b*). We summarize this result in the following corollary.

Corollary 6.2 (Asymptotic Validity of Bootstrap SIRs).
Under Assumptions 2.1, 2.2 and 5.1 and if $\ell \to \infty$ such that $\ell^3/T \to 0$ as $T \to \infty$, we have

$$\sup_{x \in \mathbb{R}} \left| P^* \left(\sqrt{T} \left(\widehat{\theta}^* - \widehat{\theta} \right)' \leq x \right) - P \left(\sqrt{T} \left(\widehat{\theta} - \theta \right)' \leq x \right) \right| \to 0$$

in probability.

Bootstrap Scheme II can be easily adopted to other interval types like e.g. the standard percentile intervals of Efron & Tibshirani (1993). However, in relative terms the simulation results were similar to the case of Hall's percentile intervals. Therefore, we focus on the latter ones.

6.3 Asymptotic results and simulation evidence

In this section we compare the coverage properties of different bootstrap and asymptotic confidence intervals for impulse responses. For this purpose, we explain the structure of our DGP in Section 6.3.1. We then determine the asymptotic distortion of the wild and pairwise bootstrap approaches in Section 6.3.2 for this DGP before presenting more detailed finite sample results in Section 6.3.3.

6.3.1 Data generating processes

The asymptotic results and the simulation evidence are obtained for a bivariate VAR in form of (2.1) and letting $p = 2$, $\nu = 0$ and

$$A_1 = \begin{pmatrix} 0.4 & 0.6 \\ -0.1 & 1.2 \end{pmatrix}, \quad A_2 = \begin{pmatrix} -0.2 & 0 \\ -0.2 & -0.1 \end{pmatrix}.$$

These DGP parameters lead to typical hump shaped impulse responses often observed in empirical applications. The moduli of the roots in the characteristic VAR polynomial are 0.717 and 0.197 which implies moderate persistence in the VAR dynamics. To control the GARCH structure in the innovation process, let $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim i.i.d.\ N(0, I_2)$ and define $w_{it} = \sigma_{it}\varepsilon_{it}$ with $\sigma^2_{it} = a_0 + a_1 w^2_{it} + b_1 \sigma^2_{i,t-1}$, $i = 1, 2$, and $a_0 = 1 - a_1 - b_1$. Hence, w_{1t} and w_{2t} are two independent univariate GARCH(1,1) processes with $E(w^2_{1t}) = E(w^2_{2t}) = 1$. The VAR innovation

Table 2: Asymptotic Variances of Elements in $\widehat{\Theta}_0$ and Coverage Probabilities of Corresponding Confidence Intervals

Case a_1 b_1	Delta Method		Wild Bootstrap		Pairwise Bootstrap	
	$\widehat{\theta}_{11,0}$	$\widehat{\theta}_{21,0}$	$\widehat{\theta}_{11,0}$	$\widehat{\theta}_{21,0}$	$\widehat{\theta}_{11,0}$	$\widehat{\theta}_{21,0}$
G0 0.00 0.00						
asymptotic variances	0.500	0.875	1.500	1.875	0.500	0.875
coverage probabilities	0.900	0.900	0.996	0.984	0.900	0.900
G1 0.05 0.94						
asymptotic variances	24.04	6.760	2.003	2.001	0.752	0.938
coverage probabilities	0.900	0.900	0.365	0.629	0.229	0.460
G2 0.05 0.90						
asymptotic variances	2.108	1.277	1.581	1.895	0.541	0.885
coverage probabilities	0.900	0.900	0.846	0.955	0.595	0.829
G3 0.50 0.00						
asymptotic variances	6.000	2.250	4.500	2.625	2.000	1.250
coverage probabilities	0.900	0.900	0.846	0.924	0.658	0.780
G4 0.30 0.60						
asymptotic variances	152.0	38.75	28.50	8.625	14.00	4.250
coverage probabilities	0.900	0.900	0.524	0.562	0.382	0.414
G5 0.20 0.75						
asymptotic variances	69.64	18.16	8.357	3.589	3.929	1.732
coverage probabilities	0.900	0.900	0.431	0.535	0.304	0.389

Note: The entries in the columns associated with *Delta method* refer to the quantities obtained from the asymptotically correct covariance matrix $\Sigma_{\widehat{\Theta}_0}$ given in Corollary 6.1. The columns headed by *Wild Bootstrap* and *Pairwise Bootstrap* show the corresponding entries for the asymptotic quantities when using $\Sigma^{PB}_{\widehat{\Theta}_i}$ and $\Sigma^{WB}_{\widehat{\Theta}_i}$, respectively. The wild bootstrap is based on $\eta_t \sim$ $i.i.d.N(0,1)$.

u_t is then defined to be a linear combination (LC) of these two processes given by

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = P \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}, \text{ where } P = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix} \text{ such that } \Sigma_u = PP' = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Thus, ρ describes the correlation between the two components in u_t and we choose $\rho = 0.5$ here to impose moderately large correlation among the two innovation processes. We label our GARCH specification as 'LC-GARCH(1,1)' in the following. It is a special case of a bivariate BEKK-GARCH(1,1,2) model, see e.g. Bauwens, Laurent & Rombouts (2006). It does not only permit to easily control the properties of u_t but also to derive asymptotic expressions of interest in a rather straightforward way. Furthermore note that due to the normality of ε_t the estimators of the VAR parameters A_1 and A_2 and of the variance parameters Σ_u are asymptotically uncorrelated. Hence, $V^{(1,2)} = 0$ such that V is block-diagonal, compare Remark 3.1.

6.3.2 Asymptotic distortions of wild and pairwise bootstrap confidence intervals

In order to simplify the interpretation of the distortions caused by the wild and pairwise bootstrap we have derived the asymptotic coverage probabilities of the corresponding bootstrap confidence intervals for the DGP introduced above. For this purpose, we compute the asymptotic covariance matrices $\Sigma_{\widehat{\Theta}_i}$ using the Delta method and exploiting that $V^{(1,2)} = 0$ in our DGP. Moreover, we derive the corresponding pairwise and wild bootstrap covariance matrices $\Sigma^{PB}_{\widehat{\Theta}_i}$

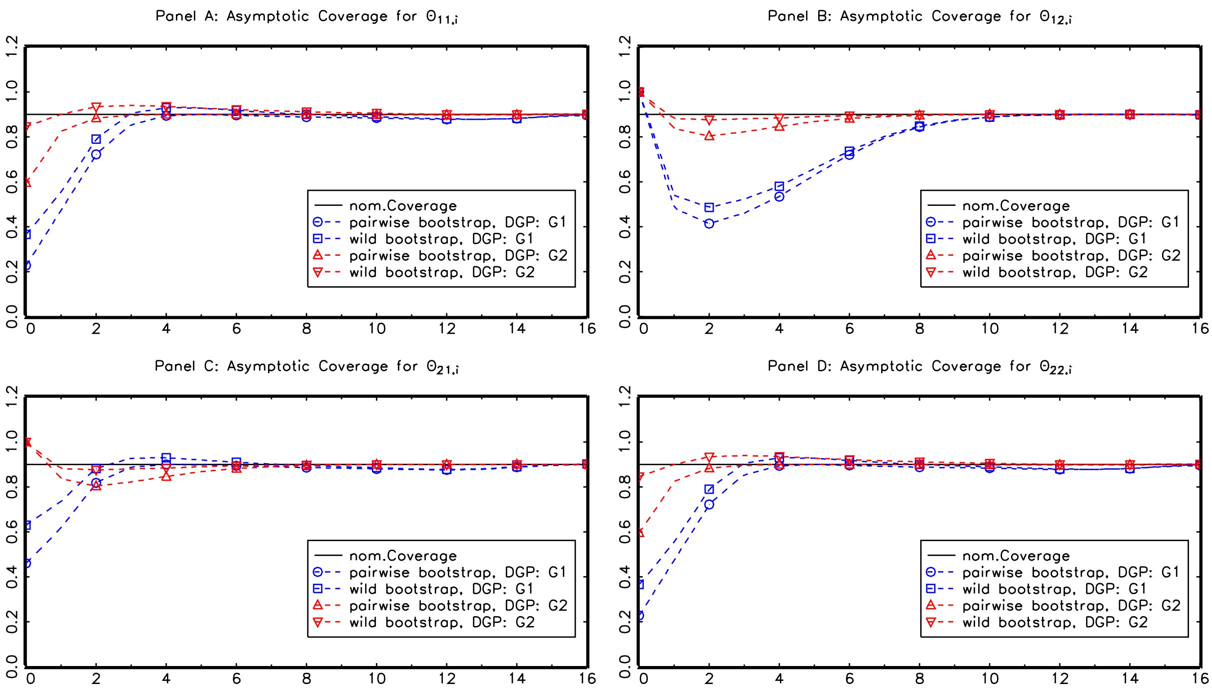

Figure 1: Asymptotic coverage probabilities of pairwise and wild bootstrap impulse response intervals. DGP: VAR(2) with LC-GARCH(1,1) innovations G1 and G2 as in Table 1.

and $\Sigma^{WB}_{\widehat{\Theta}_i}$, respectively, by extending the univariate results of Section 4.4. As described there, we only consider the wild bootstrap in relation to $\eta_t \sim \; i.i.d.N(0,1)$. To evaluate the asymptotic coverage of the boostrap methods, it is assumed that the pairwise (wild) bootstrap estimators of Θ_i are consistent and asymptotically normally distributed with variances $\Sigma^{PB}_{\widehat{\Theta}_0}$ $(\Sigma^{WB}_{\widehat{\Theta}_i})$. Details of the derivations are given in Appendix B.

Note from Table 2 that for the i.i.d. set-up (Case G0) the pairwise bootstrap correctly replicates the asymptotic variances as mentioned in Section 4.2. Hence, the asymptotic coverage probabilities of the corresponding confidence intervals are equal to the nominal level. In contrast, the wild bootstrap overestimates the asymptotic variance such that the coverage probabilities are above the nominal level. In the presence of heteroskedasticity (Cases G1 to G5), we first note that the asymptotic variances of the estimators of the elements in $\Theta_0 = P$ increase substantially. Hence, a correct confidence interval for impulse response coefficients can be expected to be much wider in case of conditional heteroskedasticity compared to an i.i.d. set-up. Moreover, we observe that both bootstrap methods typically underestimate the true asymptotic variances. As a consequence the bootstrap confidence intervals are typically too narrow and the coverage probabilities are often very low. We also note in some cases, in which the sum of the autocovariances of $\boldsymbol{u}_t^2$ is not too large (Case G2 and G3), that the wild bootstrap may overestimate the variances.

To get a more informative picture, we also report asymptotic coverage probabilities of pairwise and wild bootstrap IRF intervals at higher response horizons for DGPs G1 and G2 in Figure 1. Interestingly, we typically observe the most severe problems related to interval coverage rates for Θ_0, i.e. for the period where the shock occurs. At larger horizons i the actual asymptotic coverage converges to the nominal one. This may be explained by the fact that the

IRF variance $\Sigma_{\widehat{\Theta}_i}$ depends on the VAR slope parameters in $\boldsymbol{\beta}$ for $i > 0$. The relevant covariance matrix block $V^{(1,1)}$, however, is correctly replicated by the pairwise and wild bootstraps. Moreover, the estimation uncertainty regarding the variance parameters in $\boldsymbol{\sigma}$ becomes less important rather quickly as i increases. This follows from the fact that $C_{i,\boldsymbol{\sigma}}$ in (6.1) depends on the VMA parameter matrices Φ_i that converge exponentially fast to zero as i increases. In contrast, $C_{i,\boldsymbol{\beta}}$ may even grow for small response horizons before it decreases with a slower rate than $C_{i,\boldsymbol{\sigma}}$ for increasing responses horizons.

Due to the factor $\{E^*\left(\eta_t^4\right) - 1\} = 2$ in $V_{WB}^{(2,2)}$, compare Section 4.1, the coverage probabilities of the wild bootstrap intervals are slightly higher than those of the pairwise bootstrap. This behavior may even lead to wild bootstrap intervals with a coverage above the nominal level as in Case G2. Also note, that the asymptotic coverage is generally much closer to the nominal level for Case G2 than for Case G1. Thus, a small reduction in the GARCH coefficient b_1, and hence in GARCH persistence, strongly reduces the error in coverage probability.

6.3.3 Simulation results on impulse response interval coverage

We compare the properties of the different impulse response intervals using one DGP variant with i.i.d. errors (i.e. Case G0 with $a_1 = 0$ and $b_1 = 0$) and two variants with GARCH innovations with $a_1 = 0.05$ and $b_1 = 0.94$ (Case G1) and $a_1 = 0.05$ and $b_1 = 0.90$ (Case G2) in order to mimic typical empirical GARCH patterns.[3] These GARCH parameters, together with the normality assumption on ε_t, guarantee that Assumption 5.1 is satisfied. For each DGP we generate $M = 5000$ sets of time series data of length $T = 500$ and $T = 5000$ and construct bootstrap impulse response intervals using the standard (i.i.d.) Hall's percentile method as well as recursive- and fixed-design wild bootstrap, pairwise bootstrap and MBB versions of Hall's percentile intervals. The MBB intervals are obtained according to Bootstrap Scheme II presented in Section 6.2. We use different block lengths as described below. The nominal coverage is 90% and we use $B = 999$ bootstrap draws to construct Hall's percentile intervals. For comparison, we also report results of the Delta method confidence intervals based on Corollary 6.1. To simplify the implementation we impose that $V^{(1,2)}$ is zero in our set-up. As mentioned in Remark 3.3, $V^{(2,2)}$ is estimated by applying the VARHAC approach of Den Haan & Levin (1996) using the Akaike Information Criterion (AIC) with a maximum lag order p_{max}.

We present some typical results in Figures 2 and 3 in order to highlight our main findings. We focus on the coverage for $\theta_{21,i}$ and $\theta_{22,i}$ since the findings for $\theta_{11,i}$ and $\theta_{12,i}$ do not give further insights.

Results for $T = 500$ in Panel A an B indicate that the introduction of a persistent GARCH structure reduces the empirical coverage of all considered methods substantially. The i.i.d- and pairwise bootstrap methods are affected most strongly: the empirical coverage on impact may drop down to just above 20%. At the same time the coverage rates of both wild bootstrap variants also drop substantially. Note that both asymptotically correct methods, the residual-based MBB and the Delta method approach, also produce intervals with coverage substantially below nominal level. Although in some cases and at low horizons the MBB seems to outperform

[3] Other parameter constellations have been used for robustness checks, which are discussed later.

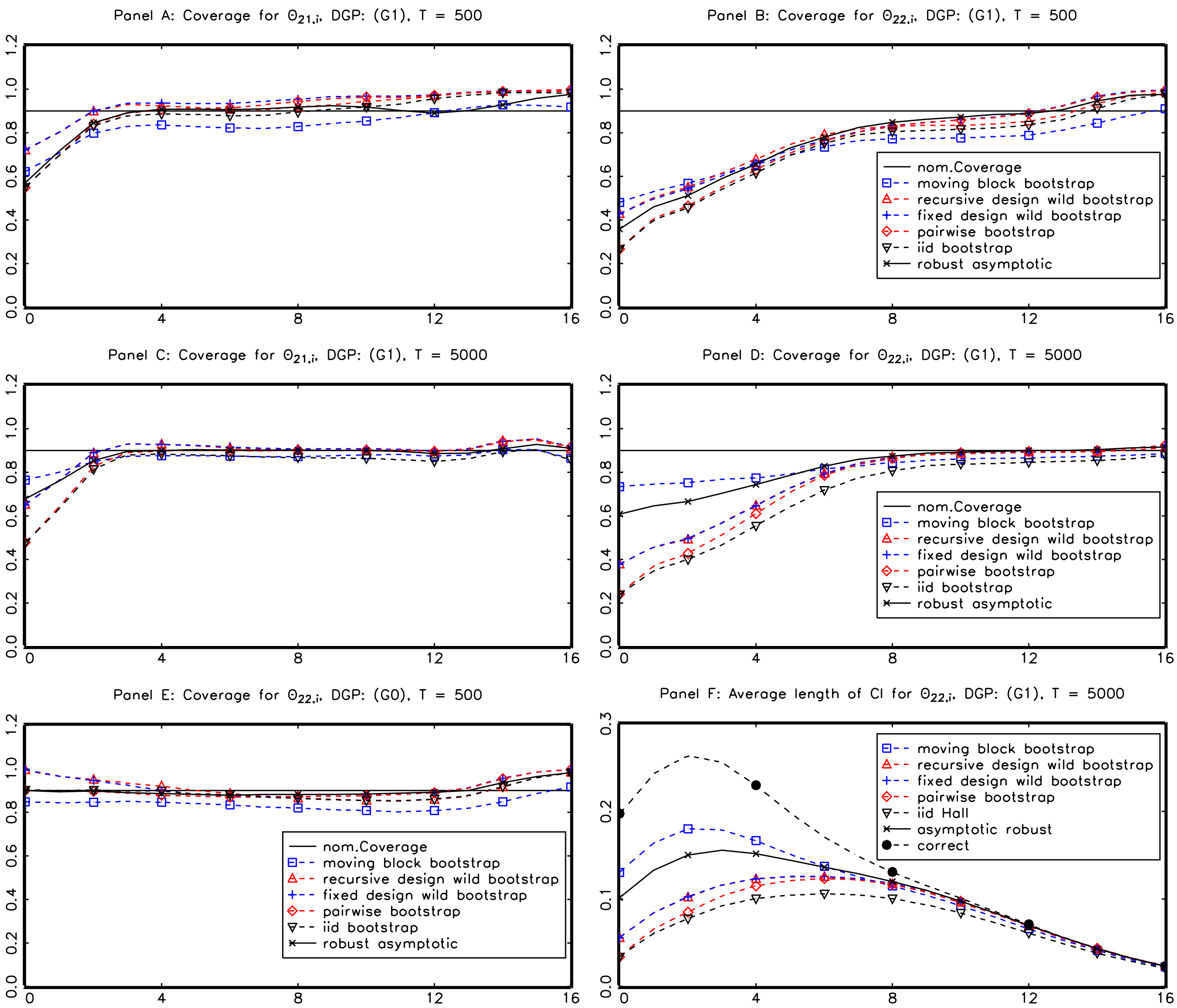

Figure 2: Empirical coverage rates of bootstrap and asymptotic impulse response intervals. Moving block bootstrap (MBB) block length and VARHAC lag order: $\ell = 50$ and $p_{max} = 8$ ($T = 500$) and $\ell = 200$ and $p_{max} = 16$ ($T = 5000$), DGP: VAR(2) with GARCH innovations G1 as in Table 1.

the other approaches marginally, even for moderately large samples the MBB intervals do not entirely solve the coverage problems induced by the persistent GARCH innovation structure. We also observe that the coverage at later horizons increase towards nominal coverage for all methods.

As a reference, we also report corresponding coverage rates for a DGP with i.i.d. innovations (Case G0) in Panel E. In this case the i.i.d.- and pairwise bootstrap procedures lead to intervals with empirical coverage rates very close to the nominal level of 90%. In contrast, both the recursive- and fixed-design wild bootstrap lead to intervals with coverage rates above the nominal level. Note that these simulation results nicely line up with those discussed in Table 2 and Figure 1. In addition we find that the MBB intervals show coverage somewhat below nominal level, which indicates a loss of efficiency as the block bootstrap is not needed in this case.

As expected, with $T = 5000$ observations (see Panels B and C of Figure 2), the inconsistent methods still produce intervals with very low coverage. In contrast, the coverage of intervals from the consistent MBB and the Delta method increase substantially. Nevertheless, the required

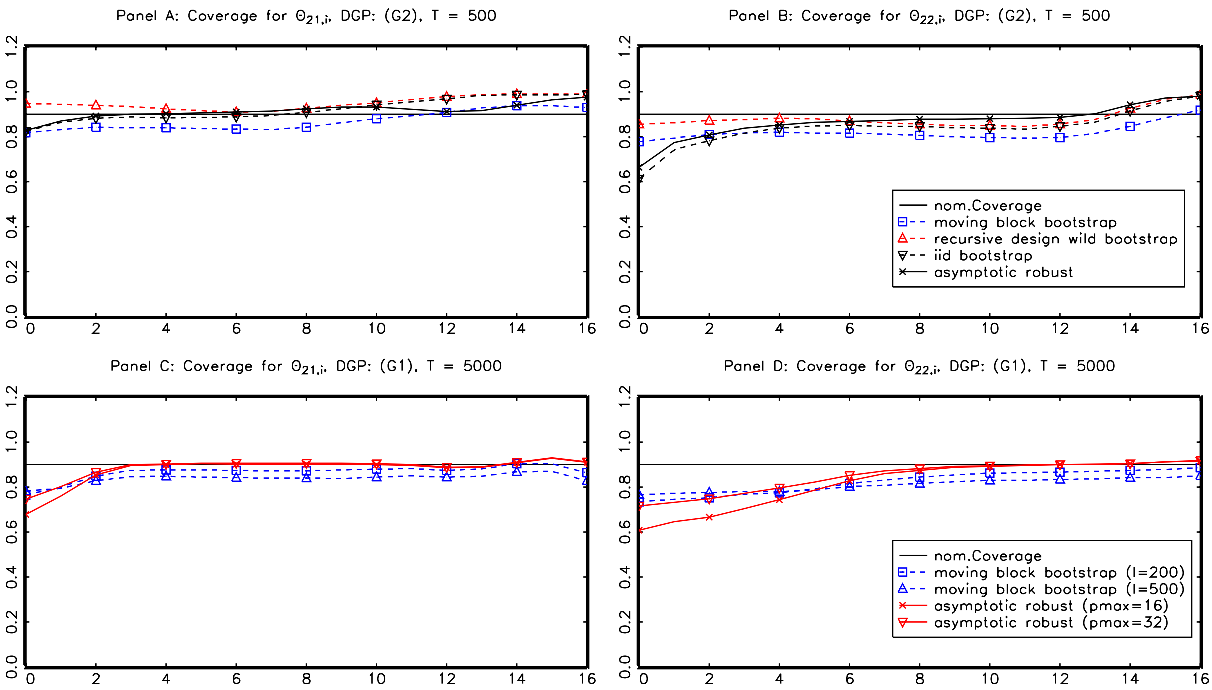

Figure 3: Empirical coverage rates of bootstrap and asymptotic impulse response intervals. $T = 5000$. Panels A and B: Moving block bootstrap block length and VARHAC lag order: $\ell = 50$ and $p_{max} = 8$, DGP: VAR(2) with GARCH innovations G1 and G2 as in Table 1.

sample size for making the MBB work reasonably well in practice seems to be fairly large if the GARCH structure is very persistent. Similar comments apply to the Delta method approach. The reason for the finite sample distortions is the downward bias of the estimators of $\Sigma_{\widehat{\Theta}_i}$. As a consequence, the confidence intervals are too narrow such that their coverage falls below the nominal level. This is illustrated in Panel F of Figure 2, where we show the different average interval lengths for G1 and $T = 5000$ together with the length of the asymptotically correct confidence intervals derived from Corollary 6.1. Obviously, the higher empirical coverage of the MBB and the Delta method intervals is due to their larger width. The wide intervals reflect the tremendous increase in estimation uncertainty when comparing Case G1 with a situation of i.i.d. innovations. Clearly, the MBB and the asymptotic Delta method approach still underestimate the true sampling variation. However, the associated variance estimates converge to the correct ones as T increases although the convergence seems to be rather slow.

We conduct a number of additional simulation experiments to address further issues and briefly summarize our findings. First, we considered Case G2 for which the GARCH parameter b_1 is reduced from 0.94 to 0.90. Panels A and B of Figure 3, which correspond to Panels A and B of Figure 2, show the empirical coverage for some of the approaches and $T = 500$. Obviously, all approaches result in much more appropriate empirical coverages compared to Case G1. Hence, a small reduction in the persistence of the GARCH process also strongly reduces the finite sample error in coverage probabilities. Nevertheless, the empirical coverage rates can still be somewhat below (or above) the nominal level on impact. In this respect, the moving block bootstrap performs reasonably well.

Panels C and D of Figure 3 demonstrate the effects of varying the block length ℓ for the MBB and the maximal lag order p_{max} used in the VARHAC approach for estimating $V^{(2,2)}$. Our results suggests that a longer block length or larger values of p_{max} lead to comparably higher coverage rates in larger samples. For instance, using $p_{max} = 32$ instead of $p_{max} = 16$ increases coverage of the confidence interval for $\theta_{22,0}$ by about 15 percentage points for Case G1 if $T = 5000$ (see Panel D of Figure 3). We generally find that the residual-based MBB leads to better empirical coverage at impact and early response horizons than the Delta method approach. Nevertheless, there are also situations in which the latter approach marginally dominates, in particular if the response horizon increases. Potentially, the Delta method may benefit from imposing $V^{(1,2)} = 0$ in our simulations.

We have conduct further experiments but for the sake of brevity we only summarize the findings without reporting detailed results. First, we try different residual correlations and look at coverage results for $\rho = 0.1$ and $\rho = 0.9$. We find that the strongest impact is on the cross-responses $\theta_{21,i}$. The larger the contemporaneous correlation, the lower is the empirical coverage for the response coefficients $\theta_{21,i}$. Second, we consider different alternative GARCH and VAR specifications. For the VAR part, we also use the bivariate VAR(1) of Kilian (1998*b*,*a*, 1999) and a bivariate VAR(5) model estimated from US-Euro interest rate spread data. Alternative GARCH specifications include various GARCH parameter combinations and conditional distributions for our LC-GARCH(1,1) and a bivariate BEKK(1,1,1) specification estimated from an interest rate spread system. We also allow ε_t to follow an asymmetric distribution, like e.g. a mixed-normal distribution that leads to a non-zero covariance matrix $V^{(1,2)}$. While we again find that the reduction in coverage rates is stronger the more persistent the GARCH equations and the more heavy-tailed the innovation distributions are, none of our alternative GARCH and VAR specifications affect the relative performance of the considered approaches in any important way.

Overall, our results highlight that the i.i.d.- and pairwise bootstrap procedures are not appropriate tools for inference on IRFs if very persistent GARCH effects are present. It is important to note that this is not merely a small sample phenomenon but also persists in very large samples. Despite being asymptotically invalid the wild bootstrap, however, performs reasonably well in moderately large samples. In the presence of conditional heteroskedasticity, using the residual-based MBB is asymptotically correct. Nevertheless, our simulation experiments suggest that the MBB as well as the asymptotic Delta method procedure work reasonably well only in fairly large samples. However, in case of less persistent GARCH effects that may be observed for weekly or monthly financial market or macroeconomic data, finite sample inference is more reliable. In any case, practitioners have to be aware of the increased estimation uncertainty that should be reflected in wider confidence intervals compared to the case of i.i.d. innovations. Essentially, the reported intervals may not fully reflect the underlying estimation uncertainty.

7 Conclusions

Our paper provides theoretical results for inference in VAR models in the presence of conditional heteroskedasticity of unknown form. We derive the joint asymptotic distribution of the LS estimators of both the VAR parameters as well as of the unconditional innovation variance parameters in the presence of conditional heteroskedasticity. The results are important for inference on quantities that are functions of both VAR and innovation covariance variance parameters, as e.g. in the case of impulse responses to orthogonalized shocks. We show that under appropriate assumptions the residual-based moving block bootstrap leads to asymptotically valid inference in this set-up while the commonly applied wild and pairwise bootstrap schemes fail in this respect.

We illustrate the performance of asymptotic and bootstrap inference under heteroskedasticity in the context of impulse responses that depend on the VAR and the innovation covariance parameter estimates. The results of our simulation study indicate that the estimation uncertainty can be rather high in case of conditional heteroskedasticity when compared to an i.i.d. set-up. Importantly, this is not merely a finite sample issue but is rather due to the asymptotic properties as can be seen from numerical evaluations of the relevant asymptotic variance expressions. Moreover, the asymptotically valid Delta method and bootstrap approaches often underestimate the true sampling variation. Furthermore, it turns out that the bootstrap schemes which are asymptotically invalid do not need to perform worse than the MBB if the sample size is small.

Our results have important implications for practical work using IRFs on time series with heteroskedasticity patterns. Practitioners should be aware of the fact that reported IRF intervals may understate the actual estimation uncertainty substantially. Therefore, interpreting the confidence intervals for IRFs should be done cautiously against this background.

An interesting extension of our framework is to consider cointegrated VAR models for variables that are integrated of order 1. One may expect that appropriate asymptotic results can also be obtained for such a set-up given the results in Cavaliere, Rahbek & Taylor (2010) and Jentsch et al. (2014). To be precise, a joint central limit theorem on the relevant estimators corresponding to Theorem 3.1 as well as a proof of the asymptotic validity of the MBB applied to residuals obtained from an estimated vector error correction model is required. This is left for future research.

References

Alter, A. & Schüler, Y. S. (2012), 'Credit spread interdependencies of European states and banks during the financial crisis', *Journal of Banking and Finance* **36**(12), 3444–3468.

Bauwens, L., Laurent, S. & Rombouts, J. V. K. (2006), 'Multivariate GARCH models: a survey', *Journal of Applied Econometrics* **21**(1), 79–109.

Benkwitz, A., Lütkepohl, H. & Neumann, M. H. (2000), 'Problems related to confidence intervals for impulse responses of autoregressive processes', *Econometric Reviews* **19**, 69–103.

Benkwitz, A., Lütkepohl, H. & Wolters, J. (2001), 'Comparison of bootstrap confidence intervals for impulse responses of German monetary systems', *Macroeconomic Dynamics* **5**(01), 81–100.

Berkane, M. & Bentler, P. (1986), 'Moments of elliptically distributed random variates', *Statistics & Probability Letters* **4**(6), 333 – 335.

Bernanke, B. S. & Blinder, A. S. (1992), 'The federal funds rate and the channels of monetary transmission', *American Economic Review* **82**(4), 901–922.

Billingsley, P. (1995), *Probability and Measure*, 3nd edn, Wiley.

Boussama, F. (1998), Ergodicité, mélange et estimation dans le modelés GARCH, PhD thesis, Université 7 Paris.

Breitung, J., Brüggemann, R. & Lütkepohl, H. (2004), Structural vector autoregressive modeling and impulse responses, *in* H. Lütkepohl & M. Krätzig, eds, 'Applied Time Series Econometrics', Cambridge University Press, Cambridge, pp. 159–196.

Brüggemann, R., Härdle, W., Mungo, J. & Trenkler, C. (2008), 'VAR modeling for dynamic loadings driving volatility strings', *Journal of Financial Econometrics* **6**(3), 361–381.

Brockwell, P. & Davis, R. (1991), *Time Series: Theory and Methods*, 2nd edn, Berlin: Springer-Verlag.

Cavaliere, G., Rahbek, A. & Taylor, A. R. (2010), 'Cointegration rank testing under conditional heteroskedasticity', *Econometric Theory* **26**(06), 1719–1760.

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

Davidson, J. (1994), *Stochastic Limit Theory*, New York: Oxford University Press.

Den Haan, W. J. & Levin, A. (1996), Inferences from parametric and non-parametric covariance matrix estimation procedures, NBER Technical Working Papers 0195, National Bureau of Economic Research, Inc.

Deo, R. S. (2000), 'Spectral tests of the martingale hypothesis under conditional heteroscedasticity', *Journal of Econometrics* **99**, 291–315.

Efron, B. (1979), 'Bootstrap methods: another look at the jackknife', *The Annals of Statistics* **7**(1), 1–26.

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

Fachin, S. & Bravetti, L. (1996), 'Asymptotic normal and bootstrap inference in structural VAR analysis', *Journal of Forecasting* **15**, 329–341.

Francq, C. & Zakoïan, J.-M. (2004), 'Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes', *Bernoulli* **10**(4), 605–637.

Francq, C. & Zakoïan, J.-M. (2010), *GARCH Models*, Wiley, Chichester.

Gonçalves, S. & Kilian, L. (2004), 'Bootstrapping autoregressions with conditional heteroskedasticity of unknown form', *Journal of Econometrics* **123**(1), 89–120.

Gonçalves, S. & Kilian, L. (2007), 'Asymptotic and bootstrapping inference for AR(∞) processes with conditional heteroskedasticity', *Econometric Reviews* **26**(6), 609–641.

Hafner, C. M. (2004), 'Fourth moment structure of multivariate GARCH models', *Journal of Financial Economerics* **1**(1), 26–54.

Hafner, C. M. & Herwartz, H. (2009), 'Testing for linear vector autoregressive dynamics under multivariate generalized autoregressive heteroskedasticity', *Statistica Neerlandica* **63**(3), 294–323.

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

Herwartz, H. & Lütkepohl, H. (2014), 'Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks', *Journal of Econometrics,* **forthcoming**.

Jentsch, C., Paparoditis, E. & Politis, D. N. (2014), 'Block bootstrap theory for multivariate integrated and cointegrated time series', *Journal of Time Series Analysis,* **forthcoming**.

Kilian, L. (1998*a*), 'Confidence intervals for impulse responses under departures from normality', *Econometric Reviews* **17**(1), 1–29.

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

Kilian, L. (1999), 'Finite sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses', *The Review of Economics and Statistics* **81**(4), 652–660.

Kilian, L. (2009), 'Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market', *American Economic Review* **99**(3), 1053–1069.

Kim, S. & Roubini, N. (2000), 'Exchange rate anomalies in the industrial countries: A solution with a structural VAR approach', *Journal of Monetary Economics* **45**(3), 561–586.

Kreiss, J.-P. (1997), 'Asymptotic properties of residual bootstrap for autoregressions', *Unpublished manuscript* .

Kreiss, J.-P. & Franke, J. (1992), 'Bootstrapping stationary autoregressive moving-average models', *Journal of Time Series Analysis* **13**(4), 297–317.

Künsch, H. (1989), 'The jackknife and the bootstrap for general stationary observations', *The Annals of Statistics* **17**(3), 1217–1241.

Lahiri, S. (2003), *Resampling Methods for Dependent Data*, New York: Springer.

Lindner, A. M. (2009), Stationarity, mixing, distributional properties and moments of GARCH(p,q)-processes, *in* T. Andersen, R. Davis, J.-P. Kreiss & T. Mikosch, eds, 'Handbook of Financial Time Series', Berlin: Springer-Verlag, pp. 43–69.

Ling, S. & McAleer, M. (2002), 'Stationarity and the existence of moments of a family of GARCH processes', *Journal of Econometrics* **106**(1), 109–117.

Ling, S. & McAleer, M. (2003), 'Asymptotic theory for a Vector ARMA-GARCH model', *Econometric Theory* **19**(2), 280–310.

Liu, R. Y. & Singh, K. (1992), Moving blocks jackknife and bootstrap capture weak dependence, *in* R. LePage & L. Billard, eds, 'Exploring the Limits of Bootstrap', New York: Wiley, pp. 225–248.

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

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

Normandin, M. & Phaneuf, L. (2004), 'Monetary policy shocks:: Testing identification conditions under time-varying conditional volatility', *Journal of Monetary Economics* **51**(6), 1217–1243.

Paparoditis, E. & Politis, D. N. (2001), Unit root testing via the continuous-path block bootstrap, Discussion Paper 2001-06, Department of Economics, University of California-San Diego (UCSD).

Paparoditis, E. & Politis, D. N. (2003), 'Residual-based block bootstrap for unit root testing', *Econometrica* **71**(3), 813–855.

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

Runkle, D. E. (1987), 'Vector autoregressions and reality', *Journal of Business and Economic Statistics* **5**(4), 437–454.

Shao, X. (2011), 'A bootstrap-assisted spectral test of white noise under unknown dependence', *Journal of Econometrics* **162**(2), 213–224.

Sims, C. (1992), 'Interpreting the macroeconomic time series facts: the effects of monetary policy', *European Economic Review* **36**, 975–1011.

Smeekes, S. & Urbain, J.-P. (2014), A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing, Research Memoranda RM/14/008, Maastricht University.

A Proofs

A.1 Proof of Theorem 3.1

We consider only the more sophisticated part (ii) as (i) can be treated as a special case. We define $\widetilde{\boldsymbol{\sigma}} = \text{vech}(\widetilde{\Sigma}_u)$, where $\widetilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^T u_t u_t'$ and due to $\sqrt{T}(\widehat{\boldsymbol{\sigma}} - \widetilde{\boldsymbol{\sigma}}) = o_P(1)$ by standard arguments, we can replace $\widehat{\boldsymbol{\sigma}}$ by $\widetilde{\boldsymbol{\sigma}}$ in the following calculations. Furthermore, by using

$$Z_{t-1} = \begin{pmatrix} y_{t-1} \\ \vdots \\ y_{t-p} \end{pmatrix} = \sum_{j=0}^{\infty} \begin{pmatrix} \Phi_j u_{t-1-j} \\ \vdots \\ \Phi_j u_{t-p-j} \end{pmatrix} = \sum_{j=1}^{\infty} \begin{pmatrix} \Phi_{j-1} u_{t-j} \\ \vdots \\ \Phi_{j-p} u_{t-j} \end{pmatrix} = \sum_{j=1}^{\infty} C_j u_{t-j}, \tag{A.1}$$

it can be shown that

$$\begin{aligned} \sqrt{T}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma} \end{pmatrix} &= \begin{pmatrix} \left\{(\frac{1}{T}ZZ')^{-1} \otimes I_K\right\} \sum_{j=1}^{\infty}(C_j \otimes I_K)\frac{1}{\sqrt{T}}\sum_{t=1}^T \left\{\text{vec}(u_t u_{t-j}')\right\} \\ \left(\frac{1}{\sqrt{T}}\sum_{t=1}^T L_K \left\{\text{vec}(u_t u_t') - \text{vec}(\Sigma_u)\right\}\right) \end{pmatrix} \quad (A.2) \\ &= A_m + (A - A_m), \end{aligned}$$

where L_K is the elimination matrix defined in Assumption 2.1(iv), A denotes the righthand-side of (A.2) and A_m is the same expression, but with $\sum_{j=1}^{\infty}$ replaced by $\sum_{j=1}^{m}$ for some $m \in \mathbb{N}$. In the following, we make use of Proposition 6.3.9 of Brockwell & Davis (1991) and it suffices to show

$$\begin{aligned} &(a) \quad A_m \overset{\mathcal{D}}{\to} \mathcal{N}(0, V_m) \text{ as } T \to \infty \\ &(b) \quad V_m \to V \text{ as } m \to \infty \\ &(c) \quad \forall\, \delta > 0: \lim_{m\to\infty}\limsup_{T\to\infty} P(|A - A_m|_1 > \delta) = 0. \end{aligned}$$

To prove (a), setting $\widetilde{K} = K(K+1)/2$, we can write

$$\begin{aligned} A_m &= \begin{pmatrix} (\frac{1}{T}ZZ')^{-1} \otimes I_K & O_{K^2p\times\widetilde{K}} \\ O_{\widetilde{K}\times K^2p} & I_{\widetilde{K}} \end{pmatrix} \begin{pmatrix} C_1 \otimes I_K & \cdots & C_m \otimes I_K & O_{K^2p\times\widetilde{K}} \\ O_{\widetilde{K}\times K^2} & \cdots & O_{\widetilde{K}\times K^2} & I_{\widetilde{K}} \end{pmatrix} \\ &\quad \times \frac{1}{\sqrt{T}}\sum_{t=1}^T \begin{pmatrix} \text{vec}(u_t u_{t-1}') \\ \vdots \\ \text{vec}(u_t u_{t-m}') \\ L_K\left\{\text{vec}(u_t u_t') - \text{vec}(\Sigma_u)\right\} \end{pmatrix} \\ &= \widehat{Q}_T R_m \frac{1}{\sqrt{T}}\sum_{t=1}^T W_{t,m} \end{aligned}$$

with an obvious notation for the $(K^2p + \widetilde{K} \times K^2p + \widetilde{K})$ matrix $\widehat{Q}_T$, the $(K^2p + \widetilde{K} \times K^2m + \widetilde{K})$ matrix R_m and the $K^2m + \widetilde{K}$-dimensional vector $W_{t,m}$. By Lemma A.2, we have that $\widehat{Q}_T \to Q$ in probability, where $Q = \text{diag}(\Gamma^{-1} \otimes I_K, I_{\widetilde{K}})$. Now, the CLT required for part (a) follows from

Lemma A.1 with

$$V_m = \begin{pmatrix} V_m^{(1,1)} & V_m^{(1,2)} \\ V_m^{(2,1)} & V^{(2,2)} \end{pmatrix} = QR_m\Omega_m R_m'Q',$$

which leads to $V^{(2,2)} = \Omega^{(2,2)}$ defined in (A.3), $V_m^{(2,1)} = V_m^{(1,2)\prime}$ and

$$\begin{aligned} V_m^{(2,1)} &= \sum_{j=1}^{m}\sum_{h=0}^{\infty} L_K\tau_{0,0,h,h+j}(C_j\otimes I_K)'(\Gamma^{-1}\otimes I_K)', \\ V_m^{(1,1)} &= (\Gamma^{-1}\otimes I_K)\left(\sum_{i,j=1}^{m}(C_i\otimes I_K)\tau_{0,i,0,j}(C_j\otimes I_K)'\right)(\Gamma^{-1}\otimes I_K)'. \end{aligned}$$

Part (b) follows from the dominated convergence theorem as Γ is invertible and $\tau_{0,i,0,j}$ is bounded by Assumptions 2.1(ii) and 2.1(iv), respectively, and due to $\sum_{i=1}^{\infty}\|C_i\otimes I_K\| < \infty$. Now, we consider (c). The second part of $A - A_m$ in (A.2) is zero and it suffices to show (c) for the first part ignoring the factor $\widehat{Q}_T$. Let $\lambda\in\mathbb{R}^{K^2p}$ and $\delta > 0$, then (c) follows with Markov inequality from

$$\begin{aligned} & P\left(\left|\sum_{j=m+1}^{\infty}\lambda'(C_j\otimes I_K)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\operatorname{vec}(u_tu_{t-j}')\right| > \delta\right) \\ \leq\; & \frac{1}{\delta^2 T}E\left(\left|\sum_{j=m+1}^{\infty}\lambda'(C_j\otimes I_K)\sum_{t=1}^{T}\operatorname{vec}(u_tu_{t-j}')\right|^2\right) \\ =\; & \frac{1}{\delta^2}\sum_{j_1,j_2=m+1}^{\infty}\lambda'(C_{j_1}\otimes I_K)\left\{\frac{1}{T}\sum_{t_1,t_2=1}^{T}E\left(\operatorname{vec}(u_{t_1}u_{t_1-j_1}')\operatorname{vec}(u_{t_2}u_{t_2-j_2}')'\right)\right\}(C_{j_2}\otimes I_K)'\lambda \\ =\; & \frac{1}{\delta^2}\sum_{j_1,j_2=m+1}^{\infty}\lambda'(C_{j_1}\otimes I_K)\tau_{0,j_1,0,j_2}(C_{j_2}\otimes I_K)'\lambda \\ \underset{m\to\infty}{\to}\; & 0 \end{aligned}$$

because of $E\left(\operatorname{vec}(u_{t_1}u_{t_1-j_1}')\operatorname{vec}(u_{t_2}u_{t_2-j_2}')'\right) = \tau_{0,j_1,0,j_2}1(t_1=t_2)$ by Assumption 2.1(ii) and by $\|V^{(1,1)}\| < \infty$. □

Lemma A.1 (CLTs for innovations)**.**

(i) Let $W_{t,m}^{(1)} = (\operatorname{vec}(u_tu_{t-1}')',\ldots,\operatorname{vec}(u_tu_{t-m}')')'$. *Under Assumption 2.1, it holds*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}W_{t,m}^{(1)} \overset{\mathcal{D}}{\to} \mathcal{N}(0,\Omega_m^{(1,1)}),$$

where $\Omega_m^{(1,1)} = (\tau_{0,i,0,j})_{i,j=1,\ldots,K^2m}$ *is a block matrix and* $\tau_{0,i,0,j}$ *is defined in (2.7).*

(ii) Let $W_{t,m}^{(2)} = L_K\{\operatorname{vec}(u_tu_t')-\operatorname{vec}(\Sigma_u)\} = \operatorname{vech}(u_tu_t')-\operatorname{vech}(\Sigma_u)$ *and define* $W_{t,m} = (W_{t,m}^{(1)\prime}, W_{t,m}^{(2)\prime})'$.

If Assumptions 2.1 and 2.2 hold, we have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} W_{t,m} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Omega_m),$$

where Ω_m *is a* $(K^2m+\widetilde{K}\times K^2m+\widetilde{K})$ *block matrix*

$$\Omega_m = \begin{pmatrix} \Omega_m^{(1,1)} & \Omega_m^{(1,2)} \\ \Omega_m^{(2,1)} & \Omega^{(2,2)} \end{pmatrix}$$

with the $(\widetilde{K}\times\widetilde{K})$ *and* $(\widetilde{K}\times K^2m)$ *matrices*

$$\Omega^{(2,2)} = \sum_{h=-\infty}^{\infty} L_K\left\{\tau_{0,0,h,h} - \operatorname{vec}(\Sigma_u)\operatorname{vec}(\Sigma_u)'\right\}L_K', \tag{A.3}$$

$$\Omega_m^{(2,1)} = \sum_{h=0}^{\infty} L_K\left(\tau_{0,0,h,h+1},\ldots,\tau_{0,0,h,h+m}\right), \tag{A.4}$$

respectively.

Proof.
(i) Let $\lambda\in\mathbb{R}^{K^2m}$ such that $\lambda'\lambda=1$, define $V_{t,m}^{(1)}=\lambda' W_{t,m}^{(1)}$ and by Cramér-Wold device, it suffices to show that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} V_{t,m}^{(1)} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\lambda'\Omega_m^{(1,1)}\lambda).$$

Noting that $\{V_{t,m}^{(1)},\mathcal{F}_t\}$ is an m.d.s., we have to check e.g. the conditions of Theorem 24.3 in Davidson (1994), i.e.

(a) $\frac{1}{T}\sum_{t=1}^{n}\left\{\left(V_{t,m}^{(1)}\right)^2 - E\left((V_{t,m}^{(1)})^2\right)\right\} = \frac{1}{T}\sum_{t=1}^{n}\left\{\left(V_{t,m}^{(1)}\right)^2 - \lambda'\Omega_m^{(1,1)}\lambda\right\} \xrightarrow{P} 0$

(b) $\frac{1}{\sqrt{T}}\max_{t=1,\ldots,T}|V_{t,m}^{(1)}| \xrightarrow{P} 0$

Representing the expression in (a) above as

$$\frac{1}{T}\sum_{t=1}^{n}\left\{\left(V_{t,m}^{(1)}\right)^2 - E\left((V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right)\right\} + \frac{1}{T}\sum_{t=1}^{n}\left\{E\left((V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right) - \lambda'\mathbf{\Omega}_m^{(1,1)}\lambda\right\} = A_1 + A_2,$$

we can show that $\left(V_{t,m}^{(1)}\right)^2 - E\left((V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right)$ is an L_1-mixingale. This follows from

$$\begin{aligned} E\left|E\left\{\left(V_{t,m}^{(1)}\right)^2 - E\left[(V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right]\middle|\mathcal{F}_{t-k}\right\}\right| &= \begin{cases} E\left|\left(V_{t,m}^{(1)}\right)^2 - E\left[(V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right]\right|, & k=0 \\ 0, & k\geq 1 \end{cases} \\ &\leq c_t\psi_k, \end{aligned}$$

with $c_t = E\left|\left(V_{t,m}^{(1)}\right)^2 - E\left[(V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right]\right|$ and $\psi_0 = 1$ and $\psi_k = 0$ for $k \geq 1$, from

$$E\left|\left(V_{t,m}^{(1)}\right)^2 - E\left[(V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right] - E\left[\left(V_{t,m}^{(1)}\right)^2 - E\left[(V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right]\middle|\mathcal{F}_{t+k}\right]\right| = 0$$

and due to $\psi_k, c_t \geq 0$ for all $k, t \geq 0$ and $\psi_k \to 0$ as $k \to \infty$. To apply the LLN in Theorem 1(a) of Andrews (1988), we have to show uniform integrability and $\overline{\lim}_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} c_t < \infty$. The first condition follows due to Proposition 7.7(a) in Hamilton (1994) from

$$E\left|\left(V_{t,m}^{(1)}\right)^2 - E\left((V_{t,m}^{(1)})^2|\mathcal{F}_{t-1}\right)\right| \leq 2E\left(V_{t,m}^{(1)}\right)^2$$

and, by applications of Minkowski and Cauchy-Schwarz inequalities, for some $r > 1$, from

$$\begin{aligned}
E\left(V_{t,m}^{(1)}\right)^{2r} &= E\left|\lambda' W_{t,m}^{(1)}\right|^{2r} = E\left[\left|\sum_{j=1}^{K^2 m}\lambda_j W_{t,m,j}^{(1)}\right|^{2r}\right] = \left(\left(E\left[\left|\sum_{j=1}^{K^2 m}\lambda_j W_{t,m,j}^{(1)}\right|^{2r}\right]\right)^{\frac{1}{2r}}\right)^{2r} \\
&= \left(\left\|\sum_{j=1}^{K^2 m}\lambda_j W_{t,m,j}^{(1)}\right\|_{2r}\right)^{2r} \leq \left(\sum_{j=1}^{K^2 m}|\lambda_j|\left\|W_{t,m,j}^{(1)}\right\|_{2r}\right)^{2r} \leq \left(\sum_{j=1}^{K^2 m}|\lambda_j|\left\|W_{t,m,j}^{(1)}\right\|_{2r}\right)^{2r} \\
&\leq \left\{\left(\sum_{j=1}^{K^2 m}|\lambda_j|^2\right)\left(\sum_{j=1}^{K^2 m}\left\|W_{t,m,j}^{(1)}\right\|_{2r}\right)\right\}^r \leq (Km^2)^r \sup_j \left\|W_{t,m,j}^{(1)}\right\|_{2r}^{2r} \\
&< \infty, \qquad\qquad (A.5)
\end{aligned}$$

by Assumption 2.1(vi), where similar arguments yield also the second condition. In the above, we use the notation $\|A\|_p = (E(|A|_p^p))^{1/p}$. Together this leads to $A_1 \to 0$ and for A_2, we get

$$A_2 = \lambda'\left\{\frac{1}{T}\sum_{t=1}^{n} E\left(W_{t,m}^{(1)}W_{t,m}^{(1)\prime}|\mathcal{F}_{t-1}\right) - \mathbf{\Omega}_m^{(1,1)}\right\}\lambda \xrightarrow{P} 0$$

because the $(K^2 m \times K^2 m)$ matrix in parentheses above converges to zero in probability by Assumption 2.1(v). Finally, part (b) follows by Markov inequality from

$$\begin{aligned}
P\left(\frac{1}{\sqrt{T}}\max_{1\leq t\leq T}|V_{t,m}^{(1)}| > \delta\right) &\leq \sum_{t=1}^{T} P\left(|V_{t,m}^{(1)}| > \delta\sqrt{T}\right) \leq \frac{1}{\delta^{2r}T^{r-1}}\left(\frac{1}{T}\sum_{t=1}^{T} E|V_{t,m}^{(1)}|^{2r}\right) \\
&= O\left(\frac{1}{T^{r-1}}\right) = o(1)
\end{aligned}$$

for any $\delta > 0$ and by the uniform boundedness in (A.5) for some $r > 1$.

(ii) To prove the CLT for the sequence $(W_{t,m}, t \in \mathbb{Z})$ under strict stationarity and α-mixing assumptions on the innovations process $(u_t, t \in \mathbb{Z})$, we use Theorem A.8 in Lahiri (2003). Similar to the proof of part (i), let $\lambda \in \mathbb{R}^{K^2 m + \widetilde{K}}$ such that $\lambda'\lambda = 1$ and define $V_{t,m} = \lambda' W_{t,m}$. By Assumption 2.2(i), the (univariate) process $(V_{t,m}, t \in \mathbb{Z})$ is strictly stationary for all $m \in \mathbb{N}$ such that $E(V_{t,m})^{2+\delta} < \infty$ holds for some $\delta > 0$ due to Assumption 2.1(vi). Furthermore, Assumption 2.2(ii) together with Theorem 14.3 in Davidson (1994) imply that its α-mixing coefficients $(\alpha_{V,m}(n), n \in \mathbb{N})$ decay at the same rate as $(\alpha_u(n), n \in \mathbb{N})$ of the process $(u_t, t \in \mathbb{Z})$. That is, we have $\sum_{n=1}^{\infty}(\alpha_{W,m}(n))^{\delta/(2+\delta)} < \infty$, which together with Assumption 2.2(iii) matches the requirements for Theorem A.8 in Lahiri (2003). It remains to evaluate the limiting variance of $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} W_{t,m}$ and to derive the asymptotics. The variance corresponding to the first part $W_{t,m}^{(1)}$ has already been established above and it remains to check the variance of $W_{t,m}^{(2)}$ and the covariance between these two. For the variance, we have

$$\begin{aligned} \mathrm{Var}(W_{t,m}^{(2)}) &= L_K \sum_{h=-(T-1)}^{T-1} \frac{1}{T} \sum_{t=\max(1,1+h)}^{\min(T,T+h)} \mathrm{Cov}\left(\mathrm{vec}(u_t u_t'), \mathrm{vec}(u_{t-h} u_{t-h}')\right) L_K' \\ &\to \sum_{h=-\infty}^{\infty} L_K \, \mathrm{Cov}\left(\mathrm{vec}(u_t u_t'), \mathrm{vec}(u_{t-h} u_{t-h}')\right) L_K' \quad \text{as } T \to \infty \\ &= \sum_{h=-\infty}^{\infty} L_K \left\{\tau_{0,0,h,h} - \mathrm{vec}(\Sigma_u)\,\mathrm{vec}(\Sigma_u)'\right\} L_K' \end{aligned}$$

and, similarly, for the covariances, we get

$$\begin{aligned} &\mathrm{Cov}(W_{t,m}^{(2)}, W_{t,m}^{(1)}) \\ = &\sum_{h=-(T-1)}^{T-1} \frac{1}{T} \sum_{t=\max(1,1+h)}^{\min(T,T+h)} L_K \, \mathrm{Cov}\left(\mathrm{vec}(u_t u_t'), (\mathrm{vec}(u_{t-h} u_{t-h-1})', \ldots, \mathrm{vec}(u_{t-h} u_{t-h-m}')')'\right) \\ \to &\sum_{h=-\infty}^{\infty} L_K \, \mathrm{Cov}\left(\mathrm{vec}(u_t u_t'), (\mathrm{vec}(u_{t-h} u_{t-h-1})', \ldots, \mathrm{vec}(u_{t-h} u_{t-h-m}')')'\right) \quad \text{as } T \to \infty \\ = &\sum_{h=0}^{\infty} L_K \left(\tau_{0,0,h,h+1}, \cdots, \tau_{0,0,h,h+m}\right), \end{aligned}$$

where we have used $E(\mathrm{vec}(u_{t-h} u_{t-h-j})') = 0$ for all $j \geq 1$ and $\tau_{0,0,h,h+j} = 0$ for all $h < 0$ and $j \geq 0$. □

Lemma A.2 (Convergence of $\frac{1}{T}ZZ'$).
Under Assumption 2.1, it holds $\frac{1}{T}ZZ' \to \Gamma$ in probability. In particular, we have $(\frac{1}{T}ZZ')^{-1} \otimes I_K \to \Gamma^{-1} \otimes I_K$ as well as $\widehat{Q}_T \to Q$ in probability, respectively.

Proof.
It holds

$$\frac{1}{T}ZZ' = \frac{1}{T}\sum_{t=1}^{T} Z_{t-1} Z_{t-1}' = \frac{1}{T}\sum_{t=1}^{T} \sum_{j_1,j_2=1}^{\infty} C_{j_1} u_{t-j_1} u_{t-j_2}' C_{j_2}'$$

with mean $\sum_{j=1}^{\infty} C_j \Sigma_u C_j' < \infty$. By arguments similar to those used in the proof of Lemma A.1 to show uniform integrable L_1-mixingales, we get the claimed result from Assumption 2.1. Compare also the proof of Theorem 3.1 in Gonçalves & Kilian (2004) for details in the univariate setup. As Γ is non-singular by positive definiteness of Σ_u and by the stability condition $det(A(z)) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$, we also get $(\frac{1}{T}ZZ')^{-1} \otimes I_K \to \Gamma^{-1} \otimes I_K$ and $\widehat{Q}_T \to Q$ in probability, respectively. □

A.2 Proof of Theorem 5.1

By Polya's Theorem and by Lemma A.3, it suffices to show that $\sqrt{T}((\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}})', (\widetilde{\boldsymbol{\sigma}}^* - \widetilde{\boldsymbol{\sigma}})')'$ converges in distribution w.r.t. measure P^* to $\mathcal{N}(0, V)$ as obtained in Theorem 3.1, where $\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}} := ((\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\widetilde{Z}^* \otimes I_K)\widetilde{\mathbf{u}}^*$ and $\widetilde{\boldsymbol{\sigma}}^* = \text{vech}(\widetilde{\Sigma}_u^*)$ with $\widetilde{\Sigma}_u^* = \frac{1}{T}\sum_{t=1}^T \widetilde{u}_t^*\widetilde{u}_t^{*\prime}$. Here, pre-sample values $\widetilde{y}_{-p+1}^*, \ldots, \widetilde{y}_0^*$ are set to zero and $\widetilde{y}_1^*, \ldots, \widetilde{y}_T^*$ is generated according to

$$\widetilde{y}_t^* = A_1\widetilde{y}_{t-1}^* + \cdots + A_p\widetilde{y}_{t-p}^* + \widetilde{u}_t^*,$$

where $\widetilde{u}_1^*, \ldots, \widetilde{u}_T^*$ is an analogously drawn version of $u_1^*, \ldots, u_T^*$ as described in Steps 2. and 3. of the bootstrap procedure in Section 5, but from $u_1, \ldots, u_T$ instead of $\widehat{u}_1, \ldots, \widehat{u}_T$. Further, we use the notation

$$\begin{aligned} \widetilde{Z}_t^* &= \text{vec}(\widetilde{y}_t^*, \ldots, \widetilde{y}_{t-p+1}^*) \quad (Kp \times 1) \\ \widetilde{Z}^* &= (\widetilde{Z}_0^*, \ldots, \widetilde{Z}_{T-1}^*) \quad (Kp \times T) \\ \widetilde{\mathbf{u}}^* &= \text{vec}(\widetilde{u}_1^*, \ldots, \widetilde{u}_T^*) \quad (KT \times 1). \end{aligned}$$

We get analogue to (A.2) the representation

$$\begin{aligned} \sqrt{T}\begin{pmatrix} \widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\sigma}}^* - \widetilde{\boldsymbol{\sigma}} \end{pmatrix} &= \begin{pmatrix} \left\{(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1} \otimes I_K\right\} \frac{1}{\sqrt{T}}\sum_{t=1}^T\sum_{j=1}^{t-1}(C_j \otimes I_K)\left\{\text{vec}(\widetilde{u}_t^*\widetilde{u}_{t-j}^{*\prime})\right\} \\ \left(\frac{1}{\sqrt{T}}\sum_{t=1}^T L_K\left\{\text{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}) - \text{vec}(u_tu_t')\right\}\right) \end{pmatrix} \\ &= \begin{pmatrix} \left\{(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1} \otimes I_K\right\} \frac{1}{\sqrt{T}}\sum_{j=1}^{T-1}(C_j \otimes I_K)\sum_{t=j+1}^T\left\{\text{vec}(\widetilde{u}_t^*\widetilde{u}_{t-j}^{*\prime})\right\} \\ \left(\frac{1}{\sqrt{T}}\sum_{t=1}^T L_K\left\{\text{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}) - \text{vec}(u_tu_t')\right\}\right) \end{pmatrix} \quad \text{(A.6)} \\ &= A_m^* + (A^* - A_m^*), \end{aligned}$$

where A^* denotes the right-hand side of (A.6) and A_m^* is the same expression, but with $\sum_{j=1}^{T-1}$ replaced by $\sum_{j=1}^m$ for some fixed $m \in \mathbb{N}$. In the following, we make use of Proposition 6.3.9 of Brockwell & Davis (1991) and it suffices to show

(a) $A_m^* \overset{\mathcal{D}}{\to} \mathcal{N}(0, V_m)$ in probability as $T \to \infty$

(b) $V_m \to V$ as $m \to \infty$

(c) $\forall\, \delta > 0 :\ \lim_{m\to\infty}\limsup_{T\to\infty} P^*(|A^* - A_m^*|_1 > \delta) = 0$ in probability.

To prove (a), we can write

$$\begin{aligned} A_m^* \;=\; & \begin{pmatrix} (\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1} \otimes I_K & O_{K^2p\times\widetilde{K}} \\ O_{\widetilde{K}\times K^2p} & I_{\widetilde{K}} \end{pmatrix}\begin{pmatrix} C_1 \otimes I_K & \cdots & C_m \otimes I_K & O_{K^2p\times\widetilde{K}} \\ O_{\widetilde{K}\times K^2} & \cdots & O_{\widetilde{K}\times K^2} & I_{\widetilde{K}} \end{pmatrix} \\ & \times\frac{1}{\sqrt{T}}\sum_{t=1}^T\begin{pmatrix} \text{vec}(\widetilde{u}_t^*\widetilde{u}_{t-1}^{*\prime}) \\ \vdots \\ \text{vec}(\widetilde{u}_t^*\widetilde{u}_{t-m}^{*\prime}) \\ L_K\left\{\text{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}) - \text{vec}(u_tu_t')\right\} \end{pmatrix} \end{aligned}$$

$$= \widetilde{Q}_T^* R_m \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \widetilde{W}_{t,m}^*$$

as $u_t^* := 0$ for $t < 0$ and with an obvious notation for the $(K^2p + \widetilde{K} \times K^2p + \widetilde{K})$ matrix $\widetilde{Q}_T^*$, the $(K^2p + \widetilde{K} \times K^2m + \widetilde{K})$ matrix R_m and the $(K^2m + \widetilde{K})$-dimensional vector $\widetilde{W}_{t,m}^*$. By Lemma A.4, we have that $\widetilde{Q}_T^* \to Q$ with respect to P^* and from Lemma A.5, we get the CLT required for part (a). As (b) follows exactly as in the proof of Theorem 3.1, it remains to show part (c), where the factor $\widetilde{Q}_T^*$ can be ignored and the second part of $A^* - A_m^*$ is zero. Let $\lambda \in \mathbb{R}^{K^2p}$ and $\delta > 0$, then we get by the Markov inequality

$$\begin{aligned} & P^*\left(\left|\sum_{j=m+1}^{T-1} \lambda'(C_j \otimes I_K) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \operatorname{vec}(\widetilde{u}_t^* \widetilde{u}_{t-j}^{*\prime})\right| > \delta\right) \\ \leq\ & \frac{1}{\delta^2} \sum_{j_1,j_2=m+1}^{T-1} \lambda'(C_{j_1} \otimes I_K) \left\{\frac{1}{T} \sum_{t_1,t_2=1}^{T} E^*\left(\operatorname{vec}(\widetilde{u}_{t_1}^* \widetilde{u}_{t_1-j_1}^{*\prime}) \operatorname{vec}(\widetilde{u}_{t_2}^* \widetilde{u}_{t_2-j_2}^{*\prime})'\right)\right\} (C_{j_2} \otimes I_K)'\lambda \\ =:\ & R_{m,T} \end{aligned}$$

and by assuming absolute summability for the cumulants of the innovations up to order eight in Assumption 5.1, it is straightforward, but tedious to show that

$$E\left(R_{m,T}\right) \underset{T\to\infty}{\to} \tfrac{1}{\delta^2} \textstyle\sum_{j_1,j_2=m+1}^{\infty} \lambda'(C_{j_1} \otimes I_K)\tau_{0,j_1,0,j_2}(C_{j_2} \otimes I_K)'\lambda$$

as well as $E\left(|R_{m,T} - E\left(R_{m,T}\right)|_2^2\right) = o(1)$, such that

$$\frac{1}{\delta^2} \sum_{j_1,j_2=m+1}^{\infty} \lambda'(C_{j_1} \otimes I_K)\tau_{0,j_1,0,j_2}(C_{j_2} \otimes I_K)'\lambda \underset{m\to\infty}{\longrightarrow} 0$$

proves part (c), which concludes the proof. □

Lemma A.3 (Equivalence of bootstrap estimators)**.**
Under the assumptions of Theorem 5.1, we have

$$\sqrt{T}\left((\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) - (\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}})\right) = o_{P^*}(1) \quad \textit{and} \quad \sqrt{T}\left((\widehat{\boldsymbol{\sigma}}^* - \widehat{\boldsymbol{\sigma}}) - (\widetilde{\boldsymbol{\sigma}}^* - \widetilde{\boldsymbol{\sigma}})\right) = o_{P^*}(1).$$

Proof.
For simplicity, we assume throughout the proof that $T = N\ell$ holds and we show only the more complicated claim $\sqrt{T}((\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) - (\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}})) = o_{P^*}(1)$. The second assertion then follows by the same arguments as well. First, we have

$$\begin{aligned} \sqrt{T}\left((\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) - (\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}})\right) &= \left((\frac{1}{T}Z^*Z^{*\prime})^{-1} \otimes I_K\right) \frac{1}{\sqrt{T}} \left\{(Z^* \otimes I_K)\mathbf{u}^* - (\widetilde{Z}^* \otimes I_K)\widetilde{\mathbf{u}}^*\right\} \\ &\quad + \left(\left\{(\frac{1}{T}Z^*Z^{*\prime})^{-1} - (\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\right\} \otimes I_K\right) \frac{1}{\sqrt{T}}(\widetilde{Z}^* \otimes I_K)\widetilde{\mathbf{u}}^* \\ &= \left((\frac{1}{T}Z^*Z^{*\prime})^{-1} \otimes I_K\right) A_1^* + A_2^* \frac{1}{\sqrt{T}}(\widetilde{Z}^* \otimes I_K)\widetilde{\mathbf{u}}^* \end{aligned}$$

with an obvious notation for A_1^* and A_2^*. As $A_2^* = o_{P^*}(1)$, boundedness in probability of $\left((\frac{1}{T}Z^*Z^{*\prime})^{-1} \otimes I_K\right)$ and of $\frac{1}{\sqrt{T}}(\widetilde{Z}^* \otimes I_K)\widetilde{\mathbf{u}}^*$ follows by very similar arguments, we focus only on the proof of $A_1^* = o_{P^*}(1)$ in the following. Similar to (A.1), we will make use of

$$
\begin{aligned}
Z_{t-1}^* &= \begin{pmatrix} y_{t-1}^* \\ \vdots \\ y_{t-p}^* \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{t-1+1} \widehat{\Phi}_j u_{t-1-j}^* \\ \vdots \\ \sum_{j=0}^{t-p+1} \widehat{\Phi}_j u_{t-p-j}^* \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{t-1} \widehat{\Phi}_{j-1} u_{t-j}^* \\ \vdots \\ \sum_{j=p}^{t-1} \widehat{\Phi}_{j-p} u_{t-j}^* \end{pmatrix} \\
&= \begin{pmatrix} \sum_{j=1}^{t-1} \widehat{\Phi}_{j-1} u_{t-j}^* \\ \vdots \\ \sum_{j=1}^{t-1} \widehat{\Phi}_{j-p} u_{t-j}^* \end{pmatrix} = \sum_{j=1}^{t-1} \widehat{C}_j u_{t-j}^*, \qquad \text{(A.7)}
\end{aligned}
$$

where $y_t^* = \sum_{j=0}^{t-1} \widehat{\Phi}_j u_{t-j}^*$, $t = 1, \ldots, T$ with $y_{p-1}^*, \ldots, y_0^* = 0$ and $\widehat{\Phi}_0 = 1$, $\widehat{\Phi}_j = 0$ for $j < 0$ as well as $\widehat{\Phi}_j = \sum_{i=1}^{\min(j,p)} \widehat{A}_i \widehat{\Phi}_{j-i}$ for $j \in \mathbb{N}$. Analogously, we have $\widetilde{Z}_{t-1}^* = \sum_{j=1}^{t-1} C_j \widetilde{u}_{t-j}^*$. Further, we get

$$
A_1^* = \frac{1}{\sqrt{T}}(Z^* \otimes I_K)\{\mathbf{u}^* - \widetilde{\mathbf{u}}^*\} + \frac{1}{\sqrt{T}}\left(\left\{Z^* - \widetilde{Z}^*\right\} \otimes I_K\right)\widetilde{\mathbf{u}}^* = A_{11}^* + A_{12}^*
$$

and, by omitting the details for A_{11}^* and continuing with the slightly more complicated expression A_{12}^*, we get

$$
\begin{aligned}
A_{12}^* &= \sum_{j=1}^{T-1} \left(\widehat{C}_j \otimes I_K\right) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \operatorname{vec}(\widetilde{u}_t^* \{u_{t-j}^{*\prime} - \widetilde{u}_{t-j}^{*\prime}\}) \\
&\quad + \sum_{j=1}^{T-1} \left(\left\{\widehat{C}_j - C_j\right\} \otimes I_K\right) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \operatorname{vec}(\widetilde{u}_t^* \widetilde{u}_{t-j}^{*\prime}) \\
&= A_{121}^* + A_{122}^*.
\end{aligned}
$$

Now, we consider A_{122}^* first. By splitting-up the sums over j and t corresponding to the bootstrap blocks, we get

$$
\begin{aligned}
&E^*(A_{122}^* A_{122}^{*\prime}) \\
&= \frac{1}{T} \sum_{r_1,r_2=1}^{N} \sum_{s_1,s_2=1}^{\ell} \sum_{v_1,v_2=0}^{N} \sum_{w_1=\max(s_1+(v_1-1)\ell,1)}^{\min(s_1+v_1\ell-1,T-1)} \sum_{w_2=\max(s_2+(v_2-1)\ell,1)}^{\min(s_2+v_2\ell-1,T-1)} \left(\left\{\widehat{C}_{w_1} - C_{w_1}\right\} \otimes I_K\right) \\
&\quad \times E^*\left(\operatorname{vec}(\widetilde{u}_{s_1+(r_1-1)\ell}^* \widetilde{u}_{s_1+(r_1-1)\ell-w_1}^{*\prime}) \operatorname{vec}(\widetilde{u}_{s_2+(r_2-1)\ell}^* \widetilde{u}_{s_2+(r_2-1)\ell-w_2}^{*\prime})'\right) \left(\left\{\widehat{C}_{w_2} - C_{w_2}\right\} \otimes I_K\right)',
\end{aligned}
$$

where the conditional expectation on the last right-hand side does not vanish for the three cases (i) $r_1 = r_2$, $v_1 = v_2 = 0$ (all in one block), (ii) $r_1 = r_2$, $v_1 = v_2 \geq 1$ (first and third and second and fourth in the same block, respectively), (iii) $r_1 \neq r_2$, $v_1 = v_2 = 0$ (first and second and third and fourth in the same block, respectively). By taking the Frobenius norm of $E^*(A_{122}^* A_{122}^{*\prime})$ and using the triangle inequality, case (i) can be bounded by

$$
\begin{aligned}
& K^2 \frac{1}{\ell} \sum_{s_1,s_2=1}^{l} \sum_{w_1=1}^{s_1} \sum_{w_2=1}^{s_2} |\widehat{C}_{w_1} - C_{w_1}|_2 |\widehat{C}_{w_2} - C_{w_2}|_2 \\
& \qquad \times \frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} |\operatorname{vec}(u^c_{t+s_1} u^{c\prime}_{t+s_1-w_1}) \operatorname{vec}(u^c_{t+s_2} u^{c\prime}_{t+s_2-w_2})'|_2 \\
= \;& O_P\left(\frac{1}{T}\right) \\
& \times \left(\frac{1}{\ell} \sum_{s_1,s_2=1}^{\ell} \sum_{w_1=1}^{s_1} \sum_{w_2=1}^{s_2} d^{-w_1-w_2} \frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} |\operatorname{vec}(u^c_{t+s_1} u^{c\prime}_{t+s_1-w_1}) \operatorname{vec}(u^c_{t+s_2} u^{c\prime}_{t+s_2-w_2})'|_2 \right) \\
= \;& o_P(1),
\end{aligned}
$$

where $\ell^3/T \to 0$ and $u^c_{t+s} := u_{t+s} - \frac{1}{T-l+1}\sum_{\tau=0}^{T-l} u_{\tau+s}$, $E|\operatorname{vec}(u^c_{t+s_1} u^{c\prime}_{t+s_1-w_1}) \operatorname{vec}(u^c_{t+s_2} u^{c\prime}_{t+s_2-w_2})'|_2 \leq \Delta < \infty$ by Assumption 5.1 have been used and that there exists a constant $d > 1$ such that

$$
\sqrt{T} \sup_{j \in \mathbb{N}} d^j |\widehat{C}_j - C_j|_2 = O_P(1)
$$

holds, cf. Kreiss & Franke (1992) for a proof of the univariate case. Cases (ii) and (iii) can be treated exactly the same. Now turn to A^*_{121}. Similar to the above, we get

$$
\begin{aligned}
E^*(A^*_{121} A^{*\prime}_{121}) \;=\;& \frac{1}{T} \sum_{r_1,r_2=1}^{N} \sum_{s_1,s_2=1}^{\ell} \sum_{v_1,v_2=0}^{N} \sum_{w_1=\max(s_1+(v_1-1)\ell,1)}^{\min(s_1+v_1\ell-1,T-1)} \sum_{w_2=\max(s_2+(v_2-1)\ell,1)}^{\min(s_2+v_2\ell-1,T-1)} (\widehat{C}_{w_1} \otimes I_K) \\
& \times E^* \Big(\operatorname{vec}(\widetilde{u}^*_{s_1+(r_1-1)\ell} (u^*_{s_1+(r_1-1)\ell-w_1} - \widetilde{u}^*_{s_1+(r_1-1)\ell-w_1})') \\
& \times \operatorname{vec}(\widetilde{u}^*_{s_2+(r_2-1)\ell} (u^*_{s_2+(r_2-1)\ell-w_2} - \widetilde{u}^*_{s_2+(r_2-1)\ell-w_2})')' \Big) (\widehat{C}_{w_2} \otimes I_K)',
\end{aligned}
$$

and again the three cases (i) – (iii) as described above do not vanish exactly. By using $\widehat{u}_t - u_t = (A_1 - \widehat{A}_1)y_{t-1} + \cdots + (A_p - \widehat{A}_p)y_{t-p} =: (B - \widehat{B})Z_{t-1}$ and $\sqrt{T}(B - \widehat{B}) = O_P(1)$, we get that the (Frobenius) norm of case (i) can be bounded by

$$
\begin{aligned}
& K^4 |B - \widehat{B}|_2^2 \frac{1}{\ell} \sum_{s_1,s_2=1}^{\ell} \sum_{w_1=1}^{s_1} \sum_{w_2=1}^{s_2} |\widehat{C}_{w_1}|_2 |\widehat{C}_{w_2}|_2 \\
& \qquad \times \frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} |\operatorname{vec}(u^c_{t+s_1} Z^{c\prime}_{t+s_1-w_1-1}) \operatorname{vec}(u^c_{t+s_2} Z^{c\prime}_{t+s_2-w_2-1})'|_2 \\
= \;& o_P(1),
\end{aligned}
$$

where $Z^c_{t+s-1} := Z_{t+s-1} - \frac{1}{T-\ell+1}\sum_{\tau=0}^{T-\ell} Z_{\tau+s-1}$ and by similar arguments as used above for showing $A^*_{122} = o_{P^*}(1)$. □

Lemma A.4 (Convergence of $\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime}$)**.**
Under the assumptions of Theorem 5.1, it holds $\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime} \to \Gamma$ in probability w.r.t. P^*. In particular, we have $(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1} \otimes I_K \to \Gamma^{-1} \otimes I_K$ as well as $\widetilde{Q}^*_T \to Q$ in probability with respect to P^*.

Proof.
Insertion of $\widetilde{Z}^*_{t-1} = \sum_{j=1}^{t-1} C_j \widetilde{u}^*_{t-j}$ leads to

$$\begin{aligned}
\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime} &= \frac{1}{T}\sum_{t=1}^{T}\widetilde{Z}^*_{t-1}\widetilde{Z}^{*\prime}_{t-1} = \frac{1}{T}\sum_{t=1}^{T}\sum_{j_1,j_2=1}^{t-1} C_{j_1}\widetilde{u}^*_{t-j_1}\widetilde{u}^{*\prime}_{t-j_2}C'_{j_2} \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{h=-(t-2)}^{t-2}\sum_{s=\max(1,1-h)}^{\min(t-1,t-1-h)} C_{s+h}\widetilde{u}^*_{t-(s+h)}\widetilde{u}^{*\prime}_{t-s}C'_s \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{h=1}^{t-2}\sum_{s=1}^{t-1-h} C_{s+h}\widetilde{u}^*_{t-(s+h)}\widetilde{u}^{*\prime}_{t-s}C'_s + \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1} C_s\widetilde{u}^*_{t-s}\widetilde{u}^{*\prime}_{t-s}C'_s \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{h=-(t-2)}^{-1}\sum_{s=1-h}^{t-1} C_{s+h}\widetilde{u}^*_{t-(s+h)}\widetilde{u}^{*\prime}_{t-s}C'_s \\
&= A^*_1 + A^*_2 + A^*_3
\end{aligned}$$

with an obvious notation for A^*_1, A^*_2 and A^*_3. In the following, we show that (a) $A^*_2 \to \Gamma$ and (b) $A^*_1 \to 0$ and $A^*_3 \to 0$ with respect to P^*, respectively. By using Proposition 6.3.9 in Brockwell & Davis (1991), we consider A^*_2 first which, for some fixed $m \in \mathbb{N}$, can be expressed as

$$\begin{aligned}
A^*_2 &= \sum_{s=1}^{T-1}\frac{1}{T}\sum_{t=s+1}^{T} C_s\widetilde{u}^*_{t-s}\widetilde{u}^{*\prime}_{t-s}C'_s \\
&= \sum_{s=1}^{m-1} C_s\left(\frac{1}{T}\sum_{t=s+1}^{T}\widetilde{u}^*_{t-s}\widetilde{u}^{*\prime}_{t-s}\right)C'_s + \sum_{s=m}^{T-1} C_s\left(\frac{1}{T}\sum_{t=s+1}^{T}\widetilde{u}^*_{t-s}\widetilde{u}^{*\prime}_{t-s}\right)C'_s \\
&= A^{*m}_2 + (A^{*m}_2 - A^*_2).
\end{aligned}$$

Under the imposed summability conditions for the cumulants in Assumption 5.1, it is straightforward to show that $\frac{1}{T}\sum_{t=s+1}^{T}\widetilde{u}^*_{t-s}\widetilde{u}^{*\prime}_{t-s} \to \Sigma_u$ for all $s = 1, \ldots, m$ and this leads to $A^{*m}_2 \to \Gamma^m = \sum_{s=1}^{m-1} C_s\Sigma_u C'_s$ with respect to P^*, respectively, as $T \to \infty$ and also to $\Gamma^m \to \Gamma$ as $m \to \infty$. Further, we have

$$\begin{aligned}
E^*\left(|A^{*m}_2 - A^*_2|_1\right) &\leq \sum_{j=m}^{T-1}\sum_{r,s=1}^{Kp}\sum_{f,g=1}^{K}|C_j(r,f)|\left(\frac{1}{T}\sum_{t=j+1}^{T}E^*\left(|\widetilde{u}^*_{t-j,f}\widetilde{u}^*_{t-j,g}|\right)\right)|C_j(s,g)| \\
&\leq \sum_{j=m}^{T-1}\sum_{r,s=1}^{Kp}\sum_{f,g=1}^{K}|C_j(r,f)|\left(\frac{1}{T}\sum_{t=1}^{T}E^*\left(|\widetilde{u}^*_{t,f}\widetilde{u}^*_{t,g}|\right)\right)|C_j(s,g)| \\
&\leq \left(\sum_{j=m}^{T-1}|C_j|_1^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{f=1}^{K}E^*\left(\widetilde{u}^{*2}_{t,f}\right)\right)
\end{aligned}$$

due to $|u^*_{t,f}u^*_{t,g}| \leq \frac{1}{2}\left(u^{*2}_{t,f} + u^{*2}_{t,g}\right) \leq \sum_{f=1}^{K}u^{*2}_{t,f}$. Again from Assumption 5.1, we get easily that the second factor on the last right-hand side is bounded in probability and this leads to

$$E^*\left(|A^{*m}_2 - A^*_2|_1\right) \leq \left(\sum_{j=m}^{\infty}|C_j|_1^2\right)O_P(1) \to 0$$

as $m \to \infty$, which completes the proof of $A_2^* \to \Gamma$. For proving (b), it suffices to consider A_1^* only and A_3^* can be treated completely analogue. Similar arguments as employed for part (a), lead to

$$\begin{aligned}
A_1^* &= \sum_{h=1}^{T-2}\sum_{s=1}^{T-1-h} C_{s+h}\left(\frac{1}{T}\sum_{t=h+1+s}^{T}\widetilde{u}_{t-(s+h)}^*\widetilde{u}_{t-s}^{*\prime}\right)C_s' \\
&= \sum_{h=1}^{m-2}\sum_{s=1}^{m-1-h} C_{s+h}\left(\frac{1}{T}\sum_{t=h+1+s}^{T}\widetilde{u}_{t-(s+h)}^*\widetilde{u}_{t-s}^{*\prime}\right)C_s' \\
&\quad + \left(\sum_{h=1}^{T-2}\sum_{s=1}^{T-1-h} - \sum_{h=1}^{m-2}\sum_{s=1}^{m-1-h}\right) C_{s+h}\left(\frac{1}{T}\sum_{t=h+1+s}^{T}\widetilde{u}_{t-(s+h)}^*\widetilde{u}_{t-s}^{*\prime}\right)C_s' \\
&= A_1^{*m} + (A_1^{*m} - A_1^*)
\end{aligned}$$

for some fixed $m \in \mathbb{N}$. Now, it is straightforward to show that $\frac{1}{T}\sum_{t=h+1+s}^{T} u_{t-(s+h)}^* u_{t-s}^{*\prime} \to 0$ w.r.t. P^* for all $h = 1, \ldots, m-2$ and for all $s = 1, \ldots, m-1-h$, which leads also to $A_1^{*m} \to 0$ w.r.t. P^*. To conclude the proof of part (b), we can split-up $A_1^{*m} - A_1^*$ to get

$$\begin{aligned}
A_1^{*m} - A_1^* &= \sum_{h=1}^{m-2}\sum_{s=m-h}^{T-1-h} \widehat{C}_{s+h}\left(\frac{1}{T}\sum_{t=h+1+s}^{T} u_{t-(s+h)}^* u_{t-s}^{*\prime}\right)\widehat{C}_s' \\
&\quad + \sum_{h=m-1}^{T-1}\sum_{s=1}^{T-1-h} \widehat{C}_{s+h}\left(\frac{1}{T}\sum_{t=h+1+s}^{T} u_{t-(s+h)}^* u_{t-s}^{*\prime}\right)\widehat{C}_s' \\
&= \Delta_1^{*m} + \Delta_2^{*m}.
\end{aligned}$$

Similar to the computations for part (a) above, we get for the first one

$$\begin{aligned}
E^*[|\Delta_1^{*m}|_1] &\leq \left(\sum_{h=1}^{m-2}\sum_{s=m-h}^{T-1-h} |C_{s+h}|_1 |C_s|_1\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{f=1}^{K} E^*\left(\widetilde{u}_{t,f}^{*2}\right)\right) \\
&\leq \left(\sum_{s=m}^{\infty} |C_s|_1\right)\left(\sum_{h=2}^{\infty} |C_h|_1\right) O_P(1) \\
&= o_P(1)
\end{aligned}$$

as $m \to \infty$ and analogue arguments lead to the same result for Δ_2^{*m}. □

Lemma A.5 (CLT for bootstrap innovations**).**
Let $m \in \mathbb{N}$ fixed and define $\widetilde{W}_{t,m}^* = (\text{vec}(\widetilde{u}_t^*\widetilde{u}_{t-1}^{*\prime})', \ldots, \text{vec}(\widetilde{u}_t^*\widetilde{u}_{t-m}^{*\prime})', L_K\{\text{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime})' - \text{vec}(u_t u_t')'\})'$. Under the assumptions of Theorem 5.1, it holds

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widetilde{W}_{t,m}^* \overset{\mathcal{D}}{\to} \mathcal{N}(0, \Omega_m)$$

in probability, where Ω_m is defined in Lemma A.1.

Proof.
We consider $\widehat{W}_{t,m}^* = (L_K\{\text{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime})' - \text{vec}(u_t u_t')'\}, \text{vec}(\widetilde{u}_t^*\widetilde{u}_{t-1}^{*\prime})', \ldots, \text{vec}(\widetilde{u}_t^*\widetilde{u}_{t-m}^{*\prime})')'$, which is just a suitably permuted version of $\widetilde{W}_{t,m}^*$, for notational convenience only in the sequel. Further,

let T be sufficiently large such that $\ell > m$. Then, the summation can be split up corresponding to the bootstrap blocking and with respect to summands with $\widetilde{u}_s^*$ and $\widetilde{u}_{s-q}^*$ lying in the same or in different blocks, respectively. By using the notation

$$(\widehat{W}_{t,m}^*)_{q+1} = \begin{cases} L_K\{\mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime})' - \mathrm{vec}(u_t u_t')'\}, & q = 0 \\ \mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_{t-q}^{*\prime})', & q \geq 1 \end{cases}$$

$$(\widehat{U}_{t,m}^*)_{q+1} = \begin{cases} L_K\,\mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_t^{*\prime}), & q = 0 \\ \mathrm{vec}(\widetilde{u}_t^*\widetilde{u}_{t-q}^{*\prime}), & q \geq 1 \end{cases},$$

this leads to

$$\begin{aligned}
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\widehat{W}_{t,m}^* &= \frac{1}{\sqrt{N}}\sum_{r=1}^{N}\left(\frac{1}{\sqrt{\ell}}\sum_{s=(r-1)\ell+1}^{r\ell}\widehat{W}_{s,m}^*\right) \\
&= \frac{1}{\sqrt{N}}\sum_{r=1}^{N}\left(\frac{1}{\sqrt{\ell}}\left\{\sum_{s=(r-1)\ell+1}^{(r-1)\ell+q}(\widehat{W}_{s,m}^*)_{q+1} + \sum_{s=(r-1)\ell+q+1}^{r\ell}(\widehat{W}_{s,m^*})_{q+1}\right\}_{q=0,\ldots,m}\right) \\
&= \frac{1}{\sqrt{N}}\sum_{r=1}^{N}\left(\frac{1}{\sqrt{\ell}}\left\{\sum_{s=(r-1)\ell+1}^{(r-1)\ell+q}(\widehat{U}_{s,m}^*)_{q+1}\right\}_{q=0,\ldots,m}\right) \\
&\quad + \sum_{r=1}^{N}\left(\frac{1}{\sqrt{T}}\left\{\sum_{s=(r-1)\ell+q+1}^{r\ell}(\widehat{U}_{s,m}^*)_{q+1} - E^*\left((\widehat{U}_{s,m}^*)_{q+1}\right)\right\}_{q=0,\ldots,m}\right) \\
&\quad + \frac{1}{\sqrt{N}}\sum_{r=1}^{N}\left(\frac{1}{\sqrt{\ell}}\left\{\sum_{s=(r-1)\ell+q+1}^{r\ell}E^*\left((\widehat{W}_{s,m}^*)_{q+1}\right)\right\}_{q=0,\ldots,m}\right) \\
&= A_1^* + A_2^* + A_3
\end{aligned}$$

with an obvious notation for A_1^*, A_2^* and A_3. In the following, we prove (a) $A_1^* \to 0$ w.r.t P^*, (b) $A_2^* \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega_m)$ in probability and (c) $A_3 \to 0$ in probability. First, we consider (a), where the summation is over the empty set for $q = 0$ and it suffices to show

$$\frac{1}{\sqrt{N}}\sum_{r=1}^{N}\frac{1}{\sqrt{\ell}}\sum_{s=(r-1)\ell+1}^{(r-1)\ell+q}\widetilde{u}_{s,f}^*\widetilde{u}_{s-q,g}^* \to 0$$

in probability w.r.t P^* for $q \in \{1, \ldots, m\}$ and $f, g \in \{1, \ldots, K\}$. By construction of the summation over s its conditional mean is zero as $\widetilde{u}_{s,f}^*$ and $\widetilde{u}_{s-q,g}^*$ lie always in different blocks, and by taking its conditional second moment, we get

$$\begin{aligned}
&\frac{1}{N}\sum_{r_1,r_2=1}^{N}\frac{1}{\ell}\sum_{s_1=(r_1-1)\ell+1}^{(r_1-1)\ell+q}\sum_{s_2=(r_1-1)\ell+1}^{(r_1-1)\ell+q}E^*(\widetilde{u}_{s_1,f}^*\widetilde{u}_{s_1-q,g}^*\widetilde{u}_{s_2,f}^*\widetilde{u}_{s_2-q,g}^*) \\
&= \frac{1}{N}\sum_{r=1}^{N}\frac{1}{\ell}\sum_{s_1,s_2=1}^{q}E^*(\widetilde{u}_{s_1+(r-1)\ell,f}^*\widetilde{u}_{s_2+(r-1)\ell,f}^*)E^*(\widetilde{u}_{s_1+(r-1)\ell-q,g}^*\widetilde{u}_{s_2+(r-1)\ell-q,g}^*)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\ell} \sum_{s_1,s_2=1}^{q} \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s_1,f} u^c_{t+s_2,f} \right) \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s_1-q,g} u^c_{t+s_2-q,g} \right) \\
&= O_P(\ell^{-1}) = o_P(1),
\end{aligned}
$$

by $\ell \to \infty$ as $T \to \infty$, which proves part (a). Next we show part (c). For $q = 0$, we have

$$
\begin{aligned}
& \frac{1}{\sqrt{N}} \sum_{r=1}^{N} \frac{1}{\sqrt{\ell}} \sum_{s=(r-1)\ell+1}^{r\ell} \left(E^* \left(\widetilde{u}^*_{s,f} \widetilde{u}^*_{s,g} \right) - u_{s,f} u_{s,g} \right) \\
&= \frac{1}{\sqrt{N}} \sum_{r=1}^{N} \frac{1}{\sqrt{\ell}} \sum_{s=1}^{\ell} \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s,f} u^c_{t+s,g} - u_{s+(r-1)\ell,f} u_{s+(r-1)\ell,g} \right)
\end{aligned}
$$

for all $f, g \in \{1, \ldots, K\}$, $f \geq g$ and mean and variance of the last right-hand side are of order $O(T^{-1/2})$ and $O(T^{-1})$, respectively, which shows (c). To prove part (b), let $\lambda \in \mathbb{R}^{K^2(m+1)}$ and the summands of A_2^* are denoted by $X^*_{r,T}$. We use a CLT for triangular arrays of independent random variables, cf. Theorem 27.3 in Billingsley (1995), and as $E^*(X^*_{r,T}) = 0$ by construction, we have to show that the following conditions are satisfied:

$$
(i) \quad \sum_{r=1}^{N} E^*(X^*_{r,T} X^{*\prime}_{r,T}) \to \Omega_m \text{ in probability}
$$

$$
(ii) \quad \frac{\sum_{r=1}^{N} E^* \left(|\lambda' X^*_{r,T}|^{2+\delta} \right)}{\left(\sum_{r=1}^{N} E^* \left((\lambda' X^*_{r,T})^2 \right) \right)^{(2+\delta)/2}} \to 0 \text{ as } T \to \infty \text{ for some } \delta > 0.
$$

To show (i), we can restrict ourselves to one entry of $X^*_{r,T} X^{*\prime}_{r,T}$ and we obtain

$$
\begin{aligned}
& \sum_{r=1}^{N} \frac{1}{T} \sum_{s_1,s_2=(r-1)\ell+q+1}^{r\ell} E^* \left(\widetilde{u}^*_{s_1,f_1} \widetilde{u}^*_{s_1-q_1,g_1} \widetilde{u}^*_{s_2,f_2} \widetilde{u}^*_{s_2-q_2,g_2} \right) \\
& \qquad - E^* \left(\widetilde{u}^*_{s_1,f_1} \widetilde{u}^*_{s_1-q_1,g_1} \right) E^* \left(\widetilde{u}^*_{s_2,f_2} \widetilde{u}^*_{s_2-q_2,g_2} \right) \\
&= \frac{1}{\ell} \sum_{s_1,s_2=q+1}^{\ell} \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s_1,f_1} u^c_{t+s_1-q_1,g_1} u^c_{t+s_2,f_2} u^c_{t+s_2-q_2,g_2} \right. \\
& \qquad \left. - \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s_1,f_1} u^c_{t+s_1-q_1,g_1} \right) \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s_2,f_2} u^c_{t+s_2-q_2,g_2} \right) \right).
\end{aligned} \tag{A.8}
$$

For $q_1, q_2 \geq 1$, the leading term of the last expression is

$$
\frac{1}{\ell} \sum_{s_1,s_2=q+1}^{\ell} \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u_{t+s_1,f_1} u_{t+s_1-q_1,g_1} u_{t+s_2,f_2} u_{t+s_2-q_2,g_2} \right).
$$

Due to the mds assumption imposed on the innovation process, its mean computes to $E(u_{t,f_1} u_{t-q_1,g_1} u_{t,f_2} u_{t-q_2,g_2})$ and its variance vanishes asymptotically. As all other summands of (A.8) are of lower order, this leads to the corresponding entry of $\tau_{0,q_1,0,q_2}$ in $\Omega_m^{(1,1)}$. Similarly, for $q_1 = 0, q_2 \geq 1$ and for $q_1 = q_2 = 0$, we get the corresponding entries of $\Omega_m^{(2,1)}$ and of $\Omega^{(2,2)}$, respectively. To conclude the proof of the CLT, we show the Liapunov condition (ii) for $\delta = 2$

and as the denominator is bounded, it suffices to consider the numerator only. For one entry, we get

$$\sum_{r=1}^{N} \frac{1}{T^2} \sum_{s_1,s_2,s_3,s_4=(r-1)\ell+q+1}^{r\ell} \left(E^*\left(\widetilde{u}^*_{s_1,f_1}\widetilde{u}^*_{s_1-q_1,g_1}\widetilde{u}^*_{s_2,f_2}\widetilde{u}^*_{s_2-q_2,g_2}\widetilde{u}^*_{s_3,f_3}\widetilde{u}^*_{s_3-q_3,g_3}\widetilde{u}^*_{s_4,f_4}\widetilde{u}^*_{s_4-q_4,g_4}\right)\right.$$
$$\left. - E^*\left(\widetilde{u}^*_{s_1,f_1}\widetilde{u}^*_{s_1-q_1,g_1}\widetilde{u}^*_{s_2,f_2}\widetilde{u}^*_{s_2-q_2,g_2}\right) E^*\left(\widetilde{u}^*_{s_3,f_3}\widetilde{u}^*_{s_3-q_3,g_3}\widetilde{u}^*_{s_4,f_4}\widetilde{u}^*_{s_4-q_4,g_4}\right)\right)$$

and by the moment condition of Assumption 5.1, the last expression can be shown to be of order $O_P(\ell^3/T) = o_P(1)$ under the assumptions. □

B Derivation of Asymptotic Coverage Probabilities of Bootstrap Confidence Intervals

This appendix describes the derivation of the asymptotic coverage probabilities of the pairwise and wild bootstrap confidence intervals for the impulse response coefficients. The coverage probabilities have been determined for the bivariate VAR(2)-LC-GARCH(1,1) specification introduced in Section 6.3.1.

First, we introduce some general notation. Let

$$\mathbf{A} := \begin{pmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{pmatrix} (Kp \times Kp) \quad \text{and} \quad U_t := \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix} (Kp \times 1)$$

be the parameter matrix and error term vector of the companion form of a K-dimensional VAR(p) process, respectively, and define $\Sigma_U = E(U_t U_t')$. For the bivariate VAR(2)-LC-GARCH(1,1) specification we obtain the specific expression

$$\Theta_0 := P = \begin{pmatrix} (\sigma^2_{u,1})^{1/2} & 0 \\ \sigma_{u,12}(\sigma^2_{u,1})^{-1/2} & \left(\sigma^2_{u,2} - \sigma^2_{u,12}(\sigma^2_{u,1})^{-1}\right)^{1/2} \end{pmatrix} \text{ with } \Sigma_u = \begin{pmatrix} \sigma^2_{u,1} & \sigma_{u,12} \\ \sigma_{u,12} & \sigma^2_{u,2} \end{pmatrix}. \quad \text{(B.1)}$$

Finally, let $w_t = (w_{1t}, w_{2t})'$.

We start with deriving the asymptotic covariance matrix $\Sigma_{\widehat{\Theta}_i}$ of the estimators of Θ_i, $i = 0, 1, 2, \dots$, given in (6.1). The limit variance $V^{(1,1)}$ is defined in (3.1). Since the mean of y_t is assumed to be zero in our case, we have $\text{vec}(\Gamma) = \left(I_{(Kp)^2} - \boldsymbol{A} \otimes \boldsymbol{A}\right)^{-1} \text{vec}(\Sigma_U)$ from Lütkepohl (2005, Eq. (2.1.39)). In our set-up, $C_j = (\Phi'_{j-1}, \Phi'_{j-2})'$, $j = 1, 2, \dots$, with $C_{-1} = 0$. Moreover, we have $\tau_{0,i,0,j} = (P \otimes P)\tau^w_{0,i,0,j}(P' \otimes P')$ with $\tau^w_{0,i,0,j} = E\left(\text{vec}(w_t w'_{t-i})\,\text{vec}(w_t w'_{t-j})'\right)$, $i, j \geq 1$. Since we assume $\varepsilon_t \sim i.i.d.\ N(0, I_2)$, $\tau^w_{0,i,0,j} = 0$ if $i \neq j$, see Francq & Zakoïan (2004, Lemma 4.1). Hence, the middle term in $V^{(1,1)}$ simplifies to $\sum_{i=1}^{\infty}(C_r \otimes I_K)(P \otimes P)\tau^w_{0,i,0,i}(P' \otimes P')(C_r \otimes I_K)'$.

We further have

$$\tau^w_{0,i,0,i} = \begin{pmatrix} \gamma_{w_1^2}(i)+1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \gamma_{w_2^2}(i)+1 \end{pmatrix},$$

where $\gamma_{w_k^2}(i) = \text{Cov}(w^2_{t,k}, w^2_{t-i,k})$, $k = 1,2$ and $i = 1,2,\ldots$, by generalizing the univariate results of Section 4.4. The expressions derived there can also be used to explicitly determine $\gamma_{w_i^2}(r)$ as a function of the GARCH parameters a_1 and b_1. The infinite summation in the middle term can be safely approximated by considering the first 100 summands. Finally, $C_{i,\boldsymbol{\beta}}$ is obtained by pre-multiplying the specific version of $G_i = \sum_{m=0}^{i-1} J(\boldsymbol{A}')^{i-1-m} \otimes \Phi_m$, $i = 1,2,\ldots$, with $(P' \otimes I_2)$, where the relevant version of P is given in (B.1). Putting the foregoing expressions together one obtains the specific version of $C_{i,\boldsymbol{\beta}} V^{(1,1)} C'_{i,\boldsymbol{\beta}}$.

Next, we have $V^{(2,2)} = L_2(P \otimes P) V_w^{(2,2)} (P' \otimes P') L_2'$, where $V_w^{(2,2)} = \Sigma_{w^2,0} + 2\sum_{h=1}^{\infty} \Sigma_{w^2,h}$ with

$$\begin{aligned} \Sigma_{w^2,h} &= E\Big(\text{vec}\left\{(w_t w_t') - E(w_t w_t')\right\} \text{vec}\left\{(w_{t-h} w_{t-h}') - E(w_{t-h} w_{t-h}')\right\}'\Big) \\ &= \begin{pmatrix} \gamma_{w_1^2}(h) & 0 & 0 & 0 \\ 0 & \gamma_{w_1,w_2}(h) & \gamma_{w_1,w_2}(h) & 0 \\ 0 & \gamma_{w_1,w_2}(h) & \gamma_{w_1,w_2}(h) & 0 \\ 0 & 0 & 0 & \gamma_{w_2^2}(h) \end{pmatrix}, \end{aligned} \tag{B.2}$$

$\gamma_{w_1,w_2}(0) = 1$, and $\gamma_{w_1,w_2}(h) = 0$ for $h > 0$. (B.2) is obtained again by generalizing the univariate results of Section 4.4. Then, one can combine all relevant specific expressions to get $C_{0,\boldsymbol{\sigma}} V^{(2,2)} C'_{0,\boldsymbol{\sigma}}$ what, finally, leads to

$$\Sigma_{\widehat{\Theta}_i} = C_{i,\boldsymbol{\beta}} V^{(1,1)} C'_{i,\boldsymbol{\beta}} + C_{0,\boldsymbol{\sigma}} V^{(2,2)} C'_{0,\boldsymbol{\sigma}}, \quad i = 0,1,2,\ldots, \tag{B.3}$$

since $V^{(1,2)} = 0$ in our set-up. Hence, we can determine $\Sigma_{\widehat{\Theta}_i}$ depending on the VAR parameters in A_1 and A_2, the correlation parameter ρ, and the GARCH parameters a_1 and b_1.

Let us now turn to the pairwise and wild bootstrap counterparts of $\Sigma_{\widehat{\Theta}_i}$, $i = 0,1,2,\ldots$, labelled as $\Sigma^{PB}_{\widehat{\Theta}_i}$ and $\Sigma^{WB}_{\widehat{\Theta}_i}$, respectively. Since the bootstrap schemes correctly replicate $V^{(1,1)}$ it is sufficient to consider the corresponding versions of $V^{(2,2)}$, i.e. $V^{(2,2)}_{PB}$ and $V^{(2,2)}_{WB}$. It is easy to see that $V^{(2,2)}_{PB} = L_2(P \otimes P)\Sigma_{w^2,0}(P' \otimes P')L_2'$. Note that an i.i.d. bootstrap scheme would lead to the same asymptotic variance expression.

For the wild bootstrap, one has $V^{(2,2)}_{WB} = L_2(P \otimes P)\tau^w_{0,0,0,0}(P' \otimes P')L_2' \left\{E^*(\eta_t^4) - 1\right\}$ with $E^*(\eta_t^4) = 3$ for $\eta_t \sim i.i.d.\ N(0,1)$ and

$$\tau^w_{0,0,0,0} = E\left[\text{vec}\left(w_t w_t'\right) \text{vec}\left(w_t w_t'\right)'\right] = \Sigma_{w_0^2} + \text{vec}(\Sigma_w)\,\text{vec}(\Sigma_w)' = \Sigma_{w_0^2} + \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

where $\Sigma_w = E(w_t w_t') = I_2$ in our case. Accordingly, one would get $V_{WB}^{(2,2)} = 0$ for the Rademacher distribution. Note again that the recursive- and fixed-design schemes are asymptotically equivalent. Replacing $V^{(2,2)}$ in (B.3) with $V_{PB}^{(2,2)}$ or $V_{WB}^{(2,2)}$ results in $\Sigma_{\widehat{\Theta}_i}^{PB}$ and $\Sigma_{\widehat{\Theta}_i}^{WB}$, respectively.

Finally, the asymptotic coverage probabilities of the pairwise and wild bootstrap confidence intervals are obtained as follows. It is assumed that the bootstrap estimators of $\widehat{\Theta}_i$, $i = 0, 1, 2, \ldots$, are consistent and asymptotically normally distributed with variances $\Sigma_{\widehat{\Theta}_i}^{PB}$ and $\Sigma_{\widehat{\Theta}_i}^{WB}$, respectively. Hence, both the asymptotically correct intervals and the bootstrap intervals are centered at the same value. Let $\sigma_{\widehat{\theta}}$ be the correct asymptotic standard deviation of the estimator of a particular impulse response coefficient θ taken from the relevant $\Sigma_{\widehat{\Theta}_i}$, $i = 0, 1, 2, \ldots$. Let $\sigma_{\widehat{\theta}}^B$ be the corresponding asymptotic bootstrap standard deviation taken from either $\Sigma_{\widehat{\Theta}_i}^{PB}$ or $\Sigma_{\widehat{\Theta}_i}^{WB}$. Then, the asymptotic coverage probability of the corresponding bootstrap interval can be simply obtained by

$$P(-\sigma_r \cdot z_{1-\alpha/2} \leq X \leq \sigma_r \cdot z_{1-\alpha/2}),$$

where X is standard normally distributed, $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$-quantile of the standard normal distribution, and $\sigma_r = \sigma_{\widehat{\theta}}^B / \sigma_{\widehat{\theta}}$.