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 and Elisa Eesti AS,
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

October 8, 2018

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.

Key Words: Heteroskedasticity, structural identification, vector autoregressive process
JEL classification: C32

[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.

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. A main advantage of identification via heteroskedasticity is that the data are in principle informative on the conditions for identification. Thus, identification can in principle 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 that purpose 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 for identification for the special case of stable VAR models with two volatility regimes of the residuals. The distribution of the residuals is assumed to be elliptically symmetric which covers the case of Gaussian VAR processes. 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. We present a sequence of tests which permits us to test for full identification of the structural form VAR model and show by simulation that the asymptotic theory is a good guide for small sample performance of the tests, if the sample size is sufficiently large. Finally, we present examples which show 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 and 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.

We use the following abbreviations and symbols throughout: DGP abbreviates data generating process, OLS and GLS stand for ordinary and generalized least squares, respectively, while ML and LR abbreviate maximum likelihood and likelihood ratio, respectively. GNP signifies gross national product, VAR stands for vector autoregressive and SVAR means structural vector autoregressive. The expression vec is the column vectorizing operator of a matrix and vech is the 'half' vectorizing operator that collects only the columns of a square matrix from the main diagonal downward in a column vector. The differencing operator is denoted by Δ and $\mathbb{E}$ is the expectation operator. A normal distribution with mean μ and covariance matrix Σ is signified as $\mathcal{N}(\mu, \Sigma)$ and iid abbreviates independently, identically distributed. The determinant of a matrix is signified as det, diag specifies a diagonal matrix and I_K is a $(K \times K)$ identity matrix.

2 The Model

Consider a 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, A_j $(j = 1, \ldots, p)$ are $(K \times K)$ VAR slope coefficient matrices, and u_t is a white noise error term with zero mean, $\mathbb{E}(u_t) = 0$, and (positive definite) covariance matrices

$$\mathbb{E}(u_t u_t') = \begin{cases} \Sigma_1 & \text{for} \quad t \in \mathcal{T}_1 = \{1, \ldots, T_1\}, \\ \Sigma_2 & \text{for} \quad t \in \mathcal{T}_2 = \{T_1 + 1, \ldots, T\}, \end{cases} \tag{2}$$

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 $T \to \infty$.

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). We also assume 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, it is assumed that

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

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.

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. In that case the covariance matrices in (2) can be decomposed as follows:

$$\Sigma_1 = BB', \quad \Sigma_2 = B\Lambda B', \tag{3}$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K)$ is a $(K \times K)$ diagonal matrix with positive diagonal elements and B is a nonsingular $(K \times K)$ matrix (see Lütkepohl (2013)). Using the matrix B, the structural shocks are obtained from the reduced form errors as $\varepsilon_t = B^{-1}u_t$, and their covariance matrices are given by

$$\mathbb{E}(\varepsilon_t \varepsilon_t') = \begin{cases} I_K & \text{for} \quad t \in \mathcal{T}_1, \\ \Lambda & \text{for} \quad t \in \mathcal{T}_2. \end{cases} \tag{4}$$

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

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + B\varepsilon_t. \tag{5}$$

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. As we only consider stable models, we further assume that the VAR

matrices A_j $(j = 1, \ldots, p)$ satisfy the usual stability condition

$$\det\left(I_K - A_1 z - \cdots - A_p z^p\right) \neq 0 \quad \text{for} \quad |z| \leq 1. \tag{6}$$

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 (3) are the eigenvalues of the matrix $\Sigma_1^{-1}\Sigma_2$ so that they satisfy the (generalized) eigenvalue equations

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

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

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

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 (5). Testing for a possible lack of identification is therefore of interest and will be discussed in the next section.

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 $\boldsymbol{\sigma}_i = \mathrm{vech}(\Sigma_i)$ $(i = 1, 2)$. If the DGP is Gaussian, maximizing $l(\boldsymbol{\vartheta}, \boldsymbol{\sigma})$ with respect to the parameters gives the 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 GLS procedure. In that case (1) is estimated with equationwise 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 $\lambda_s \neq \lambda_{s+1}$ for $s > 0$ and $\lambda_{s+r} \neq \lambda_{s+r+1}$ for $s < K - r$. Furthermore, let $\tilde{\kappa}_1$ and $\tilde{\kappa}_2$ be consistent estimators of κ_1 and κ_2, respectively. Then

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

□

The asymptotic 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, suitable sequences of null hypotheses have to be tested. If

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

does not hold, we know that $\lambda_1 \neq \lambda_K$ and we can 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 tested in this sequence of hypotheses 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 $K = 4$ we have to test

$$\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4.$$

Proposition 1 implies that this null hypothesis can be tested using $Q_4(\tilde{\kappa}_1, \tilde{\kappa}_2)$ with a $\chi^2(9)$ distribution. If the null hypothesis is false, it follows that $\lambda_1 \neq \lambda_4$ so that we can test

$$\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3 \quad \text{and} \quad \mathbb{H}_0 : \lambda_2 = \lambda_3 = \lambda_4$$

using $Q_3(\tilde{\kappa}_1, \tilde{\kappa}_2)$ statistics with a $\chi^2(5)$ distributions. If both null hypotheses are false, the conditions of Proposition 1 are satisfied for the following three null hypotheses:

$$\mathbb{H}_0 : \lambda_1 = \lambda_2, \quad \mathbb{H}_0 : \lambda_2 = \lambda_3 \quad \text{and} \quad \mathbb{H}_0 : \lambda_3 = \lambda_4.$$

They can then be tested with $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ tests based on asymptotic $\chi^2(2)$ distributions. Rejecting the latter 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)). Even if 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. Although all our DGPs are either VAR(0) or VAR(1) processes, we also fit VAR(p) models with $p > 1$ to the data. The error distributions are either Gaussian, $u_t \sim \mathcal{N}(0, \Sigma_t)$, or have t distributions with 5 degrees of freedom. The DGPs are chosen such that we can explore the possible dependence of the small sample properties on the dimension of the underlying process, the location of the volatility change point and the persistence of the process. Therefore our choice of DGPs varies these features. In particular, we use the following DGPs:

DGP1 Bivariate ($K = 2$) VAR(0) process $y_t = u_t$, with volatility change at $T_1 = \tau T$, where $\tau = 0.5$ and 0.2. The errors u_t are Gaussian, $u_t \sim \mathcal{N}(0, \Sigma_t)$, with $\Sigma_1 = I_2$ and $\Sigma_2 = \Lambda$, where

$$\Lambda = \text{diag}(\lambda_1, \lambda_2) \quad \text{with} \quad (\lambda_1, \lambda_2) = (2, 2), (2, 1).$$

DGP2 The second DGP is also a bivariate VAR(0) with the same parameter values as DGP1 but the error distribution is a multivariate t-distribution with 5 degrees of freedom. More precisely, the components of $u_t = (u_{1t}, u_{2t})'$ have independent $t(5)$ distributions for $t \leq T_1$ and $\sqrt{\lambda_i} \times t(5)$ distributions for $t > T_1$ and $i = 1, 2$.

DGP3 The third DGP is a bivariate VAR(1) process

$$y_t = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} y_{t-1} + u_t$$

with $a = 0.5$ and 0.9. For $a = 0.9$ the process has one persistent variable. The error process u_t is the same Gaussian process as for DGP1 with τ fixed at 0.5 and

$$\Lambda = \text{diag}(\lambda_1, \lambda_2) \quad \text{with} \quad (\lambda_1, \lambda_2) = (2, 2), (2, 1).$$

DGP4 The fourth DGP is a five-dimensional ($K = 5$) VAR(1) process

$$y_t = \begin{bmatrix} a_1 & 0 & 0 & 0 & 0 \\ 0 & a_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} y_{t-1} + u_t,$$

where $a_1 = 0.9$ and $a_2 = 0.5$. The error process is again Gaussian, $u_t \sim \mathcal{N}(0, \Sigma_t)$, and the volatility change occurs in the middle of the sample, $\tau = 0.5$. Moreover,

$$\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_5) \quad \text{with}$$

$$(\lambda_1, \ldots, \lambda_5) = (2, 2, 2, 2, 2) \quad \text{and} \quad (5, 4, 3, 2, 1).$$

The first set of λ_i's allows us to study the size of the tests and the second set of λ_i's is chosen to investigate the power. Although the λ_i's in the latter set of relative variances are chosen in equidistant steps, they may reflect differences in power. For example, the tests may have different power for $\mathbb{H}_0 : \lambda_1 = \lambda_2$ and $\mathbb{H}_0 : \lambda_4 = \lambda_5$ because, in relative terms, λ_1 and λ_2 are closer together than λ_4 and λ_5. In fact, λ_4 is twice as large as λ_5.

We fit VAR models with intercept to the data and, for DGP1 and DGP2, compare with the situation where the test statistics are computed directly for the u_t (the VAR(0) case). Estimation of the VAR slope coefficients is done by GLS and then the λ_i are estimated by optimizing the concentrated likelihood function (10) numerically. Alternatively, we could have obtained the $\tilde{\lambda}_i$ 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 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 as well. 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 DGPs (DGP1 - DGP3) are presented in Tables 1 - 3 and the results for the five-dimensional DGP4 are shown in Table 4. We will first discuss the results for the bivarite DGPs.

4.2.1 Bivariate DGPs

Considering the panels for $(\lambda_1, \lambda_2) = (2, 2)$ in Table 1, it can be seen that the tests are slightly oversized for small sample sizes when higher order VAR models are fitted. For the VAR(4) models and sample size $T = 100$ the rejection frequencies are well above 10% instead of the nominal 5%. The

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

		VAR(0)		VAR(4)	
(λ_1, λ_2)	T	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$
		$\tau = 0.5$			
$(2,2)$	100	0.067	0.064	0.144	0.148
	250	0.050	0.044	0.079	0.075
	500	0.050	0.048	0.060	0.062
$(2,1)$	100	0.338	0.323	0.443	0.429
	250	0.678	0.677	0.707	0.702
	500	0.946	0.946	0.948	0.946
		$\tau = 0.2$			
$(2,2)$	100	0.051	0.054	0.136	0.130
	250	0.052	0.050	0.073	0.072
	500	0.058	0.059	0.067	0.070
$(2,1)$	100	0.312	0.305	0.431	0.423
	250	0.672	0.668	0.701	0.693
	500	0.946	0.940	0.945	0.943

situation improves when the sample size gets larger and the empirical size is close to 5% in all cases when $T = 500$. If a VAR(0) is considered the size is close to the nominal level for all sample sizes.

The results in Table 1 also indicate that the location of the break date (represented by τ) does not seem to affect the properties of the tests substantially, at least if one considers break dates not very close to the beginning or end of the sample. The rejection frequencies in corresponding entries in Table 1 for $\tau = 0.2$ and $\tau = 0.5$ are in fact very similar. Thus, size and power of the tests do no seem to depend much on the break date.

Another feature that can be seen in Table 1 is that it does not matter much whether the true kurtosis parameters $\kappa_1 = \kappa_2 = 0$ are used in the test statistics or the parameters are estimated. The rejection frequencies of the corresponding test statistics $Q_2(0,0)$ and $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ are in all cases very similar.

The situation is very different in Table 2, where the true residual distribution is a t distribution. In that case the test based on $Q_2(0,0)$, which incorrectly assumes kurtosis parameters of a Gaussian distribution, is substantially oversized while the test based on $Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$ has size properties similar to those in Table 1, where the DGP is Gaussian. Comparing the results for $(\lambda_2, \lambda_2) = (2,1)$ in Tables 1 and 2 it can be seen that the power of

Table 2: Relative Rejection Frequencies of Tests for DGP2 (Nominal Significance Level 5%)

		VAR(0)		VAR(4)	
(λ_1, λ_2)	T	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$
		$\tau = 0.5$			
$(2,2)$	100	0.173	0.047	0.266	0.123
	250	0.187	0.048	0.229	0.067
	500	0.214	0.048	0.231	0.060
$(2,1)$	100	0.400	0.195	0.475	0.294
	250	0.618	0.391	0.646	0.425
	500	0.802	0.598	0.819	0.627
		$\tau = 0.2$			
$(2,2)$	100	0.175	0.046	0.280	0.149
	250	0.216	0.060	0.244	0.086
	500	0.259	0.067	0.277	0.078
$(2,1)$	100	0.406	0.235	0.481	0.329
	250	0.614	0.389	0.644	0.433
	500	0.822	0.611	0.828	0.636

the tests based on estimated kurtosis parameters is slightly smaller for the t distributed errors than for the Gaussian processes. Thus, the actual distribution may have an impact on the power of our tests. In any case, considering the size properties, the recommendation from the results in Table 2 is to use the test statistics with estimated kurtosis parameters if, as usual in practice, the true distribution is unknown.

In Table 3, the impact of the persistence of the DGP can be seen. It presents the rejection frequencies for the Gaussian DGP3 with persistence parameters $a = 0.5$ and $a = 0.9$. Clearly, the corresponding entries in the table for both values of a are very similar. Hence, the persistence of the process does not matter much for the properties of the tests. In fact, the VAR(4) results in Table 3 are not much different from the corresponding VAR(4) results in Table 1, meaning that it does not make much difference whether the true DGP is a Gaussian VAR(1) or a VAR(0). The more important issue appears to be the order of the process which is fitted to the data. Clearly, estimating more VAR parameters has a negative impact on the empirical size of the tests in small samples. More precisely, the tests become oversized if larger models are fitted and the sample size is small.

In summary, based on our specific bivariate DGPs it appears that the volatility change date and the persistence of the VAR process does not matter

Table 3: Relative Rejection Frequencies of Tests for DGP3 (Nominal Significance Level 5%)

		VAR(1)		VAR(4)	
(λ_1, λ_2)	T	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$	$Q_2(0,0)$	$Q_2(\tilde{\kappa}_1, \tilde{\kappa}_2)$
		$a = 0.5$			
$(2,2)$	100	0.067	0.063	0.119	0.120
	250	0.052	0.053	0.067	0.070
	500	0.047	0.047	0.057	0.056
$(2,1)$	100	0.362	0.348	0.440	0.430
	250	0.692	0.680	0.707	0.699
	500	0.932	0.934	0.939	0.937
		$a = 0.9$			
$(2,2)$	100	0.069	0.068	0.121	0.129
	250	0.067	0.066	0.082	0.083
	500	0.054	0.058	0.062	0.061
$(2,1)$	100	0.333	0.331	0.421	0.439
	250	0.699	0.698	0.715	0.720
	500	0.942	0.942	0.943	0.943

much for the empirical size and power. The number of lags and, hence, the size of the model affects the rejection frequencies. Larger models result in oversized tests in small samples. Rather large samples are necessary to move the empirical rejection frequencies close to the nominal significance level. 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 Five-dimensional DGP

Looking now at the results for the five-dimensional DGP4 in Table 4, it becomes clear that the tests are substantially oversized. For these large models the size also improves for increasing sample sizes, but the rejection frequencies still exceed the nominal 5% for $T = 500$. For example, for the VAR(4) with $(\lambda_1, \ldots, \lambda_5) = (2, 2, 2, 2, 2)$ the rejection frequencies of the tests for $T = 500$ are still around 17% rather than the desired 5%. Thus, for very large models, the sample sizes also have to be rather large for precise inference.

The size distortions also make it difficult to assess the power results for the

Table 4: Relative Rejection Frequencies of Tests for DGP4 (Nominal Significance Level 5%)

		VAR(1)		VAR(4)	
$(\lambda_1, \ldots, \lambda_5)$	$\mathbb{H}_0$	$Q_r(0,0)$	$Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$	$Q_r(0,0)$	$Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$
		$T = 100$			
$(2,2,2,2,2)$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	0.228	0.218	0.904	0.892
	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	0.080	0.077	0.641	0.622
	$\lambda_1 = \lambda_2 = \lambda_3$	0.033	0.034	0.362	0.341
	$\lambda_1 = \lambda_2$	0.019	0.019	0.186	0.172
$(5,4,3,2,1)$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	0.902	0.889	1.000	0.999
	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	0.466	0.447	0.892	0.883
	$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	0.706	0.677	0.951	0.942
	$\lambda_1 = \lambda_2 = \lambda_3$	0.181	0.176	0.541	0.532
	$\lambda_3 = \lambda_4 = \lambda_5$	0.454	0.428	0.761	0.755
	$\lambda_1 = \lambda_2$	0.052	0.049	0.231	0.229
	$\lambda_4 = \lambda_5$	0.256	0.247	0.430	0.423
		$T = 250$			
$(2,2,2,2,2)$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	0.098	0.102	0.306	0.315
	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	0.025	0.020	0.097	0.097
	$\lambda_1 = \lambda_2 = \lambda_3$	0.008	0.008	0.038	0.033
	$\lambda_1 = \lambda_2$	0.006	0.007	0.018	0.017
$(5,4,3,2,1)$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	0.999	0.999	1.000	1.000
	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	0.764	0.763	0.863	0.857
	$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	0.992	0.989	0.998	0.998
	$\lambda_1 = \lambda_2 = \lambda_3$	0.292	0.281	0.423	0.409
	$\lambda_3 = \lambda_4 = \lambda_5$	0.908	0.909	0.941	0.942
	$\lambda_1 = \lambda_2$	0.075	0.073	0.121	0.119
	$\lambda_4 = \lambda_5$	0.646	0.641	0.726	0.724
		$T = 500$			
$(2,2,2,2,2)$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	0.086	0.085	0.171	0.166
	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	0.021	0.021	0.047	0.043
	$\lambda_1 = \lambda_2 = \lambda_3$	0.008	0.009	0.014	0.014
	$\lambda_1 = \lambda_2$	0.006	0.006	0.011	0.010
$(5,4,3,2,1)$	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	1.000	1.000	1.000	1.000
	$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$	0.987	0.987	0.991	0.992
	$\lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$	1.000	1.000	1.000	1.000
	$\lambda_1 = \lambda_2 = \lambda_3$	0.570	0.570	0.621	0.624
	$\lambda_3 = \lambda_4 = \lambda_5$	1.000	1.000	1.000	1.000
	$\lambda_1 = \lambda_2$	0.154	0.150	0.194	0.193
	$\lambda_4 = \lambda_5$	0.916	0.915	0.921	0.920

DGPs with $(\lambda_1, \ldots, \lambda_5) = (5, 4, 3, 2, 1)$ in Table 4. It is noteworthy, however, that there are distinct power differences for the different null hypotheses. Testing equality of λ_i's which are further apart in relative terms leads to larger rejection frequencies than for null hypotheses for relatively more similar λ_i's. For example, for the VAR(1) and sample size $T = 500$, testing $\mathbb{H}_0 : \lambda_1 = \lambda_2$ leads to relative rejection frequencies of 0.154 and 0.150 for $Q_r(0, 0)$ and $Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$, respectively, whereas the corresponding numbers for testing $\mathbb{H}_0 : \lambda_4 = \lambda_5$ are 0.916 and 0.915.

In Table 4, it can also be seen that the rejection frequencies of corresponding tests based on estimated versus known kurtosis parameters ($Q_r(0, 0)$ versus $Q_r(\tilde{\kappa}_1, \tilde{\kappa}_2)$) are again very similar. Thus, even for our largest models the properties of the tests do not deteriorate if estimated rather than known kurtosis parameters are used. This result confirms the conclusion from the bivariate processes that it is a good idea to always use estimated kurtosis parameters in practice.

For five-dimensional DGPs we can also explore what happens when the tests are applied in situations not covered by our asymptotic theory. Note that Proposition 1 does not apply for null hypotheses $\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$, $\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3$ and $\mathbb{H}_0 : \lambda_1 = \lambda_2$ when $(\lambda_1, \ldots, \lambda_5) = (2, 2, 2, 2, 2)$ because, under the assumptions of Proposition 1, the λ_i's not included in the null hypothesis have to be different from those included in the null hypothesis. For large sample sizes the corresponding tests reject much less frequently than specified by the nominal size. In other words, they are undersized. For a VAR(4) and sample size $T = 250$, the rejection frequencies are still larger than the nominal size for $\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4$, while the rejection frequencies for this case are markedly lower than for the null hypothesis $\mathbb{H}_0 : \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5$. Thus, the fact that the tests over-reject for smaller sample sizes may just be a reflection of their bias for large models and small sample sizes.

Rejecting less often than specified by the nominal significance level in cases where many relative variances are equal is perhaps not a big problem in practice because the correct conclusion of equal variances would be drawn more often then assumed on the basis of the chosen significance level. On the other hand, the excessive rejection rates in small samples could be a problem in applied work. The investigator could be misled to the conclusion that there is more identifying information in the second moments than there really is.

Overall the results for the five-dimensional DGP confirm some basic conclusions from bivariate DGPs. The tests tend to be oversized for large models and small sample sizes. Rather large sample sizes are necessary for an empirical size close to the nominal size. The test statistics based on estimated

kurtosis parameters should be used in practice because the small sample properties of the tests do not suffer from using estimated kurtosis parameters and they protect against assuming an incorrect distribution.

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 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 (2) 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 less formal evidence is in fact used by Chen and Netšunajev (2016) to justify the

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

assumption of distinct relative variances. Using this assumption they test the long-run neutrality of demand shocks and find evidence against long-run neutrality. 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 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.

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

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 (2) 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 $T_1 = 1983\text{m}4$ 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 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 also allows us to 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

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

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.

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. 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 rather than known.

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. Also 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 fully identified structure.

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.

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. 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.

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.

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

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 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.

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.