# Choosing Between Different Time-Varying Volatility Models for Structural Vector Autoregressive Analysis[1]

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

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

**Abstract**

The performance of information criteria and tests for residual heteroskedasticity for choosing between different models for time-varying volatility in the context of structural vector autoregressive analysis is investigated. Although it can be difficult to find the true volatility model with the selection criteria, using them is recommended because they can reduce the mean squared error of impulse response estimates substantially relative to a model that is chosen arbitrarily based on the personal preferences of a researcher. Heteroskedasticity tests are found to be useful tools for deciding whether time-varying volatility is present but do not discriminate well between different types of volatility changes. The selection methods are illustrated by specifying a model for the global market for crude oil.

*Key Words:* Structural vector autoregression, identification via heteroskedasticity, conditional heteroskedasticity, smooth transition, Markov switching, GARCH

*JEL classification:* C32

[1] The paper was written in part while the first author was a Bundesbank Professor at the Freie Universität Berlin. Financial support was provided by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk". The authors thank the high-performance computing service at Freie Universität Berlin for providing the computational resources required for this work. All calculations were conducted using the statistical software R3.3.3. The comments of two anonymous referees are gratefully acknowledged.

# 1 Introduction

Following the seminal publication by Engle (1982), time-varying volatility in time series data has received increasing attention. It has been diagnosed in many time series and it is taken into account to improve inference, for risk analysis and for economic studies not only in univariate but also in multivariate time series models. Engle introduced ARCH (autoregressive conditional heteroskedasticity) processes for modelling time-varying volatility. In the meantime a range of alternative models for conditional as well as unconditional heteroskedasticity have been developed to capture various volatility patterns. For applied work this raises the question which model is best suited for a given volatility pattern.

So far the choice of a volatility model is often not well justified. It is sometimes dictated by convenience, the preferences of the analyst or it is based on ad hoc criteria with unclear implications for the objective of the analysis. Hence, it is clearly desirable to better understand the implications of choosing a specific model for time-varying volatility and the second moment structure of a time series variable or set of variables more generally.

An improved understanding of the choice of the model for the second moment structure of a time series is best considered in the context of a specific modelling and analysis framework. A number of studies have compared the forecasting ability of univariate volatility models (e.g., Hansen and Lunde (2005), Becker and Clements (2008), and Caporin and McAleer (2012)). The performance of multivariate volatility models has been studied with the objective of covariance forecasting or risk assessment in mind (e.g., Caporin and McAleer (2011), Laurent, Rombouts and Violante (2012), and Becker, Clements, Doolan and Hurn (2015)). Some of these studies are based on specific data sets and may be difficult to generalize. Moreover, some studies compare generalized ARCH (GARCH) type models only.

A crucial issue in comparisons of volatility models is the metric used for the comparison. In the present study, we focus on structural vector autoregressive (SVAR) analysis. In that framework alternative models for time-varying volatility have been used to support the identification of structural shocks. Therefore we focus in this study on choosing an appropriate volatility structure for vector autoregressive (VAR) models with the specific objective of SVAR analysis in mind. Specifically, we consider volatility models that have been used as tools for identifying structural shocks in SVAR analysis.

The volatility models of interest in this context range from simple exogenous jumps in the residual variance (Rigobon (2003), Lanne and Lütkepohl (2008)) or a smooth transition between different volatility states (Lütkepohl and Netšunajev (2017b)) to more sophisticated GARCH processes (Normandin and Phaneuf (2004), Bouakez and Normandin (2010)) and Markov switching mechanisms (Lanne, Lütkepohl and Maciejowska (2010), Herwartz and Lütkepohl (2014)). In some of the related literature even competing models are applied to the same data and it is unclear

which of them best describes the time-varying volatility of a given system of variables (see, e.g., Lütkepohl and Netšunajev (2017a)). In such a situation having objective criteria that facilitate the selection of a model would be desirable. At the same time, however, little is known about the consequences of selecting a specific volatility model for parameter estimation and inference.

The objective of this study is to compare different procedures that discriminate between competing volatility models and to determine which ones are most helpful in deciding on a volatility model for SVAR analysis. We investigate how the choice of a model affects the outcome of the structural analysis and use evaluation criteria for model selection that are based on inference for impulse response analysis which is often the objective of SVAR analysis.

For this purpose we perform the following experiment. We generate time series with different types of volatility changes and then fit a set of different models that allow for changing volatility. These models are compared with different specification tests and model selection criteria. The goal is to determine which tests or criteria are best suited for finding the process that has actually generated the data and to investigate the impact of model selection on the structural parameters of interest.

We find that tests for heteroskedasticity are useful tools for detecting time-varying volatility but are less useful for deciding on a specific model type. We compare three standard information criteria for selecting the specific volatility model and find that it depends on the DGP how successful they are in selecting the correct model. Overall Akaike's AIC has a slight advantage over the other criteria in selecting the correct model. The models chosen by AIC also tend to provide impulse response estimates with relatively small mean squared errors. Any one of the criteria is better in this respect than choosing some volatility model by convenience or the subjective preferences of the analyst. Hence, it is recommended to use information criteria to support the selection of a suitable volatility model. We illustrate our selection strategy with a VAR model for the global market for crude oil from Kilian (2009).

The remainder of this study is organized as follows. The next section presents the general model setup and reviews four common volatility models that have been used in the SVAR literature on identification through heteroskedasticity and which are compared in the simulation study. Section 3 gives an overview of the different tools for model selection. In Section 4 the setup of the Monte Carlo study is outlined and the results of the simulation study are discussed in Section 5. The empirical illustration is presented in Section 6 and Section 7 concludes.

# 2 The Model Setup and the Volatility Models

## 2.1 Model Setup

We assume that the data generating process (DGP) is a vector autoregressive process of order $p$ (VAR($p$)) of the form

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

where $y_t = (y_{1t}, \ldots, y_{Kt})'$ is a vector of observable variables, the $A_i$ are $K \times K$ coefficient matrices, $\nu$ is a $K \times 1$ constant term and the $u_t$ are $K$-dimensional serially uncorrelated, heteroskedastic or conditionally heteroskedastic residuals.

The process is assumed to be stable such that

$$\det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \quad \text{for} \quad |z| \leq 1, z \in \mathbb{C}.$$

This condition implies that $y_t$ has a Wold moving average representation

$$y_t = \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}.$$

The underlying structural innovations of the DGP, $\varepsilon_t$, can be retrieved from (1) by a linear transformation of the residuals $u_t$,

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

If the $u_t$ have a time-invariant covariance matrix $\Sigma_u$, and the $\varepsilon_t$ are independently and identically distributed with variances standardized to unity, i.e., $\varepsilon_t \sim \text{iid}(0, I_K)$, it follows that $B$ must be such that $BB' = \Sigma_u$. In a conventional homoskedastic setting this structural model is referred to as a **B-model** in some of the structural VAR literature (see, e.g., Lütkepohl (2005)). In this model the elements of the matrix $B$ can be interpreted as the instantaneous effects or impact effects of the structural innovations on the observed variables $y_t$.

Often it is convenient to restrict the elements of the main diagonal of $B^{-1}$ to unity and leave the variances of the structural innovations unrestricted, i.e., $\Sigma_u = B\Sigma_\varepsilon B'$, where $\Sigma_\varepsilon = \mathbb{E}(\varepsilon_t \varepsilon_t')$. This model corresponds to a structural form

$$B^{-1} y_t = \nu^* + A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + \varepsilon_t,$$

where $\nu^* = B^{-1}\nu$ and $A_i^* = B^{-1} A_i$ for $i = 1, \ldots, p$. It is occasionally referred to as the **A-model** and we will also use this terminology in the following.

For both types of models the structural impulse responses are obtained as the elements of the matrices $\Theta_i = \Phi_i B$, $i = 0, 1, \ldots$. More precisely, the $kl^{\text{th}}$ element of $\Theta_i$, denoted by $\theta_{kl,i}$, is interpreted as the response of variable $k$ to the $l^{\text{th}}$ structural shock after a propagation horizon of $i$ periods. Notice that for the B-model the structural impulse responses have a size of one standard deviation because

the structural shocks have standard deviation 1 in a conventional homoskedastic setting. On the other hand, for the A-model the size of the impulse responses is typically not equal to the standard deviation of the structural shocks. This distinction is important later because the impulse responses based on the A-model do not explicitly require an estimate of the structural innovation variances which may be an advantage when the variance is not time-invariant.

## 2.2 Volatility Models

In our comparison we consider a range of different models for heteroskedasticity or conditional heteroskedasticity which have all been used in SVAR analysis in the context of identification through heteroskedasticity.

### Exogenous Volatility Changes

In the first model for time-varying volatility, the covariance changes are assumed to occur at prespecified break dates,

$$\mathbb{E}(u_t u_t') = \Sigma_t = \Sigma_u(m) \quad \text{for} \quad t \in \mathcal{T}_m, \quad m = 1, \ldots, M, \tag{3}$$

where $\mathcal{T}_m = \{T_{m-1} + 1, \ldots, T_m\}$ $(m = 1, \ldots, M)$ are $M$ given volatility regimes of consecutive time periods. The $T_m$, for $m = 1, \ldots, M-1$, represent the time periods of volatility changes with $T_0 = 0$ and $T_M = T$. The change points $T_m$ may be predetermined by some statistical procedure.

Under the assumption of a constant instantaneous impact effects matrix $B$, for each volatility regime a decomposition

$$\Sigma_u(m) = B\Lambda_m B' \tag{4}$$

exists, where $\Lambda_1 = I_K$ and $\Lambda_m = \text{diag}(\lambda_{1,m}, \ldots, \lambda_{K,m})$ $(m = 2, \ldots, M)$ are diagonal matrices with strictly positive elements that can be interpreted as the changes of the variances of the structural innovations in regime $m$ relative to the first regime. Lanne et al. (2010) state the conditions under which the decomposition in (4) is (locally) unique and, hence, the structural parameters $B$ and $\Lambda$ are identified.

Assuming Gaussian residuals $u_t$, the log-likelihood function is

$$\log l(\boldsymbol{\beta}, \boldsymbol{\sigma}) = -\frac{KT}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log\det(\Sigma_t) - \frac{1}{2}\sum_{t=1}^{T} u_t'\Sigma_t^{-1}u_t, \tag{5}$$

where $\boldsymbol{\beta} = \text{vec}[\nu, A_1, \ldots, A_p]$ and $\boldsymbol{\sigma}$ contains all unknown covariance parameters. Thus, estimation by maximum likelihood (ML) of the model is straightforward.

In the simulation study we consider models with $M = 2$ volatility states. To find the breakpoint $T_1$, we fit a model allowing for one break point in the range of $[0.15T, 0.85T]$ and pick the model with the highest log-likelihood. Note, that this procedure adds another parameter that has to be accounted for in the calculation of information criteria, for example.

### Smooth Transition in Variances

As an alternative to a shift in the variances at some points in time, one may model the change in the residual covariance matrix as a smooth transition from a volatility regime with a positive definite covariance matrix $\Sigma_u(1)$ to a regime with $\Sigma_u(2)$ such that

$$\mathbb{E}(u_t u_t') = \Sigma_t = (1 - G(\gamma, c, s_t))\Sigma_u(1) + G(\gamma, c, s_t)\Sigma_u(2), \tag{6}$$

where the transition function $G(\gamma, c, s_t) = (1+\exp[-\exp(\gamma)(s_t - c)])^{-1}$ is a logistic function that depends on the smoothness parameter $\gamma$, the location parameter $c$ and a transition variable $s_t$. In our setup, small values of the smoothness parameter $\gamma$ imply a slow, gradual transition from one volatility regime to the other. When the smoothness parameter becomes very large, however, the transition resembles a step function with a discrete change between volatility states. A locally unique decomposition of the reduced form covariance matrices such as displayed in (4) is obtained if $\Sigma_u(1) = BB'$ and $\Sigma_u(2) = B\Lambda B'$, where the diagonal matrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_K)$ has distinct, strictly positive values $\lambda_k$ $(k = 1, \ldots, K)$. This model was proposed and used by Lütkepohl and Netšunajev (2017b) in the context of SVAR analysis.

For Gaussian $u_t$, the log-likelihood can be written as in (5). It is now a function of the transition parameters as well. An iterative procedure for estimation is discussed in detail by Lütkepohl and Netšunajev (2017b). Since the range of the smoothness and threshold parameters $\{\gamma, c\}$ can be bounded, a grid search can be performed over the relevant range of these two parameters.

### Markov Switching in Covariances

Alternatively, the volatility changes may be generated by a discrete Markov process $s_t$ $(t = 0, \pm 1, \pm 2, \ldots)$ with states $1, \ldots, M$, and transition probabilities $p_{ij} = \Pr(s_t = j | s_{t-1} = i)$ $(i, j = 1, \ldots, M)$. The distribution of $u_t$ is assumed to be Gaussian conditionally on $s_t$ and has a state dependent covariance matrix,

$$u_t | s_t \sim \mathcal{N}(0, \Sigma_u(s_t)). \tag{7}$$

Decomposing the residual covariance matrices as in (4), (local) uniqueness of $B$ holds under the conditions stated in Lanne and Lütkepohl (2008). The model was first proposed for SVAR analysis by Lanne et al. (2010). It has been used in a range of applied SVAR studies including Velinov and Chen (2015), Lütkepohl and Netšunajev (2014), and Lütkepohl and Velinov (2016).

The parameters of the model can be estimated by ML. The log-likelihood function is

$$\log l(\boldsymbol{\beta}, \boldsymbol{\sigma}, P) = \sum_{t=1}^{T} \log \left( \sum_{m=1}^{M} \Pr(s_t = m | Y_{t-1}) f(y_t | s_t = m, Y_{t-1}) \right), \tag{8}$$

where $P = [p_{ij}]$ is the $M \times M$ matrix of transition probabilities, $Y_{t-1} = (y'_{t-1}, \ldots, y'_{t-p})'$ and

$$f(y_t|s_t = m, Y_{t-1}) = (2\pi)^{-K/2}(\det \Sigma_u(m))^{-1/2} \exp\left\{-\frac{1}{2}u'_t\Sigma_u(m)^{-1}u_t\right\}.$$

Optimization of the log-likelihood is difficult because (i) it is unbounded so that there is no global maximum and a suitable local maximum has to be found, (ii) it is highly nonlinear and has multiple local optima, (iii) the label-switching problem has to be overcome, which arises if the ordering of the regimes changes in the course of the optimization. A feasible procedure for estimation of the Markov switching model based on an EM algorithm was proposed by Herwartz and Lütkepohl (2014). It is the method used for estimating Markov switching SVAR models in our simulations reported later in this study. When using the model, the number of volatility states, $M$, has to be specified by the analyst.

**Multivariate GARCH**

Multivariate GARCH processes offer yet another possibility to model time-varying volatility. Many different proposals for parametrizing multivariate GARCH processes are available in the literature (e.g., Bauwens, Laurent and Rombouts (2006), Silvennoinen and Teräsvirta (2009)). In the context of SVAR analysis the GO-GARCH model proposed by van der Weide (2002) is typically used. It specifies the volatility changes as

$$\mathbb{E}(u_t u'_t|\mathcal{F}_{t-1}) = \Sigma_{t|t-1} = B\Lambda_{t|t-1}B'. \tag{9}$$

Here $\mathcal{F}_t$ denotes the information available at time $t$,

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

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

$$\sigma^2_{k,t|t-1} = (1 - \gamma_k - g_k) + \gamma_k\varepsilon^2_{k,t-1} + g_k\sigma^2_{k,t-1|t-2}, \quad k = 1, \ldots, K, \tag{10}$$

where $\varepsilon_{k,t} = b^*_k u_t$ and $b^*_k$ is the $k^{\text{th}}$ row of $B^{-1}$ $(k = 1, \ldots, K)$. Moreover, $g_k \geq 0$, $\gamma_k > 0$, $g_k + \gamma_k < 1$ $(k = 1, \ldots, K)$. The model has been proposed and used for SVAR analysis by Normandin and Phaneuf (2004) and Bouakez and Normandin (2010), for example. Identification conditions for uniqueness of $B$ are stated in Sentana and Fiorentini (2001) and Milunovich and Yang (2013).

The setup of the model implies an unconditional residual covariance matrix $\mathbb{E}(u_t u'_t) = \Sigma_u = BB'$. Under Gaussian assumptions for the $\varepsilon_{k,t}$ the log-likelihood of the model is $\log l = \sum_{t=1}^{T} \log f_{t|t-1}(y_t)$, where the conditional densities have the form

$$f_{t|t-1}(y_t) = (2\pi)^{-K/2} \det(\Sigma_{t|t-1})^{-1/2} \exp\left(-\frac{1}{2}u'_t\Sigma^{-1}_{t|t-1}u_t\right). \tag{11}$$

The log-likelihood function is highly nonlinear which makes the maximization and, thus, ML estimation of the GARCH parameters and the impact effects matrix $B$ numerically challenging. In the simulation study we use a two-step algorithm for ML estimation described by Lanne and Saikkonen (2007). In the first step, the estimation procedure is broken down in univariate GARCH estimations to get initial estimates of the parameters of the volatility model and in the second step, a full, joint ML estimation of the parameters is performed starting from the initial estimates obtained in the first step.

### Alternative Volatility Models

The models for time-varying volatility presented so far all have been used in the SVAR literature for identifying the structural parameters. There are also other volatility models that have been used to account for heteroskedasticity or conditional heteroskedasticity in VAR and SVAR models without contributing to structural identification. For example, there are many variants of multivariate GARCH models that can be used to capture time-varying second moments in VAR models, as mentioned earlier (see Bauwens et al. (2006), Silvennoinen and Teräsvirta (2009)). Furthermore, different types of stochastic volatility models have been proposed and considered (e.g., Uhlig (1997), Primiceri (2005), Kim (2014)). In fact, Primiceri's model has become quite popular as a model for time-varying coefficients and has been used in other SVAR studies as well (see Kilian and Lütkepohl (2017, Section 18.5)). While these models are also useful to capture time-varying second moments, they have not been used for identification through heteroskedasticity and are therefore not included in our comparison, because we want to use the ability of the volatility models to generate accurate estimates of structural impulse responses as one metric for comparing model selection procedures.

### Model Selection

For a given set of time series it is often not clear which model best describes the volatility changes. Sometimes subject matter considerations may suggest that one model is more plausible than another one. In practice, the choice may be arbitrary to some extent and perhaps driven by convenience or the subjective preferences of the analyst. Having more systematic statistical procedures for deciding on the specific volatility model is desirable. It should also be noted that actual changes in the volatility observed in plotted residuals from all the models can look similar and, hence, purely subjective eye-balling criteria are problematic. In the next section some statistical tools for model selection and model comparison are considered.

# 3 Tools for Model Comparison

In this section two alternative strategies for comparing different volatility models are presented. One possibility is to choose the volatility model that optimizes some model selection criterion. The second approach is to check whether the chosen model wipes off all the volatility changes in the residuals or leaves some remaining heteroskedasticity or conditional heteroskedasticity in the residuals. Obviously, a model that accounts for all volatility changes is preferable to one that cannot fully account for the heteroskedasticity in the data. Standard model selection criteria are presented first for model comparison and then a couple of diagnostic tests for remaining GARCH are presented.

## 3.1 Model Selection Criteria

Since Gaussian ML estimation or quasi ML estimation is possible for all the models considered in the previous section, likelihood based information criteria for model comparison can be used. For example, the AIC criterion (Akaike (1974)),

$$\text{AIC} = -2\log l + 2 \times \text{number of free parameters},$$

the Hannan-Quinn criterion (Hannan and Quinn (1979)),

$$\text{HQ} = -2\log l + 2\log\log(T) \times \text{number of free parameters},$$

or the Bayesian criterion proposed by Schwarz (1978),

$$\text{BIC} = -2\log l + \log(T) \times \text{number of free parameters},$$

are standard criteria, some of which have been used also for choosing between volatility models (e.g., Lütkepohl and Netšunajev (2017a)). However, so far little is known about the suitability of these criteria for choosing between different volatility models for SVAR analysis.

Clearly, the best model chosen with one of these criteria is not necessarily a good model in the sense that it captures all the volatility changes in a given set of time series. To make sure there is no left-over heteroskedasticity one may consider applying diagnostic tests, as presented in the next subsection.

## 3.2 Diagnostic Tests

If the (conditional) heteroskedasticity is fully captured by the (conditional) covariance matrix $\Sigma_t$ ($\Sigma_{t|t-1}$), then the standardized residuals, $u_t^s = \Sigma_t^{-1/2} u_t$ ($\Sigma_{t|t-1}^{-1/2} u_t$) should be free of heteroskedasticity or conditional heteroskedasticity. Hence, standard GARCH tests can be applied to investigate whether a specific model has cleaned the residuals from volatility changes. A model is rejected for SVAR analysis if it leaves heteroskedasticity in the standardized residuals. The tests can also

be used to investigate whether there are volatility changes in the original reduced form residuals $u_t$. Tests based on the following statistics can be considered (see Lütkepohl and Milunovich (2016) for details on the theoretical underpinning).

**Univariate autocovariances** A test statistic can be based on univariate autocovariances of the sum of squared residuals,

$$Q_1(H) = T \sum_{j=1}^{H} [\tilde{\gamma}(j)/\tilde{\gamma}(0)]^2, \tag{12}$$

where $\tilde{\gamma}(j) = T^{-1} \sum_{t=j+1}^{T} \xi_t \xi_{t-j}$, and $\xi_t = u_t^{s\prime} u_t^s - T^{-1} \sum_{i=1}^{T} u_i^{s\prime} u_i^s$. The statistic has a $\chi^2(H)$ distribution if the standardized residuals are not conditionally heteroskedastic.

**Multivariate autocovariances** Alternatively, a test can be based on multivariate autocovariances,

$$Q_2(H) = T \sum_{j=1}^{H} \text{tr}[\tilde{\Gamma}(j)' \tilde{\Gamma}(0)^{-1} \tilde{\Gamma}(j) \tilde{\Gamma}(0)^{-1}], \tag{13}$$

where $\tilde{\Gamma}(j) = T^{-1} \sum_{t=j+1}^{T} \vartheta_t \vartheta_{t-j}'$ for $j = 0, 1, \ldots,$ and $\tilde{\Gamma}(j) = \tilde{\Gamma}(-j)'$ for $j < 0$. Furthermore, $\vartheta_t = \text{vech}(u_t^s u_t^{s\prime}) - T^{-1} \sum_{i=1}^{T} \text{vech}(u_i^s u_i^{s\prime})$ and vech denotes the half-vectorization operator. The statistic has an asymptotic $\chi^2\left(\frac{1}{4} H K^2 (K+1)^2\right)$ distribution under the null hypothesis of no remaining conditional heteroskedasticity.

**LM statistic** A third option is an LM statistic based on the regression

$$\eta_t = \delta_0 + D_1 \eta_{t-1} + \cdots + D_H \eta_{t-H} + w_t, \tag{14}$$

where $\eta_t = \text{vech}(u_t^s u_t^{s\prime})$, $\delta_0$ is a $\frac{1}{2}K(K+1)$-dimensional fixed vector, the $D_i$, $i = 1, \ldots, H$, are $\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1)$ parameter matrices and $w_t$ is an error term. The LM statistic for testing the null hypothesis $\mathbb{H}_0 : D_1 = \cdots = D_H = 0$ is

$$LM(H) = \tfrac{1}{2} T K (K+1) - T \text{tr}[\tilde{\Sigma}_w \tilde{\Gamma}(0)^{-1}], \tag{15}$$

where $\tilde{\Sigma}_w$ is the estimated residual covariance matrix of (14). The LM statistic also has an asymptotic $\chi^2\left(\frac{1}{4} H K^2 (K+1)^2\right)$ distribution if the standardized residuals are free of ARCH. In practice, the standardized residuals are, of course, replaced by estimated residuals in all these test statistics.

# 4 Design of Monte Carlo Comparison

## 4.1 DGPs

To investigate the relative merits of the alternative approaches of choosing between different volatility models, we perform a simulation experiment based on a range of different types of bivariate DGPs, i.e., $K = 2$. The conditional mean part of the model is a VAR(1) process of the form

$$y_t = \nu + A_1 y_{t-1} + u_t \tag{16}$$

with $\nu = 0$. For given $A_1$ and starting value $y_0 = (0, 0)'$, we generate the data by drawing the residuals $u_t$ from distributions characterized by the following types of volatility processes:

**Heteroskedastic DGP:** $u_t \sim \mathcal{N}(0, \Sigma_t)$ are generated with

$$\Sigma_t = \begin{cases} BB' & \text{for } t = 1, \ldots, 0.5T, \\ B\Lambda B' & \text{for } t = 0.5T + 1, \ldots, T, \end{cases}$$

where the specific values used for the structural parameters $B$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2)$ are discussed below. The change in volatility occurs in the middle of the sample.

**Smooth Transition DGP:** We choose our transition variable $s_t$ to be time, $s_t = t$ (i.e., $s_t = 1, \ldots, T$), and set the smoothness parameter $\gamma = -3$ such that a rather smooth transition is ensured between the volatility regimes in order to differentiate the process from a heteroskedastic DGP. Note, that holding $\gamma$ constant leads to transition functions whose exact shape depends on the sample size $T$. The location parameter $c$ is set to $0.5T$ such that the transition from one state to the other occurs in the center of the sample period. The covariance matrices $\Sigma_u(1) = BB'$ and $\Sigma_u(2) = B\Lambda B'$ depend on the specific structural parameters $B$ and $\Lambda$ used. The $u_t$ are generated by drawing from a $\mathcal{N}(0, \Sigma_t)$ distribution with

$$\Sigma_t = (1 - G(\gamma, c, t))\Sigma_u(1) + G(\gamma, c, t)\Sigma_u(2). \tag{17}$$

**Markov Switching DGP:** The Markov process has two states and a matrix of transition probabilities

$$P = \begin{bmatrix} .9 & .1 \\ .2 & .8 \end{bmatrix}.$$

The states $s_1, \ldots, s_T$, $s_t \in \{1, 2\}$, are first generated with these transition probabilities and then the $u_t$ $(t = 1, \ldots, T)$ are generated by drawing from $\mathcal{N}(0, \Sigma_{s_t})$ distributions with $\Sigma_1 = BB'$ and $\Sigma_2 = B\Lambda B'$.

**GARCH DGP:** In order to generate GARCH residuals, we first generate bivariate standard normal variates, $(e_{1t}, e_{2t})' \sim \mathcal{N}(0, I_2)$, and

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

where $\epsilon_{k,t} = e_{k,t}\sigma_{k,t|t-1}$ for $t = 1, \ldots, T$. Finally, the $u_t = B\Lambda^{1/2}_{t|t-1}e_t$ are generated, where $\Lambda_{t|t-1} = \text{diag}(\sigma^2_{1,t|t-1}, \sigma^2_{2,t|t-1})$. Thereby, the unconditional covariance matrix of $u_t$ is $\Sigma_u = BB'$. We use GARCH parameters $(\gamma_1, g_1) = (0.1, 0.85)'$, and $(\gamma_2, g_2) = (0.05, 0.92)$. The choice of the GARCH parameters ensures rather persistent volatility patterns as they are often observed in practice (see, e.g., Bouakez and Normandin (2010, Table 1)).

In our baseline specification we set

$$B = \begin{bmatrix} 1 & 0 \\ -1 & 10 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 7 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.4 \end{bmatrix}. \tag{18}$$

The choice of $B$ is motivated by two considerations. First, we restrict the upper right element of the matrix to zero ($b_{12} = 0$) such that the true structural model is recursive, as often assumed in applied work. Second, we choose a distinguishable sign pattern for the two columns of $B$ as well as distinct magnitudes of its diagonal elements. Thus, we are able to induce a unique ordering of the estimated $B$ matrices that corresponds with the true values of $B$.[2] The elements of $\Lambda$ are chosen such that the volatility regimes are clearly distinct. The magnitudes of the values for $\Lambda$ in (18) are well in line with estimates in the applied literature (see, e.g., Lütkepohl and Netšunajev (2017a, Table 2)). The VAR coefficient matrix $A_1$ is chosen such that the system is stable (stationary).

## 4.2 Fitted Models

We fit VAR(1) models with intercept and all four competing volatility models from Section 2 plus a VAR(1) model not accounting for heteroskedasticity to all the series. More precisely, the following models are fitted:

**H-VAR:** A heteroskedastic VAR(1) model with one change point in the residual covariance is fitted without assuming the change point. The change point is estimated as explained in Section 2.

[2]Since the $B$ matrix is identified only up to column sign and column permutations, some ordering of the columns is necessary and in the absence of an economically justified ordering, for data generated by the heteroskedastic DGP, the smooth transition DGP and the Markov switching DGP, we order the associated estimated models with respect to the size of the elements of $\Lambda$, where the smallest element is placed in the upper left corner and the columns of $B$ are ordered accordingly. For the models fitted to the GARCH DGP and the estimated GARCH model for all DGPs, however, we normalize the estimated $B$ to have positive entries on the main diagonal and place its largest element in the lower-right corner.

**ST-VAR:** VAR(1) models with smooth transition in the residual covariance matrices are fitted by the Gaussian ML procedure described in Section 2.

**MS-VAR:** VAR(1) models with volatility changes generated by a Markov switching mechanism with two volatility states are fitted by the ML procedure mentioned in Section 2.

**GARCH-VAR:** A VAR(1) with GO-GARCH residuals is fitted by the ML procedure discussed in Section 2.

**Plain VAR** denotes a plain VAR(1) model fitted by equationwise least squares without allowing for heteroskedasticity or conditional heteroskedasticity. For the impulse responses a recursive identification is assumed. Obviously, by construction the recursively identified structural VAR model has an advantage over the models identified by heteroskedasticity because it imposes a true restriction on the structural parameters.

We fit all these models to all series regardless of the true DGP and compare them with the model selection criteria, AIC, HQ and BIC. The plain VAR model is included to investigate whether the model selection criteria are biased against models with heteroskedasticity which may be a problem in particular for smaller samples because modelling the time-varying volatility increases the number of parameters. We also apply diagnostic tests to investigate which model is most suitable for capturing the volatility changes and to see whether the model selection criteria are able to properly detect models that do not remove the time-varying volatility. Later, we turn to impulse response functions to assess the consequences of model selection on estimates of these quantities. The number of replications for each simulation design is 500. The simulations are performed for sample sizes of $T = 100$, 200 and 500.

# 5 Results of the Monte Carlo Study

Selected Monte Carlo results are presented in the Tables 1 and 2. A range of further more detailed results is available in the online appendix which accompanies this article. The following discussion refers also to the results in the online appendix which the reader is encouraged to consult.

## 5.1 Properties of Information Criteria

For each simulation round we compute and compare the information criteria for all estimated models and select the model that minimizes the respective criterion. The relative selection frequencies for sample size $T = 200$ are shown in Table 1 for the four different DGPs. The following observations emerge from the table.

Table 1: Relative Frequencies of Volatility Models Chosen by Model Selection Criteria for Sample Size $T = 200$

| | | Relative frequencies of volatility models selected | | | | |
|---|---|---|---|---|---|---|
| DGP | Criterion | H-VAR | ST-VAR | MS-VAR | GARCH-VAR | plain VAR |
| Heteroskedastic | AIC | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | HQ | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | BIC | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Smooth | AIC | 0.54 | 0.46 | 0.00 | 0.00 | 0.00 |
| transition | HQ | 0.66 | 0.34 | 0.00 | 0.00 | 0.00 |
| | BIC | 0.78 | 0.22 | 0.00 | 0.00 | 0.00 |
| Markov | AIC | 0.03 | 0.00 | 0.97 | 0.00 | 0.00 |
| switching | HQ | 0.03 | 0.00 | 0.96 | 0.00 | 0.01 |
| | BIC | 0.03 | 0.00 | 0.88 | 0.00 | 0.09 |
| GARCH | AIC | 0.70 | 0.02 | 0.14 | 0.10 | 0.04 |
| | HQ | 0.58 | 0.01 | 0.09 | 0.06 | 0.27 |
| | BIC | 0.30 | 0.00 | 0.05 | 0.03 | 0.62 |

Note: Based on 500 replications of simulation experiment.

(1) The chances of finding the correct DGP by one of the information criteria depend on the underlying DGP. For example, the heteroskedastic DGP is reliably detected by all three criteria, whereas the GARCH DGP is only selected with very low probability for sample size $T = 200$ (Table 1). For the DGPs that pose more difficulties for the information criteria, the chances to pick the correct one increase with the sample size (see the online appendix).

(2) AIC is overall the most successful criterion in selecting the correct volatility process. In Table 1 it finds the correct DGP with the highest frequency, although the selection frequency of the correct DGP may still be low.

(3) Some DGPs are difficult to disentangle. For example, the smooth transition DGP is difficult to distinguish from a heteroskedastic DGP, as can be seen in Table 1, where all three criteria choose the heteroskedastic DGP with a higher frequency than the true smooth transition DGP. However, this finding is in line with the known difficulties of precisely estimating the parameter $\gamma$ which governs the smoothness of the transition between the two regimes of the ST-VAR (see, e.g., van Dijk, Teräsvirta and Franses (2002)).

In addition to the results shown in Table 1 and the related results for other sample sizes in the online appendix, we have also performed the simulation experiment without including the plain VAR model in the competition. Excluding the plain VAR model, i.e., selecting only from models with time-varying volatility,

in general does not lead to a higher detection rate of true models (the results are also shown in the online appendix). When the plain VAR model is excluded, the selection criteria select the H-VAR model more often instead.

Summarizing, the AIC, HQ and BIC information criteria are reliable indicators for sorting out the underlying processes in some situations: While a heteroskedastic DGP is reliably detected, the sample size must be sufficiently large to correctly find volatility patterns generated by the smooth transition DGP and the Markov switching mechanism. All three criteria have difficulties detecting the GARCH DGP even for relatively large samples of size $T = 500$ (see the online appendix).

Not surprisingly, the tendency to completely ignore the volatility pattern in the data by choosing the plain VAR model decreases with increasing sample size and vanishes for all DGPs for $T = 500$ except for the GARCH DGP. For all DGPs and sample sizes, the AIC outperforms the other criteria in its ability to correctly detect the underlying DGP.

## 5.2 Diagnostic Tests for Left-Over Heteroskedasticity

Relative rejection frequencies at the 5% significance level of the three ARCH tests described in Subsection 3.2 for all DGPs when a plain VAR model or the model selected by an information criterion is fitted are presented in the online appendix. The three alternative tests yield similar results and no systematic pattern in favour of one of the three tests is apparent. They all have power against the time-varying volatility generated by the four DGPs. Thus, the tests are useful tools to support a decision of whether time-varying volatility is present or not.

We have also applied the tests to the residuals of each of the models when fitted regardless of the DGP. Although the tests indicate that there is left-over heteroskedasticity, they are not helpful in discriminating between different DGPs. Therefore we do not show detailed results.

The overall conclusions from our simulation experiments for the tests for left-over heteroskedasticity are that they are useful tools for deciding about the presence of time-varying volatility in the residuals but not for discriminating between specific models. They confirm that the information criteria AIC, HQ and BIC all select models that capture the volatility changes in a set of time series at least to some extent. AIC is best in this respect. In other words, for models selected by AIC the rejection frequencies of the tests are close to the significance level chosen.

## 5.3 Implications for Structural Analysis

So far we have discussed the ability of information criteria to discriminate correctly between different models for time-varying volatility and point to the correct underlying DGP. In the context of SVAR analysis it is of particular interest to determine the implications of model selection for the estimation precision of structural impulse responses because these quantities are often in the focus of the analysis.

In this section we investigate the properties of the estimated structural impulse responses obtained from the models with time-varying volatility when identification through heteroskedasticity is used. We fit all the models and estimate the impulse responses also from the wrong models to see how much estimation precision is lost by fitting a false volatility model. Alternatively, we use information criteria to choose among the models allowing for time-varying volatility. The idea is that a researcher who finds time-varying volatility in the data decides to use that feature for identifying structural shocks. S/he may decide to use a specific volatility model or to apply model selection criteria for making a choice on the volatility model.

To evaluate the estimation precision of the impulse responses, we calculate the cumulated mean squared errors (MSEs) of the impulse response functions relative to the impulse response functions estimated from the actual DGP. We do not restrict the analysis to the instantaneous effects matrix, $B$, but focus on the first five estimated impulse responses. Thereby we capture at least to some extent estimation efficiency losses and gains for the VAR slope parameters as well. Given the VAR slope parameter values of our DGPs, the true impulse responses are close to zero for larger propagation horizons than five so that a propagation horizon of $h = 5$ captures the interesting part of the impulse responses for our purposes. The MSE of the impulse response functions up to horizon $h$ for variable $k$ induced by shock $l$ is calculated as

$$MSE_h(\theta_{kl,\bullet}) = \sum_{i=0}^{h-1} \left( \frac{1}{S} \sum_{s=1}^{S} (\theta_{kl,i} - \hat{\theta}_{kl,i}(s))^2 \right), \tag{19}$$

where $\hat{\theta}_{kl,i}(s)$ denotes the estimate of the structural impulse response $\theta_{kl,i}$ obtained in the $s^{\text{th}}$ replication of our simulation experiment.

Table 2 displays the results of the simulation experiment for sample size $T = 200$ and the corresponding results for $T = 100$ and $T = 500$ are given in the online appendix. The impulse response functions based on the B-model involve estimating the innovation variance given that a structural shock is of size one standard deviation in one of the volatility regimes. In contrast, the structural impulse responses based on the A-model do not directly involve an estimate of the residual variance. The corresponding MSEs relative to the MSEs of the DGP are shown in the right-hand columns of Table 2. The following main conclusions can be drawn.

(1) The estimated DGP typically results in the smallest MSEs. The smooth transition DGP is the only exception to this rule. For that DGP the impulse response estimates obtained from an H-VAR model can, in fact, be smaller than those from the true model (see Table 2). Of course, this result is driven by the very simple volatility change in the smooth transition DGP which can be captured well by an H-VAR model. The result is partly due to the moderate sample size and disappears for $T = 500$ (see the online appendix).

Table 2: MSEs of Impulse Response Functions of Fitted Models Relative to True Volatility Model (DGP) for Propagation Horizon up to $h = 5$ and Sample Size $T = 200$

| | | B-model | | | | A-model | | | |
|---|---|---|---|---|---|---|---|---|---|
| DGP | Fitted model | $\theta_{11}$ | $\theta_{21}$ | $\theta_{12}$ | $\theta_{22}$ | $\theta_{11}$ | $\theta_{21}$ | $\theta_{12}$ | $\theta_{22}$ |
| Heteroskedastic | H-VAR | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | ST-VAR | 0.99 | 1.00 | 1.01 | 1.02 | 0.99 | 1.00 | 1.00 | 1.01 |
| | MS-VAR | 1.17 | 1.18 | 1.09 | 1.06 | 1.07 | 1.04 | 1.07 | 1.11 |
| | GARCH-VAR | 3.70 | 3.30 | 43.90 | 64.78 | 1.98 | 1.93 | 1.65 | 1.94 |
| | AIC | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | HQ | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | BIC | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Smooth transition | H-VAR | 1.30 | 1.35 | 0.52 | 0.42 | 1.19 | 1.16 | 1.13 | 1.18 |
| | ST-VAR | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | MS-VAR | 3.26 | 3.36 | 0.76 | 0.54 | 1.76 | 1.52 | 2.18 | 2.91 |
| | GARCH-VAR | 2.76 | 2.77 | 6.74 | 6.86 | 1.59 | 1.57 | 1.51 | 1.62 |
| | AIC | 1.16 | 1.16 | 1.01 | 0.99 | 1.10 | 1.08 | 1.12 | 1.19 |
| | HQ | 1.20 | 1.20 | 0.97 | 0.93 | 1.11 | 1.08 | 1.13 | 1.21 |
| | BIC | 1.26 | 1.29 | 0.90 | 0.84 | 1.17 | 1.14 | 1.17 | 1.24 |
| Markov switching | H-VAR | 34.29 | 37.30 | 18.85 | 10.49 | 9.21 | 4.51 | 17.73 | 32.00 |
| | ST-VAR | 35.24 | 38.28 | 19.16 | 10.86 | 9.22 | 4.46 | 17.71 | 32.02 |
| | MS-VAR | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | GARCH-VAR | 5.02 | 5.69 | 13.62 | 16.33 | 3.65 | 3.74 | 2.63 | 2.83 |
| | AIC | 1.58 | 1.63 | 1.48 | 1.36 | 1.16 | 1.08 | 1.37 | 1.67 |
| | HQ | 1.67 | 1.75 | 1.52 | 1.34 | 1.20 | 1.11 | 1.43 | 1.78 |
| | BIC | 2.05 | 2.22 | 1.77 | 1.45 | 1.32 | 1.16 | 1.68 | 2.24 |
| GARCH | H-VAR | 1.79 | 1.64 | 1.78 | 1.90 | 1.55 | 1.54 | 1.59 | 1.67 |
| | ST-VAR | 1.84 | 1.61 | 1.77 | 1.98 | 1.51 | 1.50 | 1.50 | 1.58 |
| | MS-VAR | 2.00 | 1.35 | 2.34 | 2.47 | 1.90 | 1.72 | 2.04 | 2.56 |
| | GARCH-VAR | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | AIC | 1.82 | 1.51 | 1.89 | 1.97 | 1.56 | 1.50 | 1.64 | 1.89 |
| | HQ | 1.81 | 1.53 | 1.88 | 1.97 | 1.55 | 1.51 | 1.62 | 1.79 |
| | BIC | 1.81 | 1.58 | 1.85 | 1.94 | 1.56 | 1.52 | 1.61 | 1.76 |

Note: Based on 500 replications of simulation experiment. Model selection by criteria only among volatility models.

(2) Using model selection criteria is better in terms of MSE than choosing an arbitrary model by convenience or personal preference. For example, if the true DGP is the Markov switching DGP, trying to model the volatility changes by a GARCH process because such a model is quite common if volatility changes are detected, one would obtain much less precise estimates of the structural impulse responses than from the models recommended by the information criteria.

(3) Among the model selection criteria, AIC tends to result in the smallest MSEs and BIC delivers the largest values. Although it depends to some extent on the DGP and the sample size which criterion provides the best MSEs, if AIC does not provide the smallest MSEs, its MSEs are usually very close to the smallest ones among the information criteria. In most cases the MSEs of all three information criteria are close together.

(4) When identification through heteroskedasticity is used, the A-model does not have a clear advantage over the B-model. In particular, if the true DGP is unknown and information criteria are used for picking a volatility model, the B-model may lead to similar estimation precision for the structural impulse responses as the A-model (see, e.g., the results for the GARCH DGP in Table 2).

The main message from these results is that, if identification through heteroskedasticity is considered, it is recommended to use AIC for selecting the volatility model rather than making an arbitrary choice.

# 6 Empirical Example

To illustrate the selection of a volatility model for a structural VAR analysis we use an example from Lütkepohl and Netšunajev (2014) who consider a model for the global market for crude oil from Kilian (2009) with the following three variables: percent change in global crude oil production ($\Delta prod_t$), a log detrended index of real economic activity ($q_t$), and the log of the real price of oil ($p_t$). Thus, $y_t = (\Delta prod_t, q_t, p_t)'$. The same set of variables is also used in a structural VAR study by Kilian and Murphy (2012) without allowing for heteroskedasticity. Lütkepohl and Netšunajev (2014) fit MS-VAR(3) models to monthly data for the period 1973m2 - 2006m12. Hence, the sample size is $T = 404$. Their preferred model has three volatility states and is signified as MS(3)-VAR(3) model. They perform a structural analysis with that model.

We apply the same ordering of the variables as Kilian (2009) for the recursive identification of a plain structural VAR(3) model. The reduced-form residuals of this model are plotted in Figure 1. They display clear changes in volatility. The residuals of the oil production equation are much more volatile in the first part

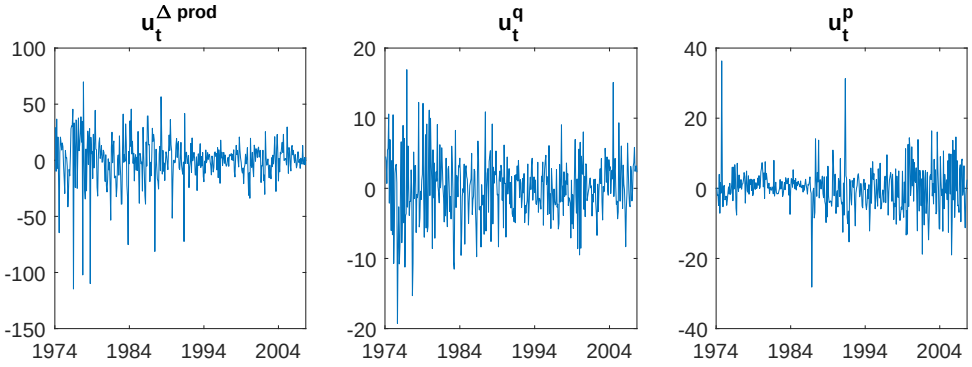


Figure 1: Reduced-form residuals of a plain VAR(3) model for the example data.

of the sample while the oil price residuals are substantially more volatile in the later sample period. Although the graphs contain a clear message already, we have performed our ARCH tests from Section 3.2 and show the results in Table 3. Clearly, all diagnostic tests have very small $p$-values and, hence, indicate time-varying residual volatility. Thus, allowing for this feature in the model is desirable.

Table 3: ARCH-Tests for Plain VAR(3) Model for Example Data

| | $p$-values | | | |
|---|---|---|---|---|
| Test | $H = 1$ | $H = 3$ | $H = 5$ | $H = 10$ |
| $Q_1(H)$ (univariate) | 0.00 | 0.01 | 0.01 | 0.01 |
| $Q_2(H)$ (multivariate) | 0.00 | 0.00 | 0.00 | 0.00 |
| $LM(H)$ | 0.00 | 0.00 | 0.01 | 0.00 |

The question is, however, whether modelling the change in volatility by a Markov switching mechanism is preferable to other volatility models. We have fitted a range of other volatility models and present the corresponding model selection criteria in Table 4, where we signify a heteroskedastic VAR(3) model with $m$ changes in variance as H($m$)-VAR(3) and the transition variable for the ST-VAR(3) model is $s_t = t$, as in the simulations.

In Table 4 we present all three criteria for all the models and it turns out that the MS(3)-VAR(3) model minimizes all the criteria simultaneously. Thus, it is clearly the preferred model. It is also interesting to note that the criteria are maximized for the plain VAR model. In other words, any one of the models accounting for changing volatility is preferred to a model that does not allow for heteroskedasticity. Of course, this result is not surprising given the residual series displayed in Figure 1. It is also worth pointing out that the MS(2)-VAR(3) model is the second best choice according to all three criteria and among the

Table 4: Comparison of Time-Varying Residual Volatility Specifications for a VAR(3) for the Example Data

| Fitted model | AIC | HQ | BIC |
|---|---|---|---|
| Plain VAR(3) | 8614.94 | 8671.96 | 8758.99 |
| H(1)-VAR(3) | 8421.17 | 8489.29 | 8593.24 |
| H(2)-VAR(3) | 8333.72 | 8401.83 | 8505.78 |
| ST-VAR(3) | 8400.88 | 8470.58 | 8576.94 |
| MS(2)-VAR(3) | 8301.76 | 8371.45 | 8477.82 |
| MS(3)-VAR(3) | **8201.40** | **8282.18** | **8405.47** |
| GARCH-VAR(3) | 8436.00 | 8507.27 | 8616.06 |

models with time-varying volatility the GARCH-VAR(3) is the least desirable model. Apparently, the volatility changes can be captured better with the Markov switching mechanism than with a GARCH type volatility process.

In Table 5 the results of the ARCH tests of Subsection 3.2 applied to the standardized residuals, that is, the estimated $\varepsilon_t = B^{-1}u_t$ divided by the respective estimated standard deviation or conditional standard deviation, are reported. The LM test and the multivariate $Q_2$ test both reject the null hypothesis of no ARCH at the 1% significance level for all but the MS(3)-VAR(3) and the GARCH-VAR(3) models at least for some lag. This result is another indication that most of the models do not properly capture all volatility changes. Although the MS(3)-VAR(3) model would also be rejected at a 5% level for lag 1 by the $Q_2$ test and for lags 1 and 10 by the LM test, it apparently takes better care of the volatility changes than its competitors. The large $p$-values of all three tests for the GARCH-VAR model are not surprising given that the residuals of this model are already prefiltered by a GARCH model. Taking this feature into account, the diagnostic tests are overall supportive of the MS(3)-VAR(3) model. As expected on the basis of our simulation results, the diagnostic tests are not as useful to discriminate between the volatility models as the model selection criteria, however, because they do not clearly point to a specific model and are biased towards the GARCH model by construction.

Once a volatility model has been decided upon, it can be used for structural analysis as in Lütkepohl and Netšunajev (2014). Since these authors use a MS(3)-VAR(3) model for their analysis, we can now draw on their results. They find that a recursive structural model is supported by the data so that an impulse response analysis can be based on a recursive model that takes into account the changes in volatility. It may still be instructive to compare the structural impulse response estimates obtained with the MS(3)-VAR(3) model to those from a plain recursive VAR(3) model which does not account for heteroskedasticity. We have plotted the impulse responses from both models in Figure 2. Obviously, most impulse responses from the two models are qualitatively similar. This is partly due to the

Table 5: ARCH-Tests on Standardized Estimated Residuals for Alternative Models for the Example Data

| Test | Model | $p$-values | | | |
|---|---|---|---|---|---|
| | | $H=1$ | $H=3$ | $H=5$ | $H=10$ |
| $Q_1(H)$ (univariate) | H(1)-VAR(3) | 0.04 | 0.20 | 0.41 | 0.81 |
| | H(2)-VAR(3) | 0.00 | 0.00 | 0.00 | 0.01 |
| | ST-VAR(3) | 0.09 | 0.36 | 0.63 | 0.94 |
| | MS(2)-VAR(3) | 0.00 | 0.00 | 0.00 | 0.02 |
| | MS(3)-VAR(3) | 0.25 | 0.38 | 0.59 | 0.49 |
| | GARCH-VAR(3) | 0.58 | 0.94 | 0.99 | 0.86 |
| $Q_2(H)$ (multivariate) | H(1)-VAR(3) | 0.00 | 0.00 | 0.08 | 0.57 |
| | H(2)-VAR(3) | 0.00 | 0.00 | 0.01 | 0.00 |
| | ST-VAR(3) | 0.00 | 0.05 | 0.40 | 0.56 |
| | MS(2)-VAR(3) | 0.00 | 0.00 | 0.00 | 0.04 |
| | MS(3)-VAR(3) | 0.03 | 0.72 | 0.32 | 0.05 |
| | GARCH-VAR(3) | 0.82 | 1.00 | 1.00 | 0.85 |
| $LM(H)$ | H(1)-VAR(3) | 0.00 | 0.00 | 0.07 | 0.12 |
| | H(2)-VAR(3) | 0.00 | 0.00 | 0.02 | 0.00 |
| | ST-VAR(3) | 0.00 | 0.05 | 0.43 | 0.09 |
| | MS(2)-VAR(3) | 0.00 | 0.00 | 0.00 | 0.02 |
| | MS(3)-VAR(3) | 0.03 | 0.78 | 0.35 | 0.04 |
| | GARCH-VAR(3) | 0.82 | 1.00 | 1.00 | 0.86 |

ordering of the shocks of the MS(3)-VAR(3) model. Recall that the shocks of this model do not have a natural order and can be reordered arbitrarily. The ordering in Figure 2 is such that the shape of the impulse responses from the MS(3)-VAR(3) model are roughly in line with that from the plain VAR model. Thereby the shocks can be given the same names as for the plain VAR model. Despite the qualitative similarity of most impulse responses the quantitative effects are partly quite distinct. For example, the oil supply shock has a much weaker impact on oil production in the MS(3)-VAR(3) model than in the plain VAR model although it boosts oil production rather persistently in both models.

# 7 Conclusions

In this study we have investigated strategies for choosing a suitable volatility model for identifying structural shocks through heteroskedasticity in a SVAR analysis. The performance of the information criteria AIC, HQ and BIC is compared. We have also considered tests for left-over heteroskedasticity in VAR models. Our investigation is based on an extensive simulation study using four DGPs for the volatility models, similar to models that have been used in the structural VAR

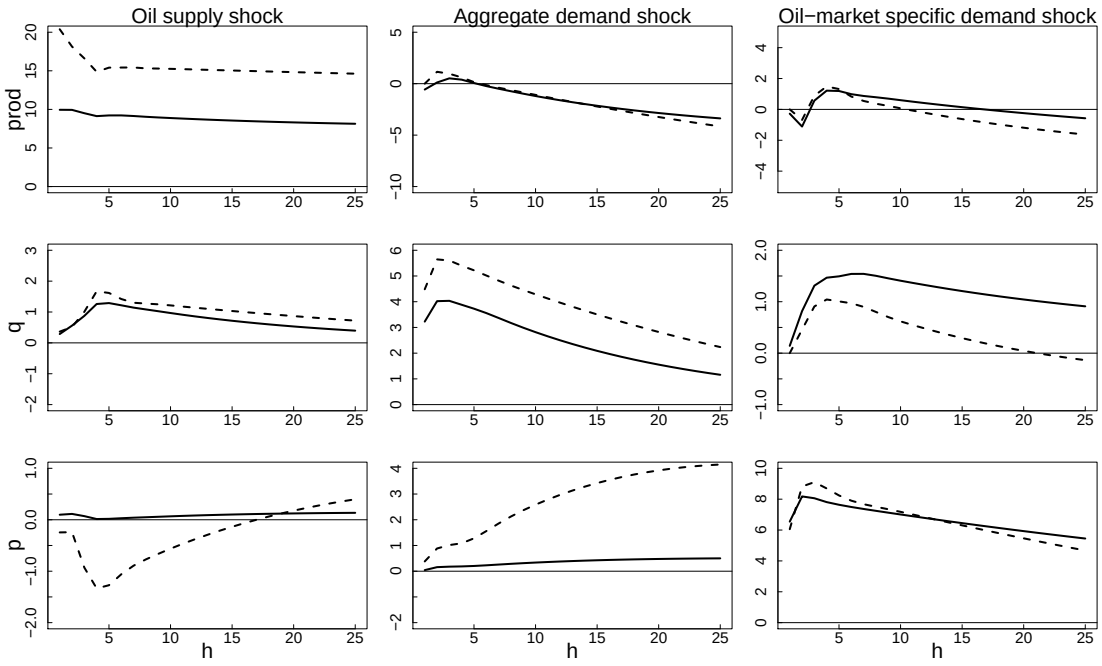


Figure 2: Comparison of impulse responses of plain VAR(3) (dashed line) and MS(3)-VAR(3) (solid line) model.

literature on identification through heteroskedasticity.

We find that among the three information criteria using AIC is preferable to HQ and BIC. AIC outperforms its competitors in its ability of correctly indicating the true model and also tends to provide the impulse response estimates with the smallest MSEs. Depending on the sample size, it may not find the true DGP very frequently. Generally it may require rather large samples of $T = 500$, for example, to detect the true DGP with some probability. For some of our DGPs a reliable detection is possible with much smaller samples, however. The heteroskedasticity tests are useful tools for indicating the presence of time-varying volatility but are not very helpful for selecting a specific model. They indicate that the models selected by the information criteria take care very well of the residual variance changes. AIC is again better than HQ and BIC in this respect. Using any one of the model selection criteria may provide substantially smaller MSEs of impulse responses than an arbitrarily chosen model.

From our results the following preferred strategy for using identification through heteroskedasticity emerges. First the data are checked for time-varying volatility by one of the (standard) diagnostic tests considered in Section 3.2. If time-varying volatility is detected, AIC should be used for selecting a suitable candidate model for the volatility process. Using any one of the information criteria is preferable to choosing the model by convenience or personal preference, as was done in some previous studies in the related SVAR literature.

We have used this strategy for modelling volatility changes in a VAR model for the global oil market and find evidence for time-varying volatility in the residuals of a VAR(3). In this example all three model selection criteria favour the same MS-VAR model for volatility changes.

Our study has obvious limitations. As usual, simulation studies are based on specific DGPs and the conclusions may not be generalizable. In the present simulation study we have considered exclusively bivariate DGPs. While this is special, it should be clear that the methods considered for structural VAR analysis are mainly relevant for low-dimensional processes because frequentist ML estimation is very difficult or impossible numerically for some of the volatility models and high-dimensional processes. For example, MS-VAR and GARCH-VAR models are difficult to estimate for higher-dimensional processes. Hence, considering the bivariate case is perhaps not a severe restriction.

Another limitation of our study is that in practice further choices have to be made by the analyst in addition to the volatility model. Notably the VAR lag order and the number of volatility states in a MS or heteroskedastic VAR model have to be specified. In our simulations we have assumed that these quantities are known whereas in practice one may also choose them by statistical tools, as in our example. This simplification may be reasonable to focus attention on the choice of volatility models. In future research it may be of interest, however, to study further aspects of joint model selection in the context of structural VAR analysis.

# References


Akaike, H. (1974). A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**(6): 716–723.

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

Becker, R. and Clements, A. E. (2008). Are combination forecasts of S&P 500 volatility statistically superior?, *International Journal of Forecasting* **24**: 122–133.

Becker, R., Clements, A. E., Doolan, M. B. and Hurn, A. S. (2015). Selecting volatility forecasting models for portfolio allocation purposes, *International Journal of Forecasting* **31**: 849–861.

Bouakez, H. and Normandin, M. (2010). Fluctuations in the foreign exchange market: How important are monetary policy shocks?, *Journal of International Economics* **81**(1): 139 – 153.

Caporin, M. and McAleer, M. (2011). Ranking multivariate GARCH models by problem dimension: An empirical evaluation, *Working Paper No. 23/2011*, University of Canterbury, Christchurch.

Caporin, M. and McAleer, M. (2012). Model selection and testing of conditional and stochastic volatility models, *in* L. Bauwens, C. Hafner and S. Laurent (eds), *Handbook of Volatility Models and Their Applications*, Wiley, New York, chapter 8, pp. 199–224.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**: 987–1007.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression, *Journal of the Royal Statistical Society. Series B* **41**(2): 190–195.

Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?, *Journal of Applied Econometrics* **20**: 873–889.

Herwartz, H. and Lütkepohl, H. (2014). Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks, *Journal of Econometrics* **183**(1): 104–116.

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

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

Kilian, L. and Murphy, D. (2012). Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models, *Journal of the European Economic Association* **10**: 1166–1188.

Kim, D. (2014). Maximum likelihood estimation for vector autoregressions with multivariate stochastic volatility, *Economics Letters* **123**: 282–286.

Lanne, M. and Lütkepohl, H. (2008). Identifying monetary policy shocks via changes in volatility, *Journal of Money, Credit and Banking* **40**(6): 1131–1149.

Lanne, M., Lütkepohl, H. and Maciejowska, K. (2010). Structural vector autoregressions with Markov switching, *Journal of Economic Dynamics and Control* **34**(2): 121–131.

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

Laurent, S., Rombouts, J. V. K. and Violante, F. (2012). On the forecasting accuracy of multivariate GARCH models, *Journal of Applied Econometrics* **27**: 934–955.

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

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

Lütkepohl, H. and Netšunajev, A. (2014). Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs, *Journal of Applied Econometrics* **29**(3): 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.

Milunovich, G. and Yang, M. (2013). On Identifying Structural VAR Models via ARCH Effects, *Journal of Time Series Econometrics* **5**(2): 117–131.

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

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies* **72**: 821–852.

Rigobon, R. (2003). Identification through heteroskedasticity, *Review of Economics and Statistics* **85**(4): 777–792.

Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics* **6**(2): 461–464.

Sentana, E. and Fiorentini, G. (2001). Identification, estimation and testing of conditionally heteroskedastic factor models, *Journal of Econometrics* **102**(2): 143 – 164.

Silvennoinen, A. and Teräsvirta, T. (2009). Multivariate GARCH models, *in* T. G. Andersen, R. A. Davis, K.-P. Kreiß and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer-Verlag, Berlin, pp. 201–229.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility, *Econometrica* **65**: 59–73.

van der Weide, R. (2002). GO-GARCH: A multivariate generalized orthogonal garch model, *Journal of Applied Econometrics* **17**(5): 549–564.

van Dijk, D., Teräsvirta, T. and Franses, P. H. (2002). Smooth transition autoregressive models — a survey of recent developments, *Econometric Reviews* **21**(1): 1–47.

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.