Testing Identification via Heteroskedasticity in Structural Vector Autoregressive Models

Helmut Lütkepohl[1]
DIW Berlin and Freie Universität Berlin,
Mohrenstr. 58, 10117 Berlin, Germany
email: hluetkepohl@diw.de

Mika Meitz
Discipline of Economics, University of Helsinki,
P. O. Box 17, FI-00014 University of Helsinki, Finland
email: mika.meitz@helsinki.fi

Aleksei Netšunajev
Tallinn University of Technology,
Akademia tee 3, 12618 Tallinn, Estonia
email: aleksei.netsunajev@ttu.ee

Pentti Saikkonen
Department of Mathematics and Statistics, University of Helsinki,
P. O. Box 68, FI-00014 University of Helsinki, Finland
email: pentti.saikkonen@helsinki.fi

July 9, 2019

[1]Part of the research for this paper was done while the first author was a Fernand Braudel Fellow in the Economics Department of the European University Institute, Florence. The second and fourth authors thank the Academy of Finland (grant number 1308628) for financial support. The paper was presented at seminars at the European University Institute, Florence, Tor Vergata University of Rome, Erasmus University Rotterdam and Leipzig University. It was also presented at the IXth Workshop in Time Series Econometrics in Zaragoza, Spain in April 2019 and at the meeting of the Econometrics Committee of the Verein für Socialpolitik at Rauischholzhausen, Germany in February 2019. We thank the participants for helpful comments.

Abstract. Tests for identification through heteroskedasticity in structural vector autoregressive analysis are developed for models with two volatility states where the time point of volatility change is known. The tests are Wald type tests for which only the unrestricted model including the covariance matrices of the two volatility states have to be estimated. The residuals of the model are assumed to be from the class of elliptical distributions which includes Gaussian models. The asymptotic null distributions of the test statistics are derived and simulations are used to explore their small sample properties. Two empirical examples illustrate the usefulness of the tests in applied work.

Key Words: Heteroskedasticity, structural identification, vector autoregressive process

JEL classification: C32

1 Introduction

Identification by heteroskedasticity of the shocks has become a standard tool in structural vector autoregressive (VAR) analysis (see, e.g., Kilian and Lütkepohl (2017, Chapter 14)). Heteroskedasticity can complement identifying restrictions based on economic theory or subject matter knowledge. The underlying idea is that if the variance of the structural shocks changes during the sample period and there is heterogeneity in the variance changes of different shocks, this feature can be used to distinguish (identify) the shocks. The objective of this study is to present a formal statistical test for the required variance heterogeneity and hence the identification of the shocks.

Of course, identifying structural shocks purely through their statistical properties implies that a further step is necessary to associate the identified shocks with economic shocks of interest. Therefore identification through heteroskedasticity is often complemented by conventional identifying restrictions, for example, on the impact effects of the shocks. In particular, if competing identifying restrictions are available, which are just-identifying in a conventional setting, identifying information from heteroskedasticity can be used as over-identifying restrictions which opens up the possibility to formally test identifying restrictions that are otherwise not testable. Such tests are not considered in this paper but are discussed, for example, in Lanne and Lütkepohl (2008), Netšunajev (2013) and Lütkepohl and Netšunajev (2017a).

A main advantage of identification via heteroskedasticity is that the data are in principle informative on the conditions for identification. Thus, identification can be investigated by statistical tests. The problem in developing such tests is that the model is typically not identified under the null hypothesis of no identification which complicates the derivation of the asymptotic distributions of standard tests. Some authors still use standard Wald and likelihood ratio (LR) tests for identification through heteroskedasticity and approximate the distribution under the null hypothesis by the usual χ^2 distributions. Examples are Lanne, Lütkepohl and Maciejowska (2010), Herwartz and Lütkepohl (2014), Lütkepohl and Velinov (2016), Velinov and

Chen (2015), Netšunajev (2013) and Lütkepohl and Netšunajev (2014). However, so far the asymptotic distributions of these tests have not been derived formally and it is unlikely that the assumed χ^2 distributions provide precise approximations to the true asymptotic distributions of the test statistics. Alternatively, some authors have proposed Bayesian methods for assessing identification in this context (e.g., Woźniak and Droumaguet (2015) and Lütkepohl and Woźniak (2017)).

In the following we will develop formal frequentist tests that can help in assessing identification through heteroskedasticity for the special case of stable VAR models with two volatility regimes of the residuals. Such simple models for the change in volatility have been considered, e.g., by Rigobon (2003), Lanne and Lütkepohl (2008, 2014), and Lütkepohl and Schlaak (2018).[2] For developing our tests, we assume that the distribution of the residuals is elliptically symmetric which covers the case of Gaussian VAR processes but also models where the residuals have t distributions or mixtures of normal distributions. We develop Wald type tests for which we can derive the asymptotic distribution under the null hypothesis of no identification. Our results shed further doubts on the previously assumed test distributions for related statistics. Of course, if the test indicates identification through heteroskedasticity, the identifying information may still be limited. In other words, there may still be weak identification. This issue in the context of identification through heteroskedasticity has been discussed by Lewis (2018) and more general results on weak identification are given by Andrews (2018). Such procedures may be used in addition or alternatively to our approach.

Our tests may only indicate that there is some identifying information through heteroskedasticity but may not suggest that the structural model is fully identified. We discuss a sequential testing procedure that can be helpful in this context. Our tests can be used at different stages of such a procedure. Then the properties of the full sequential procedure may be of interest

[2]The actual models used for empirical analysis are more sophisticated in some of these articles, where more volatility states and alternative volatility models are considered as well. We focus on the simpler case to make the problem tractable and leave more general models for future research.

which we do not consider theoretically. However, we show by simulation that the asymptotic theory is a good guide for small sample performance of the individual tests, if the sample size is sufficiently large and we also explore the properties of the sequential procedure by simulation. Finally, we present examples which illustrate the usefulness of our tests for applied work.

The remainder of this study is structured as follows. The model is set up in the following section. Section 3 presents the tests for identification and their asymptotic properties. Section 4 considers the small sample properties of the tests. Two empirical examples based on US data are discussed in Section 5. The final section concludes. The proofs of the asymptotic results for the test statistics are provided in Appendix A.

2 The Model

Consider a stable, stationary K-dimensional reduced-form VAR(p) model

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

where ν is an intercept term and A_j $(j = 1, \ldots, p)$ are $(K \times K)$ VAR slope coefficient matrices satisfying the usual stability condition

$$\det\left(I_K - A_1 z - \cdots - A_p z^p\right) \neq 0 \quad \text{for} \quad |z| \leq 1. \tag{2}$$

The error process u_t is white noise with zero mean, $\mathbb{E}(u_t) = 0$, and (positive definite) covariance matrices

$$\mathbb{E}(u_t u_t') = \begin{cases} \Sigma_1 & \text{for} \quad t \in \mathcal{T}_1 = \{1, \ldots, T_1\}, \\ \Sigma_2 & \text{for} \quad t \in \mathcal{T}_2 = \{T_1 + 1, \ldots, T\}, \end{cases} \tag{3}$$

where T signifies the sample size. Thus, the errors of the model are assumed to be heteroskedastic so that the covariance matrix changes from Σ_1 to Σ_2 at time $T_1 + 1$ which we assume to be known. Moreover, we assume that for some fixed fraction $\tau \in (0, 1)$, T_1 is the integer part of τT, i.e., $T_1 = [\tau T]$, so that the sample size for both volatility regimes goes to infinity as $T \to \infty$.

The setup in (3) is used here for convenience. For our asymptotic analysis it is in fact sufficient that $\mathcal{T}_1$ contains a fraction of τ sample periods while

$\mathcal{T}_2$ contains the remaining fraction of $1-\tau$ periods in the sample. In other words, $\mathcal{T}_1$ and $\mathcal{T}_2$ do not have to contain consecutive parts of the sample. The important condition is that the size of both sets goes to infinity with the sample size in such a way that both Σ_1 and Σ_2 can be estimated consistently.

We consider the case where the error term u_t has an elliptically symmetric distribution or briefly an elliptical distribution possessing a density

$$\frac{1}{\sqrt{\det \Sigma_t}} g(u_t' \Sigma_t^{-1} u_t),$$

where Σ_t is a symmetric positive definite matrix, $g(\cdot)$ is a positive function such that the density integrates to one and the fourth moments of the distribution exist (see, e.g., Anderson (2003, Section 2.7) for further discussion of elliptical distributions). Note that the elliptical distributions are such that all components of u_t have the same kurtosis parameter. More precisely, denoting the i^{th} diagonal element of Σ_t by σ_{it}^2, the kurtosis parameter

$$\frac{\mathbb{E}(u_{it}^4)}{3\sigma_{it}^2} - 1$$

is the same for $i = 1, \ldots, K$ (see also Anderson (2003, p. 54, Equation (36))). We explicitly allow for the possibility that the kurtosis parameter may be different for the different volatility regimes and define

$$\frac{\mathbb{E}(u_{it}^4)}{3\sigma_{it}^2} - 1 = \begin{cases} \kappa_1 & \text{for} \quad t \in \mathcal{T}_1, \\ \kappa_2 & \text{for} \quad t \in \mathcal{T}_2. \end{cases}$$

Notice, however, that the case of Gaussian residuals is obtained as a special case by choosing the kurtosis parameter equal to zero. Thus, even if the variance changes across the sample, we may have $\kappa_1 = \kappa_2$, e.g., if the sample is Gaussian. Other distribution families covered by our assumptions are t distributions and mixtures of normal distributions. We need the elliptical distributions to apply limiting results from Anderson (2003) in our derivation of the test for identification in Section 3.

The covariance matrices in (3) can be decomposed as follows:

$$\Sigma_1 = BB', \quad \Sigma_2 = B\Lambda B', \tag{4}$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_K)$ is a $(K \times K)$ diagonal matrix with positive diagonal elements and B is a nonsingular $(K \times K)$ matrix (see Lütkepohl (2013)). A standard assumption in the related structural VAR (SVAR) literature is that only the volatility of the shocks changes while the responses of the variables remain time invariant. This is accomplished by obtaining the structural shocks from the reduced form errors as $\varepsilon_t = B^{-1}u_t$, such that B is the matrix of impact effects of the shocks and the covariance matrices of the structural errors are given by

$$\mathbb{E}(\varepsilon_t \varepsilon_t') = \begin{cases} I_K & \text{for} \quad t \in \mathcal{T}_1, \\ \Lambda & \text{for} \quad t \in \mathcal{T}_2. \end{cases} \tag{5}$$

Thus, the structural errors are instantaneously uncorrelated in both volatility regimes. Replacing the reduced form errors u_t in (1) by the structural errors $B\varepsilon_t$ yields the SVAR(p) model

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B\varepsilon_t. \tag{6}$$

For the statistical results to be obtained later we assume that the structural errors ε_t or, equivalently, the reduced-form errors u_t are temporally independent.

It is well known (see, e.g., Theorem A9.9 and its proof in Muirhead (1982)) that the diagonal elements of the matrix Λ in (4) are the eigenvalues of the matrix $\Sigma_1^{-1}\Sigma_2$ so that they satisfy the (generalized) eigenvalue equations

$$\det(\Sigma_2 - \lambda_i \Sigma_1) = 0, \quad i = 1, \ldots, K, \tag{7}$$

whereas the columns of the matrix $B = [b_1 : \cdots : b_K]$ are the corresponding (generalized) eigenvectors that satisfy

$$(\Sigma_2 - \lambda_i \Sigma_1)\, b_i = 0, \quad i = 1, \ldots, K. \tag{8}$$

Furthermore, if the eigenvalues $\lambda_1, \ldots, \lambda_K$ are distinct, the matrix B is unique apart from permutations and sign reversals of its columns (see the aforementioned theorem of Muirhead (1982) or Lanne et al. (2010, Proposition 1)). In what follows we assume (without loss of generality) that the eigenvalues

$\lambda_1, \ldots, \lambda_K$ are ordered from largest to smallest so that $\lambda_1 \geq \cdots \geq \lambda_K > 0$ holds. If the matrix B is not unique we have an identification problem in the SVAR(p) model (6). Testing for a possible lack of identification is therefore of interest and will be discussed in the next section.

Clearly, our model is special in that it assumes two volatility regimes only. In practice, there may be more than two volatility regimes (see, e.g., Rigobon and Sack (2003), Lanne and Lütkepohl (2008)). In that case the identification conditions become more elaborate and more difficult to test formally (e.g., Kilian and Lütkepohl (2017, Section 14.3.1)). We leave this case for future research. Moreover, more general volatility models have been considered in the literature. For example, Lanne et al. (2010) and Herwartz and Lütkepohl (2014) use a Markov switching process to model endogenously changing volatility regimes. Unfortunately, we do not know whether the tests developed in the following can be extended to that case even in situations where very similar identification conditions are of interest. Again we have to leave extensions to such models for future research.

3 A Test Procedure for Identification of B

3.1 The Testing Problem

Given that the diagonal elements of the matrix Λ are ordered from largest to smallest, uniqueness of the matrix B obtains if $\lambda_1 > \cdots > \lambda_K$ and the possibility of sign reversals in B is eliminated. One possibility to fix the column signs to be used in this study, is to require that the first nonzero element of each column of B is positive. In order to test for lack of identification we consider the pair of hypotheses

$$\mathbb{H}_0 : \lambda_{s+1} = \lambda_{s+2} = \cdots = \lambda_{s+r} \ (= \lambda_0) \quad \text{versus} \quad \mathbb{H}_1 : \neg\mathbb{H}_0 \tag{9}$$

for $s \in \{0, \ldots, K-2\}$ and $r \in \{2, \ldots, K-s\}$. Thus, under the null hypothesis, r consecutive eigenvalues of Λ are equal to a value λ_0, implying lack of identification. The remaining eigenvalues $\lambda_1, \ldots, \lambda_s, \lambda_{s+r+1}, \ldots, \lambda_K$,

may have multiplicities larger than one, but have to be different from λ_0, the common value under $\mathbb{H}_0$.

Let $y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T$ be the available data. The reduced-form Gaussian log-likelihood function (apart from a constant and conditioning on the first p observations $y_{-p+1}, \ldots, y_0$) is given by

$$\begin{aligned} l(\boldsymbol{\vartheta}, \boldsymbol{\sigma}) &= -\frac{1}{2}\sum_{t=1}^{T_1} \log\det(\Sigma_1) - \frac{1}{2}\sum_{t=1}^{T_1} u_t(\boldsymbol{\vartheta})'\Sigma_1^{-1}u_t(\boldsymbol{\vartheta}) \\ &\quad -\frac{1}{2}\sum_{t=T_1+1}^{T} \log\det(\Sigma_2) - \frac{1}{2}\sum_{t=T_1+1}^{T} u_t(\boldsymbol{\vartheta})'\Sigma_2^{-1}u_t(\boldsymbol{\vartheta}), \end{aligned} \tag{10}$$

where $\boldsymbol{\vartheta} = \mathrm{vec}(\nu, A_1, \ldots, A_p)$, $u_t(\boldsymbol{\vartheta})$ signifies u_t in expression (1) when these quantities are interpreted as functions of the underlying parameters and $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2)$ with $\boldsymbol{\sigma}_i = \mathrm{vech}(\Sigma_i)$ $(i = 1, 2)$. Here vec denotes the usual column stacking operator and vech denotes the operator which stacks the columns of a square matrix from the main diagonal downwards. If the DGP is Gaussian, maximizing $l(\boldsymbol{\vartheta}, \boldsymbol{\sigma})$ with respect to the parameters gives the maximum likelihood (ML) estimators and, if the true distribution is not Gaussian but of a more general elliptical form, the resulting estimators are quasi-ML estimators.

Instead of ML estimation one may use a feasible generalized least squares (GLS) procedure. In that case (1) is estimated with equationwise ordinary least squares (OLS) in a first step. The residuals $\hat{u}_t$ obtained in that way are then used for estimating the covariance matrices as

$$\hat{\Sigma}_i = \frac{1}{T_i}\sum_{t\in\mathcal{T}_i} \hat{u}_t\hat{u}_t', \quad i = 1, 2,$$

where $T_2 = T - T_1$. In a further step the GLS estimator

$$\tilde{\boldsymbol{\vartheta}} = \left(\sum_{t=1}^{T} Z_{t-1}Z_{t-1}' \otimes \hat{\Sigma}_t^{-1}\right)^{-1} \left(\sum_{t=1}^{T} (Z_{t-1} \otimes \hat{\Sigma}_t^{-1}) y_t\right), \tag{11}$$

is computed, where $Z_{t-1} = (1, y_{t-1}', \ldots, y_{t-p}')'$ and $\hat{\Sigma}_t = \hat{\Sigma}_i$ for $t \in \mathcal{T}_i$ $(i = 1, 2)$. If the VAR process is stable, these estimators have standard asymptotic properties and can be used accordingly (see Lütkepohl (2005, Chapter 17)). Then the GLS residuals can be used to estimate the covariance matrices Σ_1

and Σ_2. In what follows, $\tilde{\boldsymbol{\vartheta}}$ can be any estimator of $\boldsymbol{\vartheta}$ such that $\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta} = O_p(T^{-1/2})$.

Then one readily finds that $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$ are asymptotically equivalent to their (unfeasible) counterparts based on the reduced form errors or, specifically,

$$\tilde{\Sigma}_1 = \frac{1}{T_1}\sum_{t=1}^{T_1}\tilde{u}_t\tilde{u}_t' = \frac{1}{T_1}\sum_{t=1}^{T_1}u_t u_t' + o_p(T^{-1/2}) \tag{12}$$

$$\tilde{\Sigma}_2 = \frac{1}{T-T_1}\sum_{t=T_1+1}^{T}\tilde{u}_t\tilde{u}_t' = \frac{1}{T-T_1}\sum_{t=T_1+1}^{T}u_t u_t' + o_p(T^{-1/2}), \tag{13}$$

where $\tilde{u}_t$ signifies the residuals described above, i.e., $\tilde{u}_t = y_t - \tilde{\nu} - \tilde{A}_1 y_{t-1} - \cdots - \tilde{A}_p y_{t-p}$ (cf. Proposition 3.2 in Lütkepohl (2005)). Replacing the theoretical covariance matrices Σ_1 and Σ_2 in equations (7) and (8) by the estimators $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$ we obtain the vector of eigenvalues $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K)$ and the matrix of eigenvectors $\tilde{B} = [\tilde{b}_1 : \cdots : \tilde{b}_K]$. Similarly to their theoretical counterparts, the estimated eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K$ are ordered from largest to smallest and, as they are distinct with probability one, we have $\tilde{\lambda}_1 > \cdots > \tilde{\lambda}_K > 0$ almost surely. Eliminating the possibility of sign reversals in $\tilde{B}$ in the same way as in B we therefore have a one-to-one continuous correspondence between the estimators $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$ and the elements of the matrix $\tilde{B}$ and the vector $\tilde{\boldsymbol{\lambda}}$. Thus, $\tilde{B}$ and $\tilde{\boldsymbol{\lambda}}$ can be viewed as unrestricted estimators of B and $\boldsymbol{\lambda}$.

Deriving the asymptotic properties of estimated eigenvalues is known to be a complicated problem when the theoretical eigenvalues are not distinct which is the case under our null hypothesis. In the context of principal component analysis, where the population eigenvalues satisfy equation (7) with $\Sigma_1 = I_K$, and with independent observations a complete solution to this problem is provided by Anderson (1963) (see also Anderson (2003, Sec. 11.7.3), and Muirhead (1982, Sec. 9.5 and 9.6)), whereas Anderson (2003, Sec. 13.6.3) treats the case of a general Σ_1 (again with independent observations). In what follows we adopt Anderson's approach to our problem.

For setting up our test statistics, we also need consistent estimates of the kurtosis parameters. One possible estimator is discussed in Schott (2001, p.

33),

$$\tilde{\kappa}_m = \frac{1}{3K} \sum_{k=1}^{K} \frac{z_k^m}{w_k^m} - 1, \quad m = 1, 2, \tag{14}$$

where

$$z_k^m = \frac{\sum_{t \in \mathcal{T}_m} (\tilde{u}_{kt} - \bar{u}_k^m)^4 - 6\tilde{\sigma}_k^4}{T_m - 4}$$

and

$$w_k^m = \frac{T_m}{T_m - 1} \left(\tilde{\sigma}_k^4 - \frac{z_k^m}{T_m} \right).$$

Here $\bar{u}_k^m = T_m^{-1} \sum_{t \in \mathcal{T}_m} \tilde{u}_{kt}$ is the mean of the residuals associated with the m^{th} volatility regime. Of course, if the u_t are Gaussian and this fact is known to the analyst, the kurtosis parameters can simply be replaced by zero, i.e., $\tilde{\kappa}_1 = \tilde{\kappa}_2 = 0$ in the test statistic. Similarly, if the distribution is such that $\kappa_1 = \kappa_2$ the kurtosis parameter can be estimated from the full sample using the formulas as above based on the full sample.

3.2 The Test Statistic

We base our test statistic on the eigenvalues $\tilde{\lambda}_{s+1}, \ldots, \tilde{\lambda}_{s+r}$. In principal component analysis with Gaussian iid data, the LR test for testing the equality of eigenvalues is based on the ratio of the geometric mean and arithmetic mean of the ML estimators of the eigenvalues assumed to be identical under the null hypothesis (see Anderson (1963) or Anderson (2003, Sec. 11.7.3)). Proceeding according to this pattern, we consider the test statistic

$$\begin{aligned} Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2) &= -c(\tau, \tilde{\kappa}_1, \tilde{\kappa}_2)^2 Tr \log \left(\frac{\prod_{k=s+1}^{s+r} \tilde{\lambda}_k^{1/r}}{\frac{1}{r} \sum_{k=s+1}^{s+r} \tilde{\lambda}_k} \right) \\ &= c(\tau, \tilde{\kappa}_1, \tilde{\kappa}_2)^2 \left[-T \sum_{k=s+1}^{s+r} \log(\tilde{\lambda}_k) + Tr \log \left(\frac{1}{r} \sum_{k=s+1}^{s+r} \tilde{\lambda}_k \right) \right], \end{aligned} \tag{15}$$

where $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ are consistent estimators of the kurtosis parameters and the term

$$c(\tau, \tilde{\kappa}_1, \tilde{\kappa}_2)^2 = \left(\frac{1 + \tilde{\kappa}_1}{\tau} + \frac{1 + \tilde{\kappa}_2}{1 - \tau} \right)^{-1}$$

is included to obtain a convenient limiting distribution. Since the test statistic involves unrestricted estimators only, the test is akin to a Wald test. Of course, other distance measures could also be considered. The following proposition gives the asymptotic distribution of the test statistic under the null hypothesis. It is proven in Appendix A.

Proposition 1. Let u_t have an elliptical distribution possessing a density as well as finite fourth moments with kurtosis parameters κ_i for $t \in \mathcal{T}_i$ $(i = 1, 2)$, where $\mathcal{T}_1 = \{1, \ldots, T_1 = [\tau T]\}$, $\mathcal{T}_2 = \{T_1 + 1, \ldots, T\}$ and the fraction $\tau \in (0, 1)$ is assumed to be known and fixed. Furthermore, let $\lambda_1 \geq \cdots \geq \lambda_K$ be ordered from largest to smallest and let $Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$ be the test statistic defined in equation (15) for testing the pair of hypotheses

$$\mathbb{H}_0 : \lambda_{s+1} = \lambda_{s+2} = \cdots = \lambda_{s+r} (= \lambda_0) \quad \text{versus} \quad \mathbb{H}_1 : \neg \mathbb{H}_0$$

for $s \in \{0, \ldots, K-1\}$ and $r \in \{2, \ldots, K-s\}$. Suppose that $\lambda_s \neq \lambda_{s+1}$ and $\lambda_{s+r} \neq \lambda_{s+r+1}$. Furthermore, let $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ be consistent estimators of κ_1 and κ_2, respectively. Then

$$Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2) \xrightarrow{d} \chi^2(\tfrac{1}{2}(r+2)(r-1)),$$

where $\xrightarrow{d}$ denotes convergence in distribution. □

The χ^2 limiting distribution requires that r eigenvalues are equal to λ_0 and all other eigenvalues are different from λ_0, i.e., $\lambda_s \neq \lambda_{s+1}$ and $\lambda_{s+r} \neq \lambda_{s+r+1}$. In order to ensure this condition, the following considerations may be helpful. If

$$\mathbb{H}_0 : \lambda_1 = \cdots = \lambda_K$$

does not hold, we know that $\lambda_1 \neq \lambda_K$ and Proposition 1 enables us to test

$$\mathbb{H}_0 : \lambda_1 = \cdots = \lambda_{K-1}, \quad \mathbb{H}_0 : \lambda_2 = \cdots = \lambda_K.$$

If these two null hypotheses are false, we can test all null hypotheses involving $K-2$ consecutive eigenvalues etc.. If all null hypotheses in this sequence are

false, we can finally test

$$\mathbb{H}_0 : \lambda_1 = \lambda_2, \ldots, \mathbb{H}_0 : \lambda_{K-1} = \lambda_K.$$

If all the null hypotheses are rejected, the tests support that all the structural parameters are identified via heteroskedasticity.

For example, for $K = 3$, Proposition 1 implies that we can test the null hypothesis

$$\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3$$

using $Q_3(\tilde{\kappa}_1, \tilde{\kappa}_2)$ with a $\chi^2(5)$ distribution. If the null hypothesis is false, it follows that $\lambda_1 \neq \lambda_3$ so that we can test

$$\mathbb{H}_0 : \lambda_1 = \lambda_2 \quad \text{and} \quad \mathbb{H}_0 : \lambda_2 = \lambda_3$$

using $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ statistics with a $\chi^2(2)$ distribution. Rejecting the latter two null hypotheses is evidence of a fully identified structural model via heteroskedasticity.

In the previous literature a related Wald test for equality of two eigenvalues of a similar type is sometimes used with a $\chi^2(1)$ distribution (e.g., Lanne et al. (2010), Velinov and Chen (2015)). Although somewhat different volatility models are used in these publications, Proposition 1 suggests that the $\chi^2(1)$ distribution is a poor approximation to the actual asymptotic distributions of the test statistics. An adjustment of the degrees-of-freedom (df) parameter is likely to be useful. Note that increasing the df parameter increases the correspondingly assumed p-values and, hence, may reduce the number of rejections.

4 Small Sample Properties of Tests for Identification

4.1 Experimental Design

We consider a range of DGPs to investigate the small sample properties of our tests. All DGPs have zero intercept, $\nu = 0$. We still fit VARs with

intercept. The error distributions are either Gaussian, $u_t \sim \mathcal{N}(0, \Sigma_t)$, or have t distributions with 5 degrees of freedom ($t(5)$ distributions). Some of the simulation results are just summarized in the following and numbers are provided in the Supplement for this article, where also further details of some of the DGPs are presented. We report results for the following specific DGPs.

DGP1 Bivariate ($K = 2$) VAR(0) process $y_t = u_t$, with volatility change at $T_1 = \tau T$, where $0 < \tau < 1$. The errors u_t are Gaussian or have $t(5)$ distributions with covariance matrix $\Sigma_1 = I_2$ for $t = 1, \ldots, T_1 = \tau T$ and $\Sigma_2 = \text{diag}(\lambda_1, \lambda_2)$ for $t > T_1$, where $(\lambda_1, \lambda_2) = (2, 2), (2, 1)$. The $t(5)$ distributed u_t are temporally independent and are generated as $\sqrt{\lambda_i} \times t(5)$ distributions for $t > T_1$ and $i = 1, 2$.

DGP2 Gaussian bivariate VAR(2) process

$$y_t = \begin{bmatrix} 0.190 \\ 0.523 \end{bmatrix} + \begin{bmatrix} -0.036 & -0.705 \\ -0.093 & 1.211 \end{bmatrix} y_{t-1} + \begin{bmatrix} 0.090 & 0.796 \\ -0.085 & -0.276 \end{bmatrix} y_{t-2} + u_t,$$

where the slope coefficients are the estimated values for the first example process considered in Section 5.1. The error process is Gaussian, $u_t \sim \mathcal{N}(0, \Sigma_t)$ and $\tau = 0.3$ which also corresponds to the empirical value for the example process. The covariance matrices are formed as in equation (4) with

$$B = \begin{bmatrix} 0.317 & 1.059 \\ 0.242 & -0.450 \end{bmatrix}$$

and $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ with $(\lambda_1, \lambda_2) = (0.5, 0.5), (0.5, 0.1)$. The latter choice is also inspired by the example process. Thus, DGP2 has features similar to the first example process.

Estimation of the VAR slope coefficients is done by GLS and then the λ_i are obtained as generalized eigenvalues using (7) with estimated covariance matrices $\tilde{\Sigma}_1 = T_1^{-1} \sum_{t=1}^{T_1} \tilde{u}_t \tilde{u}_t'$ and $\tilde{\Sigma}_2 = (T - T_1)^{-1} \sum_{t=T_1+1}^{T} \tilde{u}_t \tilde{u}_t'$, where $\tilde{u}_t$ are the GLS residuals. Even for the Gaussian processes we pretend that we

Table 1: Relative Rejection Frequencies of Tests for Gaussian DGP1 (Nominal Significance Level 5%)

Assumed τ		$\tau = 0.5$ (true value)		$\tau = 0.4$		$\tau = 0.3$	
(λ_1, λ_2)	T	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$
$(2,2)$	100	0.054	0.063	0.064	0.072	0.056	0.062
(size)	250	0.045	0.046	0.046	0.050	0.057	0.055
	500	0.046	0.050	0.057	0.053	0.049	0.050
$(2,1)$	100	0.331	0.342	0.258	0.262	0.174	0.190
(power)	250	0.703	0.709	0.549	0.557	0.348	0.348
	500	0.949	0.950	0.846	0.844	0.716	0.720

Note: DGP1 is a bivariate Gaussian VAR(0) process with $\tau = 0.5$, $u_t \sim \mathcal{N}(0, I_2)$ for $t = 1, \ldots, \tau T$ and $u_t \sim \mathcal{N}(0, \text{diag}(\lambda_1, \lambda_2))$ for $t = \tau T + 1, \ldots, T$.

do not know the true distribution and fit models with possibly two distinct kurtosis parameters. We also vary the sample size because it is expected to affect the properties of the tests. Specifically, $T = 100, 250, 500$ are used. The number of replications of all simulation experiments is 1000.

4.2 Simulation Results

The results for the bivariate DGP1 and DGP2 are discussed first. They are presented in Tables 1 - 3 and are also complemented by further results in the Supplement. Moreover, additional results for other processes are briefly summarized in Section 4.2.2.

4.2.1 Bivariate DGPs

In Table 1 we report results for a Gaussian DGP1 not only for the case where the change point of the volatility is specified correctly ($\tau = 0.5$) but also consider the situation of a misspecified volatility change point. In the panel for $(\lambda_1, \lambda_2) = (2, 2)$ in Table 1, it can be seen that the relative rejection frequencies of the tests in finite samples are roughly in line with the nominal size of 5%. Even for sample size $T = 100$ the relative rejection frequencies are reasonably close to 5%, regardless of the kurtosis parameters used. In other words, $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ has very similar rejection frequencies as $Q_2(0,0)$ for a Gaussian process. Furthermore, the relative rejection frequencies under the

Table 2: Relative Rejection Frequencies of Tests for DGP1 with Different Distributions (Nominal Significance Level 5%)

		Gaussian DGP		t-distributed DGP	
(λ_1, λ_2)	T	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$
$(2,2)$	100	0.054	0.063	0.173	0.047
(size)	250	0.045	0.046	0.187	0.048
	500	0.046	0.050	0.214	0.048
$(2,1)$	100	0.331	0.342	0.400	0.195
(power)	250	0.703	0.709	0.618	0.391
	500	0.949	0.950	0.802	0.598

Note: The underlying DGPs are a VAR(0) process with Gaussian and t distributed errors, respectively.

null hypothesis are not much affected by misspecifying the volatility change point (see the last four columns in Table 1).

In this respect, the situation is quite different for the power of the tests. In the lower part of Table 1, it is seen that the tests do have considerable power even for $T = 100$ if τ is specified correctly ($\tau = 0.5$). However, the power declines when the volatility change point is misspecified (see again the last four columns in Table 1). In fact, the farther away the assumed change point is from the true change point, the lower the power. It is again worth noting that it does not seem to matter much for the power of the tests whether the true kurtosis parameters are known or estimated (compare the corresponding results for $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ and $Q_2(0,0)$ in Table 1).

To explore the impact of misspecifying the kurtosis parameters we have also simulated DGP1 with t distributed errors u_t and compare the results in Table 2 to results for Gaussian errors. Clearly, if the errors are t distributed and the kurtosis parameters are mistakenly set to zero as for Gaussian processes, the corresponding test statistic $Q_2(0,0)$ is considerably oversized even in large samples. For example, for $T = 500$ the relative rejection frequency is 0.214 for $Q_2(0,0)$ in Table 2 instead of the desired 5% if the null hypothesis is true ($(\lambda_1, \lambda_2) = (2,2)$). In contrast, the relative rejection frequencies of $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ are very close to 5% if the null hypothesis is true, regardless of the error distribution. This outcome suggests that it is advisable to use the test statistics with estimated kurtosis parameters if, as usual in practice, the

Table 3: Relative Rejection Frequencies of Tests for DGP2 (Nominal Significance Level 5%)

		VAR(2)		VAR(1)	
(λ_1, λ_2)	T	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$
$(0.5, 0.5)$	100	0.076	0.100	0.059	0.065
(size)	250	0.064	0.062	0.043	0.046
	500	0.065	0.066	0.039	0.041
$(0.5, 0.1)$	100	0.330	0.360	0.255	0.272
(power)	250	0.644	0.650	0.561	0.570
	500	0.910	0.909	0.869	0.866
$(0.5, 0.1)$	100	0.253	0.243	0.234	0.194
(size-adjusted power)	250	0.613	0.622	0.584	0.580
	500	0.904	0.892	0.876	0.879

Note: DGP2 is a bivariate Gaussian VAR(2) process with $\tau = 0.3$.

true distribution is unknown. Nothing much can be gained from using the true kurtosis parameters if the distribution is known to be Gaussian.

So far we have only considered the situation where the true DGP is white noise (VAR(0)). From a practical point of view this situation is not of much interest, of course, and in Table 3 we present results for DGP2 which is a Gaussian VAR(2) process based on a real life dataset. The table presents the relative rejection frequencies for the situation that the VAR order is correctly specified to be $p = 2$ and also for the case where VAR(1) processes are fitted and, hence, the order is underspecified. Obviously, this situation may occur in practice and is therefore of interest here.

For the case of a correctly specified VAR lag order, the tests are slightly oversized in small samples due to the larger dimensional parameter space. For example, for $T = 100$ and $(\lambda_1, \lambda_2) = (0.5, 0.5)$ the test based on $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ has a relative rejection frequency of 0.100 in Table 3 which is double the nominal significance level of 0.05. We have also fitted VAR(4) processes to DGP1 and present the results in Tables S1 an S2 of the Supplement, where it can be seen that the tests are oversized for larger models with more parameters if the sample size is small ($T = 100$). As in Table 3, the relative rejection frequencies are much closer to 5% for the larger sample sizes, however.

Comparing to the corresponding results in Table 1, it can also be seen in Table 3 that the size-adjusted power declines for larger models. In Table 3

the unadjusted power of the test based on $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ is 0.360 for the VAR(2) and $T = 100$, which is similar to the power of the test for the corresponding case in Table 1 for a VAR(0). However, if we account for the fact that the test is oversized in Table 3 for the VAR(2) and compute the size-adjusted power, the relative rejection frequency is only 0.243. Thus, the results in Table 3 clearly show that the tests are oversized and the actual power of the tests is reduced for larger models in small samples when $T = 100$. Fortunately, the small sample distortions in size and power largely disappear for the larger sample sizes ($T = 250$ and 500).

If we now consider the results for the misspecified VAR(1) process in Table 3, it turns out that in this situation the tests can even be somewhat too conservative with relative rejection frequencies below the nominal 5% for larger samples. Also the size-adjusted power may be distorted and tends to be lower than for the correctly specified VAR(2) process.

In summary, based on our specific bivariate DGPs it appears that the number of lags and, hence, the size of the model affects the rejection frequencies. Larger models result in oversized tests in small samples with reduced actual power. Size and power distortions appear in small samples if the VAR order is under specified. If instead the volatility change point is misspecified, the size of the tests is not much affected but the power is reduced. If the true distribution of the DGP is not known to be Gaussian, then it always makes sense to use the test statistics based on estimated kurtosis parameters because they display very similar rejection frequencies in the Gaussian case to the test statistics based on known kurtosis parameters and their empirical size is much closer to the nominal size if the true distribution is non-Gaussian.

4.2.2 Extensions

We have also considered a number of extensions of the simulations reported so far and present further detailed results in the Supplement to this article. In Tables S1 and S2 of the Supplement we explore the impact of the location of the volatility shift within the sample. More precisely, we compare results based on DGP1 with break fractions $\tau = 0.5$ and 0.2. In other words, we

investigate the implication of the volatility change happening closer to the beginning of the sample. The results for both Gaussian and t distributed processes are very similar for $\tau = 0.5$ and $\tau = 0.2$. Thus, even if the break point is not in the middle of the sample this seems to have no substantial impact on the small sample properties of our tests. Note that in this case we have always fitted processes with correctly specified break dates.

In Table S3 of the Supplement, the properties for two Gaussian VAR(1) processes with different persistence are compared. One of the processes has an autoregressive root much closer to the unit circle than the other process. Again this does not affect the small sample properties of our tests much.

Finally, in Table S4 of the Supplement results for a 5-dimensional Gaussian VAR(1) process are presented which indicate that the tests tend to reject too often under the null hypothesis for larger models. This, of course, is in line with our previous finding in Table 3. The magnitude of the distortions is quite substantial for such a large model and even for sample size $T = 500$ there is still some bias. We have not found satisfactory small sample corrections for our tests in large models and leave the issue for future research.

4.3 Sequential Testing

So far we have explored the properties of our tests when they are applied to a single null hypothesis. However, in practice it is tempting to use them sequentially for higher-dimensional processes, as discussed in Section 3.2. Although we emphasize that our asymptotic results do not relate to this situation, we have also considered the possibility to apply the tests sequentially in a simulation exercise. To this end, we have generated the following 3-dimensional Gaussian VAR(0) process.

DGP3 3-dimensional ($K = 3$) Gaussian VAR(0) process $y_t = u_t$, with volatility change at $T_1 = 0.5T$, i.e., $\tau = 0.5$. The errors $u_t \sim \mathcal{N}(0, \Sigma_1 = I_3)$ for $t = 1, \ldots, \tau T_1$ and $u_t \sim \mathcal{N}(0, \Sigma_2 = \text{diag}(\lambda_1, \lambda_2, \lambda_3))$ for $t = T_1 + 1, \ldots, T$, where $(\lambda_1, \lambda_2, \lambda_3) = (2, 2, 2), (3, 2, 1)$ and $(3, 2, 2)$.

Simulation results for this case are provided in Table 4. In that table the

Table 4: Relative Rejection Frequencies of Tests for 3-dimensional Gaussian VAR(0) DGP (Nominal Individual Significance Level 5%)

T	$(\lambda_1, \lambda_2, \lambda_3)$	Hypotheses	$Q_r(0,0)$	$Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$
100	$(2,2,2)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	0.073	0.081
	(size)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.020	0.022
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.008	0.012
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0.001	0
	$(3,2,1)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	0.548	0.570
	(power)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.114	0.128
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.226	0.240
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0.008	0.013
	$(3,2,2)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	0.146	0.153
	(power)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.035	0.043
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.026	0.027
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0	0
250	$(2,2,2)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	0.056	0.053
	(size)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.019	0.018
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.010	0.009
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0	0
	$(3,2,1)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	0.934	0.936
	(power)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.270	0.272
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.612	0.616
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0.112	0.110
	$(3,2,2)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	0.252	0.264
	(power)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.103	0.106
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.031	0.031
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0	0
500	$(2,2,2)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	0.055	0.054
	(size)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.018	0.017
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.009	0.010
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0	0
	$(3,2,1)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	1	1
	(power)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.509	0.515
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.932	0.929
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0.480	0.451
	$(3,2,2)$	$\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$	0.530	0.527
	(power)	$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$	0.309	0.305
		$\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01} : \lambda_2 = \lambda_3$	0.036	0.038
		$\mathbb{H}_{01}$ and $\mathbb{H}_{02}\|\neg\mathbb{H}_{01}$ and $\mathbb{H}_{03}\|\neg\mathbb{H}_{01}$	0.004	0.004

Note: The notation $\mathbb{H}_{0i}|\neg\mathbb{H}_{0j}$ means that the null hypothesis $\mathbb{H}_{0i}$ is tested conditionally on $\mathbb{H}_{0j}$ being rejected.

notation $\mathbb{H}_{0i}|\neg\mathbb{H}_{0j}$ means that the null hypothesis $\mathbb{H}_{0i}$ is tested conditionally on $\mathbb{H}_{0j}$ being rejected. Thus, for example, $\mathbb{H}_{02}|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$ means that the null hypothesis $\mathbb{H}_{02} : \lambda_1 = \lambda_2$ is only tested for those simulations in which $\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$ has been rejected.

The results for $T = 100$ and $(\lambda_1, \lambda_2, \lambda_3) = (2, 2, 2)$ confirm that the test is slightly oversized for larger models. It rejects the true hypothesis $\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$ in roughly 8% of the cases if $Q_3(\tilde{\kappa}_1, \tilde{\kappa}_2)$ is used with a nominal significance level of 5%. All other tests in the same panel are conditional on this outcome and, not surprisingly, false conclusions regarding the identification of the process are rare. In particular, the false conclusion that the process is fully identified via heteroskedasticity is never reached when $Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$ is used. As one would expect the same is true if the sample size is increased and $(\lambda_1, \lambda_2, \lambda_3) = (2, 2, 2)$.

However, in the panels for $(\lambda_1, \lambda_2, \lambda_3) = (3, 2, 1)$, a full identification is also found rarely if $T = 100$, although the model is clearly identified. More precisely, $Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$ finds full identification only in 1.3% of the cases for $T = 100$. Fortunately, the relative frequency goes up to 45.1% for $T = 500$. Of course, even that means that full identification is not found in more than half of the cases. In other words, the tests do have some power in the sequential procedure, but it is rather small even for larger samples. Note, however, that some identifying information through heteroskedasticity is always found in this case because, for $T = 500$ and $(\lambda_1, \lambda_2, \lambda_3) = (3, 2, 1)$, the false null hypothesis $\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$ is always rejected.

It is perhaps also of interest to consider the case $(\lambda_1, \lambda_2, \lambda_3) = (3, 2, 2)$ which corresponds to a partially identified model. In this case it is in fact rather unlikely that full identification will be diagnosed even in small samples. In other words, even for $T = 100$ the three hypotheses $\mathbb{H}_{01}$, $\mathbb{H}_{02}|\neg\mathbb{H}_{01}$, and $\mathbb{H}_{03}|\neg\mathbb{H}_{01}$ are never rejected jointly. Also the true conditional null hypothesis $\mathbb{H}_{03}|\neg\mathbb{H}_{01}$ alone is not rejected very often. More precisely, the relative rejection frequency in the testing sequence of $Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$ for $T = 100$ is 0.027. In fact, this rejection frequency increases when the sample size increases because the false null hypothesis $\mathbb{H}_{01} : \lambda_1 = \lambda_2 = \lambda_3$ is rejected more often when

the sample size increases. Thus, the actual rejection frequency for $\mathbb{H}_{03}|\neg\mathbb{H}_{01}$ moves closer to the nominal 5%. On the other hand, the rejection frequency of the false null hypothesis $\mathbb{H}_{02}|\neg\mathbb{H}_{01} : \lambda_1 = \lambda_2$ is also not rejected very often for sample size $T = 250$. For this case, the relative rejection frequency of $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ is 0.106.

The overall conclusion for the sequential testing procedure is that for our still relatively small 3-dimensional example process the test is valuable but one needs to be careful drawing conclusions regarding full and partial identification of the underlying model because the overall procedure is obviously less powerful than the individual tests and it is generally more prone to error. Given the results for the individual tests, one may speculate that it becomes even more difficult to reach correct conclusions if the process is larger (higher-dimensional or larger lag order).

5 Empirical Examples

We present two empirical examples to illustrate the use of our tests for identification. The first one reconsiders a bivariate model for US data originally proposed by Blanchard and Quah (1989) and the second one has been used to analyze the interaction between US monetary policy and the stock market.

5.1 Blanchard-Quah Model

Blanchard and Quah (1989) identify demand and supply shocks in a bivariate macro model for US economic growth and unemployment by assuming that the demand shocks have no lung-run effects on output. Their model has become a textbook example for identification by restrictions on the long-run effects of the structural shocks (see, e.g., Breitung, Brüggemann and Lütkepohl (2004), Lütkepohl (2005, Chapter 9), Kilian and Lütkepohl (2017, Chapter 10)). Chen and Netšunajev (2016) use seasonally adjusted quarterly data for the period 1970q1 - 2007q4 and use identification through heteroskedasticity to investigate the validity of the long-run neutrality of demand shocks in a VAR(2) model for $y_t = (\Delta gnp_t, U_t)$, where gnp_t denotes the log of

Table 5: Estimated Relative Variances of Blanchard-Quah Model

Relative variance	Estimate	Standard deviation
λ_1	0.457	0.154
λ_2	0.152	0.041

Table 6: Identification Test for Blanchard-Quah Model

$\mathbb{H}_0$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$	degrees of freedom	p-value
$\lambda_1 = \lambda_2$	8.600	2	0.014

GNP and U_t is the unemployment rate. They model volatility changes by a smooth transition in the reduced form error covariance matrices. Their estimated change in the variances turns out to be a decline in the error variances around 1983q1 which is roughly the time where the Great Moderation starts in the US (see also Figure 1 of Chen and Netšunajev (2016)). Therefore it is plausible to use the VAR model (1) with a change in the residual covariance matrix in period 1983q1.

We have used the data from Chen and Netšunajev (2016) and estimated a VAR(2) model with error covariance change as in expression (3) with $T_1 =$ 1982q4. Since we have a sample size of $T = 152$, the corresponding sample fraction of the break is $\tau = 0.34$. The estimated relative variances (λ_i's) together with estimated standard errors are presented in Table 5. Both $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are smaller than one so that the second part of the sample clearly is associated with lower residual volatility.

The estimated λ_i's are clearly distinct and, based on the standard errors in Table 5, one may expect that they are significantly different. This informal evidence is in fact used by Chen and Netšunajev (2016) to justify the assumption of distinct relative variances. Using our test statistic $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$, we can now formally test the null hypothesis $\mathbb{H}_0 : \lambda_1 = \lambda_2$. The associated p-value is given in Table 6. It is clearly smaller than 5% so that $\mathbb{H}_0$ is rejected at a common level of significance. Thereby we support the assumption underlying the analysis of Chen and Netšunajev (2016). Note that we use the test statistic with estimated kurtosis parameters to avoid the assumption of a Gaussian error distribution.

5.2 A US Monetary Macro Model

Our second example is based on a benchmark study by Bjørnland and Leitemo (2009) who investigate the interaction between US monetary policy and the stock market using a structural VAR analysis. The relation between US monetary policy and the stock market has been investigated in a number of other articles as well (e.g., Park and Ratti (2000), Cheng and Jin (2013)). Bjørnland and Leitemo consider a five-dimensional system of variables, $y_t = (q_t, \pi_t, c_t, \Delta sp_t, r_t)'$, where q_t is the linearly detrended log of an industrial production index, π_t denotes the annual change in the log of consumer prices (CPI index), c_t is the annual change in the log of the World Bank (non energy) commodity price index, sp_t is the log of the real S&P500 stock price index deflated by the consumer price index to measure the real stock prices, Δsp_t are the corresponding returns, and r_t denotes the Federal Funds rate.

Bjørnland and Leitemo (2009) identify monetary policy and stock market shocks by zero restrictions on the impact effects and the long-run effects. These restrictions are controversial and have been questioned by other authors. Notably, Lütkepohl and Netšunajev (2017a, 2017b) consider identification through heteroskedasticity to investigate the validity of the Bjørnland-Leitemo identifying assumptions.

Lütkepohl and Netšunajev (2017a, 2017b) use monthly US data for the period 1970m1 - 2007m6 and more sophisticated volatility models than our simple shift in the covariance matrices. However, the smooth transition models used by Lütkepohl and Netšunajev (2017b) indicate that considering a VAR model such as (1) with error covariances (3) and a shift date in 1984 may provide a reasonable approximation (see in particular Figure 1a of Lütkepohl and Netšunajev (2017b)). Therefore we use their data and fit a VAR(3) model with a shift in the error covariance matrix after time $T_1 =$ 1983m4 which again roughly corresponds to splitting the data at the time when the Great Moderation started. The total sample size in this case is $T = 450$ and, hence, the fraction of the first volatility regime is $\tau = 0.37$.

The estimated relative variances together with estimated standard errors

Table 7: Estimated Relative Variances of US Monetary Macro Model

Relative variance	Estimate	Standard deviation
λ_1	0.939	0.155
λ_2	0.873	0.152
λ_3	0.577	0.089
λ_4	0.318	0.052
λ_5	0.054	0.005

Table 8: Identification Tests for US Monetary Macro Model

$\mathbb{H}_0$	$Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$	degrees of freedom	p-value
$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	75.328	14	2.060e−10
$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	13.565	9	0.138
$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	65.565	9	1.120e−10
$\lambda_1 = \lambda_2 = \lambda_3$	2.671	5	0.751
$\lambda_2 = \lambda_3 = \lambda_4$	9.997	5	0.075
$\lambda_3 = \lambda_4 = \lambda_5$	47.474	5	4.548e−9
$\lambda_1 = \lambda_2$	0.054	2	0.973
$\lambda_2 = \lambda_3$	1.737	2	0.420
$\lambda_3 = \lambda_4$	3.565	2	0.168
$\lambda_4 = \lambda_5$	28.654	2	5.995e−7

are shown in Table 7. Again the second volatility regime is associated with lower volatility because all relative variances are smaller than one. However, given the large estimated standard errors of some of the relative variances, it is clearly not obvious from Table 7 that the λ_i's are all distinct, although one may expect that some of the differences may be statistically significant.

To investigate the statistical significance of differences in the λ_i's formally we use again our tests with estimated kurtosis parameters. Since our set of variables includes a stock market index, an assumption of Gaussian model errors may be questionable and, hence, it is reasonable to allow for distributions with more kurtosis. Some test results are presented in Table 8.

The null hypothesis that all five λ_i's are identical is very strongly rejected at any conventional significance level. Thus, there is strong evidence that there is some additional identifying information in the second moments of the process. This result already suffices to indicate that there is further identifying information that may enable the researcher to test the conventional restrictions imposed by Bjørnland and Leitemo (2009). Of course,

it may also be of interest to continue with a more detailed analysis which of the shocks are identified by heteroskedasticity. Therefore it is important to note that, if the five λ_i are not all equal, we can also test that the first four or last four relative variances are identical. The null hypothesis $\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$ results in a p-value of 0.138 and, hence, at conventional significance levels, it cannot be rejected. In contrast, the hypothesis $\mathbb{H}_0 : \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$ is strongly rejected.

Given these results, we cannot be sure that the conditions for our tests hold for null hypotheses $\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3$ and $\mathbb{H}_0 : \lambda_2 = \lambda_3 = \lambda_4$. Recall that Proposition 1 requires that λ_4 is different from λ_3 to test the former hypothesis and λ_1 is different from λ_2 to test the latter hypothesis using the asymptotic distribution given in the proposition. Thus, the corresponding p-values in Table 8 may be unreliable. On the other hand, taking them at face value, they are consistent with the first four λ_i's being equal. In contrast, our test of $\mathbb{H}_0 : \lambda_3 = \lambda_4 = \lambda_5$ has a p-value smaller than 0.001 and hence the hypothesis is strongly rejected. Note that this test is justified by Proposition 1 and the result is consistent with the previous tests.

Using the arguments of the previous paragraph, Proposition 1 only provides a basis to test the final null hypothesis in Table 8, $\mathbb{H}_0 : \lambda_4 = \lambda_5$. Also this hypothesis is clearly rejected at any common significance level, thereby providing support for λ_5 being different from all other λ_i's. On the other hand, our tests do not support that $\lambda_1, \lambda_2, \lambda_3$ and λ_4 are different. Thus, strictly speaking, our test supports only identification of the last shock. Since the last shock does not have an economic label, only a more detailed analysis of the properties of the shock and its impulse responses can disclose which economic shock is behind the fifth shock in our statistical model. In principle it is of course no problem if the shocks are only partially identified. In fact, in many studies, researchers have only been interested in one or two shocks in a larger set of possible shocks. For example, they may just be interested in the effects of monetary policy shocks. If a single shock is identified, also the corresponding impulse responses etc. are identified and can be estimated and evaluated.

It may be worth noting that Lütkepohl and Netšunajev (2017b) use their model to test hypotheses regarding identifying zero restrictions on the impact and long-run effects of the shocks which are not overidentifying in a conventional structural VAR model and, hence, would not be testable without additional identifying information. Such tests become feasible, of course, if heteroskedasticity provides at least some identifying information. In fact, Lütkepohl and Netšunajev (2017b) reject most of the restrictions of interest in their study implying that heteroskedasticity apparently provides sufficient information for the tests to have power. Our tests enable the researcher to assess in more detail how much additional identifying information can be expected from heteroskedasticity and ideally also which hypotheses can reasonably be tested. Finally, we remind the reader that Lütkepohl and Netšunajev (2017b) considered a different volatility model so that our results strictly speaking do not apply to their model. Clearly, it would be of interest to have identification tests similar to our new tests for more sophisticated volatility models as well.

6 Conclusions

In this study we have developed frequentist tests for identification through heteroskedasticity in structural vector autoregressive models. We consider VAR models with two volatility states. The change point of the volatility is assumed to be known. The tests are Wald type tests such that only the unrestricted model has to be estimated. The model errors are assumed to be from the class of elliptical distributions. This class of distributions includes the Gaussian distribution as well as t and mixed normal distributions. We propose test versions where the kurtosis of the distribution is assumed to be known and also allow for the possibility that the kurtosis is estimated.

The asymptotic null distributions of the test statistics are derived and are shown to be χ^2 distributions although the models are not identified under the null hypothesis. We have also explored the small sample properties of the tests by Monte Carlo simulations and we have found that the tests are

oversized for large models when the sample size is small. However, for larger samples and smaller models, size and power of the tests is quite reasonable and the properties of the tests do not depend on the timing of the volatility break. However, misspecification of the volatility change point is found to reduce power and underspecification of the lag order may lead to size and power distortions. The small sample properties are very little affected by estimating the kurtosis parameters. Thus, in practice we recommend to use the test versions which are based on estimated kurtosis parameters.

Two empirical examples are considered to illustrate the usefulness of the tests. The first example considers a bivariate model for US data. Our tests support the assumption of earlier studies that the model is identified by heteroskedasticity. The second example is based on a five-dimensional model for US data. It has been used to analyze the interaction between US monetary policy and the stock market. We find that there is some identifying information from heteroskedasticity but there is little support for a full identification.

There are a number of desirable extensions of our tests. First, it would be useful if tests for more than two volatility regimes could be developed. Moreover, the volatility model is very special. It assumes that the change in volatility is extraneously generated. Other models have been used in the literature on identification through heteroskedasticity. It is desirable to have tests for identification also for other related models.

A Proof of Proposition 1

We assume that the u_t have an elliptical distribution possessing a density as well as finite fourth moments as in Proposition 1. Such assumptions are needed because we are using limiting results for elliptical distributions from Anderson (2003).

We study $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K$, the eigenvalues of $\tilde{\Sigma}_1^{-1}\tilde{\Sigma}_2$, and follow the pattern of proof in Anderson (2003, Sections 13.6.1 and 13.6.2). As in Anderson (2003, eqn. (9) on p. 550), for the theoretical developments that follow it will be

convenient to transform the estimators $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$ and consider the matrices

$$\tilde{\Omega}_1 = B^{-1}\tilde{\Sigma}_1 B'^{-1} \text{ and } \tilde{\Omega}_2 = B^{-1}\tilde{\Sigma}_2 B'^{-1}.$$

(As before, we here assume that the first nonzero element on each column of B is positive.) With this transformation, the asymptotic distributions of $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ below will depend only on Λ and not on B (note also that the theoretical counterparts of $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ are $B^{-1}\Sigma_1 B'^{-1} = I_K$ and $B^{-1}\Sigma_2 B'^{-1} = \Lambda$). Furthermore, as $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K$ are the eigenvalues of $\tilde{\Sigma}_1^{-1}\tilde{\Sigma}_2$, they are also the eigenvalues of $\tilde{\Omega}_1^{-1}\tilde{\Omega}_2$ or, equivalently, the eigenvalues of $\tilde{\Omega}_1^{-1/2}\tilde{\Omega}_2\tilde{\Omega}_1^{-1/2}$. Thus, as far as asymptotic properties of the eigenvalues $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K$ or their functions are concerned, we can use the matrices $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ instead of $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$.

From (12) and (13) it follows that the asymptotic distributions of $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ can be derived by using the (independent) errors u_t in place of the residuals in the definitions of $\tilde{\Sigma}_1$ and $\tilde{\Sigma}_2$. For simplicity, denote $T_2 = T - T_1$ and note that, due to the assumption $T_1 = [\tau T]$ for some $\tau \in (0, 1)$, both $T_1 \to \infty$ and $T_2 \to \infty$ when $T \to \infty$. From Theorem 3.6.2 in Anderson (2003, p. 102), we can thus conclude that $T_1^{1/2}(\tilde{\Omega}_1 - I_K) = \tilde{Z}_1$ and $T_2^{1/2}(\tilde{\Omega}_2 - \Lambda) = \tilde{Z}_2$, say, converge jointly in distribution as $T \to \infty$ to the matrices $Z_1 = [z_{1,ij}]$ and $Z_2 = [z_{2,ij}]$ (i, $j = 1, \ldots, K$). Here Z_1 and Z_2 are independent, their elements are jointly normally distributed, and their functionally independent elements are statistically independent. Furthermore, their elements have mean zero and covariance structure given by

$$Cov[\text{vec}(Z_1)] = (1 + \kappa_1)(I_{K^2} + \mathbf{K})(I_K \otimes I_K) + \kappa_1 \text{vec}(I_K)\text{vec}(I_K)'$$

and

$$Cov[\text{vec}(Z_2)] = (1 + \kappa_2)(I_{K^2} + \mathbf{K})(\Lambda \otimes \Lambda) + \kappa_2 \text{vec}(\Lambda)\text{vec}(\Lambda)',$$

where $\mathbf{K}$ ($K^2 \times K^2$) is a commutation matrix, $\otimes$ denotes the Kronecker product and vec signifies the column stacking operator. The Gaussian case is obtained as a special case by choosing $\kappa_1 = \kappa_2 = 0$. In what follows, the null hypothesis is assumed to hold unless otherwise stated.

As in Tyler (1983, p. 413, the paragraph following equations (1)), we can describe the elements of $Cov[\text{vec}(Z_1)]$ as follows. The distinct off-diagonal elements of $Cov[\text{vec}(Z_1)]$ are uncorrelated with each other and uncorrelated with the diagonal elements, and each of them has variance $1+\kappa_1$. All diagonal elements have variance $2+3\kappa_1$ and the covariance between any two diagonal elements is κ_1. In the special case where $\Lambda = \lambda_0 I_K$ the same description clearly applies to the elements of $Cov[\text{vec}(Z_2)]$ with κ_1 replaced by κ_2, provided the variances and covariances are multiplied by λ_0^2, and by the definition of the commutation matrix the same is true when Z_2 is replaced by the matrix $[z_{2,ij}]_{i,j=s+1}^{s+r}$ and Λ is replaced by $\Lambda_2 = \lambda_0 I_r$.

Theorem 1 of Amemiya (1990) implies that $T^{1/2}(\tilde{\lambda}_{s+1}-\lambda_0,\ldots,\tilde{\lambda}_{s+r}-\lambda_0)$ converges in distribution to an $(r\times 1)$ random vector consisting of the eigenvalues of the matrix $U=[u_{ij}]_{i,j=1}^{r}=[(1-\tau)^{-1/2}z_{2,ij}-\lambda_0\tau^{-1/2}z_{1,ij}]_{i,j=s+1}^{s+r}$. The elements of U are jointly normally distributed with mean zero and covariances given in the following equations where $c(\tau,\kappa_1,\kappa_2)^2=\left(\frac{1+\kappa_1}{\tau}+\frac{1+\kappa_2}{1-\tau}\right)^{-1}$ and $i,j=s+1,\ldots,s+r$:

$$\begin{aligned}
\mathbb{E}[u_{ij}^2] &= \frac{(1+\kappa_2)\lambda_0^2}{1-\tau}+\frac{(1+\kappa_1)\lambda_0^2}{\tau} = \lambda_0^2 c(\tau,\kappa_1,\kappa_2)^{-2} \text{ for } i\neq j\\
\mathbb{E}[u_{ii}^2] &= \frac{(2+3\kappa_2)\lambda_0^2}{1-\tau}+\frac{(2+3\kappa_1)\lambda_0^2}{\tau}\\
&= 2\lambda_0^2 c(\tau,\kappa_1,\kappa_2)^{-2}+\lambda_0^2\left(\frac{\kappa_2}{1-\tau}+\frac{\kappa_1}{\tau}\right)\\
\mathbb{E}[u_{ii}u_{jj}] &= \lambda_0^2\left(\frac{\kappa_2}{1-\tau}+\frac{\kappa_1}{\tau}\right) \text{ for } i\neq j.
\end{aligned}$$

Distinct off-diagonal elements of U are independent of each other and the off-diagonal and diagonal elements of U are independent.

Now define the (infeasible) test statistic

$$Q_r(\kappa_1,\kappa_2)=c(\tau,\kappa_1,\kappa_2)^2\left[-T\sum_{k=s+1}^{s+r}\log(\tilde{\lambda}_k)+Tr\log\left(\frac{1}{r}\sum_{k=s+1}^{s+r}\tilde{\lambda}_k\right)\right]$$

for which we have

$$\begin{aligned}
Q_r(\kappa_1,\kappa_2) &\xrightarrow{d} \frac{c(\tau,\kappa_1,\kappa_2)^2}{\lambda_0^2}\sum_{i<j}u_{ij}^2+\frac{c(\tau,\kappa_1,\kappa_2)^2}{2\lambda_0^2}\left[\sum_{i=s+1}^{s+r}u_{ii}^2-\frac{1}{r}\left(\sum_{i=s+1}^{s+r}u_{ii}\right)^2\right]\\
&\overset{def}{=} Q_{1,r}^*(\kappa_1,\kappa_2)+Q_{2,r}^*(\kappa_1,\kappa_2).
\end{aligned}$$

Here $Q^*_{1,r}(\kappa_1,\kappa_2)$ and $Q^*_{2,r}(\kappa_1,\kappa_2)$ are independent and $Q^*_{1,r}(\kappa_1,\kappa_2)$ has a χ^2 distribution with $\frac{1}{2}r(r-1)$ degrees of freedom. As to $Q^*_{2,r}(\kappa_1,\kappa_2)$, defining $\boldsymbol{w}_s$ as

$$\boldsymbol{w}_s = \frac{c(\tau,\kappa_1,\kappa_2)}{\sqrt{2}\lambda_0}(u_{s+1,s+1},\ldots,u_{s+r,s+r})',$$

and the $(r\times r)$ projection matrix P_r as $P_r = I_r - \frac{1}{r}\mathbf{1}_r\mathbf{1}_r'$, where $\mathbf{1}_r = (1,\ldots,1)'$ is an $(r\times 1)$ vector, we have

$$Q^*_{2,r}(\kappa_1,\kappa_2) = \boldsymbol{w}_s' P_r \boldsymbol{w}_s.$$

Hence, it follows that the random vector $\boldsymbol{w}_s$ is normally distributed with zero mean and covariance matrix (see the above expressions of $\mathbb{E}[u_{ii}^2]$ and $\mathbb{E}[u_{ii}u_{jj}]$ $(i\neq j)$)

$$\begin{aligned} Cov[\boldsymbol{w}_s] &= I_r + \frac{c(\tau,\kappa_1,\kappa_2)^2}{2}\left(\frac{\kappa_2}{1-\tau}+\frac{\kappa_1}{\tau}\right) I_r \\ &\quad + \frac{c(\tau,\kappa_1,\kappa_2)^2}{2}\left(\frac{\kappa_2}{1-\tau}+\frac{\kappa_1}{\tau}\right)(\mathbf{1}_r\mathbf{1}_r' - I_r) \\ &= I_r + \frac{c(\tau,\kappa_1,\kappa_2)^2}{2}\left(\frac{\kappa_2}{1-\tau}+\frac{\kappa_1}{\tau}\right)\mathbf{1}_r\mathbf{1}_r'. \end{aligned}$$

Thus, we have $P_r Cov[\boldsymbol{w}_s] = P_r$ and we find that $Q^*_{2,r}(\kappa_1,\kappa_2)$ has a χ^2 distribution with $r-1$ degrees of freedom. This fact can be justified by a well-known result of quadratic forms of normal random vectors (see, e.g., result (vii) in Rao (1973, p. 188)).

From the preceding discussion we can now conclude that $Q_r(\kappa_1,\kappa_2) \xrightarrow{d} Q^*_{1,r}(\kappa_1,\kappa_2) + Q^*_{2,r}(\kappa_1,\kappa_2)$, where $Q^*_{1,r}(\kappa_1,\kappa_2)$ and $Q^*_{2,r}(\kappa_1,\kappa_2)$ are independent and have χ^2 distributions with degrees of freedom $\frac{1}{2}r(r-1)$ and $r-1$. Therefore, the infeasible test statistic $Q_r(\kappa_1,\kappa_2)$ has an asymptotic χ^2 distribution with $\frac{1}{2}(r+2)(r-1)$ degrees of freedom, and the same is true for its feasible version $Q_r(\tilde{\kappa}_1,\tilde{\kappa}_2)$, where $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ are consistent estimators of κ_1 and κ_2, respectively. This proves Proposition 1.

References

Amemiya, Y. (1990). A note of the limiting distribution of certain characteristic roots, *Statistics and Probability Letters* **9**: 465–470.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis, *Annals of Mathematical Statistics* **34**: 122–148.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd edn, John Wiley, New York.

Andrews, I. (2018). Valid two-step identification-robust confidence sets for GMM, *Review of Economics and Statistics* **100**: 337–348. DOI: 10.1162/rest_a_00682.

Bjørnland, H. C. and Leitemo, K. (2009). Identifying the interdependence between US monetary policy and the stock market, *Journal of Monetary Economics* **56**: 275 – 282. DOI: http://dx.doi.org/10.1016/j.jmoneco.2008.12.001.

Blanchard, O. J. and Quah, D. (1989). The dynamic effects of aggregate demand and supply disturbances, *American Economic Review* **79**: 655–673.

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

Chen, W. and Netšunajev, A. (2016). On the long-run neutrality of demand shocks, *Economics Letters* **139**: 57 – 60.

Cheng, L. and Jin, Y. (2013). Asset prices, monetary policy, and aggregate fluctuations: An empirical investigation, *Economics Letters* **119**: 24–27.

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

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

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

Lanne, M. and Lütkepohl, H. (2014). A statistical comparison of alternative identification schemes for monetary policy shocks, *in* J. Knif and B. Pape (eds),

Contributions to Mathematics, Statistics, Econometrics, and Finance - Essays in Honour of Professor Seppo Pynnönen, Acta Wasaensia 296, Statistics 7, University of Vaasa, Vaasa, Finland, pp. 137–152.

Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010). Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics and Control* **34**: 121–131.

Lewis, D. J. (2018). Robust inference in models identified via heteroskedasticity, *Technical Report Staff Report No. 876*, Federal Reserve Bank of New York.

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

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

Lütkepohl, H. and Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs, *Journal of Applied Econometrics* **29**: 479–496.

Lütkepohl, H. and Netšunajev, A. (2017a). Structural vector autoregressions with heteroskedasticy: A review of different volatility models, *Econometrics and Statistics* **1**: 2–18.

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

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

Lütkepohl, H. and Velinov, A. (2016). Structural vector autoregressions: Checking identifying long-run restrictions via heteroskedasticity, *Journal of Economic Surveys* **30**: 377–392.

Lütkepohl, H. and Woźniak, T. (2017). Bayesian inference for structural vector autoregressions identified by Markov-switching heteroskedasticity, *Technical report*, DIW Berlin.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Hoboken, NJ.

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

Park, K. and Ratti, R. A. (2000). Real activity, inflation, stock returns, and monetary policy, *Financial Review* **35**: 59–78.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications, 2nd ed.*, John Wiley, New York.

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

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

Schott, J. R. (2001). Some tests for the equality of covariance matrices, *Journal of Statistical Planning and Inference* **94**: 25–36.

Tyler, D. E. (1983). Robustness and efficiency properties of scatter matrices, *Biometrika* **70**: 411–420.

Velinov, A. and Chen, W. (2015). Do stock prices reflect their fundamentals? New evidence in the aftermath of the financial crisis, *Journal of Economics and Business* **80**: 1–20.

Woźniak, T. and Droumaguet, M. (2015). Assessing monetary policy models: Bayesian inference for heteroskedastic structural VARs, *Technical report*, University of Melbourne.