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*

July 16, 2015

Abstract

We consider a framework for asymptotically valid inference in stable vector autoregressive (VAR) models with conditional heteroskedasticity of unknown form. A joint central limit theorem for the LS estimators of both the VAR slope parameters as well as the unconditional innovation variance parameters is obtained from a weak vector autoregressive moving average model set-up recently proposed in the literature. 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 responses. 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 innovations. 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 impulse response coefficients. 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, Mixing, Residual-based moving block bootstrap, Pairwise bootstrap, Wild bootstrap

[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

In many econometric models for financial and macroeconomic time series the innovations may be serially uncorrelated but not identically and independently distributed (i.i.d.). An important case of deviation from independence is conditional heteroskedasticity. Indeed, conditional heteroskedasticity and other departures from the i.i.d. assumption have been documented in many empirical examples in the literature, see for instance Gonçalves & Kilian (2004). 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. Often, these time series are analyzed within vector autoregressive (VAR) models which are a popular econometric tool to summarize the dynamic interaction between the variables included in the 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)) draw conclusions based on statistics obtained from estimated VAR models. 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 bootstrap methods. The deviation from i.i.d. innovations, as e.g. the presence of conditional heteroskedasticity, invalidates a number of standard inference procedures 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 if innovations are only serially uncorrelated but not independent.

The existing time series literature makes some suggestions for valid inference under conditional heteroskedasticity. For instance, Gonçalves & Kilian (2004, 2007) consider inference on AR parameters in univariate autoregressions. Imposing an appropriate martingale difference sequence (mds) assumption on the innovations, they show that wild and pairwise bootstrap approaches are asymptotically valid 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 asymptotic approximations based on robust standard errors. Also using an mds assumption, Hafner & Herwartz (2009) focus on Wald tests for Granger-causality within VAR models. They find that wild bootstrap methods provide more reliable inference than heteroskedasticity-consistent asymptotic inference.

The aforementioned papers do not study the implications of the presence of conditional heteroskedasticity on inference on a number of VAR statistics that are also functions of the innovation covariance matrix. Popular statistics of interest in this context are structural impulse responses, 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 unconditional innovation covariance matrix. In the following we refer to the vector autoregressive slope parameter matrices simply as the 'VAR parameters', while the unconditional innovation covariance matrix is referred to as 'variance parameters'. Their joint distribution is well explored in the case of i.i.d. innovations, see e.g. Lütkepohl (2005, Chapter 3).

Although joint asymptotic inference in case of non-i.i.d. innovations is discussed in the framework of weak vector autoregressive moving average (weak VARMA) models, see Boubacar Maïnassara & Francq (2011), the implications of a departure from i.i.d. innovations for inference on statistics like e.g. structural impulse responses in VAR models is not well understood in the econometric literature.

To fill this gap, we follow Boubacar Maïnassara & Francq (2011) and allow for quite general deviations from an i.i.d. set-up. To be precise, the innovations of our stable VAR are assumed to be serially uncorrelated and to satisfy a mixing condition. Hence, we consider a weak VAR such that the results obtained in the paper cannot only be used in the case of conditional heteroskedasticity but also cover more general innovation processes.

In this paper, we contribute to the literature in a number of directions. First, we state a joint central limit theorem for the estimators of the VAR and variance parameters, which follows from Boubacar Maïnassara & Francq (2011), and compare this result to asymptotic results obtained under an appropriate mds assumption. This links our results to the related time series literature that allows for conditional heteroskedasticity of unknown form, see e.g. Gonçalves & Kilian (2004, 2007) and Hafner & Herwartz (2009).

Second, we analyze the theoretical properties of different bootstrap approaches commonly used in a VAR context with conditional heteroskedasticity. Our result here indicates that neither a wild bootstrap (recursive-design and fixed-design) nor a pairwise bootstrap approach work under mixing conditions only as they cannot mimic the proper limiting distribution. Adding an additional mds assumption is sufficient to ensure that bootstrap inference on the VAR parameters is consistent. However, the wild and pairwise bootstraps do not lead to asymptotically valid inference on (functions of) the unconditional innovation covariance matrix in this set-up. 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. innovations in case of inference on (functions of) the innovation covariance matrix.

Third, we prove that a residual-based moving block bootstrap (MBB) results in asymptotically valid joint inference on the VAR and variance parameters if suitable mixing assumptions are imposed. 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. 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. As the residual-based MBB is valid under suitable mixing assumptions alone, it is not only suitable in the presence of conditional heteroskedastitcity but is generally applicable in a weak VAR context.

Finally, we illustrate the importance and implications of the theoretical results by studying inference on structural impulse responses. This type of impulse responses are of major importance in typical VAR applications. We provide simulation evidence on the finite-sample properties of corresponding first-order asymptotic approximations and of various bootstrap approaches. We draw two main conclusions from our simulation study. First, applied researchers have to be aware that estimation uncertainty may dramatically increase if non-i.i.d. innovations are present. Second, in many situations the true sampling variation of the impulse response estimators is understated by most of the inference procedures. This, in turn, may lead to (bootstrap) confidence intervals for impulse response coefficients being too narrow at short horizons. Accordingly, applied researchers should interpret their results with caution.

The remainder of the paper is structured as follows. Section 2 provides the modeling framework and asymptotic results for the LS estimators of the VAR and unconditional variance parameters. We show the invalidity of the wild and pairwise bootstrap schemes in Section 3. The residual-based MBB scheme and its asymptotic properties are presented in Section 4. Section 5 contains a discussion on structural impulse response analysis and presents the simulation results. Finally, Section 6 concludes. The proofs are deferred to the appendix.

2 Modeling Framework and Asymptotic Inference

2.1 Notation and preliminaries

Let $(u_t, t \in \mathbb{Z})$ be a K-dimensional white noise sequence defined on a probability space $(\Omega, \mathcal{F}, P)$ such that each $u_t = (u_{1t}, \ldots, u_{Kt})'$ 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 $(y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T)$ of sample size T plus p pre-sample values from the following DGP for the K-dimensional time series $y_t = (y_{1t}, \ldots, y_{Kt})'$,

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

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 $A_1, \ldots, A_p$ and the unconditional innovation covariance matrix $\Sigma_u = E(u_t u_t')$. Our results can be easily generalized to a set-up with a non-zero intercept vector. We introduce the following notation, where the dimensions of the defined quantities are given in parentheses:

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

where 'vec' denotes the column stacking operator. The parameter $\boldsymbol{\beta}$ is estimated by $\widehat{\boldsymbol{\beta}} = \operatorname{vec}(\widehat{A}_1, \ldots, \widehat{A}_p)$ via multivariate LS, i.e. $\widehat{\boldsymbol{\beta}} = ((ZZ')^{-1}Z \otimes I_K)\mathbf{y}$, see e.g. Lütkepohl (2005, p. 71). 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

$$y_t = \sum_{j=0}^{\infty} \Phi_j u_{t-j}, \quad t \in \mathbb{Z}, \tag{2.3}$$

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 $\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j$, $i = 1, 2, \ldots$. Further, we define the $(Kp \times K)$ matrices $C_j = (\Phi'_{j-1}, \ldots, \Phi'_{j-p})'$ and the $(Kp \times Kp)$ matrix $\Gamma = \sum_{j=1}^{\infty} C_j \Sigma_u C'_j$. The standard estimator of Σ_u is

$$\widehat{\Sigma}_u = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_t \widehat{u}'_t, \tag{2.4}$$

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 column-wise the elements on and below the main diagonal of a square matrix.

2.2 Assumptions and asymptotic inference

We obtain the asymptotic properties of our estimators following the results of Boubacar Maïnassara & Francq (2011). To this end, we use the sufficient part of their assumptions on the process $(y_t, t \in \mathbb{Z})$ in addition to the stability condition for the DGP (2.1):

Assumption 2.1 (mixing innovations)**.**

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

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

(iii) *Let $\alpha_u(m) = \sup_{A \in \mathcal{F}^0_{-\infty}, B \in \mathcal{F}^\infty_m} |P(A \cap B) - P(A)P(B)|$, $m = 1, 2, \ldots$, denote the α-mixing coefficients of the process $(u_t, t \in \mathbb{Z})$, where $\mathcal{F}^0_{-\infty} = \sigma(\ldots, u_{-2}, u_{-1}, u_0)$, $\mathcal{F}^\infty_m = \sigma(u_m, u_{m+1}, \ldots)$. We have for some $\delta > 0$,*

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

and that $E|u_t|^{4+2\delta}_{4+2\delta}$ is uniformly bounded, where $|A|_p = (\sum_{i,j} |a_{ij}|^p)^{1/p}$ for some matrix $A = (a_{ij})$.

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

$$\tau_{0,a,b,c} = E\left(\text{vec}(u_t u'_{t-a}) \, \text{vec}(u_{t-b} u'_{t-c})'\right), \tag{2.5}$$

use $\widetilde{K} = K(K+1)/2$ and assume that the $(K^2 m + \widetilde{K} \times K^2 m + \widetilde{K})$ matrix Ω_m defined in (A.5) *exists and is positive definite for all $m \in \mathbb{N}$.*

Here, the common i.i.d. assumption for the process $(u_t, t \in \mathbb{Z})$ is replaced by the less restrictive mixing condition in Assumption 2.1(iii). In particular, Assumption 2.1 covers a large class of dependent, but uncorrelated stationary innovation processes and allows for conditional heteroskedasticity. The assumption on Ω_m in (iv) corresponds to assumption A8 in Theorem 3 in Boubacar Maïnassara & Francq (2011). It ensures that the asymptotic variance matrix V of our estimators, introduced in Theorem (2.1) below, is positive definite.

Following, Francq & Raïssi (2007) a process (2.1) satisfying Assumption 2.1 is called a weak VAR process. It is a special case of the weak VARMA process analyzed by Boubacar Maïnassara & Francq (2011) and Dufour & Pelletier (2011). We state the following central limit theorem (CLT) for our estimators.

Theorem 2.1 (Unconditional CLT). *Under Assumption 2.1, 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 '$\overset{\mathcal{D}}{\to}$' denotes convergence in distribution,

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

with

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

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

$$V^{(2,2)} = L_K\left(\sum_{h=-\infty}^{\infty}\left\{\tau_{0,0,h,h} - \operatorname{vec}(\Sigma_u)\operatorname{vec}(\Sigma_u)'\right\}\right)L_K'$$

and 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 any $(K \times K)$ matrix A.*

Theorem 2.1 follows from Theorems 1 and 2 of Boubacar Maïnassara & Francq (2011). The asymptotic variance matrix of the estimators in Boubacar Maïnassara & Francq (2011) is based on a sandwich form which can be re-written and simplified for our VAR set-up in order to obtain V as in Theorem 2.1. For details see Appendix A.1. Note that Francq & Raïssi (2007, Proposition 2) have already derived asymptotic normality for $\widehat{\boldsymbol{\beta}}$ using Assumptions 2.1(i)-(iii).

Remark 2.1. One may also write

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

with $\boldsymbol{u}_t^2 = \text{vech}(u_t u_t')$. 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 $(\boldsymbol{u}_t^2, t \in \mathbb{Z})$. If the error terms were i.i.d., we obviously have $V^{(2,2)} = \text{Var}(\boldsymbol{u}_t^2) = L_K \tau_{0,0,0,0} L_K' - \boldsymbol{\sigma}\boldsymbol{\sigma}'$.[1]

Remark 2.2. Implementing asymptotic inference based on Theorem 2.1 requires estimation of V. Boubacar Maïnassara & Francq (2011, Section 4) have proposed an estimator $\widehat{V}$ based on the sandwich form representation of V. This estimator involves a VARHAC-type approach and is consistent under appropriate assumptions, see Boubacar Maïnassara & Francq (2011, Theorem 3). We apply this estimator in our simulations, cf. Appendix A.1.

Remark 2.3. The results in Theorem 2.1 are obtained under the assumption 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. 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.

Remark 2.4. It is possible to derive a CLT for $\widehat{\boldsymbol{\beta}}$ alone based on an appropriate mds assumption, i.e. a vector-valued analogue of Gonçalves & Kilian (2004, Assumption A) which comprises a number of regularity conditions in addition to the plain mds-Assumption 2.2 below. In fact, one does not have to impose strict stationarity in full strength or a mixing condition as in Assumption 2.1(iii). However, a type of fourth-order stationarity is required. For details see Brüggemann, Jentsch & Trenkler (2014, Sections 2-3). Note, however, that strict stationarity and the mixing condition are required to prove the joint CLT for $\widehat{\boldsymbol{\beta}}$ *and* $\widehat{\boldsymbol{\sigma}}$ in Theorem 2.1. A CLT for mds is not applicable here. This is due to the fact that $\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}$ includes $\text{vech}(u_t u_t')$ and is therefore not an mds. In contrast, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ is an mds as it only contains terms of the form $\text{vech}(u_t u_{t-j}')$ with $j \geq 1$.

The sub-matrices $V^{(1,1)}$ and $V^{(2,1)}$ in the asymptotic variance matrix V in Theorem 2.1 simplify if we additionally impose the following mds assumption on the innovations.

Assumption 2.2 (mds innovations)**.**
$E(u_t|\mathcal{F}_{t-1}) = 0$ almost surely, where $\mathcal{F}_{t-1} = \sigma(u_{t-1}, u_{t-2}, \ldots)$ is the σ-field generated by $(u_{t-1}, u_{t-2}, \ldots)$.

The following corollary directly follows from Brüggemann et al. (2014, Theorem 3.1(ii)).

Corollary 2.1. *Under Assumptions 2.1 and 2.2, we have*

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

[1] The result $V^{(2,2)} = 2D_K^+(\Sigma_u \otimes \Sigma_u)D_K^{+\prime}$ implied by Boubacar Maïnassara & Francq (2011, Remark 3) requires an additional normality assumption on u_t, also compare Lütkepohl (2005, Proposition 3.4). Then, one has $\tau_{0,0,0,0} = (I_{K^2} + C_{KK})(\Sigma_u \otimes \Sigma_u) + \text{vec}(\Sigma_u)\,\text{vec}(\Sigma_u)'$, see Ghazal & Neudecker (2000). Note, $D_K^+ = (D_K' D_K)^{-1} D_K'$ with D_K being the $(K^2 \times 0.5K(K+1))$ duplication matrix which is defined such that $\text{vec}(A) = D_K\,\text{vech}(A)$ holds for any symmetric $(K \times K)$ matrix A. Moreover, C_{KK} is the $(K^2 \times K^2)$ dimensional commutation matrix which is defined such that $C_{KK}\,\text{vec}(A) = \text{vec}(A')$ holds for any $(K \times K)$ matrix A.

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

Remark 2.5. V is in general not block-diagonal either under the additional mds Assumption 2.2. Hence, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\sigma}}$ are in general asymptotically correlated in contrast to a set-up in which $u_t \sim i.i.d.(0, \Sigma_u)$. A corresponding finding on asymptotically correlated estimators for mean and variance parameters has been obtained by Francq & Zakoïan (2004) and Ling & McAleer (2003) in relation to (vector) ARMA-GARCH processes.

From their results one may conclude that V is block-diagonal, i.e. $V^{(2,1)} = 0$, if u_t in (2.1) follows a stable vector GARCH process with a conditional error that has a spherically symmetric distribution. In fact, a spherical symmetry assumption would lead to $\tau_{0,0,h,h+j} = 0$ for all $h \geq 0$ and $j \geq 1$, see also Hafner (2004, Lemma 1) and Francq & Zakoïan (2004, Lemma 4.1).

3 Asymptotic Invalidity of the Wild and Pairwise Bootstraps

Since the finite sample properties of asymptotic VAR inference approaches can be rather poor, the use of bootstrap methods is often advocated, see e.g. Kilian (1998*b*,*a*, 1999) in the context of 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 view of the literature on bootstrapping autoregressions with non-i.i.d. innovations, cf. Gonçalves & Kilian (2004, 2007) for the univariate case and mds innovations, it is clear that a residual-based i.i.d. resampling scheme does not work in general. In contrast, the following schemes have been proposed and widely used in the literature to overcome the problem of bootstrap inconsistency: (a) recursive-design wild bootstrap, (b) fixed-design wild bootstrap, (c) pairwise bootstrap.

By imposing only mixing conditions on the innovation process, it is rather easy to see that the bootstrap procedures (a) - (c) cannot mimic the proper limiting distribution in Theorem 2.1, neither for $\widehat{\boldsymbol{\beta}}$ nor for $\widehat{\boldsymbol{\sigma}}$. However, an appropriate (additional) mds error term assumption is sufficient to ensure that bootstrap procedures (a), (b), and (c) correctly mimic the limiting distribution for $\widehat{\boldsymbol{\beta}}$. Wild and pairwise bootstraps are tailor-made for the dependence structure inherent to mds sequences. Heuristically, they are capable to approximate properly the true limiting distribution for $\widehat{\boldsymbol{\beta}}$ in Corollary 2.1, but not that in Theorem 2.1. This is because the bootstraps (a) - (c) are just sufficient to capture exclusively the summand corresponding to $h = 0$ in $V^{(1,1)}$ in Theorem 2.1 which leads exactly to the $V^{(1,1)}$ expression in Corollary 2.1. This result follows from Hafner & Herwartz (2009) and by extending the results of Gonçalves & Kilian (2004) to the multivariate case. Hence, in a mds set-up, the bootstrap schemes are asymptotically valid for inference that only refers to the VAR parameters, including e.g. the case of forecast error impulse responses (FEIRs).

In view of these results, it seems natural to check whether the bootstrap consistency of schemes (a) - (c) for $\sqrt{T}(\boldsymbol{\beta} - \boldsymbol{\beta})$ also extends to the joint CLT of $\sqrt{T}(\boldsymbol{\beta} - \boldsymbol{\beta})$ *and* $\sqrt{T}(\widehat{\boldsymbol{\sigma}} - \boldsymbol{\sigma})$

in the framework of Corollary 2.1, that is, under an mds assumption. We will show below that all three procedures actually fail to mimic the proper distribution in Corollary 2.1.[2] In order to do so, it is advisable to inspect the asymptotic variance of $\widehat{\boldsymbol{\sigma}}$. Here, it suffices to consider the notationally simpler univariate case $K = 1$ and to compare $\sqrt{T}(\widehat{\sigma}_u^2 - \sigma_u^2)$, where $\sigma_u^2 = E(u_t^2)$ and $\widehat{\sigma}_u^2 = \frac{1}{T}\sum_{t=1}^T \widehat{u}_t^2$, to its bootstrap analogues. Furthermore, all bootstrap schemes will not be able to replicate the covariance block $V^{(2,1)}$ in Corollary 2.1 either.

3.1 Recursive-design and fixed-design wild bootstrap

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

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

both approaches coincide here and lead to 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 $E^*(\eta_t^4) < \infty$, where $E^*(\cdot) = E(\cdot|X_1, \ldots, X_n)$ denotes as usual the expectation conditional on $X_1, \ldots, X_n$. From $E^*(\eta_t^2) = 1$, we get $E^*(\sqrt{T}(\widehat{\sigma}_{WB}^{2*} - \widehat{\sigma}_u^2)) = 0$ and

$$E^* \left(\sqrt{T}\left(\widehat{\sigma}_{WB}^{2*} - \widehat{\sigma}_u^2\right)\right)^2 = \frac{1}{T}\sum_{t=1}^T \widehat{u}_t^4 \left(E^*\left(\eta_t^4\right) - 1\right),$$

where the right-hand side converges in probability to

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

This 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 will 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 2.1. The latter fact has already been observed in Kreiss (1997) for linear processes. Similarly, it is easy to show that $V_{WB}^{(2,1)} = 0 \neq V^{(2,1)}$ holds in general, i.e. the wild bootstrap estimates the potentially non-zero limiting covariances always as being zero.[3]

3.2 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

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

[2]In Brüggemann et al. (2014, Section 3.3) we show that the block wild bootstrap recently proposed by Shao (2011) is not consistent either in this set-up.

[3]There may exist distributions for η_t with $E^*(\eta_t^4)$ such that $V_{WB}^{(2,2)} = V^{(2,2)}$. However, $E^*(\eta_t^4)$ depends on unknown features of the innovation process and it is not automatically assured that $E^*(\eta_t) = 0$ and $E^*(\eta_t^2) = 1$ in such a case. Moreover, no wild bootstrap with (η_t) being an i.i.d. sequence is able to produce a non-zero $V_{WB}^{(2,1)}$. Hence, the wild bootstrap will generally be inappropriate.

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

$$E^*\left(\sqrt{T}\left(\widehat{\sigma}_{PB}^{2*} - \widehat{\sigma}_u^2\right)\right)^2 = \frac{1}{T}\sum_{s=1}^T \widehat{u}_s^4 - \left(\frac{1}{T}\sum_{s=1}^T \widehat{u}_s^2\right)^2,$$

where the right-hand side converges in probability to $V_{PB}^{(2,2)}$ and we have

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

Thus, the pairwise bootstrap in inconsistent in general. Observe here that the pairwise bootstrap is equivalent to an i.i.d. bootstrap applied to the residuals. Similarly, one can show that

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

Finally, note that the pairwise bootstrap leads to asymptotically valid inference on the innovation variance if $u_t \sim i.i.d.\ (0, \sigma_u)$ in contrast to the wild bootstrap approaches.

3.3 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:

$$u_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = a_0 + a_1 u_{t-1}^2 + b_1 \sigma_{t-1}^2, \text{ with } a_0 = 1 - a_1 - b_1 \text{ and } \varepsilon_t \sim i.i.d.\ N(0,1). \quad (3.1)$$

From Francq & Zakoïan (2010, Chapter 2) and using some algebra one obtains

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

where the first-order autocorrelation of u_t^2 is given by $\rho_{u^2}(1) = \frac{a_1\{1 - b_1(a_1+b_1)\}}{1 - 2a_1b_1 - b_1^2}$, $\operatorname{Var}(u_t^2) = E(u_t^4) - \sigma_u^4$ with $\sigma_u^4 = 1$ and $E(u_t^4) = \frac{1-(a_1+b_1)^2}{1-(a_1+b_1)^2 - a_1^2(\kappa_\varepsilon - 1)}\kappa_\varepsilon$ with $\kappa_\varepsilon = E(\varepsilon_t^4) = 3$.

From the previous subsections we obtain $V_{PB}^{(2,2)} = \operatorname{Var}(u_t^2)$ for the pairwise bootstrap and $V_{WB}^{(2,2)} = E(u_t^4)(E^*(\eta_t^4) - 1)$ for the wild bootstraps. 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, there can be a tremendous asymptotic downward bias with respect to $V^{(2,2)}$. Nevertheless, the bias depends quite importantly on a_1 and b_1. E.g. in Cases G2 and G3, $V_{WB}^{(2,2)}$ is relatively close to $V^{(2,2)}$. 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.

4 Residual-Based Moving Block Bootstrap

Block bootstrap methods have been used to capture dependencies in time series data. 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, whereas Jentsch, Paparoditis & Politis (2014) provide theory for residual-based block bootstraps in multivariate integrated and co-integrated models. Here, we propose to use the moving block bootstrap techniques for the residuals obtained from a fitted VAR(p) model to approximate the limiting distribution of $\sqrt{T}((\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})',(\widehat{\boldsymbol{\sigma}}-\boldsymbol{\sigma})')'$ derived in Theorem 2.1.

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

$$u_{j\ell+s}^* = \widehat{u}_{j\ell+s}^* - E^*(\widehat{u}_{j\ell+s}^*) = \widehat{u}_{j\ell+s}^* - \frac{1}{T-\ell+1}\sum_{r=0}^{T-\ell}\widehat{u}_{s+r} \tag{4.1}$$

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

$$\widehat{\boldsymbol{\beta}}^* = \mathrm{vec}(\widehat{A}_1^*,\ldots,\widehat{A}_p^*) = ((Z^* Z^{*\prime})^{-1} Z^* \otimes I_K)\mathbf{y}^*, \tag{4.2}$$

where Z^* and $\mathbf{y}^*$ are defined analogously to Z and $\mathbf{y}$ in (2.2), respectively, but based on

$y^*_{-p+1}, \dots, y^*_0, y^*_1, \dots, y^*_T$. Further, we define the bootstrap analogue of $\widehat{\Sigma}_u$ as

$$\widehat{\Sigma}^*_u = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}^*_t \widehat{u}^{*\prime}_t, \tag{4.3}$$

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 $\widehat{\boldsymbol{\sigma}}^* = \operatorname{vech}(\widehat{\Sigma}^*_u)$.

Remark 4.1. Contrary to a bootstrap scheme that uses i.i.d. resampling of the residuals, the standard centering $\widehat{u}_t = \widetilde{u}_t - \frac{1}{T}\sum_{s=1}^{T} \widetilde{u}_s$, $t = 1, \dots, T$, 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 (4.1) 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. Even if an intercept term ν is included in the VAR(p) model, we will have $E^*(u^*_t) \neq 0$ in general for the MBB and the centering in Step 3 is advisable. However, the effect of not properly centered residuals vanishes asymptotically in both cases and we expect only a slight loss in performance in practice.

Remark 4.2. In Bootstrap Scheme I we rely on pre-whitening the data which should be much more efficient than drawing from blocks of y_t directly. As for the wild bootstrap approach one may also consider a fixed-design MBB rather than relying on the recursive structure in Step 4. In contrast to the wild bootstrap framework of Gonçalves & Kilian (2004), a fixed-design MBB cannot shown to be asymptotically valid under weaker error term assumptions than the recursive-design MBB. Therefore, we do not consider the fixed-design MBB in the following.

Remark 4.3. As pointed out by one referee, it is also possible to use a block-of-blocks bootstrap in which the MBB is applied directly to the sample formed of tupels $\{(y_t, Z_{t-1}), t = 1, \dots, T\}$ leading to bootstrap tupels $\{(y^*_t, Z^*_{t-1}), t = 1, \dots, T\}$, see e.g. Politis & Romano (1992*a*,*b*). Based on heuristic arguments, the use of

$$\widehat{\boldsymbol{\beta}}^* = \left(\frac{1}{T}\sum_{t=1}^{T} Z^*_{t-1} Z^{*\prime}_{t-1}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \operatorname{vec}(y^*_t Z^{*\prime}_{t-1}) \quad \text{and} \quad \widehat{\boldsymbol{\sigma}}* = \operatorname{vech}\left(\frac{1}{T}\sum_{t=1}^{T} \widehat{u}^{**}_t \widehat{u}^{**\prime}_t\right),$$

where $\widehat{u}^{**}_t := y^*_t - (\widehat{A}^*_1, \dots, \widehat{A}^*_p) Y^*_{t-1}$ and $\widehat{\boldsymbol{\beta}} = \operatorname{vec}(\widehat{A}^*_1, \dots, \widehat{A}^*_p)$, should lead to valid bootstrap approximations in the set-ups of Theorem 2.1 and of Corollary 2.1. We leave a rigorous proof for future research. Simulation results, which are available upon request from the authors, show that the block-of-blocks bootstrap performs almost identical to the residual-based MBB.

We make the following assumption.

Assumption 4.1 (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 $j = 2, \dots, 8$ and*

$a_1, \ldots, a_j \in \{1, \ldots, K\}$, $\mathbf{a} = (a_1, \ldots, a_j)$ that

$$\sum_{h_2,\ldots,h_j=-\infty}^{\infty} |\mathrm{cum}_{\mathbf{a}}(0, h_2, \ldots, h_j)| < \infty \tag{4.4}$$

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 4.1 is implied by

$$\sum_{m=1}^{\infty} m^{n-2} (\alpha_u(m))^{\delta/(2n-2+\delta)} < \infty \tag{4.5}$$

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 8th moment of a GARCH process exists for given GARCH parameters and the conditional distribution, compare Ling & McAleer (2002) and Lindner (2009).

Now, we state

Theorem 4.1 (Residual-based MBB consistency).
Under Assumptions 2.1 and 4.1 and if $\ell \to \infty$ such that $\ell^3/T \to 0$ as $T \to \infty$, we have

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

in probability, where P^* denotes the probability measure induced by the residual-based MBB and $\bar{K} = K^2 p + (K^2 + K)/2$. 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 $i = 1, \ldots, d$.

The proof of bootstrap consistency in Theorem 4.1 is provided in Appendix A.2. Note that no mds structure is required for the consistency of the MBB, which is achieved under suitable mixing and moment conditions alone. However, the covariance matrix V in the form of Corollary 2.1 will be recovered by Bootstrap Scheme I when the mds condition is additionally imposed.

5 Inference on Structural Impulse Responses

To illustrate the implications of our results we focus on structural impulse responses as they are often used in empirical VAR studies. Note in this respect that the presence of conditional heteroskedasticity has been exploited in the VAR context for structural identification of shocks, see e.g. Rigobon (2003), Normandin & Phaneuf (2004) and Herwartz & Lütkepohl (2014), but the implications for inference on impulse responses have not been analyzed in detail yet. 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 impulse responses 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 derive the asymptotic distribution of the impulse response estimators under Assumption 2.1 by relying on the Delta method. Then, 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.

5.1 Asymptotic distribution of structural impulse responses

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. 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 IRF 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 2.1.

Corollary 5.1 (CLT for Structural Impulse Responses)**.**
Under Assumption 2.1 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

$$\Sigma_{\widehat{\Theta}_i} = C_{i,\boldsymbol{\beta}} V^{(1,1)} C'_{i,\boldsymbol{\beta}} + C_{i,\boldsymbol{\sigma}} V^{(2,2)} C'_{i,\boldsymbol{\sigma}} + C_{i,\boldsymbol{\beta}} V^{(2,1)'} C'_{i,\boldsymbol{\sigma}} + C_{i,\boldsymbol{\sigma}} V^{(2,1)} C'_{i,\boldsymbol{\beta}} \tag{5.1}$$

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(p) process, see e.g. Lütkepohl (2005, Section 2.1.1),* $H = \dfrac{\partial\,\mathrm{vec}(P)}{\partial \boldsymbol{\sigma}'}$, *and* $V^{(i,j)}$, $i, j = 1, 2$, *are as defined in Theorem 2.1.*

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 (5.1) that are present whenever the off-diagonal blocks in V are non-zero. If Assumption 2.2 is additionally imposed, then the expressions for $V^{(1,1)}$ and $V^{(2,1)}$ from Corollary 2.1 may be used.

5.2 Bootstrap inference on structural impulse responses

As a valid alternative to first-order asymptotic approximation based on Corollary 5.1 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 4 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 4.1 corresponding to arguments in Kilian (1998*b*). We summarize this result in the following corollary.

Corollary 5.2 (Asymptotic Validity of Bootstrap SIRs).
Under Assumptions 2.1 and 4.1 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. We leave for future research studying the performance of percentile-t-intervals as well as the use of bias-adjustments to improve bootstrap accuracy.

5.3 Asymptotic results and simulation evidence

In this section we compare the coverage properties of different bootstrap and asymptotic confidence intervals for impulse responses. First, the structure of our DGP is explained in Section 5.3.1. Then, we determine the asymptotic distortion of the wild and pairwise bootstrap approaches in Section 5.3.2 before presenting more detailed finite sample results in Section 5.3.3.

5.3.1 Data generating processes

We consider a bivariate VAR in the form of (2.1) 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 parameter values 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. Moreover, let $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})' \sim i.i.d.\ N(0, I_2)$ and define $w_{it} = \sigma_{it}\varepsilon_{it}$ with $\sigma_{it}^2 = a_0 + a_1 w_{it}^2 + b_1 \sigma_{i,t-1}^2$, $i = 1, 2$, and $a_0 = 1 - a_1 - b_1$. Hence, w_{1t} and w_{2t} are two independent univariate GARCH(1,1) processes with $E(w_{1t}^2) = E(w_{2t}^2) = 1$. The VAR innovation 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.

5.3.2 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. To this end, 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 $\Sigma_{\widehat{\Theta}_i}^{PB}$

and $\Sigma^{WB}_{\widehat{\Theta}_i}$, respectively, by extending the univariate results of Section 3.3. 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 and wild bootstrap estimators of Θ_i are consistent and asymptotically normally distributed with variances $\Sigma^{PB}_{\widehat{\Theta}_0}$ and $\Sigma^{WB}_{\widehat{\Theta}_i}$, respectively. Details of the derivations are given in Appendix B of Brüggemann et al. (2014).

Recall from Section 3.2 that the pairwise bootstrap correctly replicates the asymptotic variances in the i.i.d. set-up (Case G0). Hence, the asymptotic coverage probabilities of the corresponding confidence intervals are equal to the nominal level as can be seen in Table 2. In contrast, the wild bootstrap overestimates the asymptotic variance such that the coverage probabilities are above the nominal level. In the presence of conditional heteroskedasticity (Cases G1 to G5) the asymptotic variances of the estimators of the elements in $\Theta_0 = P$ increase substantially. Thus, a correct confidence interval for an impulse response coefficient 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 corresponding bootstrap confidence intervals are typically too narrow and the coverage probabilities are often very low. However, in cases, in which the sum of the autocovariances of $\boldsymbol{u}_t^2$ is not too large (Case G2 and G3), the wild bootstrap may overestimate the variances. This is due to the factor $\{E^*\left(\eta_t^4\right) - 1\} = 2$ appearing in $V^{(2,2)}_{WB}$, compare Section 3.1.

To get a more informative picture, we also report asymptotic coverage probabilities of the intervals associated with the pairwise and wild bootstrap at higher response horizons for Cases G1 and G2 in Figure 1. Interestingly, the deficiencies in coverage accuracy of the impulse response intervals are concentrated at short horizons[4] at which the impulse response estimators are dominated by $\hat{\boldsymbol{\sigma}}$ or even exclusively depend on $\hat{\boldsymbol{\sigma}}$ as in case $h = 0$. Hence, it is mainly the underestimation of $V^{(2,2)}$, the estimation uncertainty in $\boldsymbol{\sigma}$, that deteriorates coverage accuracies in the presence of conditional heteroskedasticity. We observe a similar pattern in the finite sample simulations discussed in the next subsection, i.e. the coverage deficiencies vanish for later response horizons.

Finally, 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.

5.3.3 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.[5] These GARCH parameters, together with the normality assumption on ε_t, guarantee that Assumption 4.1 is satisfied. For each DGP we generate $M = 5000$ sets of time series data of length $T = 500$ and $T = 5000$ and construct

[4] This is in line with earlier simulation results in the literature which show that ignoring estimation uncertainty in the variance of the errors or misspecifying the unconditional distribution of the bootstrap error term undermines coverage accuracy only at short horizons, see Kilian & Kim (2011, footnote 5) and Kilian (1998*a*), respectively.

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

impulse response intervals using Hall's percentile method which is outlined in Steps 3 and 4 of Bootstrap Scheme II presented in Section 5.2. The *i.i.d.* bootstrap, recursive- and fixed-design wild bootstraps, pairwise bootstrap as well as the MBB bootstrap are considered. 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 5.1. As mentioned in Remark 2.2, we use the estimator of Boubacar Maïnassara & Francq (2011, Section 4) which involves a VARHAC-type approach to estimate the asymptotic variance matrix V.

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 and 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 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 result in intervals with coverage rates above the nominal level. Note that these results nicely line up with the findings 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 C and D 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 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 5.1. 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.

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 $T = 500$. We find that 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 on impact and for 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.

We have conducted further experiments but for the sake of brevity we only summarize the findings without reporting detailed results. First, we try the residual correlations $\rho = 0.1$ 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 structural impulse responses 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.

6 Conclusions

We have provided theoretical results for inference in stable VAR models with uncorrelated innovations satisfying a mixing assumption. This weak VAR framework covers general deviations from independence including the important case of conditional heteroskedasticity of unknown form. We have considered joint inference on the VAR parameters and the unconditional innovation variance parameters in order to provide asymptotically valid inference tools regarding quantities that are functions of both sets of parameters, as e.g. in the case of impulse responses to orthogonalized shocks. In particular we show that under our weak VAR assumptions a residual-based moving block bootstrap leads to asymptotically valid inference while the commonly applied wild and pairwise bootstrap schemes fail in this respect.

In the context of impulse responses to structural shocks, our simulation results indicate that estimation uncertainty may increase dramatically in the presence of conditional heteroskedasticity compared to an i.i.d. set-up. Note that this is not merely a finite sample issue but is rather due to the asymptotic properties. We point out that the asymptotically valid Delta method and bootstrap approaches often underestimate the true sampling variation at shorter impulse response horizons. Moreover, the bootstrap schemes which are asymptotically invalid do not need to perform worse than the MBB if the sample size is small. Our results highlight that practitioners should be aware of the fact that reported impulse response intervals may understate the actual estimation uncertainty substantially in the presence of conditional heteroskedasticity. Therefore, interpreting the confidence intervals for structural impulse responses should be done cautiously against this background.

There are two interesting extensions of our framework. The first one 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 2.1 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. Second, it is worthwhile to extend the current framework to sieve vector autoregressions in which the true lag order is not finite. Hence, a counterpart to Inoue & Kilian (2002) is required but allowing for conditional heteroskedasticity instead of i.i.d. errors as considered in Inoue & Kilian (2002). This line of further research is also linked to Gonçalves & Kilian (2007) who allow for conditional heteroskedasticity in infinte-order univariate autoregressions.

Acknowledgements

We would like to thank the co-managing editor and two anonymous referees of this journal. We also thank Silvia Gonçalves, Jens-Peter Kreiss, Ulrich K. Müller as well as participants of the conferences "Recent Developments for Bootstrap Methods in Time Series" (Copenhagen, September 2013), ESEM (Toulouse, August 2014), CFE-ERCIM (Pisa, December 2014) and of the econometric meeting of the German Economic Association (Rauischolzhausen, February 2015) for very helpful comments. The research was supported by the Deutsche Forschungsgemeinschaft (DFG) through the SFB 884 'Political Economy of Reforms' and the project BR 2941/1-2.

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.

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.

Boubacar Maïnassara, Y. & Francq, C. (2011), 'Estimating structural VARMA models with uncorrelated but non-independent error terms', *Journal of Multivariate Analysis* **102**(3), 496–505.

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.

Brüggemann, R., Jentsch, C. & Trenkler, C. (2014), Inference in VARs with Conditional Heteroskedasticity of Unknown Form, Working Paper Series of the Department of Economics, University of Konstanz 2014-13, Department of Economics, University of Konstanz.

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.

Dufour, J. M. & Pelletier, D. (2011), 'Practical methods for modelling weak VARMA processes: Identification, estimation and specification with a macroeconomic application', *Discussion Paper, McGill University, CIREQ and CIRANO* .

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. & Raïssi, H. (2007), 'Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors', *Journal of Time Series Analysis* **28**(3), 454–470.

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.

Ghazal, G. A. & Neudecker, H. (2000), 'On second-order and fourth-order moments of jointly distributed random matrices: a survey', *Linear Algebra and its Applications* **321**(1-3), 61 – 93. Eighth Special Issue on Linear Algebra and Statistics.

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,* **183**, 104–116.

Inoue, A. & Kilian, L. (2002), 'Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models', *International Economic Review* **43**(2), 309–332.

Jentsch, C., Paparoditis, E. & Politis, D. N. (2014), 'Block bootstrap theory for multivariate integrated and cointegrated time series', *Journal of Time Series Analysis,* **36**(3), 416–441.

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.

Kilian, L. & Kim, Y. J. (2011), 'How reliable are local projection estimators of impulse responses?', *Review of Economics and Statistics* **93**, 1460–1466.

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.

Politis, D. & Romano, J. (1992*a*), 'A generale resampling scheme for triangular arrays of α-mixing random variables wirth application to the problem of spectral density estimation', *The Annals of Statistics* **20**, 1985–2007.

Politis, D. & Romano, J. (1992*b*), A nonparametric resampling procedure for multivariate confidence regions in time series analysis, *in* C. Page & R. LePage, eds, 'Computing Science and Statistics', Proceedings of the 22nd Symposium on the Interface, Springer-Verlag, New York, pp. 98–103.

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.

Table 1: Moments for GARCH(1,1) model (3.1)

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 $\eta_t \sim \; i.i.d.N(0,1)$.

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 5.1. The columns headed by *Wild Bootstrap* and *Pairwise Bootstrap* show the corresponding entries for the asymptotic quantities when using $\Sigma_{\widehat{\Theta}_i}^{PB}$ and $\Sigma_{\widehat{\Theta}_i}^{WB}$, respectively. The wild bootstrap is based on $\eta_t \sim \; i.i.d.N(0,1)$.

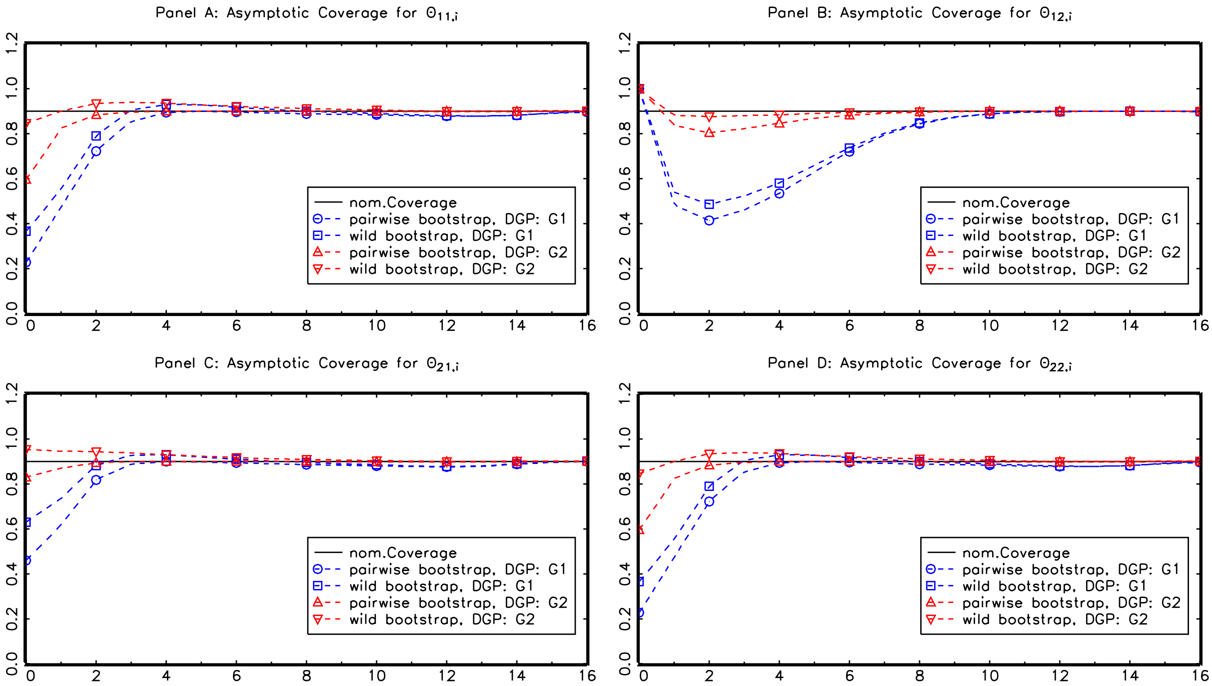

Figure 1: Asymptotic coverage probabilities of pairwise and wild bootstrap impulse response intervals. DGP: VAR(2) with LC-GARCH(1,1) innovations G1 and G2 as in Table 1.

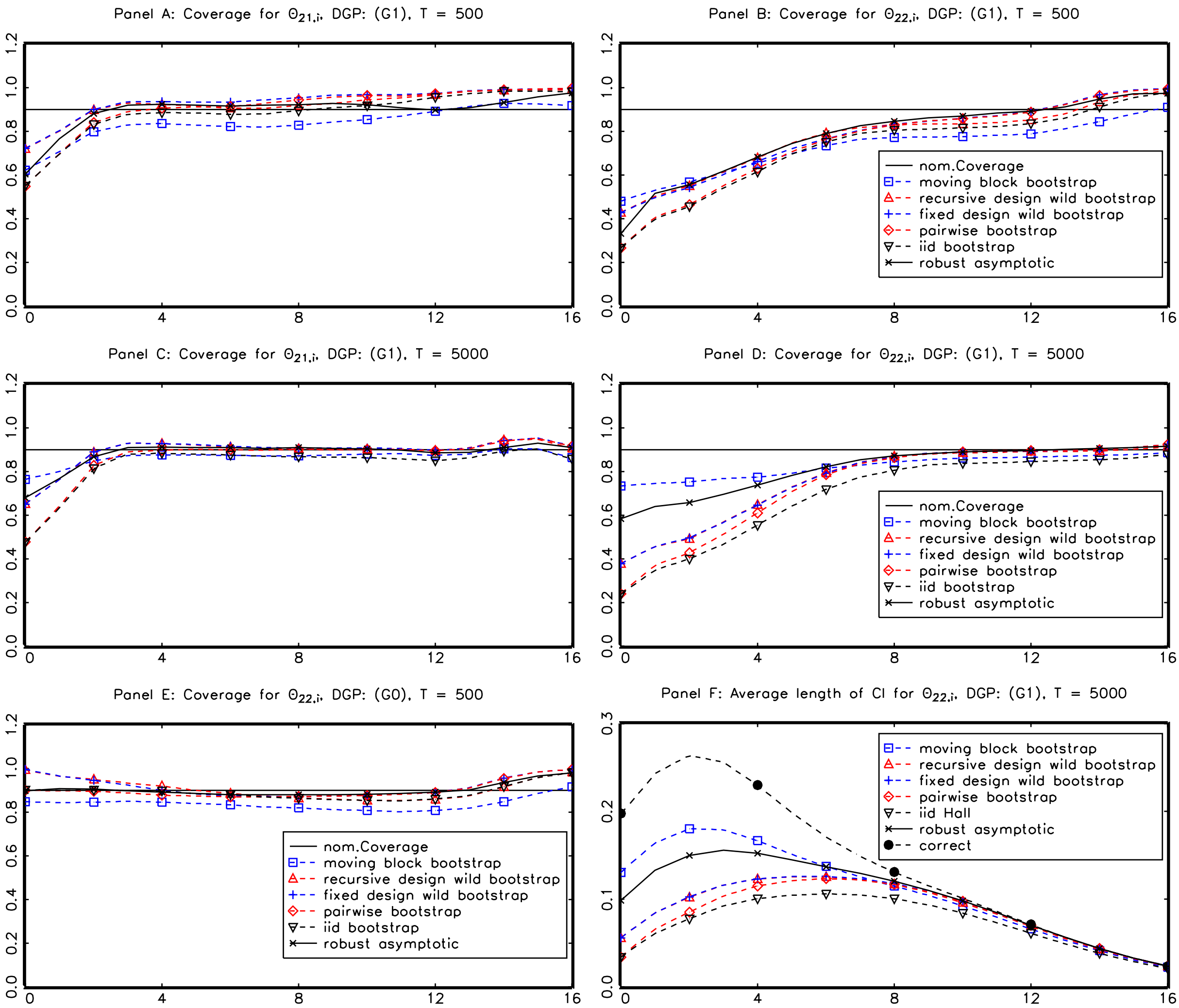

Figure 2: Empirical coverage rates of bootstrap and asymptotic impulse response intervals. Moving block bootstrap (MBB) block length and VARHAC lag order: $\ell = 50$ and $p_{max} = 8$ ($T = 500$) and $\ell = 200$ and $p_{max} = 16$ ($T = 5000$), DGP: VAR(2) with GARCH innovations G1 as in Table 1.

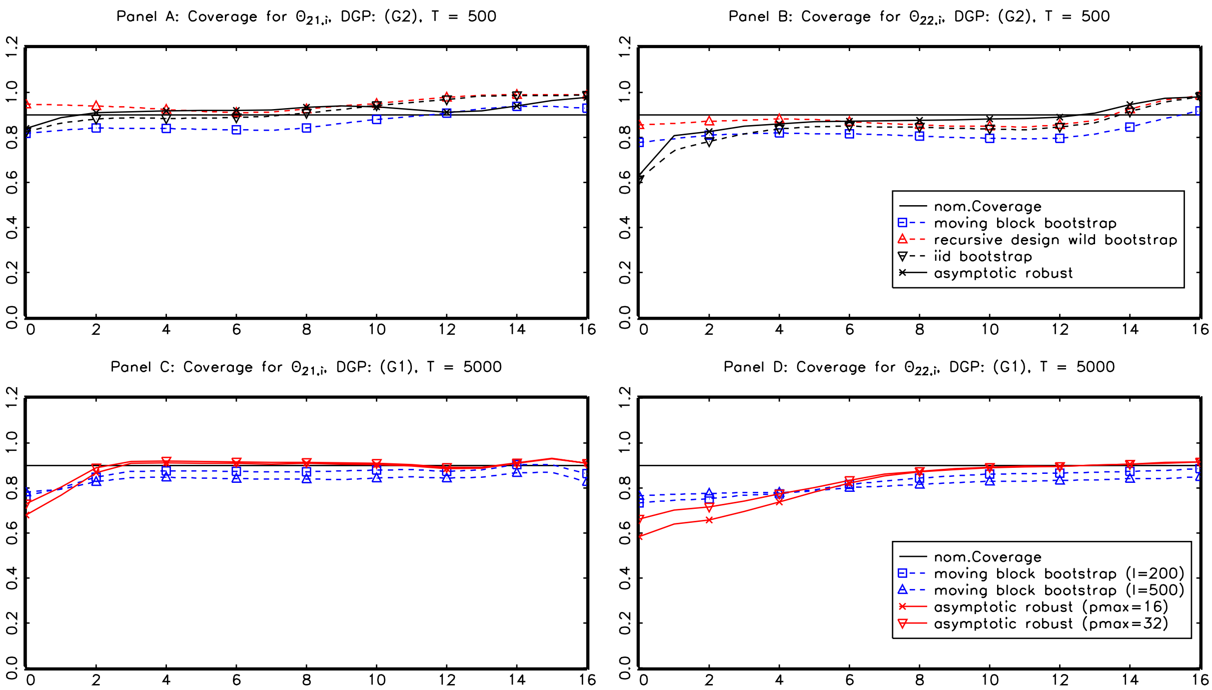

Figure 3: Empirical coverage rates of bootstrap and asymptotic impulse response intervals. Panels A and B: DGP: VAR(2) with GARCH innovations G2 as in Table 1, $T = 500$. Panels C and D: Moving block bootstrap block length ℓ and VARHAC lag order p_{max}. DGP: VAR(2) with GARCH innovations G1 as in Table 1, $T = 5000$.

A Proofs

A.1 Proof of Variance Matrix in Theorem 2.1

Asymptotic normality follows from Theorems 1 and 2 of Boubacar Maïnassara & Francq (2011) with Assumption 2.1. For the limiting variance derived there, we obtain for the VAR set-up

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

with

$$J = \begin{pmatrix} J_{11} & 0 \\ 0 & J_{22} \end{pmatrix},$$

$J_{11} = 2E(Z_{t-1}Z_{t-1}^{\prime}) \otimes \Sigma_u^{-1} = 2\Gamma \otimes \Sigma_u^{-1}$, $J_{22} = D_K^{\prime}(\Sigma_u^{-1} \otimes \Sigma_u^{-1})D_K$. Moreover,

$$B = \sum_{h=-\infty}^{\infty} \begin{pmatrix} \mathrm{Cov}\left(\Upsilon_t^{(1)}, \Upsilon_{t-h}^{(1)}\right) & \mathrm{Cov}\left(\Upsilon_t^{(1)}, \Upsilon_{t-h}^{(2)}\right) \\ \mathrm{Cov}\left(\Upsilon_t^{(2)}, \Upsilon_{t-h}^{(1)}\right) & \mathrm{Cov}\left(\Upsilon_t^{(2)}, \Upsilon_{t-h}^{(2)}\right) \end{pmatrix} = \sum_{h=-\infty}^{\infty} \mathrm{Cov}(\Upsilon_t, \Upsilon_{t-h}) \qquad \text{(A.1)}$$

with

$$\Upsilon_t = \begin{pmatrix} \Upsilon_t^{(1)} \\ \Upsilon_t^{(2)} \end{pmatrix} = \begin{pmatrix} 2(Z_{t-1} \otimes I_K)\Sigma_u^{-1}u_t \\ D_K^{\prime}\left(\mathrm{vec}(\Sigma_u^{-1}) - \mathrm{vec}(\Sigma_u^{-1}u_t u_t^{\prime}\Sigma_u^{-1})\right) \end{pmatrix}. \qquad \text{(A.2)}$$

By inserting $Z_{t-1} = \sum_{j=1}^{\infty} C_j u_{t-j}$ into (A.2), determining the covariances in (A.1), and using

$$J^{-1} = \begin{pmatrix} 0.5(\Gamma^{-1} \otimes \Sigma_u) & 0 \\ 0 & D_K^{+}(\Sigma_u \otimes \Sigma_u)D_K^{+\prime} \end{pmatrix}$$

we obtain the expression for V in Theorem 2.1. □

A.2 Proof of Theorem 4.1

By Polya's Theorem and by Lemma A.1 below, 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 2.1, where $\widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}} := ((\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1}\widetilde{Z}^* \otimes I_K)\widetilde{\mathbf{u}}^*$, $\widetilde{\boldsymbol{\sigma}}^* = \mathrm{vech}(\widetilde{\Sigma}_u^*)$ with $\widetilde{\Sigma}_u^* = \frac{1}{T}\sum_{t=1}^{T} \widetilde{u}_t^* \widetilde{u}_t^{*\prime}$ and $\widetilde{\boldsymbol{\sigma}} = \mathrm{vech}(\widetilde{\Sigma}_u)$ with $\widetilde{\Sigma}_u = \frac{1}{T}\sum_{t=1}^{T} u_t u_t^{\prime}$. Here, pre-sample values $\widetilde{y}_{-p+1}^*, \dots, \widetilde{y}_0^*$ are set to zero and $\widetilde{y}_1^*, \dots, \widetilde{y}_T^*$ is generated according to

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

where $\widetilde{u}_1^*, \dots, \widetilde{u}_T^*$ is an analogously drawn version of $u_1^*, \dots, u_T^*$ as described in Steps 2. and 3. of the bootstrap procedure in Section 4, but from $u_1, \dots, u_T$ instead of $\widehat{u}_1, \dots, \widehat{u}_T$. Further, we use the notation

$$\begin{aligned} \widetilde{Z}_t^* &= \mathrm{vec}(\widetilde{y}_t^*, \dots, \widetilde{y}_{t-p+1}^*) \quad (Kp \times 1) \\ \widetilde{Z}^* &= (\widetilde{Z}_0^*, \dots, \widetilde{Z}_{T-1}^*) \quad (Kp \times T) \end{aligned}$$

$$\widetilde{\mathbf{u}}^* = \text{vec}(\widetilde{u}_1^*, \ldots, \widetilde{u}_T^*) \quad (KT \times 1).$$

By using (A.4) below, we get the representation

$$\begin{aligned}
\sqrt{T}\begin{pmatrix} \widetilde{\boldsymbol{\beta}}^* - \widetilde{\boldsymbol{\beta}} \\ \widetilde{\boldsymbol{\sigma}}^* - \widetilde{\boldsymbol{\sigma}} \end{pmatrix} &= \begin{pmatrix} \left\{(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1} \otimes I_K\right\} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{j=1}^{t-1} (C_j \otimes I_K) \left\{\text{vec}(\widetilde{u}_t^* \widetilde{u}_{t-j}^{*\prime})\right\} \\ \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} L_K \left\{\text{vec}(\widetilde{u}_t^* \widetilde{u}_t^{*\prime}) - \text{vec}(u_t u_t')\right\}\right) \end{pmatrix} \\
&= \begin{pmatrix} \left\{(\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime})^{-1} \otimes I_K\right\} \frac{1}{\sqrt{T}} \sum_{j=1}^{T-1} (C_j \otimes I_K) \sum_{t=j+1}^{T} \left\{\text{vec}(\widetilde{u}_t^* \widetilde{u}_{t-j}^{*\prime})\right\} \\ \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} L_K \left\{\text{vec}(\widetilde{u}_t^* \widetilde{u}_t^{*\prime}) - \text{vec}(u_t u_t')\right\}\right) \end{pmatrix} \quad \text{(A.3)} \\
&= A_m^* + (A^* - A_m^*),
\end{aligned}$$

where L_K is the elimination matrix defined in Theorem 2.1, A^* denotes the right-hand side of (A.3) and A_m^* is the same expression, but with $\sum_{j=1}^{T-1}$ replaced by $\sum_{j=1}^{m}$ for some fixed $m \in \mathbb{N}$, $m < T$. In the following, we make use of Proposition 6.3.9 of Brockwell & Davis (1991) and it suffices to show

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

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

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

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.2, we have that $\widetilde{Q}_T^* \to Q$ with respect to P^*, where $Q = \text{diag}(\Gamma^{-1} \otimes I_K, I_{\widetilde{K}})$. Now, the CLT required for part (a) follows from Lemma A.3 with

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

which leads to $V^{(2,2)} = \Omega^{(2,2)}$ defined in (A.8), $V_m^{(2,1)} = V_m^{(1,2)\prime}$ and

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

Part (b) follows from summability of C_j and uniform boundedness of $\sum_{h=-\infty}^{\infty} \tau_{0,i,h,h+j}$ for $i, j \in \mathbb{N}$ which is implied by the cumulant condition of Assumption 4.1. 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 $\lambda \in \mathbb{R}^{K^2 p}$ and $\delta > 0$, then we get by the Markov inequality

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

By the cumulant condition in Assumption 4.1, it is straightforward, but tedious to show that

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

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

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

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

Lemma A.1 (Equivalence of bootstrap estimators).
Under the assumptions of Theorem 4.1, we have

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

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

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

$$= \left((\frac{1}{T} Z^* Z^{*\prime})^{-1} \otimes I_K \right) A_1^* + A_2^* \frac{1}{\sqrt{T}} (\widetilde{Z}^* \otimes I_K) \widetilde{\mathbf{u}}^*$$

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. We will make use of

$$\begin{aligned} Z_{t-1}^* &= \begin{pmatrix} y_{t-1}^* \\ \vdots \\ y_{t-p}^* \end{pmatrix} = \begin{pmatrix} \sum_{j=0}^{t-1+1} \widehat{\Phi}_j u_{t-1-j}^* \\ \vdots \\ \sum_{j=0}^{t-p+1} \widehat{\Phi}_j u_{t-p-j}^* \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{t-1} \widehat{\Phi}_{j-1} u_{t-j}^* \\ \vdots \\ \sum_{j=p}^{t-1} \widehat{\Phi}_{j-p} u_{t-j}^* \end{pmatrix} \\ &= \begin{pmatrix} \sum_{j=1}^{t-1} \widehat{\Phi}_{j-1} u_{t-j}^* \\ \vdots \\ \sum_{j=1}^{t-1} \widehat{\Phi}_{j-p} u_{t-j}^* \end{pmatrix} = \sum_{j=1}^{t-1} \widehat{C}_j u_{t-j}^*, \end{aligned} \quad \text{(A.4)}$$

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

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

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 4.1 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.2 (Convergence of $\frac{1}{T}\widetilde{Z}^*\widetilde{Z}^{*\prime}$)**.**
Under the assumptions of Theorem 4.1, 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

$$
\begin{aligned}
A_2^* &= \sum_{s=1}^{T-1}\frac{1}{T}\sum_{t=s+1}^{T} C_s\widetilde{u}^*_{t-s}\widetilde{u}^{*\prime}_{t-s}C'_s \\
&= \sum_{s=1}^{m-1} C_s\left(\frac{1}{T}\sum_{t=s+1}^{T}\widetilde{u}^*_{t-s}\widetilde{u}^{*\prime}_{t-s}\right)C'_s + \sum_{s=m}^{T-1} C_s\left(\frac{1}{T}\sum_{t=s+1}^{T}\widetilde{u}^*_{t-s}\widetilde{u}^{*\prime}_{t-s}\right)C'_s \\
&= A_2^{*m} + (A_2^{*m} - A_2^*).
\end{aligned}
$$

Under the imposed summability conditions for the cumulants in Assumption 4.1, 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 $A_2^{*m} \to \Gamma^m = \sum_{s=1}^{m-1} C_s\Sigma_u C'_s$ with respect to P^*, respectively, as $T \to \infty$ and also to $\Gamma^m \to \Gamma$ as $m \to \infty$. Further, we have

$$
\begin{aligned}
E^*\left(|A_2^{*m} - A_2^*|_1\right) &\leq \sum_{j=m}^{T-1}\sum_{r,s=1}^{Kp}\sum_{f,g=1}^{K}|C_j(r,f)|\left(\frac{1}{T}\sum_{t=j+1}^{T}E^*\left(|\widetilde{u}^*_{t-j,f}\widetilde{u}^*_{t-j,g}|\right)\right)|C_j(s,g)| \\
&\leq \sum_{j=m}^{T-1}\sum_{r,s=1}^{Kp}\sum_{f,g=1}^{K}|C_j(r,f)|\left(\frac{1}{T}\sum_{t=1}^{T}E^*\left(|\widetilde{u}^*_{t,f}\widetilde{u}^*_{t,g}|\right)\right)|C_j(s,g)| \\
&\leq \left(\sum_{j=m}^{T-1}|C_j|_1^2\right)\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{f=1}^{K}E^*\left(\widetilde{u}^{*2}_{t,f}\right)\right)
\end{aligned}
$$

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 4.1, we get easily that the second factor on the last right-hand side is bounded in probability and this leads to

$$
E^*\left(|A_2^{*m} - A_2^*|_1\right) \leq \left(\sum_{j=m}^{\infty}|C_j|_1^2\right)O_P(1) \to 0
$$

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

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

for some fixed $m \in \mathbb{N}$. Now, it is straightforward to show that $\frac{1}{T}\sum_{t=h+1+s}^{T} u^*_{t-(s+h)}u^{*\prime}_{t-s} \to 0$ 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

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

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

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

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

Lemma A.3 (CLT for bootstrap innovations)**.**
Let $m \in \mathbb{N}$ fixed and define $\widetilde{W}^*_{t,m} = (\text{vec}(\widetilde{u}^*_t\widetilde{u}^{*\prime}_{t-1})', \ldots, \text{vec}(\widetilde{u}^*_t\widetilde{u}^{*\prime}_{t-m})', L_K\{\text{vec}(\widetilde{u}^*_t\widetilde{u}^{*\prime}_t)' - \text{vec}(u_t u_t')'\})'$. Under the assumptions of Theorem 4.1, it holds

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

in probability, where Ω_m is a block matrix

$$\Omega_m = \begin{pmatrix} \Omega_m^{(1,1)} & \Omega_m^{(1,2)} \\ \Omega_m^{(2,1)} & \Omega^{(2,2)} \end{pmatrix} \tag{A.5}$$

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

$$\Omega_m^{(1,1)} = \left(\sum_{h=-\infty}^{\infty} \tau_{0,i,h,h+j} \right)_{i,j=1,\dots,m} \tag{A.6}$$

$$\Omega_m^{(2,1)} = \sum_{h=-\infty}^{\infty} L_K \left(\tau_{0,0,h,h+1}, \dots, \tau_{0,0,h,h+m} \right), \tag{A.7}$$

$$\Omega^{(2,2)} = \sum_{h=-\infty}^{\infty} L_K \left\{ \tau_{0,0,h,h} - \operatorname{vec}(\Sigma_u) \operatorname{vec}(\Sigma_u)' \right\} L_K', \tag{A.8}$$

respectively.

Proof.
We consider $\widehat{W}^*_{t,m} = (L_K\{\operatorname{vec}(\widetilde{u}^*_t \widetilde{u}^{*\prime}_t)' - \operatorname{vec}(u_t u_t')'\}, \operatorname{vec}(\widetilde{u}^*_t \widetilde{u}^{*\prime}_{t-1})', \dots, \operatorname{vec}(\widetilde{u}^*_t \widetilde{u}^{*\prime}_{t-m})')'$, 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

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

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

this leads to

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

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

$$\begin{aligned}
&\frac{1}{N}\sum_{r_1,r_2=1}^{N}\frac{1}{\ell}\sum_{s_1=(r_1-1)\ell+1}^{(r_1-1)\ell+q}\sum_{s_2=(r_1-1)\ell+1}^{(r_1-1)\ell+q}E^*(\widetilde{u}^*_{s_1,f}\widetilde{u}^*_{s_1-q,g}\widetilde{u}^*_{s_2,f}\widetilde{u}^*_{s_2-q,g})\\
=\;&\frac{1}{N}\sum_{r=1}^{N}\frac{1}{\ell}\sum_{s_1,s_2=1}^{q}E^*(\widetilde{u}^*_{s_1+(r-1)\ell,f}\widetilde{u}^*_{s_2+(r-1)\ell,f})E^*(\widetilde{u}^*_{s_1+(r-1)\ell-q,g}\widetilde{u}^*_{s_2+(r-1)\ell-q,g})\\
=\;&\frac{1}{\ell}\sum_{s_1,s_2=1}^{q}\left(\frac{1}{T-\ell+1}\sum_{t=0}^{T-\ell}u^c_{t+s_1,f}u^c_{t+s_2,f}\right)\left(\frac{1}{T-\ell+1}\sum_{t=0}^{T-\ell}u^c_{t+s_1-q,g}u^c_{t+s_2-q,g}\right)\\
=\;&O_P(\ell^{-1})=o_P(1)
\end{aligned}$$

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 $X^*_{r,T}X^{*\prime}_{r,T}$. With $(q_1,q_2)\in\{0,\ldots,m\}^2$, we obtain

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

$$= \frac{1}{\ell} \sum_{s_1,s_2=q+1}^{\ell} \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s_1,f_1} u^c_{t+s_1-q_1,g_1} u^c_{t+s_2,f_2} u^c_{t+s_2-q_2,g_2} \right. \tag{A.9}$$
$$\left. - \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s_1,f_1} u^c_{t+s_1-q_1,g_1} \right) \left(\frac{1}{T-\ell+1} \sum_{t=0}^{T-\ell} u^c_{t+s_2,f_2} u^c_{t+s_2-q_2,g_2} \right) \right).$$

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

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

Due to strict stationarity of the innovation process, its mean computes to

$$\left(\sum_{h=0}^{\ell-q_1-1} \frac{\ell - \min(q_1+h, q_2)}{\ell} + \sum_{h=-(\ell-q_2-1)}^{-1} \frac{\ell - \min(q_1, q_2-h)}{\ell} \right) E(u_{0,f_1} u_{-q_1,g_1} u_{-h+q_2-q_1,f_2} u_{-h-q_1,g_2})$$

which converges to

$$\sum_{h=-\infty}^{\infty} E(u_{0,f_1} u_{-q_1,g_1} u_{-h+q_2-q_1,f_2} u_{-h-q_1,g_2}) = \sum_{h=-\infty}^{\infty} E(u_{0,f_1} u_{-q_1,g_1} u_{-h,f_2} u_{-h-q_2,g_2})$$

as $\ell \to \infty$ with $T \to \infty$ by assumption and q_1, q_2 remain fixed. Its variance vanishes asymptotically and, as all other summands of (A.9) are of lower order. This leads to the corresponding entry of $\sum_{h=-\infty}^{\infty} \tau_{0,q_1,h,h+q_2}$ 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 Lyapunov condition (ii) for $\delta = 2$ and as the denominator is bounded, it suffices to consider the numerator only. For one entry, we get

$$\sum_{r=1}^{N} \frac{1}{T^2} \sum_{s_i=(r-1)\ell+q_i+1,\ i=1,\dots,4}^{r\ell} \left(E^* \left(\widetilde{u}^*_{s_1,f_1} \widetilde{u}^*_{s_1-q_1,g_1} \widetilde{u}^*_{s_2,f_2} \widetilde{u}^*_{s_2-q_2,g_2} \widetilde{u}^*_{s_3,f_3} \widetilde{u}^*_{s_3-q_3,g_3} \widetilde{u}^*_{s_4,f_4} \widetilde{u}^*_{s_4-q_4,g_4} \right) \right.$$
$$\left. - E^* \left(\widetilde{u}^*_{s_1,f_1} \widetilde{u}^*_{s_1-q_1,g_1} \widetilde{u}^*_{s_2,f_2} \widetilde{u}^*_{s_2-q_2,g_2} \right) E^* \left(\widetilde{u}^*_{s_3,f_3} \widetilde{u}^*_{s_3-q_3,g_3} \widetilde{u}^*_{s_4,f_4} \widetilde{u}^*_{s_4-q_4,g_4} \right) \right)$$

and by the moment condition of Assumption 4.1, the last expression can be shown to be of order $O_P(\ell^3/T) = o_P(1)$ under the assumptions. □